



Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft

NACHRICHTEN



Jahrgang 34 • 4/2002

Hochleistungsrechnen

**Im Inhaltsverzeichnis verwendete
Abkürzungen:**

HIK	Hauptabteilung Informations- und Kommunikationstechnik
IFP	Institut für Festkörperphysik
IKET	Institut für Kern- und Energietechnik
IMK	Institut für Meteorologie und Klimaforschung
IRS	Institut für Reaktorsicherheit

Herausgeber:

Forschungszentrum Karlsruhe GmbH
Postfach 3640 · D-76021 Karlsruhe
Telefon-Nr. (07247) 82-0

Redaktion:

Dr. Klaus Körting

Redaktionsbeirat:

Prof. Dr. J. Blümer, IK-IEKP
Dr. W. Breitung, IKET
Prof. Dr. E. Dinjus, ITC-CPV
Dr. J. Gspann, IMT
Dr. J. Hoffmann, ÖA
Dr. R. Krieg, IRS
Dr. W. Pfleging, IMF-I
Prof. Dr. U. Schurath, IMK (Vorsitzender)
Dr. K.-F. Weibezahn, IMB

Grafik und Satz:

Stolz Grafisches Atelier · Karlsruhe

Layout:

Tassilo Schnitzer

Druck:

Wilhelm Stober GmbH · Eggenstein

Nachdruck mit Genehmigung des
Forschungszentrums Karlsruhe GmbH
unter Nennung der Gesellschaft und
des Autors gestattet. Beleg erbeten.

Die NACHRICHTEN
können kostenlos über die
Hauptabteilung Bibliothek und Medien
des Forschungszentrums bezogen
werden.

Printed in the Federal Republic of
Germany

Gedruckt auf chlorfrei gebleichtem
Papier



„Einer der wichtigsten Erfolgsfaktoren
für die Spitzenforschung und die In-
dustrie wird künftig der schnelle und
standortunabhängige Zugriff auf welt-
weit verteilte Daten und Ressourcen
der Informationstechnik sein.“

M. Kunze und H. Marten auf Seite 280.

Inhalt / Contents

- **Einführung in das Hochleistungsrechnen – Vergangenheit, Gegenwart und Zukunft** 269
Introduction to Supercomputing – Past, Presence, and Future

F. Schmitz, HIK

At the Forschungszentrum Karlsruhe, use of computers for scientific-technical applications started in 1959 with the Z22 and was pushed in particular by the nuclear engineering institutes. In 1960, the Z22 correctly predicted criticality of the FR2 research reactor. Due to constantly increasing requirements above all in the non-nuclear fields, so-called supercomputers have entered the environment since 1983. From 2006, these classical supercomputers will presumably be replaced by (low-cost) SMP-based clusters. This future development may be reflected by the term “GRID”.

- **Das Virtuelle Rechenzentrum Karlsruhe – Beispiel einer engen Kooperation zwischen Großforschung und Universität** 275

The Karlsruhe Virtual Computing Centre – an Example for a Close Cooperation of a Large Research Centre with an University

K. Mickel, HIK; W. Juling, Universität Karlsruhe

Nowadays it's very difficult, to find IT-experts, and it's very expensive to buy and to support one or more high-performance computers (HPCs). Therefore the both large computer centres of the Forschungszentrum Karlsruhe and of the University of Karlsruhe have established a Virtual Computer Centre (VCC). In this very close partnership, every partner is responsible for operation and support of only one HPC architecture, but the scientists and students of both institutions can use the different HPCs of both partners without any restriction. In this way it becomes possible, to specialize the IT-services on both sides and to save a lot of money.

- **Grid-Computing und e-Science** 280
Grid-Computing and e-Science

M. Kunze, H. Marten, HIK

A Grid is an environment that provides access and management for the whole range of computing resources needed to solve complex computing and data handling problems. A Grid is a well understood and standardized set of interfaces and Web-Services for directory services, brokerage and usage. New high-bandwidth networking technology allows for the synergetic usage of resources and thus yields in a new class of e-Science applications. The LHC Computing Grid project serves as an example to discuss the technical realization.

■ **Mathematische Programmbibliotheken im Wandel der Zeit**

284

Mathematical Program Libraries in the Changing Times

I. Bischoff-Gauß, R. Pietschmann, F. Schmitz, HIK

Program Libraries have been important since the beginning of electronic data processing. With regard to technology and applied science libraries offered subroutines for simple elementary functions up to complex Fast Fourier Transforms, and also utility programs for using I/O-devices. Mathematical Libraries especially are the key to efficient running application programs, because the programmers put all their skill and knowledge of hardware features in these collection of algorithms and computer programs. Modern libraries use interfaces to hardware-optimized basic libraries enabling programs to run efficiently on a variety of computers.

■ **Monte-Carlo-Simulation des Teilchentransports und Anwendung in der Fusionsneutronik**

289

Monte Carlo Simulation of Particle Transport and Application in Fusion Neutronics

U. Fischer, Y. Chen, IRS

The Monte Carlo method for simulating particle transport phenomena on the microscopic level is described. A programme system is presented which enables to calculate three-dimensional dose rate distributions for any complex geometry based on the Monte Carlo technique. Application calculations are described which haven been performed for the ITER experimental fusion reactor by making use of the FZK Linux cluster in the parallel mode.

■ **Numerische Simulation von Gasedetonationen im ITER-FEAT-Fusionsreaktor**

295

Numerical Simulation of Gaseous Detonations in the ITER-FEAT Fusion Reactor

R. Redlinger, IKET; L. Obholz, HIK

The parallelized version of the finite difference code DET3D enables one to simulate 3-dimensional gaseous detonations on very fine numerical meshes. Thus, calculations with discretized geometry models of, say, 10 million numerical cells present no problem. In the present paper, this is illustrated by calculations which were performed for a safety study of the planned ITER-FEAT fusion reactor.

■ **Hochtemperatursupraleiter – Festkörpersimulation** 302
High-Temperature Superconductors – Solid-State Simulation

K.-P. Bohnen, R. Heid, IFP

The role of electron-phonon coupling has been a long-standing issue since the discovery of high-temperature superconductors (HTSC) in 1986. New theoretical developments together with the increasing power of supercomputers make it now possible to calculate the lattice dynamics and the electron-phonon coupling very accurately even for complex systems like the HTSCs. We discuss in this article the most important steps in this approach with respect to CPU-time and memory requirements for the system $\text{YBa}_2\text{Cu}_3\text{O}_7$.

■ **Computeralgebra: Schlüssel zu Präzisionsrechnungen in der Teilchenphysik** 305
Computer Algebra, a Decisive Tool for Precision Calculations in Particle Physics

J.H. Kühn, Institut für Theoretische Teilchenphysik, Universität Karlsruhe
M. Steinhauser, II. Institut für Theoretische Physik, Universität Hamburg

During the past two decades computer algebra has developed into a decisive tool for precision calculations of cross sections and decay rates of current interest in particle physics. In many cases analytic, computer aided evaluations are superior to numerical calculations. After a brief history of these developments a few, typical packages will be presented, similarities and differences described. Parallelization as well as automatized generation of these complicated and extensive programs are central current developments.

■ **Modellierung vom Feinsten** 311
Modelling of Fine Particles

N. Riemer, H. Vogel, IMK

Aerosol particles are harmful for human health and have an impact on the climate on the regional and global scale. They are composed of numerous substances which reflect the diversity of their sources and locations. To investigate the impact of aerosol particles on the system atmosphere, three dimensional models are a powerful tool. The comprehensive mesoscale model system KAMM/DRAIS has therefore been extended to calculate aerosol dynamical processes. Since the representation of the aerosol phase requires considerable computer resources the application of super computers is essential.

■ **Das Karlsruher Simulationsmodell der mittleren Atmosphäre –
Modellbeschreibung und Anwendungen**

317

**The Karlsruhe Simulation Model of the Middle Atmosphere –
Model Description and Applications**

W. Kouker, T. Reddmann, R. Ruhnke, IMK

Despite international agreements on the protection of the stratospheric ozone layer, the recovery of the ozone layer is expected to be delayed as a result of the greenhouse gas effect. At the IMK, the Karlsruhe Simulation Model of the Middle Atmosphere (KASIMA) has been developed to simulate stratospheric processes in order to enhance the reliability of the future prediction of the stratosphere. The model results of KASIMA are compared to measurements which are obtained by different field campaigns. Here, studies investigating the transport processes of the stratosphere at different scales as well as the stratospheric chemistry leading to the ozone hole above the Antarctica are presented.

■ **Berichtigung
Correction**

321

Einführung in das Hochleistungsrechnen – Vergangenheit, Gegenwart und Zukunft

F. Schmitz, HIK

Einleitung

Zur Entwicklung des ersten deutschen Eigenbau-Forschungsreaktors FR2 wurde 1959 als erster Rechner die Z22 im damaligen Institut für Neutronenphysik und Reaktortechnik (INR) installiert. Dies war die Keimzelle für den kontinuierlichen Rechnerausbau, und von hier aus begann 1968 die Gründung der DVZ (Datenverarbeitungszentrale). Diese stellte die damals leistungsfähigsten Rechner von IBM den Benutzern unterschiedlicher Fachrichtungen zur Verfügung. 1983 konnten durch eine gemeinsame Beschaffung mit der Universität Karlsruhe die Wissenschaftler im damaligen Kernforschungszentrum Karlsruhe einen Hochleistungsrechner nutzen. Dies war der Anfang einer bis heute fortgeführten Kooperation, die 1994 in der Bildung des Virtuellen Rechenzentrums mündete.

Hochleistungsrechnen – Höchstleistungsrechnen, was ist denn das?

Wird der eigene PC oder die Workstation zu klein für die Anwendung, so versucht man auf Rechner einer anderen Leistungsklasse auszuweichen. Dies sind in der Regel größere Compute-Server, die sowohl einen höheren Speicherbedarf als auch die Nachfrage nach stärkeren Prozessoren befriedigen können.

Übersteigt der Bedarf nun wiederum die gängigen Ausstattungen von Compute-Servern, besteht die Möglichkeit, bei den sogenannten Hochleistungsrechen-

zentren zu rechnen. Diese sind besser ausgestattet und bieten dem Benutzer State-of-the-art-Leistung in den Bereichen Compute-Leistung, Hauptspeicherbedarf, I/O-Leistung und Kommunikationsbandbreite. Darüber hinaus gibt es in Deutschland von einem Ausschuss des Wissenschaftsrates (dem „Nationalen Koordinierungsausschuss zur Beschaffung und Nutzung von Höchstleistungsrechnern“) genehmigte und von Bund und dem jeweiligen Bundesland finanzierte Zentren, die zum Zeitpunkt der Installation die jeweilige Spitzentechnologie zur Verfügung stellen. Dies sind momentan das HLRS in Stuttgart, HLRB in München und das NIC in Jülich. Darüber hinaus existieren weitere Rechenzentren (DWD in Offenbach, HLRE in Hamburg, IPP in Garching und HRZN Hochleistungsrechner Nord), die erst kürzlich mit leistungsfähiger Hardware ausgestattet wurden.

In den Vereinigten Staaten von Amerika hat man verstärkt in die

ASCI-Projekte investiert um von den teureren Rechnern mit Spezialprozessoren auf preisgünstigere Plattformen umzuschwenken. Bei diesen Projekten wird Wert auf Commodity-Komponenten gelegt. Vektorprozessoren werden hier nicht mehr eingesetzt.

Um jetzt Hochleistungsrechnen gegen Höchstleistungsrechnen „sinnvoll“ abzugrenzen, kann man momentan (2002) bei etwa 15 Mio. € Hardware-Investition eine Art Trennlinie ziehen [4]. Zu bedenken ist, dass 1967 eine Telefunken-Digitalrechenanlage TR440 13 Mio. DM und 1985 eine IBM 3090/200 sogar 20 Mio. DM gekostet hat.

Vom Hochleistungsrechner zum Grid

Eine kurze Einführung

Konrad Zuse, John V. Atanasoff, Clifford Berry, Howard Aiken und Grace Hopper schufen ab 1936 die Voraussetzungen für die heutige Datenverarbeitung (Abb. 1).



Abb.1: ENIAC I, Electronical Numerical Integrator And Calculator (Gewicht: 30 Tonnen).

Mit der Erfindung des Transistors 1947 und der Entwicklung des ICs 1958 begann die Miniaturisierung, die Supercomputer aber auch PCs erst möglich gemacht hat.

1964 begann mit der CDC 6600 die Ära der Supercomputer mit 1 Megaflop/s (millions floating point operations per second, 1 Megaflop/s entspricht 10^6 Gleitkomma-Operationen pro Sekunde, auch Megaflops), fünf Jahre später folgte der Nachfolgerechner CDC 7600 mit einer Spitzenleistung von 10 Megaflop/s. 1971 brachte Intel mit dem 4004 den ersten Mikroprozessor der Welt heraus. Zwei Jahre später erschien mit dem Intel 8080 der erste 8-Bit-Allzweckmikroprozessor und 1975 der erste PC (Name erst 1981 von IBM geprägt) auf den Markt: Der MITS Altair 8800. Hier sind auch Apple I (1976), Commodore Pet 2001 (1977), Sinclair ZX-80 (1980), IBM PC (1981) zu nennen (www.weller.to).

1974 wurde der erste „massiv parallele“ Rechner mit 64 PEs (processing elements) im NASA Ames Research Center installiert. Er fand Einsatz in den Bereichen Strömungssimulation, Bildverarbeitung und Wettervorhersage. Diese Array-Prozessoren, die noch in verschiedenen Rechnern (CM-2, MasPar) eingebaut wurden, werden nicht mehr hergestellt, da alle PEs zu einem Takt immer nur die gleiche Instruktion, jedoch auf anderen Daten ausführen. Mehr durchgesetzt haben sich seit 1969 die Vektorprozessoren, die nach dem Fließbandprinzip die Datenströme abarbeiten und dadurch nach einer start-up Zeit pro Takt und Pipeline ein Ergebnis liefern. 1976 startete Cray mit einem 133-Megaflop/s-Rechner (Abb. 2), der von Los Alamos National Laboratory für M\$ 8.8 gekauft wurde und über 8 Megabyte Hauptspeicher verfügte, eine erfolgreiche Firmen-

geschichte, die im März 1996 von der Übernahme durch SGI (Silicon Graphics Inc.) beendet wurde. Von diesem Rechner wurden weltweit 35 Exemplare verkauft, auch nach Deutschland. 1979 erhielt die Max-Planck-Gesellschaft IPP die erste Cray in Deutschland. Nach 1980 wurde von CDC die Cyber 205 und ab 1983 von Fujitsu die VP-Linie (in Deutschland von Siemens ab 1986 vermarktet) als Gegengewicht zu den Cray-Rechnern am Markt positioniert. Etwas später stiegen auch NEC und Hitachi in diesen Markt ein. Auch versuchte sich IBM mit einem Vektorzusatz zur IBM 3090 ein Stück vom Kuchen zu sichern. Dies war jedoch genauso erfolglos wie die Gründung der Tochterfirma ETA von CDC, die mit den Nachfolgerechner der Cyber 205, der ETA 10, im April 1989 wegen technischer Probleme scheiterte. Convex versuchte mit seinen Cray-kompatiblen Rechnern Kunden im Hochschulbereich anzusprechen. Letztendlich werden aber die Parallelrechner die durch die hohe CPU- und I/O-Leistung dominierenden Vektorrechner verdrängen können, da nur so der Leistungshunger vieler Anwendungen zu stillen ist. Die Applikationen müssen vom Anwender parallelisiert werden und dadurch steigt die Komplexität, da neben CPU-Leistung auch Bandbreite und start-up time des Kommunikationsnetzwerkes sowie die I/O-Leistung zu berücksichtigen sind. Jedoch war im Jahr 2002 der leistungsfähigste Rechner ein Vektor-Parallelrechner mit einer Spitzenleistung von 40 Tflop/s (Earth Simulator in Japan, SX-6 von NEC). Dieser Rechner ist leistungsfähiger (www.top500.org) als die 12

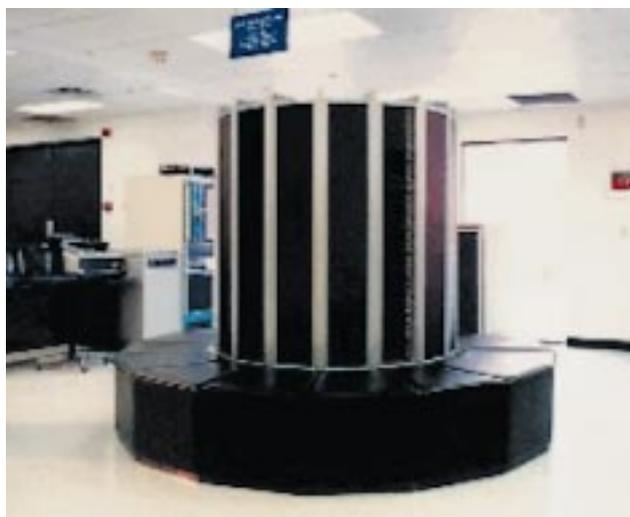


Abb. 2: 133-Megaflop/s-Rechner von Cray (1976).

nächstplatzierten Rechner, die überwiegend in den USA installiert sind.

Bei den Parallelrechnern, deren beschleunigte Entwicklung mit dem Sinken der Preise für Hardware einhergeht, begann der kommerzielle Erfolg 1985 mit dem Aufkommen der Transputern von Inmos (T400, T800), welche sogar als Zusatzkarten für PCs erhältlich waren. Man fing an, bis auf Ausnahmen (CM-5, KSR-1, VPP-Linie, Cray, ...), preiswerte massiv parallele Systeme auf der Basis von Standardhardware zu fertigen. Viele Ansätze verschwanden (Supremum, iPSC/860, ip-Systems), weil sie im Hard- oder Softwarebereich ihre Schwächen hatten. Geblieben sind die großen Vektor-Parallel-Rechner, die SMP-Cluster (IBM, SGI, HP, SUN) und die PC-Cluster.

Vektorrechner haben heute schon die Einzelprozessorleistung, die man etwa 2005 von Intel-basierten Prozessoren erwarten kann. Die Applikationsperformance auf einem Prozessor wird aber erst 2006/2007 gleichziehen können. Dann wird auch die Ära der Vektorrechner nach fast 40 Jahren der Vergangenheit angehören, denn für ein Nischendasein als Spezialrechner dürften Vektor-Parallel-Rechner zu kostspielig sein.

Will man all die weltweit verfügbaren Ressourcen allgemein verfügbar machen (so wie „Strom aus der Steckdose“), dann bedarf es einer geeigneten Vernetzung und der Standardisierung von Schnittstellen. Dies sind die ersten Schritte zu einem weltweiten Netzwerk (Grid), in dem Datenspeicher (Daten-Grid), Rechenleistung (Com-

pute-Grid) und anderes angeboten werden.

Geschichtliche Entwicklung im Forschungszentrum

Im Umfeld der Reaktortechnik lassen sich wesentliche Grundlagen zunächst nur durch aufwendige numerische Rechenmodelle simulieren und erforschen. Zur Unterstützung der Wissenschaftler bei solchen numerischen Rechnungen wurde bereits 1959 mit dem Röhrenrechner Z22 (Abb. 3) ein für damalige Verhältnisse leistungsfähiger Rechner im Institut für Neutronenphysik und Reaktortechnik betrieben. 1968 wurde schließlich mit der Datenverarbeitungszentrale (DVZ) das zentrale Rechenzentrum des Forschungszentrums gegründet, das seither ohne Unterbrechung für die Wissenschaftler des Zentrums leistungsfähige Rechner betreibt. In Kooperation mit der Universität Karlsruhe konnte dann ab 1983 der Vektorrechner Cyber 205 für große rechenintensive Programme genutzt werden. Im Herbst

1987 begann dann die Ära der Fujitsu-Vektorrechner im Forschungszentrum. Der erste Rechner, Modell VP50 mit 7 ns Zykluszeit, hatte 64 Megabyte Hauptspeicher und erbrachte eine Rechenleistung von etwa 138 Megaflop/s für die Matrixmultiplikation. Auch mit der IBM 3090 wurde 1990 versucht, durch 6 Prozessoren und einer zusätzlichen Vektoreinheit pro Prozessor die Leistungsfähigkeit zu steigern. Jedoch ergab sich nur dann ein Vorteil gegenüber dem VP50, wenn man parallelisierte. Da die IBM 3090/600VF jedoch auch im interaktiven Betrieb genutzt wurde, konnte sie sich nie für das Hochleistungsrechnen etablieren. Im Mai 1991 erfolgte die Ablösung des VP50 durch das leistungsfähigere Modell VP400-EX (1988 wurde gemeinsam von Universität und Forschungszentrum ein S600/20 beschafft, jedoch ein VP400-EX als Interimslösung an der Universität installiert). Nach der Abschaltung des VP400-EX im Herbst 1994 wurde ein Rechner der Firma Cray, eine J916 mit

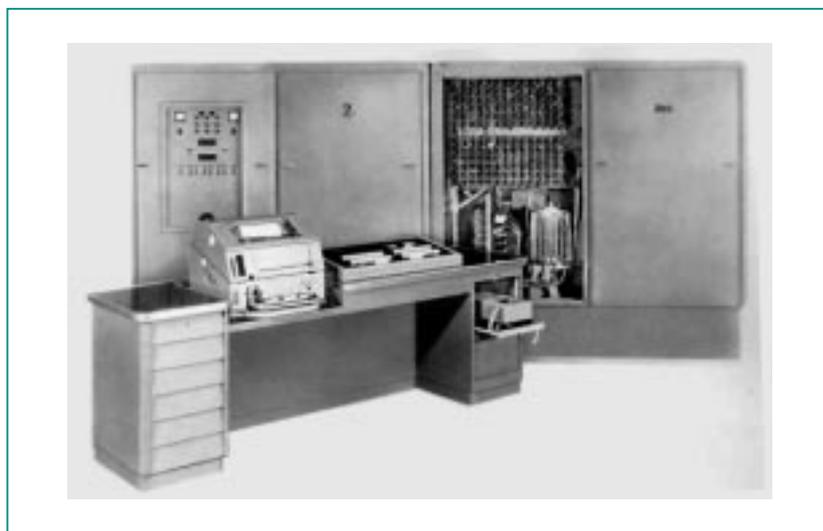


Abb. 3: Röhrenrechner Z22 von 1959.

16 Prozessoren und 4 Gigabyte Hauptspeicher als Ersatz beschafft. Nach anfänglichen Hardwareproblemen wurde der SMP-Rechner jedoch recht schnell von den Benutzern akzeptiert und musste erst im Herbst 1998 abgeschaltet werden, nachdem die

Auslastung des Rechners stark nachgelassen hatte. Seit Frühjahr 1996 befindet sich ein VPP300 im Forschungszentrum (Abb. 4), welcher, wie der IBM/RS6000 SP an der Universität, im Rahmen des Virtuellen Rechenzentrums (VRZ) betrieben wird. 1998 wurde das

erste kleine Linux-Cluster mit zwei Doppelprozessorknoten installiert und einem Institut zur Verfügung gestellt. Das Cluster wurde sukzessive ausgebaut und im Sommer 2000 durch eine hpcline der Firma Siemens-Nixdorf ersetzt. Diese Aktivitäten sind die Keim-

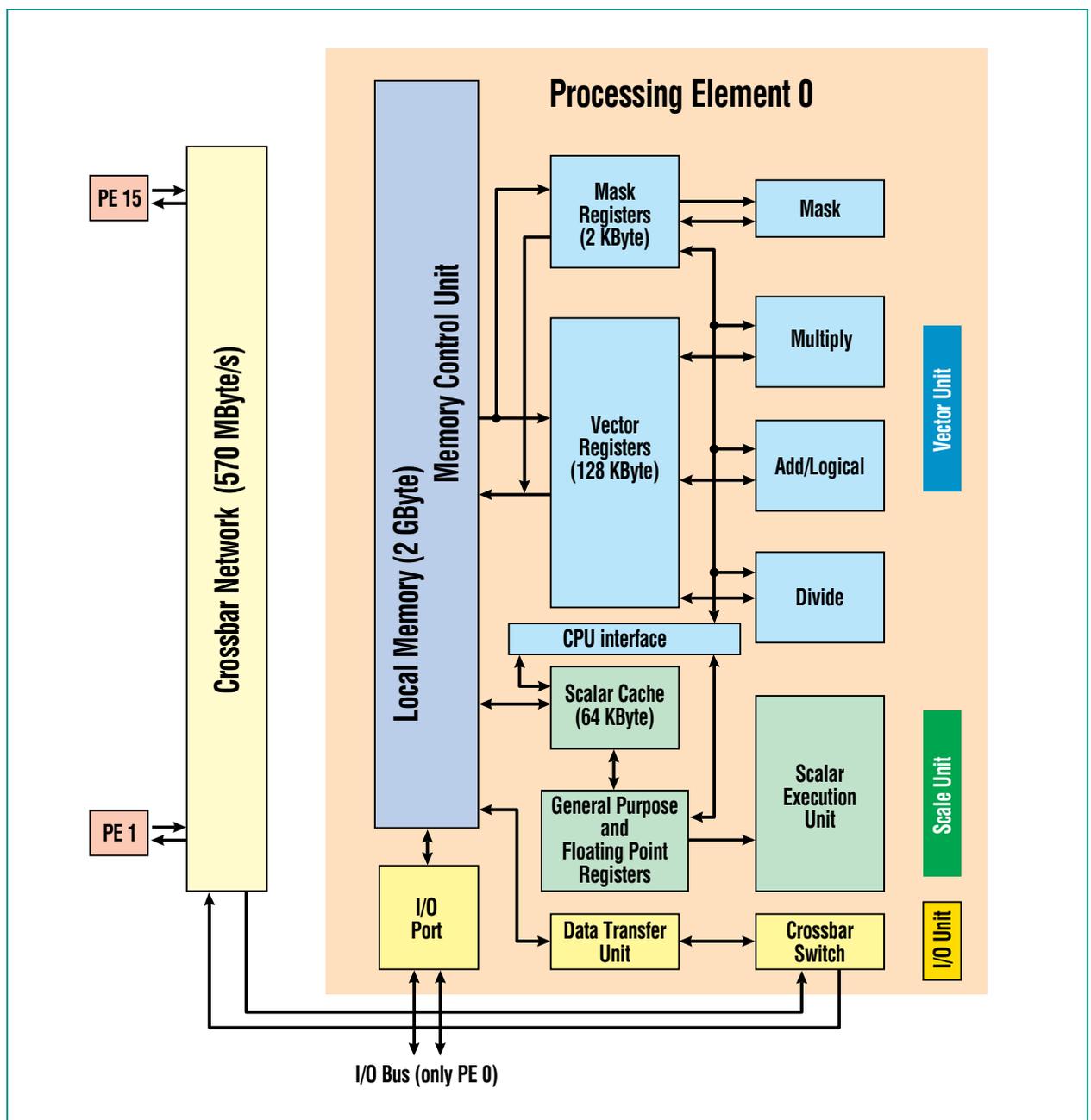


Abb. 4: Schaltbild des VPP300.

zelle des Grid-Computings im Forschungszentrum Karlsruhe. Außerdem wurde im Jahr 2002 ein Athlon-System 1800+ mit 10 Doppelprozessoren und Gigabit-Netzwerk in Betrieb genommen. Zusätzlich zum VPP300 wurde dessen Nachfolgerechner VPP5000 im Frühjahr 2000 mit vier Prozessoren installiert und ein Jahr später auf 8 Prozessoren – zwei davon mit 16 Gigabyte Hauptspeicher – aufgerüstet. Mit 9.6 Gigaflop/s Spitzenleistung verfügt dieser Rechner über den zweitschnellsten Prozessor weltweit.

Nur mit geringen Widerständen ging dann Ende 1999 im Forschungszentrum auch der letzte IBM kompatible Rechner mit dem Betriebssystem MVS in den Ruhestand. Seitdem sind ausschließlich Unix-Rechner im Einsatz. Die dadurch entstandenen Lücken wurden aber bereits vorsorglich durch AIX-Rechner (Unix-Computer der Firma IBM) geschlossen. 1993 stellte die damalige Hauptabteilung Datenverarbeitung und Instrumentierung (HDI) – heute Hauptabteilung Informations- und Kommunikationstechnik (HIK) – als zentraler Diensteanbieter drei AIX-Workstations zur Verfügung, die dann später durch ein Cluster mit vier Power2-Prozessoren abgelöst wurden. Dieses System wurde im Sommer 2002 stillgelegt. Ende 1999 beschaffte das damals neu gegründete Institut INT eine IBM Nighthawk (acht Power3-SP-Knoten mit jeweils acht 222-MHz-Prozessoren) welche von HIK betrieben wird. Dieser Rechner wurde durch zwei Vorrechner und zwei 4-Wege-Knoten als Compute-Server ergänzt.

Heutiger Stand

Nun, was hat das Rechenzentrum des Forschungszentrums Karlsruhe heute den Wissenschaftlern zu bieten?

Dies sind die folgenden Systeme:

- Ein VPP5000 mit 8 Prozessoren und 80 Gigabyte verteiltem Speicher. Jeder Prozessor liefert die Spitzenleistung von 9.6 Gigaflop/s, die etwa für die Matrixmultiplikation fast erreicht wird. Der lokale RAID-5-Plattenplatz liegt bei netto 730 Gigabyte. Damit lassen sich Datenübertragungsraten von 100-140 Megabyte/s realisieren. Zusätzlich stehen über einen Fileserver 2 Terabyte RAID-5-Platten netto zur Verfügung.
- Ein SP-Cluster aus 64 Power3 Nighthawk-Knoten der Firma IBM mit entsprechendem Switch. Jeder Knoten verfügt über 8 Prozessoren, die über einen Crossbar auf einen gemeinsamen Hauptspeicher von 4 Gigabyte zugreifen. Die Gesamtleistung beträgt maximal 60 Gigaflop/s. Zusätzlich stehen zwei Vorrechner und weitere zwei Knoten mit jeweils vier Prozessoren, aber höherer Taktung, zur Verfügung. Dieser Rechner wurde vom Institut für Nanotechnologie gekauft und im Rechenzentrum aufgestellt.
- Das SP-Cluster der Universität Karlsruhe mit 256 Prozessoren (Winterhawk und Nighthawk Knoten) kann im Rahmen des Virtuellen Rechenzentrums (www.rz.uni-karlsruhe.de/Uni/RZ/VRZ) genutzt werden.

- Ein PC-Cluster der Firma FSC (Fujitsu Siemens Computers) mit 16 Doppelprozessorboards, Intel PII mit 700 MHz und jeweils 1 Gigabyte Hauptspeicher. Verbunden sind die Rechenknoten mit einem preiswerten Fast-Ethernet. Die Leistungsfähigkeit des Gesamtsystems liegt bei etwa 22 Gigaflop/s, wovon aber, anders als beim Vektorrechner VPP5000, nur 20% von normalen Anwendungen erreicht werden. Hier besteht noch erheblicher Entwicklungsbedarf.
- Ein lokaler Lieferant baute ein System mit 10 Athlon 1800+ Doppelprozessoren, Gigabit-Netzwerk und bis zu 3 Gigabyte Hauptspeicher. Dieses System ist seit Frühjahr 2002 in Betrieb.

Zukünftige Entwicklungen

Die momentan verfügbaren Hardwareressourcen im Forschungszentrum zeichnen sich sowohl durch ihre Vielfalt als auch durch die Exklusivität aus. Rechner mit einem Hauptspeicher von 16 Gigabyte für eine einzige nicht parallelisierte Anwendung sind selten, werden aber von den Benutzern nachgefragt. Auch Rechner mit großzügigen SMP-Fähigkeiten und einem schnellen Netzwerk sind für die großen Probleme der Physiker unabdingbar. Speziell auf dem Gebiet der preiswerten Komplettsysteme der Linux-Cluster versucht HIK die Bedürfnisse der Benutzer zu befriedigen. Neben Anwendungen aus der Neutronen-, Kern- und Reaktorphysik sowie Thermo- und Fluidodynamik, Strukturmechanik, Elektromagnetik werden die Rechner sehr stark durch Anwendungen aus der

Festkörperphysik und der Meteorologie genutzt.

Diese Anwendungen benötigen auch weiterhin starke Prozessoren mit den Zugriffsmöglichkeiten auf hochwertigen Speicher. Damit wird die Klasse der SMP-basier-

ten Systeme weiterhin im Forschungszentrum seine Berechtigung haben, ob darunter Vektorprozessoren zu finden sind, kann man bezweifeln. Eine Integration von Systemen des Rechenzentrums in ein mögliches HGF-weites „GRID“ wäre denkbar.

Danksagung

Der Autor dankt den Herren Sanitz und Schommers (HIK) für die Durchsicht des Manuskripts.

Literatur

- [1] H. Borgwaldt,
D. Sanitz,
E.G. Schlechtendahl,
KfK Nachrichten
Jahrgang 16 · 2/84
S. 109-116
- [2] W. Höbel, D. Sanitz
Nachrichten Jahrgang 28
2-3/96 S. 85-96
- [3] *Proceedings of the SC2001*
Denver
- [4] *Stellungnahme des Wissenschaftsrats zur Anmeldung des Landes Baden-Württemberg auf Errichtung eines Höchstleistungsrechner-Systems in Karlsruhe und Stuttgart zum 32. Rahmenplan Mannheim, 17.5.2002*

Das Virtuelle Rechenzentrum Karlsruhe – Beispiel einer engen Kooperation zwischen Großforschung und Universität

K. Mickel, HIK; W. Juling, Universität Karlsruhe

Einleitung

Seit 1993 veröffentlichen die Universitäten Mannheim und Tennessee gemeinsam jeweils im Juni und im November jedes Jahres die Liste TOP500, die einen aktuellen Überblick über die 500 leistungstärksten Computer der Welt gibt und auf deren Erscheinen die Fachwelt stets mit großer Ungeduld wartet [1]. Die theoretisch erreichbare Spitzenleistung von Hoch- und Höchstleistungsrechnern wird in dieser Liste in der hier üblichen Einheit Flop/s (floating point operations/second) angegeben; dieses Maß gibt an, wie viele Rechenschritte der jeweilige Computer pro Sekunde theoretisch maximal durchführen kann. Im Juni 1993 lag die summierte Leistung aller damals in der Liste TOP500 aufgeführten 500 Hochleistungsrechner bei $1,7 \times 10^{12}$ Flop/s, also bei 1,7 Teraflop/s. Nur neun Jahre später, im Juni

2002, betrug die summierte Spitzenleistung der 500 leistungstärksten Computer der Welt bereits 220 Teraflop/s; der auf Platz 1 stehende leistungstärkste Computer der Welt, der japanische „Earth Simulator“, leistet alleine 35,8 Teraflop/s. Aus Abbildung 1 ist zu entnehmen, dass die Rechenleistung der Höchstleistungsrechner dieser Welt sich alljährlich etwa um den Faktor 1,79 erhöht. Knapp die Hälfte der in der TOP500-Liste aufgeführten Computer ist dabei wissenschaftlichen Einrichtungen zuzurechnen, woraus man erkennen kann, dass auch und gerade der Rechenzeitbedarf der Wissenschaft ungestüm zunimmt, ohne dass ein Ende dieser Entwicklung absehbar wäre. Das gilt ganz besonders für naturwissenschaftlich-technisch orientierte Einrichtungen wie die Universität und das Forschungszentrum Karlsruhe.

Universität und Forschungszentrum – zwei namhafte Forschungseinrichtungen in Karlsruhe

Die Universität Karlsruhe (14.600 Studierende, etwa 4.000 Beschäftigte) ist die älteste technisch orientierte Universität in Deutschland. Nur 10 km von ihr entfernt befindet sich das Forschungszentrum Karlsruhe; hier arbeiten etwa 3.500 Mitarbeiter auf den Gebieten Struktur der Materie, Erde und Umwelt, Gesundheit, Energie sowie Schlüsseltechnologien. Die Institute des Forschungszentrums und der Universität arbeiten seit jeher auf zahlreichen Gebieten sehr eng zusammen, und es gibt in großem Umfang gemeinsame wissenschaftliche Arbeitsgruppen und Institute. Als sich das wissenschaftliche Rechnen in den 60er-Jahren zu einer

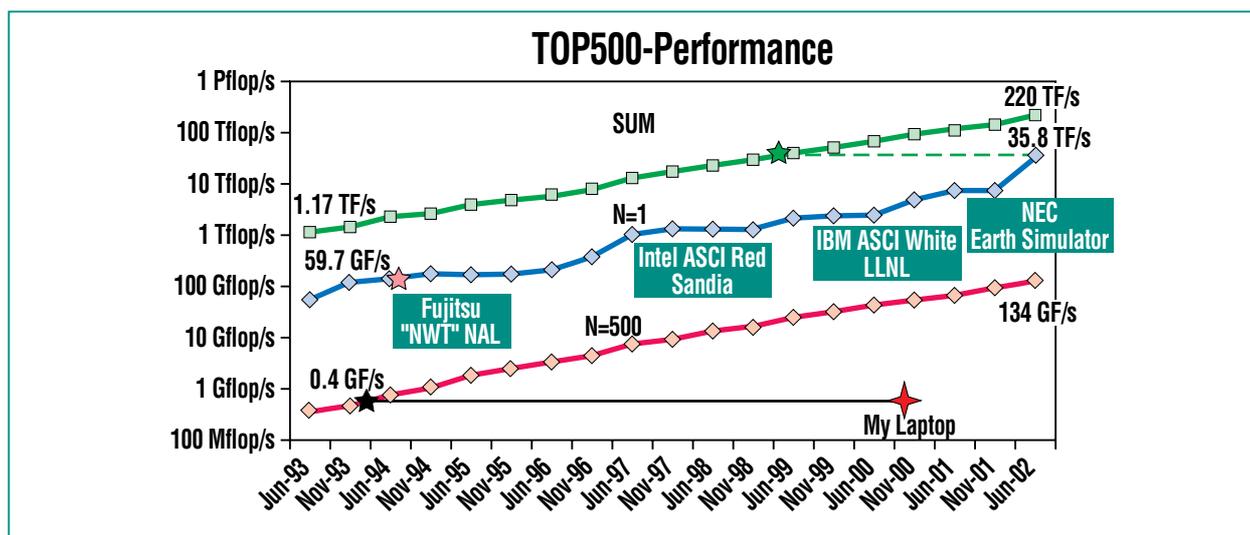


Abb. 1: Entwicklung der Leistungsfähigkeit der 500 weltweit leistungstärksten Höchstleistungsrechner von 1993 bis 2002. Die mit N = 1 markierte Kurve zeigt die Entwicklung der Leistungsfähigkeit des jeweils erstplatzierten, N = 500 die des jeweils letztplatzierten Computers in den einzelnen Listen. „SUM“ gibt die summierte Leistungsfähigkeit aller 500 Höchstleistungsrechner der jeweiligen TOP500-Liste an. (Quelle: <http://www.top500.org>)

zunehmend wichtigeren Komponente naturwissenschaftlicher und technischer Forschung entwickelte – heute gilt es neben Theorie und Experiment als dritte eigenständige wissenschaftliche Disziplin – wurden in beiden Einrichtungen große wissenschaftliche Rechenzentren eingerichtet, die bis heute zu den leistungsfähigsten ihrer Art in Deutschland zählen. Diese beiden Rechenzentren bieten seither in ihren jeweiligen Einrichtungen den Anwendern die ganze breite Palette der Informations- und Kommunikationstechnik an, ohne die eine naturwissenschaftlich-technische Forschungseinrichtung heutzutage nicht mehr konkurrenzfähig sein kann. Beide damals gegründeten Rechenzentren arbeiteten in der Folge und bis heute auf allen Gebieten der Informations- und Kommunikationstechnik und insbesondere bei Betrieb und Betreuung von Höchstleistungsrechnern sehr eng und vertrauensvoll zusammen, so dass beispielsweise schon immer Bedarfsspitzen je nach aktueller Ressourcen- und Bedarfsverteilung wechselseitig ausgeglichen werden konnten.

Die Idee des Virtuellen Rechenzentrums – doppelter Nutzen bei halbem Aufwand

Die systemtechnische Betreuung von Höchstleistungsrechnern ist überaus personalaufwändig, wengleich der Aufwand erfreulicherweise nicht proportional mit der erreichbaren Rechenleistung der Systeme ansteigt. Hier kommt jedoch erschwerend hinzu, dass Höchstleistungsrechner in unter-

schiedlichen „Architekturen“ angeboten werden, was den Betreuungsaufwand deutlich erhöht. Die einzelnen Architekturen, zum Beispiel „Vektorrechner“ oder „SMP-Rechner“ (SMP = symmetric multiprocessor), sind jeweils für bestimmte Anwendungsfelder besonders gut geeignet. Die wissenschaftlichen Rechenzentren des Forschungszentrums Karlsruhe und der Universität Karlsruhe stellten ihren Anwendern bis 1995 jeweils Höchstleistungsrechner dieser beiden Architekturen zur Verfügung, es wurden also in beiden Rechenzentren sowohl Vektorrechner als auch SMP-Rechner betreut und betrieben. Angesichts der nur beschränkt verfügbaren Haushaltsmittel und der sehr angespannten Situation auf dem Arbeitsmarkt für hoch qualifizierte IT-Spezialisten entstand die damals fast revolutionäre Idee, die beiden genannten Rechenzentren zu einem „Virtuellen Rechenzentrum“ zusammenzuführen, ohne die realen Rechenzentren in ihrem Bestand anzutasten. Der Grundgedanke war, dass jedes der beiden Rechenzentren sich auf Betrieb und Betreuung jeweils genau einer Rechnerarchitektur beschränken sollte, so dass auf diese Weise der jeweilige Betreuungsaufwand deutlich vermindert werden könnte. Die Anwender beider Institutionen, also von Universität und Forschungszentrum, sollten jedoch unabhängig von ihrem eigenen Arbeitsplatz über ein Hochleistungsdatenetz ungehinderten Zugriff auf die Ressourcen beider Rechenzentren haben. Die im Frühjahr 1996 getroffene Vereinbarung besagte, dass im Rechenzentrum der Universität fortan

schwerpunktmäßig SMP-Rechner betrieben werden sollten und dass das Rechenzentrum des Forschungszentrums sich auf Vektorrechner konzentrieren sollte. Anfängliche politische, organisatorische und finanztechnische Schwierigkeiten zwischen den beiden Partnern Universität und Forschungszentrum (bzw. zwischen Land Baden-Württemberg und Bund) konnten geregelt werden, so dass seit Mitte 1996 die Anwender beider Einrichtungen ungehinderten Zugriff auf die Computer-Ressourcen beider Rechenzentren haben und so die für ihre jeweiligen wissenschaftlichen Aufgaben am besten geeignete Rechnerarchitektur nutzen können.

Die Struktur des Virtuellen Rechenzentrums – Technik und Organisation

Der Schwerpunkt der Zusammenarbeit der beiden realen Rechenzentren des Forschungszentrums und der Universität im gemeinsamen Virtuellen Rechenzentrum liegt bei der gegenseitigen Nutzungsmöglichkeit der vorhandenen Höchstleistungsrechner. Von Beginn an hatte man sich darauf geeinigt, dass Computer in SMP-Architektur ihren Platz im Rechenzentrum der Universität haben sollten, Vektorrechner hingegen im Rechenzentrum des Forschungszentrums. Im Zuge des ständig wachsenden Ressourcen-Bedarfs der Wissenschaftler wurden die installierten Höchstleistungsrechner zeitlich mehrfach erneuert und deutlich verstärkt. Heute ist in der Universität ein SMP-Computer

IBM RS/6000 SP installiert, im Forschungszentrum ein Vektorrechner das Typs Fujitsu VPP5000. Abbildung 2 zeigt symbolisch die Struktur des Virtuellen Rechenzentrums: Auf der linken Seite findet sich die im Rechenzentrum des Forschungszentrums vorhandene Technik, auf der rechten das System des Rechenzentrums der Universität. Verbunden sind beide Rechenzentren über eine 155-Megabit/s-ATM-Strecke in Glasfasertechnik. Über die ATM-Verbindung ist von den angedeuteten Workstations beider Einrichtungen auch ein Zugriff auf die Ressourcen des Rechenzentrums der jeweils anderen Einrichtung möglich.

Fujitsu VPP5000 im Rechenzentrum des Forschungszentrums [2]

Im Rechenzentrum des Forschungszentrums Karlsruhe wird ein Hochleistungs-Vektor-Parallel-Rechner VPP5000 des Herstellers Fujitsu betrieben. Dieser Computer besteht in der hier vorhandenen Konfiguration aus acht Prozessorelementen (PE), von denen eines für interaktives Arbeiten freigegeben ist. Jedes einzelne Prozessorelement verfügt über eine Skalar- und eine Vektoreinheit sowie jeweils über einen eigenen SDRAM-Arbeitspeicher (SDRAM: Synchronous Dynamic Random Access Memory, ein moderner und besonders schneller

RAM-Typ). Dieser SDRAM-Speicher weist bei sechs Prozessorelementen eine Größe von je acht Gigabyte auf, bei den restlichen beiden eine Größe von je 16 Gigabyte. Jedes dieser Prozessorelemente erreicht eine vektorielle Spitzenleistung von 9,6 Gigaflop/s (Milliarden Gleitpunktoperationen pro Sekunde). Die einzelnen Prozessorelemente sind über ein Kommunikationsnetzwerk in Crossbar-Technologie miteinander verbunden (bidirektional 1,6 Gigabyte/s). Das Gesamtsystem verfügt somit über insgesamt 80 GigaByte Hauptspeicher und eine theoretische Spitzenleistung („peak performance“) von 76,8 Gigaflop/s. Die VPP5000 wird mit dem UNIX-Betriebssystem

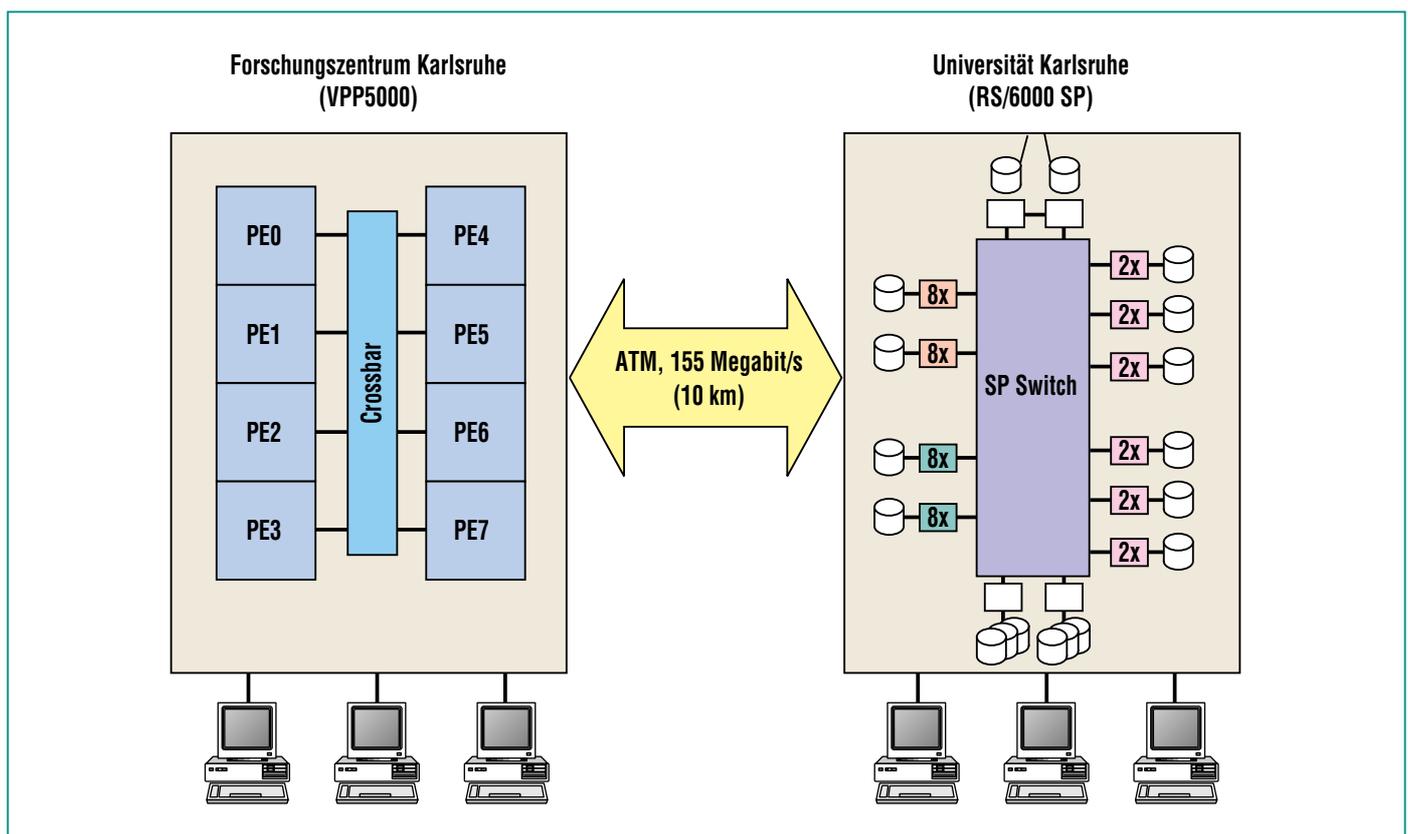


Abb. 2: Die Struktur des Virtuellen Rechenzentrums. Über die ATM-Vernetzung zwischen beiden realen Rechenzentren ist ein Zugriff aller Wissenschaftler auf die Ressourcen beider Rechenzentren möglich.

UXP/V betrieben, das auf UNIX System V Rel. 4 basiert. Auf jedem Prozessorelement läuft ein eigenes UXP/V, das – je nach Prozessorelement – interaktiven oder Batch-Betrieb ermöglicht. Benutzer können von anderen Systemen, z. B. von Workstations, über TCP/IP mit dem VPP5000-System interaktiv kommunizieren und beispielsweise Batch-Jobs absetzen, die dann über NQS (Network Queuing System) auf der VPP5000 gestartet werden. Aus der Architektur der VPP5000 ergibt sich, dass dieser Hochleistungsrechner besonders gut für serielle, jedoch vektorisierte Anwendungen geeignet ist, aber eben nicht für parallelisierte.

IBM RS/6000 SP im Rechenzentrum der Universität [3]

Im Rechenzentrum der Universität Karlsruhe ist dagegen ein Hochleistungsrechner des Typs IBM RS/6000 SP installiert. Bei diesem modernen Computer in SMP-Architektur handelt es sich um einen Parallelrechner mit 116 Mehrprozessorknoten mit je zwei oder acht Prozessoren (112 WinterHawk-II-2-Wege-SMP-Knoten mit je zwei POWER3-II Prozessoren, 375 MHz und zwei Gigabyte Hauptspeicher, sowie vier NightHawk-II-8-Wege-SMP-Knoten mit je acht POWER3-II Prozessoren, 375 MHz und acht Gigabyte Hauptspeicher). Insgesamt besteht das System somit aus 256 Prozessoren, die eine theoretische Gesamtspitzenleistung von 384 Gigaflop/s erbringen. Das System verfügt außerdem mit 256 Gigabyte über einen der europaweit größten Hauptspeicher, so

dass es sich besonders gut für Programme mit extrem großen Speichieranforderungen eignet. In aller Regel werden hier Paralleljobs auf bis zu 64 Prozessoren bearbeitet. Es lassen sich aber auch ohne weiteres Jobs mit gleichzeitiger Nutzung von bis zu 160 oder in Spezialfällen von bis zu 256 Prozessoren rechnen. Die IBM RS/6000 SP läuft vollständig integriert unter AIX 4.3.3, PSSP 3.1, Parallel Environment 2.4 und dem Distributed Computing Environment (DCE). Zur Steuerung des Batchbetriebes wird der IBM LoadLeveler eingesetzt. Aufgrund seiner Architektur eignet sich dieser Hochleistungsrechner besonders gut für sogenannte parallele Programme, die mithilfe des Tools MPI (message passing interface) parallelisiert wurden, für parallele Anwendungen, welche auf einem gemeinsamen Hauptspeicher arbeiten, aber auch ohne weiteres für normale serielle Anwendungen.

Die Organisation

Geleitet und koordiniert wird das Virtuelle Rechenzentrum Karlsruhe von einem virtuellen Direktorium, das jedoch aus zwei sehr realen Personen besteht, nämlich den Leitern der beiden beteiligten wissenschaftlichen Rechenzentren. Das Direktorium hat die Aufgabe, neue und laufende Vorhaben im Umfeld des Virtuellen Rechenzentrums zu planen und zu beobachten, sowie die beiderseitigen Investitionsplanungen jeweils schon im Vorfeld abzustimmen und zu koordinieren. Diesem Ziel dient auch, dass die beiden Leiter dem IT-Planungsausschuss der jeweils anderen Insti-

tution als beratende Mitglieder angehören. Die gegenseitige Inanspruchnahme von Ressourcen hat im Übrigen dem Grundsatz der vollen Kostendeckung zu entsprechen. Dabei wird angestrebt, dass die von den beiden Partnern im Jahresmittel erbrachten Kooperations- und Sachleistungen in etwa gleichwertig sind, so dass von finanziellen Ausgleichsleistungen abgesehen werden kann.

Zusammenfassung und Ausblick

Der effiziente Einsatz vorhandener Mitarbeiter und Hochleistungsrechner ist angesichts teurer Ressourcen und kaum noch verfügbarer IT-Spezialisten zu einer zwingenden Notwendigkeit geworden. Die Rechenzentren des Forschungszentrums Karlsruhe und der Universität Karlsruhe haben deshalb schon 1996 ein „Virtuelles Rechenzentrum“ gegründet, um durch eine sehr enge Kooperation einen besonders wirtschaftlichen Einsatz von Menschen und Maschinen zu erreichen. Der Grundgedanke des Virtuellen Rechenzentrums Karlsruhe besteht darin, gleichartige und bei beiden Partnern nachgefragte Dienste nur noch von einem Partner betreuen zu lassen, sie aber dennoch bei beiden Partnern anzubieten. Diese Vorgehensweise erlaubt es den beteiligten Partnern, sich auf die Betreuung jeweils einer Computerarchitektur zu spezialisieren, wodurch die angebotenen Dienste deutlich wirtschaftlicher zu erbringen sind. Die andauernden Fortschritte in der Informationstechnologie ermöglichen es beiden Einrichtungen, auf diesem Gebiet seit mehreren Jah-

ren eine sehr enge und vertrauensvolle Kooperation einzugehen und mit Leben zu füllen, die beispielhaft ist für die zukünftig geplanten weltumspannenden Grid-Aktivitäten verschiedener wissenschaftlicher Disziplinen. So konzipiert und realisiert beispielsweise die weltweite Community der Hochenergie- und Elementarteilchenphysik derzeit ein global an-

gelegtes Virtuelles Rechenzentrum, welches nach gegenwärtigem Planungsstand aus sieben über die ganze Welt verteilten und durch extrem leistungsfähige interkontinentale Datennetze verbundenen großen realen Rechenzentren bestehen wird. Eines dieser sieben sehr großen Rechenzentren entsteht gegenwärtig im Forschungszentrum Karls-

ruhe. Auf diese Weise kann das Forschungszentrum Karlsruhe die einst auf lokaler Ebene geborene Idee des Virtuellen Rechenzentrums nun im globalen Maßstab fortführen.

Literatur

[1] <http://www.top500.org>

[2] <http://hikwww2.fzk.de/hik/orga/horas/hlr/vpp/VPP5000.html>

[3] <http://www.rz.uni-karlsruhe.de/Uni/RZ/Hardware/SP2>

Grid-Computing und e-Science

M. Kunze, H. Marten, HIK

Einleitung

Einer der wichtigsten Erfolgsfaktoren für die Spitzenforschung und Industrie wird zukünftig darin bestehen, jederzeit schnell und standortunabhängig auf weltweit verteilte Daten und IT-Ressourcen zugreifen zu können. Die Grundlage dazu bildet eine hochkomplexe, weltweit vernetzte Informationsinfrastruktur mit differenzierten Zugangs- und Administrationsmechanismen, für die sich allgemein der Begriff „Grid-Computing“ etabliert hat [1]. Der Ausdruck Grid-Computing wurde dabei geprägt durch die Assoziation mit dem elektrischen Stromnetz (Power Grid). In der Tat war die ursprüngliche Motivation für Grid-Computing, dass Rechenkapazität in gleicher Weise universell und transparent zur Verfügung gestellt werden soll, wie dies heute für elektrische Energie der Fall ist. Allerdings ist dieser Vergleich nicht voll zutreffend, da es beim elektrischen Strom nur wenige anzu-passende Parameter gibt (Spannung und Frequenz), wohingegen beim Austausch von Rechnerleistung wesentlich mehr Grundeigenschaften zu berücksichtigen sind: So können sich Zielrechner z.B. unterscheiden in Architektur, Betriebssystem, Anbindung an das Netzwerk und vielem mehr. Gerade in dieser technischen Vielfalt liegen viele grundsätzliche Probleme, die noch nicht alle vollständig untersucht und gelöst sind. Ausgehend von dieser ursprünglichen Sichtweise wird das Grid-Computing heute vielfach umfassender definiert als die sichere, flexible, koordinierte und gemeinsame Nutzung von Ressourcen innerhalb virtueller Organisationen

[4]. Als Ressourcen sind in diesem Zusammenhang neben Rechenkapazität auch Daten und Instrumente, sowie im weitergehenden Sinne Personen zu verstehen. Virtuelle Organisationen werden hierbei dynamisch gebildet aus einer Menge von global verteilten Institutionen. Beispiele für Virtuelle Organisationen sind Application Service Provider oder Konsortien von Firmen die gemeinsam ein neues Produkt entwickeln wollen (z.B. ein Auto oder Flugzeug), aber auch internationale wissenschaftliche Kooperationsprojekte: Es ist abzusehen, dass kooperative Problemlösungsumgebungen zukünftig in vielen Bereichen von Industrie und Wissenschaft eine signifikante Rolle spielen werden und damit einen wesentlichen Einfluss auf die Entwicklung globaler Informationsinfrastrukturen, insbesondere das Internet, ausüben.

e-Science

Der Begriff e-Science (im Sinne von „enhanced“ Science) wurde geprägt von Dr. John Taylor, dem Vorsitzenden des britischen Wissenschaftsrates: „e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it.“ Die e-Science ermöglicht eine neuen Qualität der wissenschaftlichen Zusammenarbeit weltweit agierender Kollaborationen: Die grundlegende Voraussetzung dafür ist die allgemeine Verfügbarkeit umfangreicher Datensammlungen im Internet, sowie der Zugang zu sehr großen verteilten Computer-Ressourcen und hoch performanten Visualisierungstools an den Arbeitsplät-

zen der beteiligten Wissenschaftler.

Das World Wide Web hat Standards geschaffen für den weltweiten Zugang zu Information jeglicher Art. Zur Unterstützung der e-Science ist eine weitergehende Standardisierung und die Bildung einer ungleich mächtigeren Infrastruktur zwingende Voraussetzung. Neben den Informationen auf Webseiten benötigen die Wissenschaftler einfachen Zugang zu teuren Instrumenten, zu verteilten Teraflop-Maschinen und Farmen preiswerter Personalcomputer, sowie Hochgeschwindigkeitsnetze, um Terabytes an Informationen in verteilten Datenbanken abzufragen.

Das Grid definiert eine Architektur, die alle diese Dinge zusammenbringt und damit die Vision der e-Science in greifbare Nähe rückt: e-Science umfasst als Konzept die Anwendung der Grid-Technologie für vielfältige Disziplinen, wie etwa für Hochenergie- und Astrophysik, Biowissenschaften und Medizin oder Geo- und Umweltforschung, unter Einbeziehung von Experimentgroßgeräten (Beschleuniger, Satelliten, Tomographen u. a.) einerseits und den Aufbau von regional verteilten Massendaten-Archiven und deren Bereitstellung für differenzierte Informationsanalysen andererseits. In diesem Sinne wäre eine Konsolidierung von Datenbankformaten über Grenzen von Arbeitsgruppen hinweg denkbar etwa im Bereich der Genomforschung oder Astronomie. So könnten z.B. im Falle der Beobachtung einer künftigen Supernova-Explosion Wissenschaftler die Aufzeichnungen aus der Astronomie mit den

Beobachtungen der Gravitationswellen-Experimente sowie mit den Messdaten von Neutrino-detektoren aus der Teilchenphysik korrelieren: Der erkenntnistheoretische Mehrwert einer solch umfassenden Methodik liegt auf der Hand.

Grid-Architekturen

Eines der schwierigsten Probleme beim Entwurf von Grid-Systemen liegt in der allgemeinverbindlichen Definition von Protokollen und Diensten, die eine Interoperabilität von Systemen und Anwenderprogrammen überhaupt erst garantieren. Als Sammelbegriff für die benötigte standardisierte Vermittlungs- und Verwaltungssoftware wird allgemein der Begriff „Middleware“ verwendet. Die Middleware stellt die untere Schicht des Grid-Computing dar, auf welche verteilte Anwendungen aufbauen können. Beispiele für derartige Anwendungen sind Simulationsrechnungen in der Hochenergie- und Astroteilchenphysik sowie in der Medizin oder der Zugriff auf verteilt gespeicherte sehr große Datenbanken. Die Middleware zur Konstruktion eines Grid muss dabei folgendes leisten [3]:

- Koordination und Koallokation von Ressourcen, ohne einer zentralen Kontrollinstanz unterworfen zu sein
- Verwendung offener Standardprotokolle und Schnittstellen
- Realisierung nichttrivialer komplexer Dienstqualitäten durch Zusammenschalten von globalen Ressourcen

In der Praxis führen diese Forderungen zur Einrichtung von sog. Web-Services für die zu verwendenden Ressourcen, die global sichtbare Funktionalitäten bieten wie z.B. Verzeichnisdienste zum Aufspüren von Ressourcen, Resource-Broker zur dynamischen Vermittlung, sowie Scheduler zum Buchen und zur Inanspruchnahme von Ressourcen. Die Sicherheitsproblematik kann gelöst werden durch den Einsatz von PKI¹⁾-gestützten Verfahren zur Authentifizierung auf der Basis von ISO-konformen Zertifikaten.

Ein lokales Computercluster mit einem zentral verwalteten Batch-System bildet nach diesen Vorgaben allerdings noch kein Computing Grid (Ein solches Angebot würde analog in etwa dem Fall entsprechen, dass ein Kunde seine Elektrogeräte zum Betrieb in ein bestimmtes Kraftwerk transportieren müsste). Die genannten Kriterien werden aber realisiert in den aktuellen groß angelegten Grid-Projekten, wie z.B. GriPhyN, PPDG, EU DataGrid, CrossGrid oder LHC Computing Grid: Alle diese Projekte integrieren dynamisch Ressourcen aus verschiedenen Institutionen, von denen jede mit ihren eigenen Betriebsregelungen und Sicherheitsmechanismen aufwartet; als gemeinsame Basis für die Middleware wird das am Argonne Lab entwickelte Globus Toolkit [2] verwendet zur Koallokation von Ressourcen und der Adressierung von Problemen im Bereich Sicherheit, Zuverlässigkeit und Durchsatz. Die neueste Entwicklung geht dabei in Richtung standardisierter Web-Services, sog. Grid-Services, welche

es erlauben, eine verteilte Applikation aus normierten, weltweit verstreuten Komponenten mit allgemein anerkannten und offenen Schnittstellen aufzubauen. Unter Einhaltung des künftigen Industriestandards „Open Grid Services Architecture“ (OGSA) [5] spielt es dabei auch keine Rolle, ob zusammen geschaltete Komponenten und Dienste aus der Unix-Welt oder aus der Windows Welt stammen. Generell wird so auch eine Zusammenarbeit zwischen Applikationen ermöglicht, die auf verschiedener Middleware-Technologie basieren, wie z.B. UNICORE, DotNet usw.

Netzwerke

Eine zwingende Voraussetzung zur Realisierung von Grid-Computing ist die Existenz einer Breitbandvernetzung mit Quality-of-Service-Garantie. Mit dem vom DFN-Verein betriebenen G-WiN steht der Wissenschaft in Deutschland eine der weltweit modernsten Kommunikationsinfrastrukturen zur Verfügung. Hierdurch entsteht ein nachhaltiger Standortvorteil, der vor dem Hintergrund einer zunehmenden Globalisierung über die Wissenschaft hinaus von sehr großer Bedeutung ist. Den Kern des G-WiN bildet ein bundesweites Hochgeschwindigkeitsdatennetz auf modernster Glasfasertechnologie. Die Teilnehmer des G-WiN können Anschlüsse mit bis zu 2,5-Gigabit/s-Kapazität, später auch bis zu 10 Gigabit/s erhalten, wobei die etwa 700 deutschen Wissenschaftseinrichtungen Zugang erhalten über 29 Knotenpunkte. Der

¹⁾ Public Key Infrastructure

Rahmen seiner Untersuchungen ein immenses Aufkommen an Daten, die ausgewertet und verarbeitet werden müssen. Dazu sollen weltweit freie oder ungenutzte Prozessor- und Speicher-Kapazitäten von Zehntausenden von Computern zu einem Daten-Netzwerk (einem sog. DataGrid) zusammengefasst werden. Das Grid Computing stellt einen der neuesten Ansätze dar, um den enormen Bedarf an Rechen- und Speicher-Kapazitäten solch umfassender Projekte zu decken und nutzt das Internet und die darin enthaltenen Computerressourcen. Tausende Wissenschaftler von Universitäten auf der ganzen Welt werden bei der Analyse der gesammelten Daten zusammenarbeiten. Hierbei ist die Idee, die Daten nicht wie bisher an einem Ort vorzuhalten und zu verarbeiten, sondern diese auf mehrere Zentren zu verteilen und transparent zu nutzen [6]: In der MONARC-Studie [7] wurde gezeigt, dass ein hierarchisches System mit verschiedenen Ebenen (Tiers) den Anforderungen gerecht wird. In diesem Modell hat CERN als Tier-0 die Aufgabe, Ex-

perimentdaten zu speichern und zu rekonstruieren. Die Analyse sowie die Erzeugung simulierter Ereignisse wird vor allen Dingen in regionalen Tier-1-Zentren stattfinden, gefolgt von nationalen Tier-2-Zentren, Institutsrechnern (Tier-3) und Arbeitsplatzrechnern (Tier-4).

Neben CERN, RAL, IN2P3 und vielen anderen hochrangigen Forschungszentren wird sich auch das Forschungszentrum Karlsruhe an diesem Projekt beteiligen: Die Hauptabteilung Informatik- und Kommunikationstechnik (HIK) wird als Kompetenzzentrum zur Weiterentwicklung und Anwendung der Grid-Technologie beitragen; auf der Ressourcenseite wird mit dem „Grid-Computing Centre Karlsruhe (GridKa)“ in den nächsten Jahren ein beachtliches Potenzial an Computerleistung und Speicherplatz geschaffen, welches Zugang zu einer neuen Qualität wissenschaftlichen Arbeitens eröffnen wird [8,9]. Obwohl die Datennahme der vier Großexperimente ATLAS, ALICE, CMS und LHCb bei LHC erst im Jahre

2007 beginnt, werden schon jetzt Prototypen für die Datenverarbeitung gebaut, sog. Testbeds, mit jährlich wachsender Funktionalität und Komplexität, um die kritischen Parameter des weltweiten Computing-Modells zu untersuchen und zu optimieren. Darüber hinaus bietet GridKa schon jetzt Ressourcen zur Auswertung der Daten für vier weitere in Produktion befindliche Experimente der Kern- und Teilchenphysik (BaBar am SLAC, D0 und CDF am FermiLab, COMPASS am CERN). Es ist das Ziel der Aufbauarbeiten, Ressourcen für alle Experimente möglichst gemeinsam bereitzustellen, um durch Synergie-Effekte Einsparungen bei den Gesamtkosten zu erreichen; an den durch GridKa versorgten Experimenten nehmen in Deutschland insgesamt mehr als 400 Wissenschaftler aus 41 Instituten teil. Eine weitere wichtige Aufgabe besteht darin, dass GridKa weltweit als zentrale Instanz zur Ausstellung von Grid-Computing Zertifikaten für die deutsche Kern- und Teilchenphysik anerkannt ist.

Literatur

- [1] I. Foster, C. Kesselman, *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann Publ. 1999.
- [2] *The Globus Project*, <http://www.globus.org/>
- [3] I. Foster, *What is the Grid?*, <http://www.gridtoday.com/02/0722/100136.html>
- [4] I. Foster, C. Kesselman, S. Tuecke, *The Anatomy of the Grid: Enabling Scalable Virtual Organizations*, *International J. Supercomputer Applications*, 15(3), 2001.
- [5] I. Foster, C. Kesselman, J. Nick, S. Tuecke, *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration*. June 22, 2002. <http://www.globus.org/research/papers/ogsa.pdf>
- [6] Report of the Steering Group of the LHC Computing Review, CERN/LHCC/2001-004, http://lhc-computing-review-public.web.cern.ch/lhc-computing-review-public/Public/Report_final.PDF
- [7] *MONARC Phase 2 Report*, CERN/LCB 2000-001, <http://monarc.web.cern.ch/MONARC/docs/phase2report/Phase2Report.pdf>
- [8] P. Braun-Munzinger et al., *Requirements for a Regional Data and Computing Centre in Germany (RDCCG)*, <http://grid.fzk.de/grid/LHCComputing-1july01.pdf>
- [9] H. Marten et al., *A Grid Computing Centre at Forschungszentrum Karlsruhe*, <http://grid.fzk.de/grid/RDCCG-answer-v8.pdf>

Mathematische Programmbibliotheken im Wandel der Zeit

I. Bischoff-Gauß, R. Pietschmann, F. Schmitz, HIK

Einleitung

Mathematische (Unter-)Programmbibliotheken sind Bestandteil eines vielfältigen Angebots an Programmen und Programmpaketen, welches unter dem Sammelbegriff „Mathematische Software“ für die technisch-wissenschaftlichen Anwender der Datenverarbeitung verfügbar ist. Der Begriff „Mathematical Software“ wurde 1969 von John R. Rice geprägt und als „Computer programs which implement widely applicable mathematical procedures“ definiert [2]. Das heutige Spektrum umfaßt einerseits Programmbibliotheken zur Bearbeitung von Aufgabenstellungen aus der Numerischen Mathematik (Funktionsberechnungen, lineare und nichtlineare Gleichungssysteme, Anfangs- und Randwertprobleme bei gewöhnlichen und partiellen Differentialgleichungen, schnelle Fourier-Transformationen etc.), aus der Statistik (beschreibende Statistik, Regression und Korrelation, Varianzanalyse, Testverfahren etc.), andererseits aber auch die sog. Computer-Algebra-Systeme, welche Ergebnisse sowohl in symbolischer als auch in numerischer Form liefern können. Zusätzlich bieten beide Arten ausführliche Online-Hilfen sowie Möglichkeiten der grafischen Darstellung und Visualisierung.

Offensichtliche Gründe für die Bereitstellung mathematischer Software sind: Minimierung des Programmieraufwandes, Nutzung von in Programme gegossenem Spezialwissen nach dem neuesten Stand der Wissenschaft, Produktion von gesicherten Ergeb-

nissen. Unter diesen Aspekten gehören Programm-Bibliotheken unverzichtbar zum Software-Angebot eines technisch-wissenschaftlichen Rechenzentrums [3].

Historie

Geht man in der Geschichte des Rechenzentrums zurück bis etwa 1969/70, so findet man drei Quellen mathematischer Unterprogramme vor: Die SYS1.FORTLIB als Bestandteil des FORTRAN-Laufzeitsystems, welche neben Funktionen wie Sinus, Tangens etc. auch Ein-/Ausgabe- und sonstige Dienstprogramme bereitstellte. Die GFK.FORTLIB als Sammlung von FORTRAN- und Assembler-Programmen für mathematische und allgemeinere Aufgabenstellungen, welche von Mitarbeitern des Forschungszentrums der gesamten Benutzerschaft zur Verfügung gestellt wurden. Und das Scientific Subroutine Package (SSP), wohl die erste Sammlung dieser Art von über 250 mathematischen und statistischen Unterprogrammen – die meisten davon in Ausführungen für einfache und doppelte Genauigkeit – entwickelt von IBM und im FORTRAN-Quellcode verfügbar (was der Verbreitung förderlich war, weil jeder Benutzer sich die Programme auf Lochkarten ausstanzen und sie bei Bedarf auch modifizieren konnte).

Ab 1972 bot das Rechenzentrum die kostenpflichtigen Lizenzprogramme SLMATH (für FORTRAN) und PLMATH (für PL/1) als Nachfolger für SSP an, wobei PLMATH – in Ermangelung anderer Quellen für die Programmiersprache PL/1 – erst 1992 abgemietet wurde.

Einige Entwickler von SSP gründeten 1971 in den USA die Firma IMSL (International Mathematical and Statistical Libraries), welche ein Jahr später die erste rein kommerzielle mathematische Programmbibliothek für FORTRAN an IBM-Kunden auslieferte (komplett in je einer Version für einfache und doppelte Genauigkeit) [2]. Ab 1973 wurde die gleiche Bibliothek auch für Kunden anderer Rechnerhersteller verfügbar, was ein Novum darstellte, konnten doch jetzt Quellprogramme mit IMSL-Aufrufen zwischen verschiedenen Plattformen ausgetauscht werden, sofern nur die IMSL-Bibliothek vorhanden war. Allerdings war der Quellcode der Unterprogramme nicht verfügbar. Das Rechenzentrum lizenzierte ab 1978 diese Bibliothek und bietet sie bis heute seinen Benutzern an. Sie zeichnete sich in der Vergangenheit schon aus durch die Qualität der Algorithmen, ein umfangreiches Anwendungsspektrum, eine weite Verbreitung und gute Dokumentation in Form von Handbüchern, die in späteren Versionen auch Online verfügbar waren und durch komfortable Online-Hilfen ergänzt wurden. Ab 1991 wurde von IMSL auch eine Bibliotheksversion für die Programmiersprache C angeboten.

Mit der HARWELL Subroutine Library (HSL), einem nicht kommerziell vertriebenen Produkt von HARWELL Laboratory in England, verschaffte das Rechenzentrum 1980 seinen Benutzern wieder Zugriff auf Programme im Quellcode, welcher mit der Auflage des Zitierens in wissenschaftlichen Veröffentlichungen für Forschungszwecke frei verfügbar war. Im ma-

thematischen Teil deckte die HSL nahezu den Bereich der IMSL ab, im Statistikbereich nur einen kleinen Ausschnitt. Einige Dienstprogramme waren im IBM-Assembler-Code geschrieben (dies schränkte die Portabilität ein). Mit dem Übergang vom Betriebssystem MVS auf Unix 1999 endete die HARWELL-Ära im Forschungszentrum.

Eine neue Qualität bezüglich der Zuverlässigkeit von Ergebnissen numerischer Berechnungen brachte das IBM-Produkt ACRITH, die „High Accuracy Arithmetic Library“, mit dem Konzept der „verifizierenden Algorithmen“ bzw. der vom Programm „verifizierten Ergebnisse“: Statt eines einzigen Ergebnisses werden jetzt zwei Zahlen zurückgegeben, die das einschließende Intervall für das – unbekannte – exakte Ergebnis definieren. Die theoretische Basis für die gesicherte Fehlerschranken liefernden Algorithmen – die Anwendung der Intervall-Arithmetik sowie von Fixpunktsätzen der Analysis, der gerichteten Rundung von Zwischenergebnissen, die Verwendung eines langen Akkumulators für die exakte Berechnung von Skalarprodukten etc. – wurde schon Anfang der siebziger Jahre wesentlich von der Universität Karlsruhe gelegt. Die Firma IBM hat diese Arbeiten in das Produkt ACRITH überführt und vermarktet und sogar einen Rechner – die IBM 4361 – gebaut, bei dem die wesentlichen Arithmetik-Funktionen in Hardware implementiert worden waren. Das Rechenzentrum hat diese Bibliothek 1986 für die Benutzer verfügbar gemacht und damit – hoffentlich – ins Be-

wusstsein gerückt, nicht jedem Ergebnis bisheriger Rechnungen blindlings zu vertrauen. Die Methoden haben sich langfristig nicht durchgesetzt, vielleicht weil die Rechenzeiten gegenüber herkömmlichen Algorithmen deutlich länger waren und ACRITH nicht das gesamte Spektrum anderer Bibliotheken abdeckte.

1987 wurde im Rechenzentrum der erste Vektorrechner installiert, der Siemens-Fujitsu VP50. Um die Möglichkeiten dieser neuen Rechnergeneration – das Verarbeiten von Vektoren als Ganzes oder in Teilen gegenüber der bisherigen elementweisen Abarbeitung – nutzen zu können, mussten Programme „vektoriert“ werden. Diese Aufgabe übernahmen die Compiler der Hersteller, jedoch waren auch die Programmierer gefordert, ebenso die Anbieter von Bibliotheken. Die Herstellerbibliothek „Scientific Subroutine Library 2“ (SSL2) – sie war schon auf dem Skalarrechner im Einsatz – wurde in einer für die spezielle Hardware der VP50 optimierten und um Vektoralgorithmen erweiterten Version mitgeliefert. Der Effizienzgewinn der diese Bibliothek nutzenden Programme ging allerdings zu Lasten ihrer Portabilität.

Eine Alternative zur Effizienzsteigerung ohne Portabilitätseinbußen boten die BLAS-Programme in Verbindung mit einer plattformunabhängigen Bibliothek. BLAS („Basic Linear Algebra Subroutines“) ist das Ergebnis eines von ACM-SIGNUM geförderten Standardisierungsprojektes für Mathematische Software, in einer ersten Version 1979 bereitgestellt und in ACM Transactions On Mathematical Software (TOMS) do-

kumentiert als ein Satz von Fortran-Unterprogrammen für die elementaren Operationen mit Vektoren und Matrizen. Man hatte sich auf Namen, Parameterlisten und funktionale Beschreibungen geeinigt, ein bedeutender Schritt für die Entwicklung von Programmen der linearen Algebra.

Der Software-Hersteller IMSL paßte 1987 seine Bibliotheken an diese Entwicklung an: entsprechender Programm-Code wurde durch zum Standard konforme BLAS-Aufrufe ersetzt. Hardware-Hersteller optimierten die BLAS-Bibliothek für ihre speziellen Architekturen, so IBM mit der 3090-Vector-Facility und Siemens-Fujitsu mit dem VP400, die 1990 bzw. 1991 im Rechenzentrum installiert wurden. Dadurch konnten die Benutzer durch Einbinden dieser optimierten BLAS-Bibliotheken an Stelle der IMSL-BLAS oft in signifikantem Maße Rechenzeit einsparen.

Ein neuer Rechnertyp wurde 1995 im Rechenzentrum etabliert: der Parallel-Rechner CRAY-J916 unter dem Betriebssystem UNICOS. Die Möglichkeit, ein Programm auf mehreren Prozessoren gleichzeitig – also „parallel“ – auszuführen (schon auf der IBM3090/600VF möglich, aber nicht effizient realisierbar), war eine neue Herausforderung an Mathematiker und Entwickler mathematischer Software. Hierzu bot der Hersteller CRAY neben einer optimierten BLAS-Bibliothek und einer auf BLAS basierenden Bibliothek LAPACK, die die Funktionalität der bekannten Bibliotheken LINPACK und EISPACK abdeckte, auch eine eigene mathematische Bibliothek SCILIB an, deren wesentli-

che Programme die Vorteile der Parallel-Vektorarchitektur nutzen. Die Firma IMSL stellte für die CRAY-Rechner ebenfalls die Architektur nutzende Versionen ihrer Bibliothek bereit.

Die rasant gesteigerte Rechenleistung der Computer-Hardware zwingt die Software-Entwickler zu neuen Verfahren der Anpassung und Optimierung, die bisher sehr zeitaufwendig waren und nur von hochqualifizierten Spezialisten geleistet werden konnten. Ziel des seit 1998 laufenden Forschungsprojekts ATLAS (Automatically Tuned Linear Algebra Software) ist, empirische Methoden in Form des Software-Pakets ATLAS bereit zu stellen, um Programme der linearen Algebra – in erster Linie die BLAS-Bibliothek, aber auch aus LAPACK – für unterschiedliche Hardware-Plattformen zu optimieren [4].

Mathematische Bibliotheken der Gegenwart

Die Anzahl weltweit angebotener freier und kommerzieller mathematischer Bibliotheken unterscheidet sich dramatisch von der Anzahl der im Forschungszentrum zentral installierten Bibliotheken. Hier einige interessante Seiten im Netz:

www.netlib.org/index.html

math.nist.gov

www.netlib.org/lapack/index.html

www.visual-numeric.de/produkte/index.html#IMSL

www.nag.co.uk/numeric/numerical_libraries.asp

Im Rechenzentrum sind folgende mathematischen Unterprogramm-bibliotheken zentral verfügbar:

BLAS (Basic Linear Algebra Subprograms)

Die BLAS-Bibliothek ist eine aus Public-Domain Routinen bestehende Sammlung von Basisprogrammen zur effizienten Lösung von Problemen der Linearen Algebra. Es werden Routinen für Vektor-Vektor, Vektor-Matrix und Matrix-Matrix Operationen angeboten. Nach und nach hat sich diese Sammlung als Standard etablieren können, da die im Quelltext verfügbaren Routinen von allen größeren Herstellern als hardwareoptimierte BLAS-Version (-> ATLAS) zur Verfügung gestellt werden. Auch die Anbieter von Softwarepaketen bedienen sich der BLAS Routinen. Vom Rechenzentrum wurden diese Routinen erstmals auf einem Vektorrechner VP50 und einer IBM 3090/600 VF zur Verfügung gestellt.

LAPACK (Linear Algebra PACKage)

Eine neue Art von numerischer Unterprogramm-bibliothek löste ab 1992 die Bibliotheken EISPACK (seit 1976) und LINPACK (seit 1979) ab. LAPACK bietet eine Vielzahl von sogenannten 'driver routines', die sich wieder in 'simple' und 'expert' unterteilen. Es wird aber auf die optimierten BLAS Bibliotheken des jeweiligen Herstellers zugegriffen.

ScaLAPACK (Scalable LAPACK)

Durch Hardware-designänderungen von Hochleistungsrechnern wurde der Bedarf nach parallelen Bibliotheken geweckt. 1995 wurde die erste Version für die sogenannte MIMD-Architektur (multiple instruction streams, multiple data streams, Flynn Einteilung) zur Verfügung gestellt. Für große Probleme eröffnet diese Bibliothek die Möglichkeit, verteilten Speicher und schnelle Prozessoren über leistungsstarke Netze effizient zu nutzen. Leider steigt durch einen hohen Speicherbedarf die Anzahl der zu nutzenden Prozessoren so stark an, dass bei einer Eigenwertberechnung die Anwendung nur verteilt auf einem VPP300 mit viel lokalem Speicher gerechnet werden kann.

IMSL

Die seit 1972 angebotene IMSL Bibliothek wird laut Benutzerbefragung auch wegen des hohen Preises vor der NAG Bibliothek (Numerical Algorithms Group) bevorzugt. Deshalb wird bis heute die IMSL auf den zentralen Rechnern angeboten, obwohl die NAG moderner ist. Die IMSL bietet mit ihren mathematischen und statistischen Teilen Fortran- als auch C-Programmierern eine einfache Schnittstelle an. Über ein MPI-Interface können verteilte Systeme gemeinsam an der Lösung einer Aufgabe arbeiten.

Bibliotheken von morgen

Bedingt durch die Annäherung der Leistungsdaten von Hochleistungsprozessoren und die in ho-

her Stückzahl produzierten Commodity-Komponenten und den Verdrängungswettbewerb der Generalisten gegen die Spezialisten im Hochleistungsrechnerbau (IBM, SUN, HP, Intel, Cray, Thinking Machines usw.) sowie durch Leistungssteigerung mittels paralleler Strukturen ist auch ein Umdenken bei numerischen Bibliotheken notwendig.

Gerade diese parallelen Systeme können aber eine wesentlich stärkere Beschleunigung von Anwendungen bewirken, als der Intel-Mitbegründer Gordon Moore noch 1968 vorhergesagt hatte.

Adaptive Algorithmen in der Numerik haben den Vorteil, dass etwa ein Berechnungsgebiet nicht mehr wie bei herkömmlichen Verfahren a priori statisch festgelegt wird, sondern dynamisch während des Berechnungsvorganges zugeordnet wird. Dabei werden Größen wie Rechenleistung, Netzbandbreite, Latenz, usw. berücksichtigt. So ergibt sich eine dynamische Lastverteilung und optimale Nutzung der Ressourcen. Der Güte des Ergebnisses als auch dem Ressourcenmanagement gilt dabei das Hauptaugenmerk. Jedoch gibt es momentan nur Lösungen für ganz bestimmte Fragestellungen. An-

sätze sind erkennbar bei adaptiver FEM-Software, Bildanalysesoftware und in der Bioinformatik.

Beispiel einer IMSL-Nutzung

Am Institut für Meteorologie und Klimaforschung wurde mit Hilfe umfangreicher numerischer Simulationen am VPP5000 eine Windstatistik für den Raum Baden-Württemberg berechnet (www.lfu.baden-wuerttemberg.de/lfu/uis/energie.html). Dabei wurde in einem horizontalen Raster mit einer Gitterweite von 1 km die mittlere Windgeschwindigkeit in

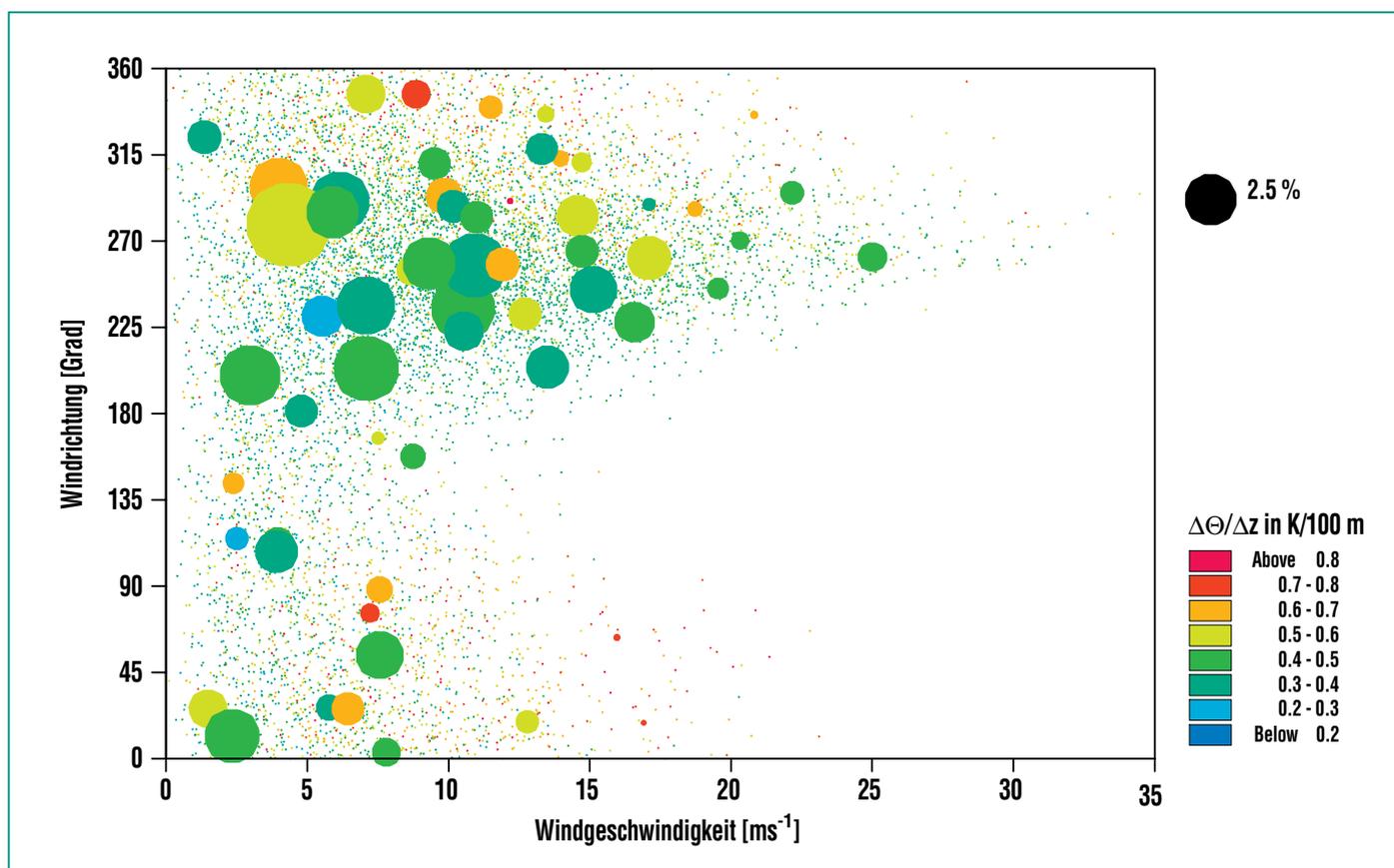


Abb. 1: Clusteranalyse für die Parameter Windrichtung, Windgeschwindigkeit und Temperaturschichtung (gefüllte Kreise). Die Größe der Kreise ist ein Maß für die Häufigkeit, die Punkte stellen die gesamten Beobachtungen dar.

verschiedenen Höhen bestimmt. Zur Berechnung ist eine Zusammenfassung der Großwetterlagen erforderlich. Diese erfolgt mit Hilfe einer Clusteranalyse, die mit der IMSL-Bibliothek durchgeführt wird, wobei die Prozeduren CDIST, CLINK und CNUMB zur Anwendung kommen. Im verwendeten numerischen Simulationsmodell KAMM [1] wird die großräumige Wetterlage charakterisiert durch die horizontalen Windkomponenten u und v in 700 hPa und einem vertikalen Temperaturgradienten $\delta\theta/\delta z$. Als Eingangsdaten dienen Tagesmittelwerte der NCEP/NCAR Reanalysedaten über einen Zeitraum von 25

Jahren. Es werden hierarchische Cluster nach folgendem Schema erzeugt: im ersten Schritt wird jede Beobachtung (u, v in 700 hPa, $\delta\theta/\delta z$) als ein Cluster betrachtet. Es wird dann der Abstand zwischen den Clustern berechnet und die am nächsten zusammenliegenden Cluster zu einem neuen Cluster zusammengefasst. Diese Prozedur wird wiederholt, bis die gewünschte Anzahl von Clustern erreicht ist. Im vorliegenden Falle werden die ca. 9000 Beobachtungen zu 60 Clustern zusammengefasst. Die Verteilung der Cluster in der dreidimensionalen Häufigkeitsverteilung für die Parameter Windrichtung, Windgeschwindigkeit

und Temperaturschichtung ist in der Abb. 1 wiedergegeben. Für jeden Clusterpunkt wird eine numerische Simulation durchgeführt, woraus anschließend mit der entsprechenden Gewichtung die mittleren Windverhältnisse bestimmt werden.

Literatur

- [1] G. Adrian, F. Fiedler, *Contrib. Phys. Atmos*, 64, 27-48, 1991
- [2] Cowell W. R., *Sources and Development of Mathematical Software*. Prentice-Hall (1984), Englewood Cliffs, New Jersey.
- [3] J.-Fr. Hake, P. Jansen, *Mathematische Software: Ergebnisse und Trends. Interner Bericht, Forschungszentrum Jülich* (1992).
- [4] R.C. Whaley., A. Petitet, J. J. Dongarra, *Automated Empirical Optimization of Software and the ATLAS Project* (2000). Quelle: Internet.

Monte-Carlo-Simulation des Teilchentransports und Anwendung in der Fusionsneutronik

U. Fischer, Y. Chen, IRS

Einführung

In der Plasmakammer eines Fusionsreaktors verschmelzen die leichten Atomkerne der Wasserstoffisotope Deuterium (d) und Tritium (t) bei hohen Temperaturen ($\cong 100$ Millionen Grad Celsius) zu Heliumkernen. Pro Fusionsreaktion wird ein Neutron emittiert, das mit 14.1 MeV den Löwenanteil der freigesetzten Bindungsenergie in Höhe von 17.6 MeV davon trägt. Die 14-MeV-Neutronen verlassen die Plasmakammer ungehindert und treten mit den Atomkernen der umgebenden Materialien in Wechselwirkung. Die Beschreibung des Ausbreitungsprozesses der Neutronen- und Photonenstrahlung („Teilchentransport“) im Fusionsreaktor ist Gegenstand der Fusionsneutronik. Das Monte-Carlo-Verfahren [1] ermöglicht es, den Teilchentransport auf mikroskopischer Ebene in beliebig komplexer Geometrie zu beschreiben und eignet sich damit speziell für die Fusionsneutronik. Im folgenden Beitrag wird die Monte-Carlo-Technik zur Beschreibung des Teilchentransports skizziert. Sodann wird ein Programmsystem vorgestellt, mit dessen Hilfe es möglich ist, dreidimensionale Dosisratenverteilungen nach Reaktorabschaltung zu bestimmen. Schließlich werden Anwendungsbeispiele präsentiert, die auf dem HPC Linux-Cluster des FZK im Parallelbetrieb gerechnet wurden.

Fusionsneutronik und Monte Carlo

Die kernphysikalischen Wechselwirkungsprozesse in den Materialien der das Plasma umgeben-

den Reaktorkomponenten führen zur Abbremsung und schließlich Absorption der schnellen 14-MeV-Neutronen. Dabei werden die Materialien aktiviert, es wird hochenergetische γ -Strahlung emittiert, die sich weiter ausbreitet, sowie geladene Teilchen, die in den Materialien schnell abgebremst werden und diese dabei aufheizen. Art und Umfang der stattfindenden Wechselwirkungsprozesse werden bestimmt durch die räumliche Materialverteilung und die zugehörigen nuklidspezifischen Wechselwirkungswahrscheinlichkeiten. Daraus ergibt sich dann die räumliche und energetische Verteilung der Neutronen- und Photonenfelder.

Die Ausbreitung der Neutronen und Photonen lässt sich aus mikroskopischer Sicht beschreiben. Das Monte-Carlo-Verfahren simuliert individuelle Teilchenschicksale wie sie analog in der physikalischen Realität ablaufen würden. Ausgehend von der „Geburt“ eines Teilchens, z. B. eines Neutrons in einer (d,t)-Fusionsreaktion, wird der ganze „Lebensweg“ des Teilchens bis zu seinem „Tod“ (z. B. Absorption) verfolgt. Dabei wird die probabilistische Natur der kernphysikalischen Wechselwirkungsprozesse ausgenutzt. Die Geburt eines Teilchens, seine Energie, seine Flugrichtung, seine weiteren Wechselwirkungen, die dadurch bedingte(n) Emission(en) neuer Teilchen, deren Energie, Flugrichtung und weitere Wechselwirkungen etc., all diese Ereignisse und Eigenschaften sind probabilistischer Natur und können „ausgewürfelt“, d.h. mit Hilfe generierter Zufallszahlen simuliert werden

(daher die Bezeichnung „Monte-Carlo-Verfahren“).

Das Monte-Carlo-Verfahren eignet sich zur Lösung komplexer Transportprobleme, die sich nicht oder nur schwer mit deterministischen Methoden berechnen lassen. Dies gilt insbesondere für geometrisch komplexe Materialanordnungen wie sie bei Fusionsreaktoren des Tokamak-Typs gegeben sind. Die Geometrie kann sehr flexibel dargestellt werden, im allgemeinen durch Schnittflächen räumlicher Körper (Kugel, Quader, Zylinder, Torus, etc.) im dreidimensionalen Raum. Bei der rechnerischen Verfolgung der Teilchenschicksale ist es lediglich erforderlich, die Schnittpunkte der Teilchenbahnen mit den die geometrischen Zellen begrenzenden Flächen zu berechnen. Die Grenzen der Anwendungsmöglichkeit werden bestimmt durch die Zahl der für ein Problem zu berücksichtigenden Teilchenschicksale, mithin durch die benötigte Rechenzeit. Der Aufwand an Rechenzeit wächst naturgemäß mit der Komplexität der behandelten Geometrie und ebenso mit der Zahl und Komplexität der zu berechnenden physikalischen Größen. Die in den letzten Jahren ständig steigende Rechenleistungen einschließlich der Verfügbarkeit von massiven Parallelrechnern haben die Anwendungsmöglichkeiten für Monte-Carlo-Rechnungen beträchtlich erweitert.

Berechnung von 3D-Dosisratenverteilungen für ITER

Der technische Entwurf für den Experimentalreaktor ITER [2] wur-

de in einer langjährigen internationalen Zusammenarbeit zwischen der EU, Japan, Russland und den USA ausgearbeitet. Mit ITER soll der Nachweis der wissenschaftlichen und technologischen Machbarkeit der Kernfusion als künftiger Energiequelle erbracht werden. Während des Betriebs von ITER muss gewährleistet sein, dass das Wartungspersonal Zugang zum Kryostaten hat, um dort befindliche Systemkomponenten zu warten und zu reparieren. Aufgrund der hohen Aktivierung der Reaktorkomponenten in diesem Bereich ist ein Zugang nur nach Abschaltung des Reaktors möglich. Eine wichtige Aufgabe ist es, die Dosisratenverteilung zu bestimmen, so dass fest-

gelegt werden kann wann und wo Zutritt möglich ist. Dies erfordert einerseits ein geeignetes Instrumentarium zur Berechnung der dreidimensionalen zeitabhängigen Dosisratenverteilung und andererseits ein geeignetes nukleares Experiment zu dessen Validierung. Ein solches Experiment wurde am Neutronengenerator in Frascati (FNG) als Kooperation zwischen ENEA Frascati, der TU Dresden und dem Forschungszentrum Karlsruhe durchgeführt [3]. Zur rechnerischen Analyse wurde am FZK ein Programmsystem erstellt, das eine weitgehend automatisierte Berechnung der dreidimensionalen Dosisratenverteilung nach Reaktorabschaltung erlaubt [4].

Programmsystem

Kernstücke des Programmsystems sind das Monte-Carlo-Programm MCNP [5], mit dem die Transportrechnungen für die Neutronen und die Zerfallsgammaquanten durchgeführt werden, sowie das Inventarprogramm FISPACT [6], mit dem die Quellterme der Zerfallsgammaquanten berechnet werden. Abb. 1 zeigt schematisch den Programmablauf. Unter Benutzung eines dreidimensionalen Geometriemodells werden zunächst die Neutronenflussspektren mit MCNP für alle Materialzellen berechnet, die nach Bestrahlung eine γ -Strahlenquelle darstellen können. Die Neutronenflussspektren werden über ein Interface an FISPACT übergeben zur Berechnung der Quellver-

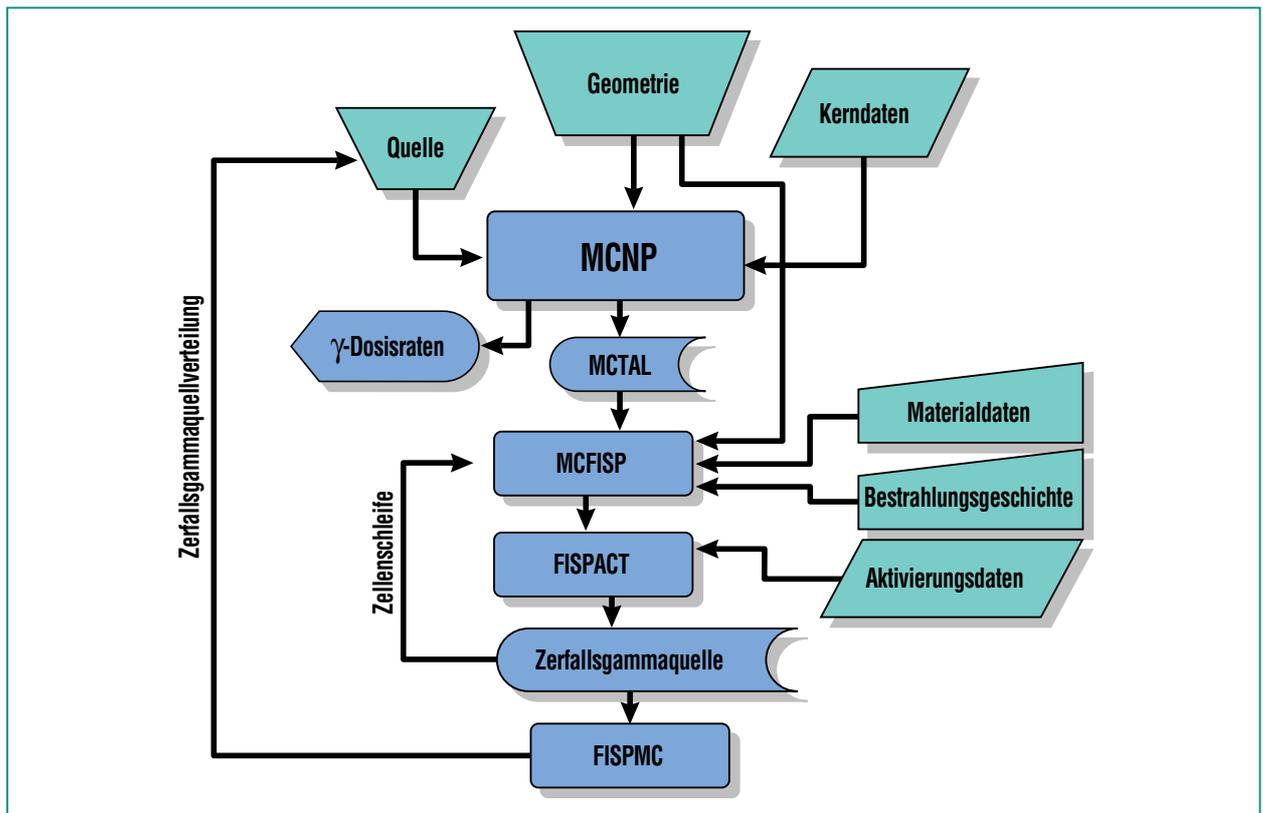


Abb. 1: Programmschema für die Berechnung dreidimensionaler Nachzerfallsdosisraten.

teilung der Zerfallsgammaquanten. Ein weiteres Interface stellt diese Verteilung dann MCNP für die Transportrechnung der Zerfallsgammaquanten zur Verfügung. Die rechnerische Verfolgung der Zerfallsgammaquanten mit MCNP ermöglicht es, die durch die Materialaktivierung während der Bestrahlung bedingte γ -Strahlendosis zu jedem Zeitpunkt nach Reaktorabschaltung an jedem Ort innerhalb des behandelten Geometriemodells zu bestimmen.

Validierung

Das Programmsystem wurde anhand des oben erwähnten Bestrahlungsexperimentes validiert. Eine ITER-typische Materialanordnung aus Stahl und wasseräquivalentem Material wurde zu diesem Zweck mit insgesamt $1.95 \cdot 10^{15}$ 14 MeV-Neutronen am Neutronengenerator Frascati be-

strahlt. Die Nachzerfallsdosisrate wurden in einem Hohlraum der Materialanordnung in einem Zeitraum von ≈ 1 h bis 20 Tage nach der Bestrahlung gemessen [3]. Die berechneten Nachzerfallsdosisraten sind in Abb. 2 mit den von einem Team der TU Dresden gewonnen Messwerten verglichen. Es zeigt sich insgesamt eine recht befriedigende Übereinstimmung. Die Abweichungen liegen bei maximal $\pm 15\%$ und sind im wesentlichen auf Unsicherheiten in den zu Grunde liegenden Aktivierungswirkungsquerschnitten zurückzuführen.

Dosisratenverteilung im Bereich der Handhabungsöffnung von ITER

Zum Nachweis seiner Eignung zur Behandlung großer und komplexer Systeme wurde das Programmsystem eingesetzt, um die

Dosisratenverteilung von ITER nach Reaktorabschaltung im Bereich der zentralen Handhabungsöffnung zu bestimmen. Abb. 3 zeigt hierzu einen Vertikalschnitt durch das von ITER Garching entwickelte MCNP-Torussektormodell. Abb. 4 zeigt einen Horizontalschnitt durch den Bereich der Handhabungsöffnung mit Materialzellen und zugeordneten Zellnummern. Die sehr aufwändige Berechnung der Neutronflussspektren für insgesamt 1663 Materialzellen wurde im Parallelbetrieb unter PVM („Parallel Virtual Machine“) [7] auf dem HPC-Linux-Cluster des FZK durchgeführt. Zur Bestimmung der Quellverteilung der Zerfallsgammaquanten wurde eine ITER-typische Bestrahlungsgeschichte mit einer 20-jährigen Betriebszeit bei einer mittleren Verfügbarkeit von 2.7% zu Grunde gelegt.

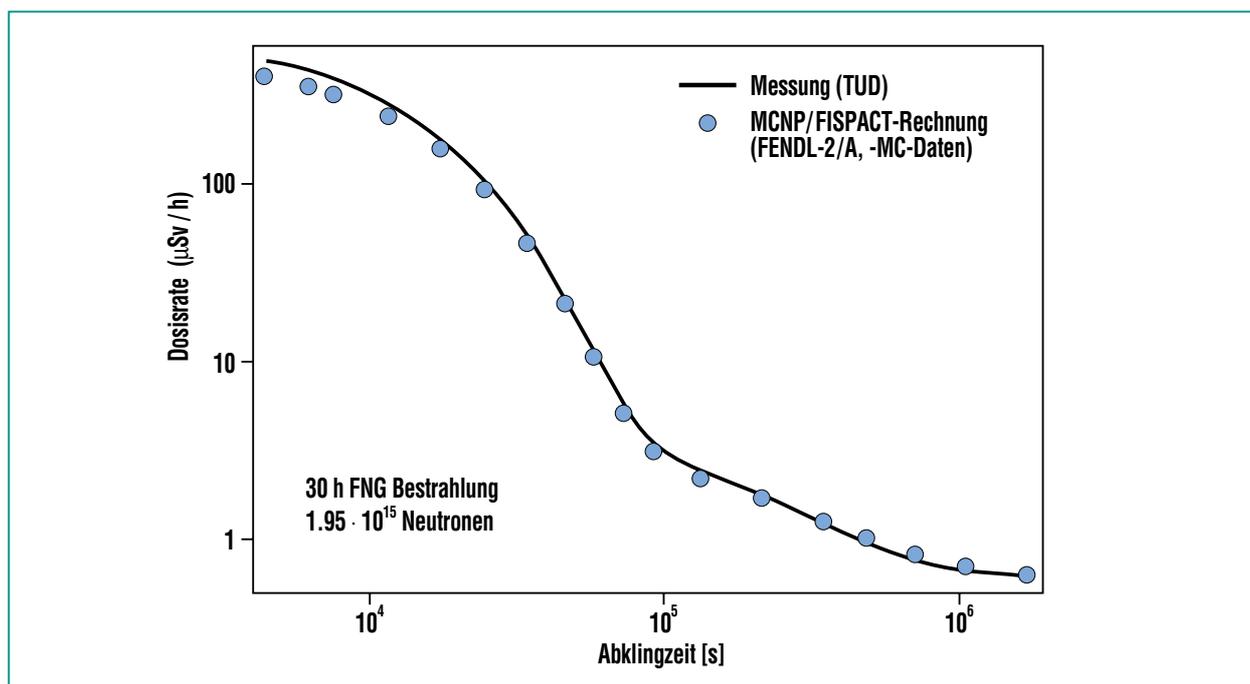


Abb. 2: Vergleich berechneter und gemessener Nachzerfallsdosisraten in einem Validierungsexperiment am Neutronengenerator Frascati.

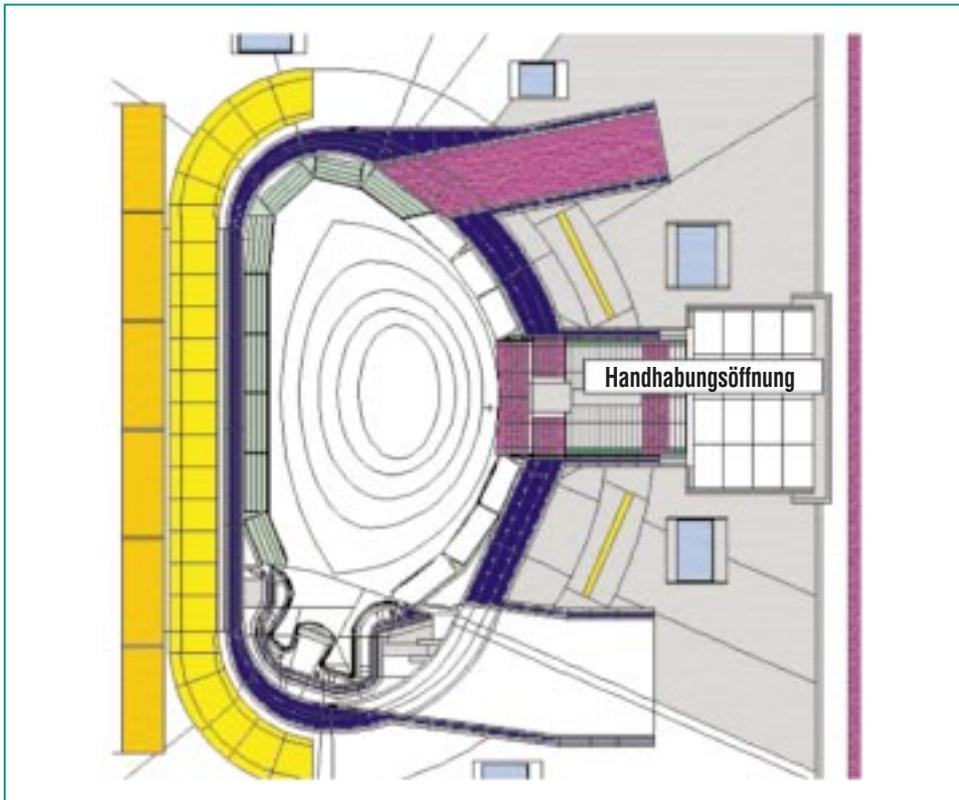


Abb. 3: MCNP-Modell von ITER (Vertikalschnitt) mit Handhabungsöffnung.

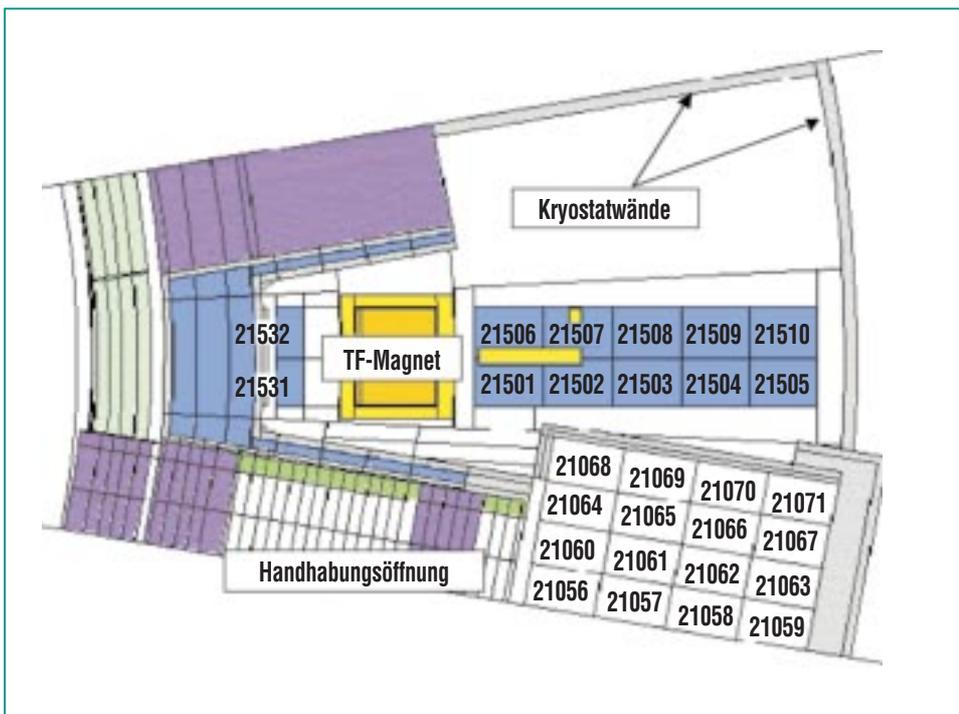


Abb. 4: Horizontalschnitt durch den Bereich der Handhabungsöffnung.

Abb. 5 zeigt die Nachzerfallsdosisraten die sich für die Zellen außerhalb der zentralen Handhabungsöffnung ergeben. Es zeigt sich, dass der Bereich um die Zellen #21502-21510 durch die Magnetfeldspulen recht gut abgeschirmt ist, so dass die Dosisrate dort innerhalb eines Tages unter den für das Wartungspersonal festgelegten Grenzwert von $100 \mu\text{Sv/h}$ fällt. Der Bereich um die Zellen #21531, 21532, ist hingegen nur durch Blanket und Vakuumgefäß abgeschirmt, so dass die Dosisleistung innerhalb des betrachteten Zeitraums von 30 Tagen über $100 \mu\text{Sv/h}$ bleibt. Wartungs- und Reparaturarbeiten sind unter den angenommenen Randbedingungen in diesem Bereich folglich nicht möglich.

Monte-Carlo-Rechnungen im Parallelmodus

Monte-Carlo-Simulationsrechnungen können prinzipiell parallel durchgeführt werden, da die Teilchenschicksale voneinander unabhängig sind. MCNP kann z. B. standardmäßig im Parallelmodus unter PVM [7] eingesetzt werden. Dabei wird die gleiche Monte-Carlo-Simulationsrechnung parallel auf mehreren Prozessoren durchgeführt. Die Gesamtzahl der zu berücksichtigenden Teilchenschicksale wird vom „Master“ auf die zur Verfügung stehenden Prozessoren aufgeteilt, indem diesen verschiedene Zufallszahlensequenzen zugewiesen werden. Die Ergebnisse der Simulationsrechnungen auf den verschiedenen Prozessoren werden nach Beendigung vom Master eingesammelt und zum Gesamtergebnis aufbereitet. Dieses ist völlig identisch

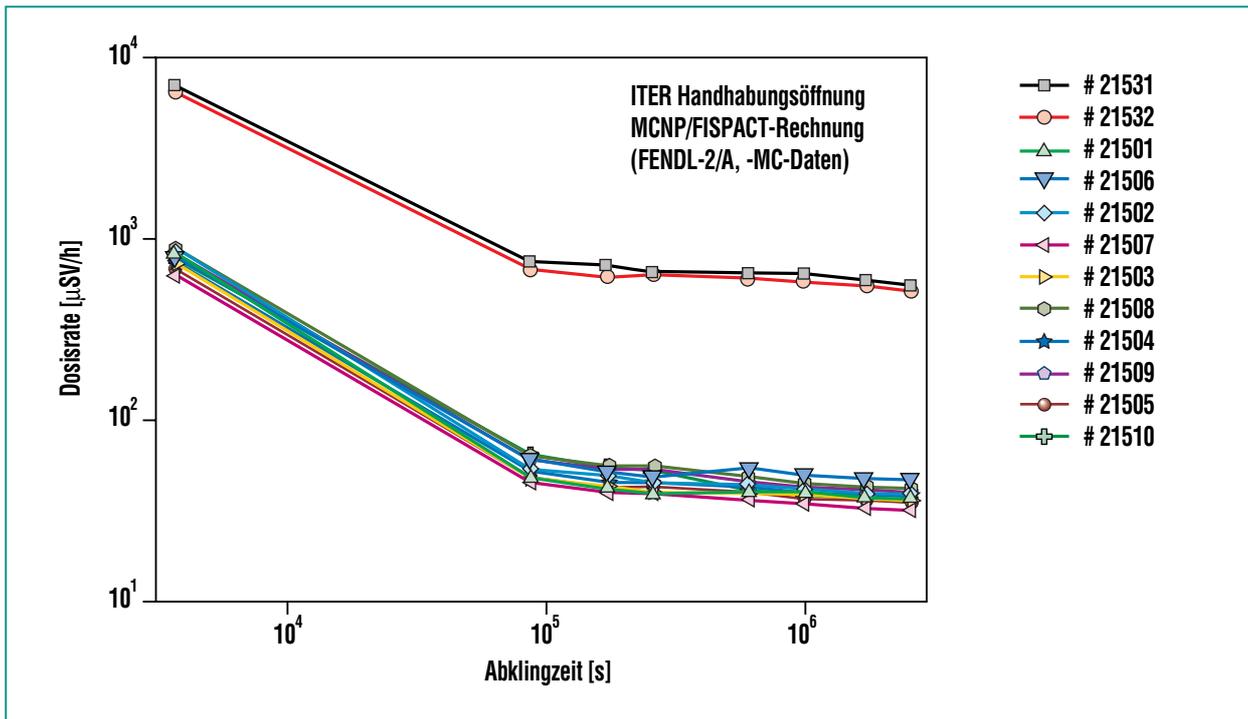


Abb. 5: Berechnete Nachzerfallsdosisraten im Bereich der Handhabungsöffnung.

mit dem Ergebnis das sich bei einer sequentiellen Rechnung auf einem Prozessor unter Verwendung der gleichen Zufallszahlensequenz ergeben würde. Auf dem benutzen HPC Linux-Cluster skaliert der Zeitgewinn dabei fast linear mit der Anzahl der genutzten Prozessoren, da sich die Kommunikation zwischen den Prozessoren auf die Zeiten beschränkt, bei denen die Teilergebnisse vom Master eingesammelt und ausgewertet werden. Die Häufigkeit dieser Kommunikationsprozesse wird durch die „Rendezvouszahl“

gesteuert, die Zahl aller – auf allen Prozessoren – bis zum Zeitpunkt des „Einsammelns“ berechneter Teilchenschicksale. Tab. 1 zeigt exemplarisch den Einfluss der Rendezvouszahl auf die Rechenzeit.

Zusammenfassung und Ausblick

Das Monte-Carlo-Verfahren ermöglicht es, den Teilchentransport auf mikroskopischer Ebene in beliebig komplexer Geometriedarstellung zu beschreiben. Mit dem

am Los Alamos National Laboratory LANL entwickelten Monte-Carlo-Programm MCNP und neueren Fusionskernbibliotheken steht ein Instrumentarium zur Verfügung, das sich vorzüglich zur Lösung neutronenphysikalischer Fragestellungen der Fusionstechnologie eignet. Die Lösung komplexer Probleme erfordert allerdings einen hohen Rechenaufwand, der unter Nutzung von Parallelrechnern wie dem HPC Linux-Cluster innerhalb eines überschaubaren Zeitrahmens geleistet werden kann. Damit war

Zahl der Prozessoren	1	15	15	15	11
Rendezvouszahl [Millionen]	–	1,0	0,5	0,1	0,8
Rechenzeit in Minuten („wall clock time“)	870	86	74	130	98

Tab. 1: Vergleich sequentieller und paralleler MCNP-Rechnungen auf dem HPC Linux-Cluster bei einer Gesamtzahl von 4,2 Millionen berechneter Teilchenschicksale.

es auch möglich, ein auf dem Monte-Carlo-Verfahren basierendes Programmsystem zu erstellen, mit dessen Hilfe drei-dimensionale Dosisratenverteilungen nach Reaktorabschaltung in beliebig komplexen Geometrien bestimmt werden können. Das Programmsys-

tem wurde anhand eines Bestrahlungsexperimentes am Neutronengenerator in Frascati validiert und anschließend auf den ITER-Reaktor angewandt. Zur Berechnung der Dosisratenverteilung in großen, die Reaktoranlage umgebenden Gebäuden, ist es not-

wendig, das Programmsystem zu erweitern. Eine Kopplung mit drei-dimensionalen deterministischen Rechenprogrammen ist dazu in Arbeit.

Literatur

- [1] I. Lux, L. Koblinger, *Monte Carlo Particle Transport Methods: Neutron and Photon Calculations*, CRC Press, Boca Raton 1991
- [2] ITER – International Thermonuclear Experimental Reactor, <http://www.iter.org>
- [3] P. Batistoni, M. Angelone, L. Petrizzi, M. Pillon, H. Freiesleben, D. Richter, K. Seidel, S. Unholzer, Y. Chen, U. Fischer, *Experimental Validation of Shutdown Dose Rates*, ENEA Report, June 2001
- [4] Y. Chen, U. Fischer, *Rigorous MCNP Based Shutdown Dose rate Calculations: Computational Scheme, Verification Calculations and Application to ITER*, 6th International Symposium on Fusion Nuclear Technology (ISFNT-6), San Diego, California, April 7-12, 2002
- [5] J.F. Briesmeister (ed.), *MCNP – A General Monte Carlo N-Particle Transport Code, Version 4C*, Los Alamos National Laboratory, Report LA-13709-M, April 2000.
- [6] R.A. Forrest, J.-Ch. Sublet, *FISPACT 99: User Manual*, UKAEA Fusion, Report UKAEA FUS 407, December 1998.
- [7] A. Geist et al, *PVM: A Users' Guide and Tutorial for Networked Parallel Computing*, The MIT Press, Cambridge, Massachusetts, 1994, <http://www.netlib.org/pvm3/book/pvm-book.html>

Numerische Simulation von Gasetonationen im ITER-FEAT-Fusionsreaktor

R. Redlinger, IKET; L. Obholz, HIK

Einleitung

Die Erzeugung von Energie durch Kernfusion ist eine der möglichen Zukunftstechnologien, die gegenwärtig in internationaler Kooperation im ITER-FEAT-Projekt (International Thermonuclear Experimental Reactor-Fusion Energy Advanced Tokamak) auf ihre technologische und wirtschaftliche Machbarkeit hin untersucht wird [1]. Grob gesagt besteht der Hauptteil (das Vakuumgefäß für das Plasma) des geplanten Fusionsreaktors aus 18 Scheiben der in Abb. 1 gezeigten Gestalt.

Für die Lizenzierung des Reaktors sind auch umfangreiche Sicher-

heitsanalysen nötig [2]. Eines der betrachteten Unfallszenarien (der sogenannte „ex-vessel loss of coolant accident“) führt auf die Möglichkeit einer Wasserstoff-Luft-Reaktion im Reaktor, wobei für die Auslegung des Reaktors insbesondere die Drucklasten als Folge dieser Verbrennungsvorgänge von Interesse sind. Detonationen liefern konservative, obere Schranken für diese Drücke, und die numerische Simulation einer derartigen (postulierten) Wasserstoff-Luft-Detonation im geplanten ITER-FEAT-Fusionsreaktor mithilfe einer parallelisierten Version des am IKET entwickelten Rechencodes DET3D soll im folgenden dargestellt werden.

Beschreibung des Rechencodes

Da Detonationen sehr schnelle Vorgänge sind – die übliche Geschwindigkeit der Detonationswelle bei Wasserstoff-Luft-Detonationen bewegt sich im Bereich von 1500-2000 m/s – ist es bei der physikalischen Modellbildung nicht nötig, Effekte wie molekulare Diffusion, Turbulenz, Strahlung oder Wärmeleitung in Betracht zu ziehen. Es genügt, die Eulergleichungen der Gasdynamik für eine Mischung von N idealen, chemisch reagierenden Gasen zu lösen. Bezeichnet ρ^k die Dichte der k -ten Komponente und $\rho = \sum_k \rho^k$ die Gesamtdichte, p den Gesamtdruck, $u = (u_1, u_2, u_3)$ die Geschwindigkeit in 3 Raumdimensionen und $e = 1/2 u^2 + \varepsilon$ die totale spezifische Energie mit ε als spezifischer innerer Energie, so lauten diese Gleichungen in cartesianischen Koordinaten x_j wie folgt ($k = 1, 2, \dots, N$; untere Indizes t bzw. x_j bedeuten Ableitung nach der Zeit bzw. nach der räumlichen Koordinate x_j ; über j wird jeweils von 1 bis 3 summiert):

$$\rho^k_t + (\rho^k u_j)_{x_j} = S^k$$

Massenerhaltung

$$(\rho u_i)_t + (\rho u_i u_j)_{x_j} + p_{x_i} = 0$$

Impulserhaltung
($i = 1, 2, 3$)

$$(\rho e)_t + [(\rho e + p) u_j]_{x_j} = 0$$

Energieerhaltung

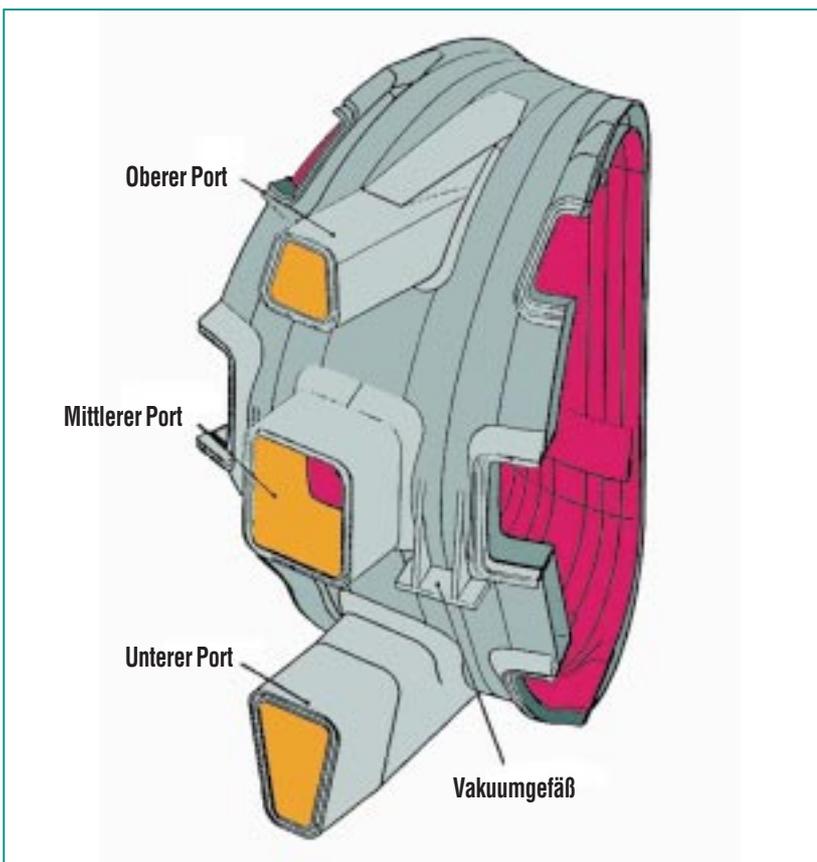


Abb. 1: Eine 20°-„Scheibe“ des ITER-FEAT-Vakuumgefäßes zur Einschließung des Fusionsplasmas.

Der Quellterm S^k in der Gleichung der Massenerhaltung modelliert dabei Änderungen der Gaskomponentendichten, die von den chemischen Reaktionen herrühren. Im Code DET3D stellen die Terme S^k die Summen von elementaren chemischen Reaktionen dar, deren Reaktionsgeschwindigkeit durch ein Arrhenius-Gesetz beschrieben wird. Sowohl die Anzahl N der Gaskomponenten als auch die Form und Anzahl der elementaren Reaktionen (inklusive der Arrhenius-Konstanten) kann vom Benutzer des Codes frei gewählt werden. Die thermodynamischen Eigenschaften der einzelnen Gase, d.h. die spezifischen inneren Energien und die (intern vom Code verwendeten) spezifischen Wärmen, werden als Polynome der Temperatur angegeben, die durch Interpolation von Messdaten [3] erhalten wurden.

Als numerischen Löser benützt DET3D ein modernes Verfahren, das Stoßvorgänge automatisch modelliert („shock capturing“), siehe etwa [4]. Das implementierte Verfahren ist explizit und benützt finite Differenzen, wobei der Benutzer die Wahl zwischen Lösern 1. und 2. Ordnung hat.

Bezüglich der Validierung des Codes an Experimenten und bekannten analytischen Lösungen sowie seiner Anwendung auf Sicherheitsstudien für Kernkraftwerke sei der Leser auf [5] verwiesen.

Parallelisierung

Zur Parallelisierung von DET3D wurde die Methode der Gebietszerlegung gewählt, d.h. das quaderförmige, 3-dimensionale Re-

chengebiet wird in einzelne Blöcke (kleinere Quader) aufgeteilt, wobei der Benutzer vor Beginn der Rechnung die Anzahl der Unterteilungen in allen 3 Koordinatenrichtungen vorgibt. Eine Zerlegung $Z = (2, 1, 2)$ bedeutet z.B. 2 Abschnitte in x-, 1 Abschnitt in y- und 2 Abschnitte in z-Richtung, das heißt $2 \times 1 \times 2 = 4$ Blöcke.

Jeder so erhaltene Block hat dann während der Rechnung „seinen“ zu ihm gehörenden Prozessor. Im eben angeführten Beispiel würde man die Rechnung also auf 4 Prozessoren durchführen.

Implementiert wurde diese Methode der Parallelisierung mit Hilfe von MPI (Message Passing Interface), einem Standard-Programmpaket, das von einer Gruppe von Hard- und Softwarefirmen sowie Benutzern entwickelt wurde [6].

Theoretisch könnte man somit bei N Prozessoren die Rechenzeit im günstigsten Fall um einen Faktor N verkleinern. In der Praxis ist das natürlich aus mehreren Gründen nicht durchführbar. Genannt seien hier:

- Nicht alle Teile des Codes sind parallelisierbar.
- Die Prozessoren müssen immer wieder synchronisiert werden. Bei DET3D z.B. jeweils nach Beendigung eines Zeitschritts. Da die einzelnen Prozessoren aber i.a. nicht dieselbe Rechenlast haben (die Blöcke sind ja meistens nicht gleich groß wählbar) und deshalb unterschiedlich lange für einen Zeitschritt brauchen, führt das zu Zeitverlusten.

- Die einzelnen Prozessoren müssen miteinander kommunizieren, d.h. Daten austauschen. Dieser Datenaustausch ist aber eine zusätzliche Aufgabe des parallelisierten Codes, die bei der Rechnung auf nur einem Prozessor gar nicht vorkommt.

Tab. 1 macht diesen Effekt anhand einer Testrechnung auf der SP-Power-3-SMP-Maschine des FZK-Rechenzentrums deutlich (eine ITER-Rechnung wie im nächsten Abschnitt beschrieben mit 500 Zeitschritten): Die theoretisch mögliche Verkürzung der Rechenzeit bei N Prozessoren um den Faktor N wird nur für sehr kleine Werte von N wenigstens annähernd erreicht. Für größere Werte entfernt man sich immer mehr davon.

Man beachte in Tab. 1 auch, dass es dabei durchaus auf die Art der Gebietszerlegung ankommt. Dieser Effekt ist schon bei 4 Prozessoren und den Zerlegungen (1, 4, 1) und (2, 2, 1) sichtbar. Überraschender ist der Vergleich der Rechnungen mit 32 und 36 Prozessoren: Die Rechnung mit 36 Prozessoren dauert wegen der ungünstigen Zerlegung (6, 6, 1) deutlich länger als die Rechnung auf 32 Prozessoren mit der günstigeren (4, 4, 2)-Zerlegung.

Neben der Verkürzung der Rechenzeit sei noch ein weiterer, sehr wichtiger Vorteil der parallelisierten Rechnung erwähnt: Da das Startfile in kleinen Stücken auf mehrere Prozessoren verteilt wird, lassen sich so wesentlich größere Probleme in Angriff nehmen als dies auf nur einem Prozessor möglich wäre, siehe z.B. die in Abschnitt 5 erwähnten Rechnungen.

Anzahl Prozessoren	Zerlegung			Rechenzeit [s]	Beschleunigungsfaktor
	x	y	z		
1	1	1	1	132732	1.00
2	1	2	1	68576	1.94
4	1	4	1	38004	3.49
4	2	2	1	35171	3.77
9	3	3	1	19280	6.88
16	4	4	1	10634	12.48
25	5	5	1	8372	15.85
32	4	4	2	5575	23.81
36	6	6	1	5969	22.24
50	5	5	2	4112	32.28
72	6	6	2	3568	37.20

Tab. 1: Einfluss der Anzahl der Prozessoren auf die Rechenzeit. Testrechnung mit ca. 1.25 Mio numerischer Zellen.

Simulation einer Gasedetonation im ITER-FEAT Fusionsreaktor

Der parallelisierte Code DET3D wurde dazu benutzt, Druckverteilungen im ITER-FEAT-Reaktor zu berechnen, die durch eine postulierte Wasserstoff-Luft-Detonation entstehen [7]. Als Unfallszenario wird dabei der Bruch eines Kühlrohres angenommen. Dadurch kommt (nach einigen Minuten) der Dampf des Kühlsystems mit dem heißen Beryllium des Vakuumgefäßes in Kontakt, was zur chemischen Erzeugung von Wasserstoff führt. Dieser Wasserstoff seinerseits kann dann später mit der Luft reagieren, die nach Ausströmen des Dampfes durch das gebrochene Kühlrohr in das Vakuumgefäß eindringt [8].

Alle Rechnungen wurden auf einem IBM-RS/6000-Parallelrech-

ner vom Typ SP-Power-3-SMP unter Verwendung von 9 Prozessoren durchgeführt. Als Gebietszerlegung wurde (3, 1, 3) gewählt.

Zunächst wurde anhand vom ITER-Team bereitgestellter Informationen ein 3-dimensionales Modell des Vakuumgefäßes im Code implementiert, das die komplexe Geometrie des Reaktor-Torus' plus Ports analytisch exakt nachbildet. Das diskretisierte numerische Gitter wird dann vom Code nach Eingabe der gewünschten numerischen Maschenweite selbst erstellt.

Im betrachteten Unfallszenario wird nun angenommen, dass insgesamt 5 kg Wasserstoff verbrannt werden. Um für die bei dieser Verbrennung auf das Vakuumgefäß ausgeübten Drucklasten limitierende obere Schranken zu bekommen, wurden folgende Anfangsbedingungen gewählt:

- stöchiometrische Mischung von Wasserstoff und Luft in einem keilförmigen Bereich des Reaktors; ansonsten Luft.
- Zündung in der Spitze des Keils an der inneren Toruswand.
- Anfangsdruck 1 bar, Anfangstemperatur 140° Celsius.
- keine internen Strukturen in den Ports.

Insgesamt ergibt das einen hochexplosiven, keilförmigen Gasbereich von ca. 290 m³, während das gesamte Volumen des Vakuumgefäßes plus Ports ca. 2250 m³ beträgt.

Nach Zündung wird dieser Keilbereich von einer kugelförmigen Detonationswelle durchlaufen. Die Welle hat eine Geschwindigkeit von ca. 1800 m/s und an der Front einen Spitzendruck von ca. 13 bar. Die Situation nach 1 ms veranschaulicht Abb. 2.

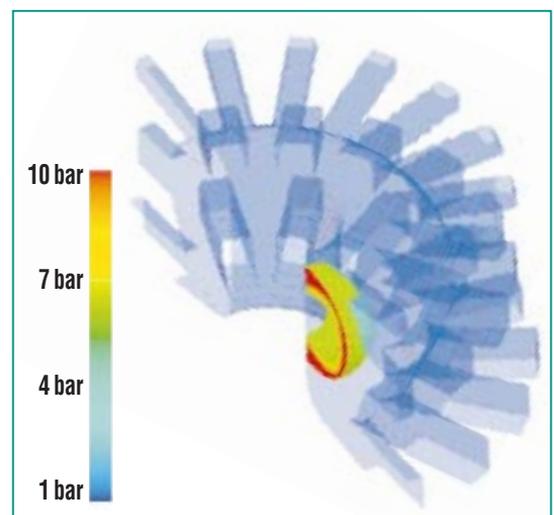


Abb. 2: Druck-Konturplot der kugelförmigen Detonationswelle nach 1 ms. Zündung war im Mittelpunkt der „Detonationskugel“. Gezeigt wird die Hälfte des Reaktormodells; Maschenweite 15 cm, ca. 2.2 Mio. Zellen.

Außerhalb des Wasserstoff-Luft-Keils breitet sich die Detonationswelle als expansive Druckwelle in den Rest des Reaktors aus, wobei Gasgeschwindigkeit und -druck sehr schnell abnehmen, da dort keine chemischen Reaktionen mehr stattfinden. Ca. 25 ms nach der Zündung treffen die beiden (symmetrischen) Druckwellen auf der gegenüberliegenden Torusseite aufeinander, wobei durch Überlagerung im Gasbereich noch Spitzendrücke von ca. 5 bar erreicht werden können, an den Wänden durch Reflektion sogar um die 10 bar (siehe z.B. Abb. 4d).

Abb. 3 zeigt die Ausbreitung der nach rechts laufenden Detonationswelle im Reaktor in Momentaufnahmen von 1 bis 6 ms anhand eines vertikalen Schnitts durch die Symmetrieebene des Torus. Bei 1 ms ist die Welle (zumindest im

Zentrum des stöchiometrischen Wasserstoff-Luft-Keils) noch annähernd kugelförmig. Die Öffnungen der Ports werden nach ca. 2 ms erreicht, die Portenden nach etwas mehr als 3 ms. Bei 6 ms sieht man schon die reflektierte, nach links laufende Detonationswelle; der gesamte Wasserstoff ist jetzt verbrannt, und es handelt sich nun nur noch darum, die folgenden Stoßwellen und Reflektionen mit dem Code zu simulieren.

Verkleinern der numerischen Zellgröße

Da die Reaktorgeometrie im Code analytisch in absoluten Einheiten (Meter) implementiert wurde, kann man durch Änderung einer einzigen Eingabegröße, nämlich der numerischen Maschenweite, immer „bessere“, automatisch

vom Code erzeugte, diskretisierte Modelle dieser Geometrie bei der Rechnung verwenden. Der Einfluss der gewählten Maschenweite auf die erhaltenen Rechenergebnisse lässt sich so auf sehr einfache Weise studieren.

Ein prinzipielles Problem bei CFD (Computational Fluid Dynamics)-Rechnungen mit diskretisierten Modellen und Gleichungen ist ja immer die Frage, wie nahe die berechneten Werte an der exakten Lösung liegen. Theoretisch gilt Konvergenz, wenn die Maschenweite gegen Null geht, doch sind über den absoluten Fehler im konkreten Fall bei gegebener Maschenweite im allgemeinen nur sehr vage Aussagen möglich. Die Art, wie die Modell-Geometrie in DET3D implementiert wird, ermöglicht nun folgende Vorgehensweise: Man verfeinert die Ma-

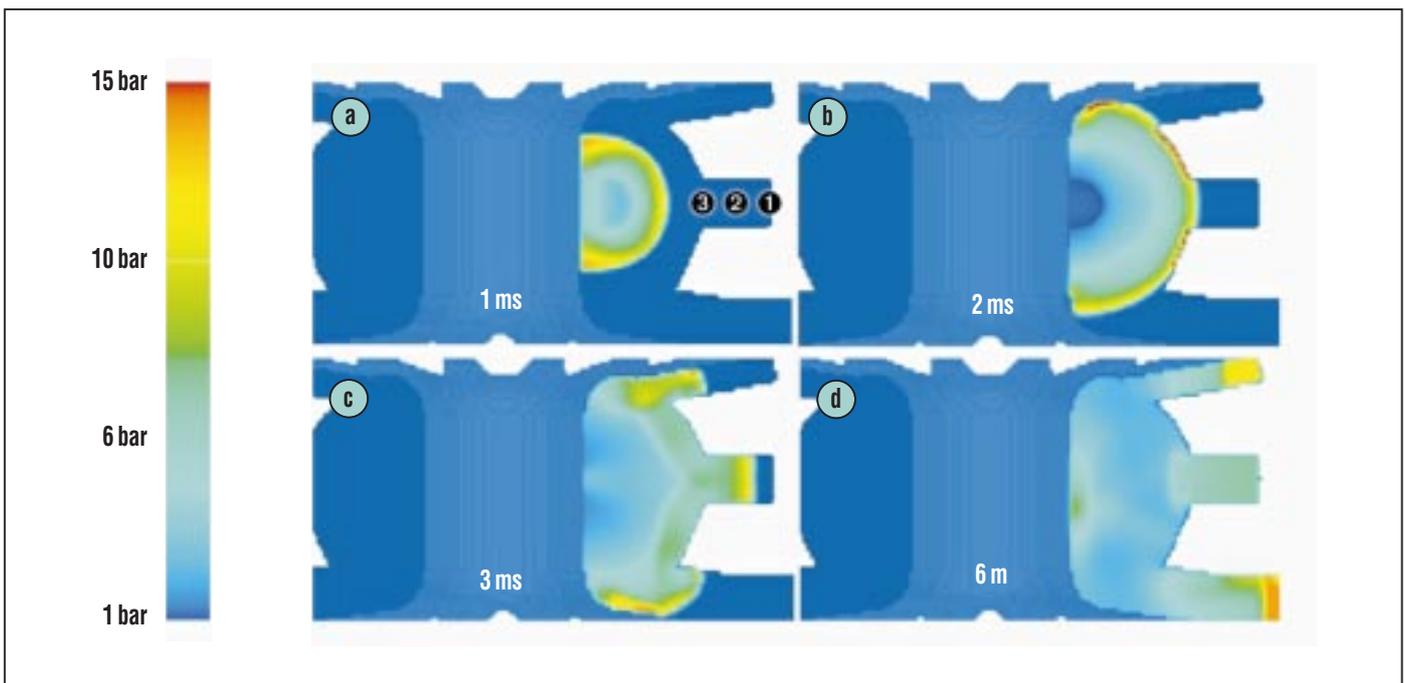


Abb. 3: Ausbreitung der Detonationswelle zu verschiedenen Zeiten nach der Zündung. Die Ziffern im hauptsächlich betroffenen mittleren Port geben die Positionen der Drucksimulationen in Abb. 4 an. Der vertikale Schnitt durch den Torus ist auf der linken Seite nicht ganz vollständig wiedergegeben.

schenweite solange, bis sich die berechneten Druckdaten nicht mehr wesentlich ändern, und akzeptiert dann diese Werte als realistisch. Diese Methode wurde auf das ITER-Szenario angewendet, und zwar für die Maschenweiten 15.0, 11.6 und 8.5 cm.

Abb. 4 zeigt einen Vergleich der Druckverläufe an den drei in Abb. 3a mit 1, 2, 3 markierten Positionen im hochexplosiven Wasser-

stoff-Luft Keil, sowie zusätzlich zum Vergleich einen Druckverlauf im inerten Luft-Bereich, und zwar an einer Position, die dem Zündpunkt an der inneren Torusseite gegenüberliegt.

Man sieht, dass sich die berechneten Daten für die Maschenweite 8.5 cm nicht mehr stark von denen für 11.6 cm unterscheiden. Noch deutlicher wird diese Übereinstimmung, wenn man nicht den

(Über-)Druck selbst, sondern sein Integral über die Zeit betrachtet, den sogenannten Impuls. Es ist bekannt, dass für die Strukturanalyse, d.h. die Untersuchung der Wirkung der Druckwellen auf die Wände, dieser Impuls von entscheidender Bedeutung ist [9], [10].

Plottet man nun diesen Impuls an verschiedenen Positionen, so erhält man für alle 3 Rechnungen je-

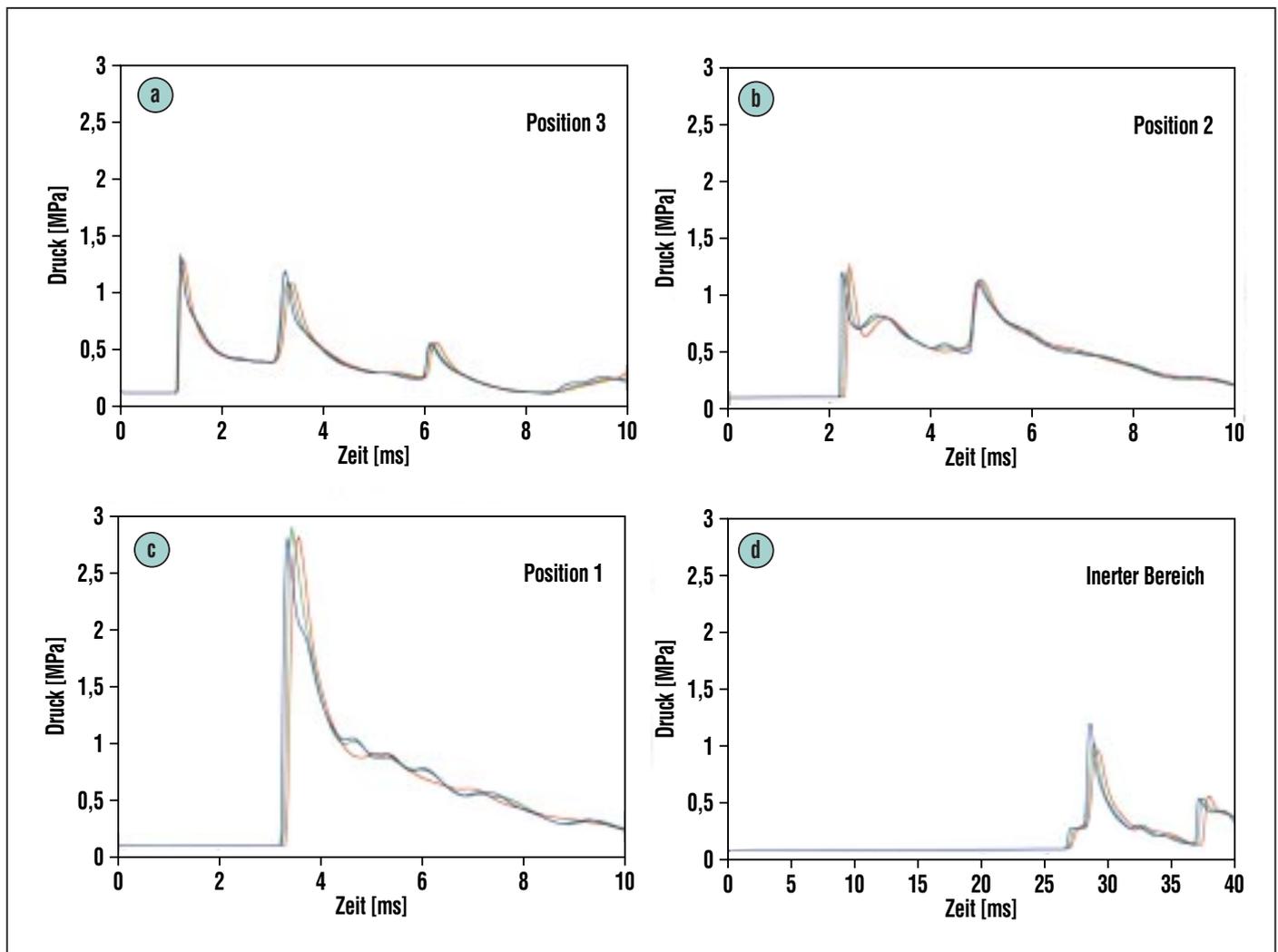


Abb. 4: Druckverläufe an den Positionen 1, 2, 3 der Abb. 3a für Rechnungen verschiedener Maschenweite. Die erste Druckspitze kennzeichnet jeweils die Ankunft der Detonationswelle, die späteren stammen von Reflektionen dieser Welle an den Wänden. Zum Vergleich noch ein Druckverlauf im inerten Bereich: Innere Toruswand, diametral gegenüber der Zündung. Maschenweite 15.0 cm: rot, 11.6 cm: grün, 8.5 cm: blau.

weils praktisch dieselbe Kurve. Ein typisches Beispiel dieser Fast-Übereinstimmung zeigt Abb. 5. Dies bedeutet, dass schon die 15-cm-Rechnung brauchbare Werte

für den Impuls liefert, und dass die mit dieser Rechnung erhaltenen Ergebnisse deshalb für eine Sicherheitsanalyse akzeptiert werden können [7].

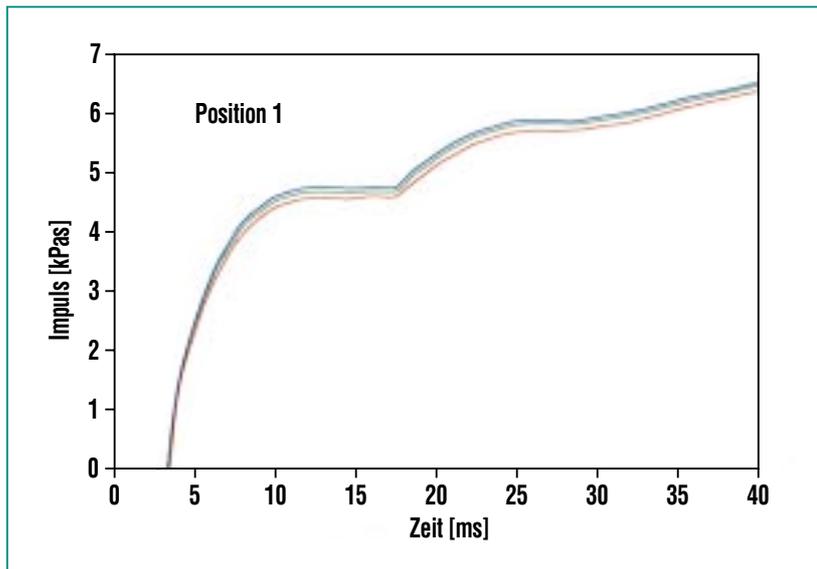


Abb. 5: Impulsverlauf an der Position 1 der Abb. 3a für Rechnungen verschiedener Maschenweite. Maschenweite 15.0 cm: rot, 11.6 cm: grün, 8.5 cm: blau.

Zusammenfassung

Die parallelisierte Version des Finite-Differenzen-Codes DET3D ermöglicht die 3-dimensionale Simulation von Gasedetonationen auf sehr feinen numerischen Gittern. Rechnungen auf diskretisierten Geometriemodellen mit z.B. 10 Millionen numerischer Zellen sind damit ohne weiteres möglich. Illustriert wird dies im vorliegenden Bericht am Beispiel einer Sicherheitsanalyse für den geplanten ITER-FEAT-Fusionsreaktor.

Literatur

- [1] W. Bahm, G. Dammertz, M. Glugla, G. Janeschitz, P. Komarek, A. Mack, *Int. Zeitschrift für Kernenergie (atw)* 47 (2002), 96-106
- [2] M.J. Gaeta, B.J. Merrill, H.-W. Bartels, C.R. Laval, L. Topilski, *Fusion Technology* 32 (1997), 23-34
- [3] *JANAF Thermochemical Tables, J. Phys. and Chem. Ref. Data* 14 (1985), Supplement
- [4] A. Harten, P.D. Lax, B. van Leer, *SIAM Review* 25 (1983), 35-62
- [5] R. Redlinger, *Nachrichten-Forschungszentrum Karlsruhe* 32 (2000), 243-249
- [6] Wegen näherer Informationen zu MPI sei auf die MPI-homepage www-unix.mcs.anl.gov/mpi verwiesen.
- [7] W. Baumann, W. Breitung, B. Kaup, R. Redlinger, J.R. Travis, *Wissenschaftliche Berichte – Forschungszentrum Karlsruhe, FZKA 6584* (2001), iv + 28 pp.
- [8] W. Baumann, W. Breitung, B. Kaup, G. Necker, P. Royl, J.R. Travis, *Wissenschaftliche Berichte – Forschungszentrum Karlsruhe, FZKA 6663* (2001), v + 37 pp.
- [9] W. Breitung, R. Redlinger, *Kerntechnik* 59 (1994), 162-170
- [10] W. Breitung, R. Redlinger, *Nuclear Technology* 111 (1995), 420-425

Hochtemperatursupraleiter – Festkörpersimulation

K.-P. Bohnen, R. Heid, IFP

Einleitung

Auf dem Gebiet der Festkörperphysik hat das Verständnis der Erscheinung der Supraleitung seit seiner Entdeckung vor über 30 Jahren stets eine herausragende Rolle gespielt. Das Verschwinden des elektrischen Gleichstromwiderstands in bestimmten Festkörpern unterhalb einer „Sprungtemperatur“ T_c verbunden mit speziellem magnetischem Verhalten hat wegen möglicher Anwendungen weltweit großes Interesse gefunden. Unglücklicherweise waren bis in die achtziger Jahre nur Supraleiter mit Sprungtemperaturen unterhalb 23 K bekannt (Tieftemperatursupraleiter), so dass ihr Einsatz wegen der hohen Anforderungen an die Kryotechnik (Kühlung mit flüssigem Helium) nur sehr begrenzt möglich war. Mit der Entdeckung der Hochtemperatursupraleiter im Jahre 1986 durch Bednorz und Müller [1] änderte sich dies sehr schnell, da bald eine Menge von Materialien entdeckt wurde, deren Sprungtemperatur zwischen 80 und 140 K liegt und daher der Kühlung mit flüssigem Stickstoff zugänglich ist.

Zum Verständnis des Phänomens der Supraleitung in den Tieftemperatursupraleitern wurde der entscheidende Beitrag mit der Entdeckung der Bedeutung der Wechselwirkung zwischen Elektronen und Gitterschwingungen (Elektron-Phonon-Wechselwirkung) 1957 geliefert [2]. Die auf dieser Wechselwirkung aufgebaute Theorie der Supraleitung ist unter dem Namen Eliashberg-Theorie bekannt. Sie erlaubt, ausgehend von den Eigenschaften

des Normalleiterzustandes, die Berechnung der Sprungtemperatur T_c und anderer Eigenschaften des Supraleiters.

In der Vergangenheit wurden die für diese Theorie benötigten Informationen über den Normalleiterzustand meist aus Messungen entnommen, so dass keine echte Möglichkeit der Vorhersage gegeben war. Dies hat sich in den letzten Jahren deutlich geändert. Die verlässliche Berechnung der Elektronenstruktur, der Gitterdynamik und der Elektron-Phonon-Kopplung ist dank verbesserter numerischer Verfahren und verbesserter Leistungsfähigkeit der Rechner inzwischen für einfache Systeme auf Workstations oder PC-Clustern möglich. Dies wurde kürzlich sehr erfolgreich für den neu entdeckten Supraleiter MgB_2 und andere Diboride gezeigt [3].

Beim Verständnis der Hochtemperatursupraleiter (HTSL) ist man von einer vergleichbaren Situation noch weit entfernt (für einen Überblick über Supraleitung in den HTSL sei auf einen Beitrag in den FZKA-Nachrichten [4] verwiesen). Dies beruht zunächst darauf, dass die für die Supraleitung entscheidende Kopplung bisher nicht bestimmt werden konnte. Erschwerend kommt hinzu, dass die Hochtemperatursupraleiter eine komplexe Kristallstruktur mit vielen Atomen in der Einheitszelle besitzen und daher nur mit sehr aufwändigen numerischen Verfahren behandelt werden können. Es ist daher bisher nicht bekannt, welche Rolle die Elektron-Phonon-Wechselwirkung in diesen Systemen spielt. In der vorliegenden Arbeit soll an einem typischen

Hochtemperatursupraleiter (HTSL) gezeigt werden, dass mit Hochleistungsrechnern Normalleitereigenschaften auch für diese Systemklasse erfolgreich simuliert werden können, so dass der Beitrag der Elektron-Phonon-Wechselwirkung zur Supraleitung parameterfrei bestimmt werden kann. Für diese Simulation wurde $YBa_2Cu_3O_7$ gewählt, da dieses der am Besten studierte HTSL ist, für den auch detaillierte experimentelle Untersuchungen der Gitterdynamik vorliegen.

Die verwendeten numerischen Verfahren werden im Hinblick auf Speicherplatz und Rechenzeitbedarf im Folgenden genauer diskutiert.

Berechnung der Gitterstruktur, elektronischen Bandstruktur, Gitterdynamik und Elektron-Phonon-Kopplung

Das Verfahren zur Bestimmung der Gitterstruktur beruht auf einer Minimalisierung der Gesamtenergie des betrachteten Festkörpers, wobei wegen der periodischen Anordnung der Atome des Kristalls nur eine Einheitszelle (die Grundeinheit, aus der der Festkörper aufgebaut werden kann) zu betrachten ist. In Abb. 1 ist diese Einheitszelle für das System $YBa_2Cu_3O_7$ dargestellt. Freie Parameter für die Simulation sind die Achsenlängen a, b, c und die Position der Atome innerhalb der Zelle (für $YBa_2Cu_3O_7$ fünf weitere Freiheitsgrade). Der numerisch aufwändigste Schritt besteht in der Lösung eines verallgemei-

nernten Eigenwertproblems (Eigenwert ϵ)

$$H\psi = \epsilon S\psi \quad (1)$$

wobei die Matrix H nichtlinear von ψ abhängt und S eine symmetri-

sche Matrix ist. Die Größe der Matrizen wird von der gewünschten Genauigkeit bestimmt.

Der benötigte Speicherplatz variiert zwischen 300 und 700 Megabytes und die zu diagonalisieren-

den Matrizen haben eine Dimension zwischen 726 und 1772. Neben der Lösung des Eigenwertproblems (1) wird bei der Energieberechnung als weiterer wichtiger numerischer Schritt eine Fast-Fourier-Transformation (FFT) mehrfach eingesetzt, um verschiedene Beiträge zu H möglichst ökonomisch zu berechnen. In Tab. 1 sind typische Rechenzeiten für eine Energieberechnung bei vorgegebener Struktur auf der VPP5000 bei verschiedener Matrixgröße angegeben. Neben den Rechenzeiten wurde auch die Zahl der ausgeführten Floating-Point-Operationen pro Sekunde angegeben, die ein Maß für die Ausnutzung der Rechenleistung des verwendeten Prozessors darstellt. Die verbesserte Nutzung der verfügbaren Rechenleistung der VPP5000 mit wachsender Systemgröße führt dazu, dass die Rechenzeit nur moderat anwächst.

Wegen der großen Zahl der zu optimierenden Strukturparameter beträgt der Gesamtaufwand für die Strukturoptimierung ein Vielfaches der in Tab. 1 angegebenen Rechenzeiten. Die berechneten Atompositionen sind in guter Übereinstimmung mit Messdaten [5].

Zur Berechnung der Gitterschwingungen ist die Kenntnis der 2. Ableitung der Energie des Systems nach den Atomkoordinaten nötig. Dabei ist der zentrale, rechenintensive Schritt die Lösung eines inhomogenen linearen Gleichungssystems. Der Rechenaufwand steigt mit der 4. Potenz der Anzahl der Atome, so dass mit wachsender Atomzahl die Leistungsfähigkeit moderner Workstations rasch überfordert ist. In

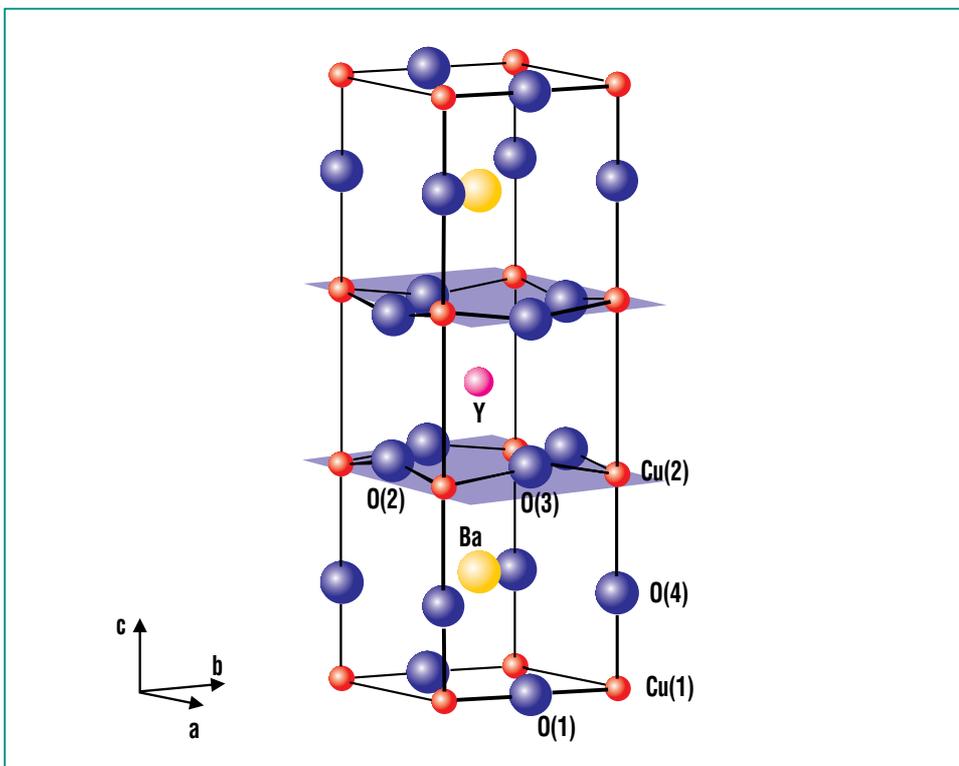


Abb. 1: Kristallstruktur von $YBa_2Cu_3O_7$.

Matrix-dimension	Struktur		Gitterdynamik	
	Zeit [Std.]	MFLOPS	Zeit [Std.]	MFLOPS
N_D				
726	7	645	29	4200
1125	13	960	33	4100
1360	17	1144	37	4000
1772	29	1387	56	3800

MFLOPS = 10^6 Floating Point Operations / sec

Tab. 1: Typische Rechenzeiten und Leistungsdaten der VPP5000 für die Berechnung der Gitterenergie (Struktur) und Gitterdynamik für $YBa_2Cu_3O_7$ für unterschiedliche Matrixdimensionen N_D .

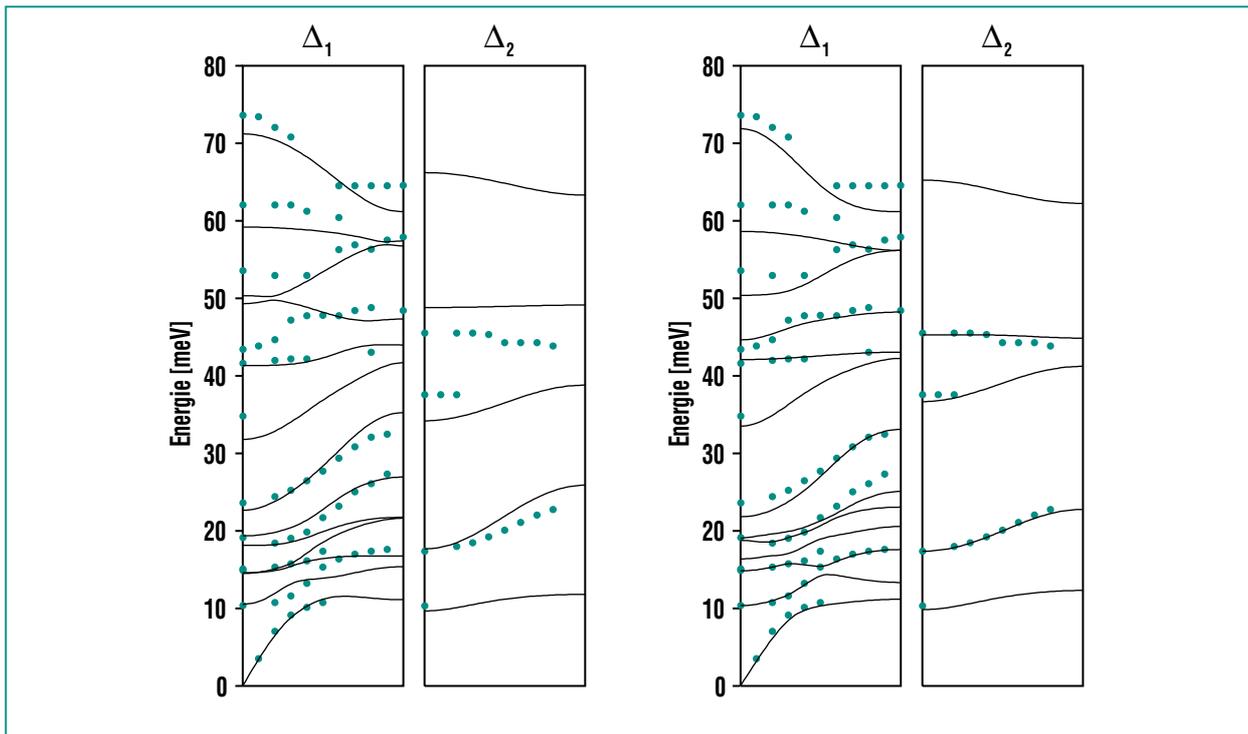


Abb. 2: Frequenzen atomarer Eigenschwingungen in einer ausgewählten Kristallrichtung für verschiedene Symmetrien (Δ_1, Δ_2). Im linken Teil der Abbildung sind Ergebnisse dargestellt, die mit moderaten Anforderungen an Speicherplatz und Rechenzeit erzielt wurden. Die Ergebnisse sehr genauer und damit sehr aufwändiger Rechnungen sind im rechten Teilbild dargestellt.

Tab. 1 sind Rechenzeiten für verschiedene Genauigkeiten für die VPP5000 für einen hochsymmetrischen Punkt in der Brillouin-Zone (BZ) angegeben. Für eine verlässliche Interpolation des gesamten Phononenspektrums benötigt man etwa 20 Punkte in der BZ.

Neben den sehr großen Anforderungen an die Rechenzeit wird in unserem Beispiel Speicherplatz von mehr als 5 Gigabyte pro Prozessor benötigt. In Abb. 2 sind die Frequenzen für Gitterschwingungen der Δ_1 - und Δ_2 -Symmetrie längs der (100)-Kristallrichtung für zwei verschiedene Rechnungen dargestellt, die sich bezüglich ihrer Genauigkeit unterscheiden. Die zweite Rechnung wurde so

bezüglich der Systemressourcen reduziert, dass die Speicherplatzanforderung unter 2 Gigabyte liegt und die Rechenzeit ungefähr einen Faktor 3-4 geringer ist als für die aufwändigsten Rechnungen.

Wie man sieht, ist die Übereinstimmung mit den experimentellen Ergebnissen von Pintschovius et al. [6] für die sehr genaue Rechnung gut. Die Absolutwerte für die Frequenzen stimmen mit den gemessenen Daten gut überein. Sehr befriedigend ist auch die Tatsache, dass der bisher unverstandene starke Abfall der höchsten Mode als Funktion des Wellenvektors q längs der (100)-Richtung in den Simulationen korrekt wiedergegeben wird.

Auf der Basis dieser Ergebnisse lässt sich der Einfluss der Elektron-Phonon-Wechselwirkung für die Supraleitung parameterfrei bestimmen. Erste Ergebnisse unserer Rechnungen deuten darauf hin, dass die Elektron-Phonon-Wechselwirkung in den HTSL eine vernachlässigbare Rolle spielt. Für eine detaillierte Diskussion wird der Leser auf Ref. 7 verwiesen.

Zusammenfassung

Die vorliegenden Rechnungen zur Elektronenstruktur, Gitterdynamik und Elektron-Phonon-Wechselwirkung haben gezeigt, dass auch komplexe Systeme wie $\text{YBa}_2\text{Cu}_3\text{O}_7$ mit modernen Simu-

lationsverfahren der Festkörperphysik erfolgreich behandelt werden können. Im Hinblick auf Speicherplatz und Rechenzeitbedarf sind Hochleistungsrechner zur Bewältigung der numerischen Behandlung zwingend notwendig.

Danksagung

Die Autoren danken Prof. Dr. H. Wühl sowie dem Institut für Technische Physik für die Überlassung der Abb. 1 und den Herren Dr. L. Pintschovius und Dr. W. Reichardt für zahlreiche Diskussionen.

Literatur

- [1] J.G. Bednorz, K.A. Müller, *Z. Phys. B.* 64, 189 (1986)
- [2] J.R. Schrieffer, *Theory of Superconductivity*, W.A. Benjamin, N.Y. (1964)
- [3] R. Heid, K.-P. Bohnen, B. Renker, *Adv. in Solid State Phys.* 42, 293 (2002)
- [4] W. Fietz, W. Schauer, H. Wühl, C. Meingast, L. Pintschovius, S. Schuppler, H. Winter, *Nachrichten – Forschungszentrum Karlsruhe, Jahrgang 31, 4 / 1999*, 233
- [5] E. Kaldis, *Handbook on the Physics and Chemistry of Rare Earths, Vol. 31*, K.A. Gschneidner, jr., L. Eyring, M.B. Maple (eds.), Elsevier Sci. (2001)
- [6] L. Pintschovius, W. Reichardt, in „*Neutron Scattering in Layered Copper-Oxide Superconductors*“, Albert Furrer (eds.), Kluwer Academic Publishers, Dordrecht (1998), 165
- [7] K.-P. Bohnen, R. Heid, R. Krauss, in *Vorbereitung*

Computeralgebra: Schlüssel zu Präzisionsrechnungen in der Teilchenphysik

J.H. Kühn, Institut für Theoretische Teilchenphysik, Universität Karlsruhe
 M. Steinhauser, II. Institut für Theoretische Physik, Universität Hamburg

Computeralgebra hat sich in den letzten beiden Dekaden zum entscheidenden Hilfsmittel für präzise Berechnungen von sogenannten Streuquerschnitten und Zerfallsraten entwickelt. Diese Größen charakterisieren die Stoßwahrscheinlichkeiten und Lebensdauern von Elementarteilchen und werden in aktuellen Experimenten der Elementarteilchenphysik gemessen. Die in diesem Zusammenhang verwendeten Methoden sind nicht nur unverzichtbar für die Überprüfung der gegenwärtigen Theorie der Elementarteilchen und damit der fundamentalen Naturgesetze bei den kleinsten derzeit zugänglichen Abständen, sie haben vielfach auch in andere Bereiche der Mathematik, der Naturwissenschaften und der Technik Eingang gefunden.

Gerade bei Anwendungen in der Teilchenphysik, der Berechnung

sogenannter Feynmandiagramme, zeigt sich besonders deutlich die Komplementarität computeralgebraischer und numerischer Methoden. Feynmandiagramme repräsentieren in grafischer Weise eine grundlegende Konsequenz relativistischer Quantenfeldtheorien: die quantenmechanische Unschärferelation erlaubt für sehr kurze Zeit den Übergang in Zustände, welche nach den Gesetzen der klassischen Mechanik, hier der Energieerhaltung, unzugänglich sind. Wegen der Äquivalenz von Energie und Masse, einer Folge der Einstein'schen Relativitätstheorie, erscheint diese Fluktuation in einer Vielfalt von Zuständen, die im Allgemeinen aus den verschiedensten Kombinationen von Teilchen und Antiteilchen bestehen. So steht das in Abb. 1 (a) abgebildete Diagramm für die Fluktuation eines Muons in ein fast tausendmal schwereres W-Boson, ein Photon und ein Neutrino.

Ein äußeres elektromagnetisches Feld, repräsentiert durch die an das W-Boson gekoppelte Wellenlinie, dient dazu, den Beitrag dieser Konfiguration zum magnetischen Moment des Muons auszumessen. Gleichzeitig stehen diese Diagramme für Rechenvorschriften, welche unmittelbar aus der störungstheoretischen Behandlung einer Quantenfeldtheorie folgen.

Die konkrete Formulierung der zugrunde liegenden Algorithmen, also die Erzeugung der Feynmandiagramme und deren Umsetzung in algebraische Ausdrücke, werden heute vielfach mittels Computeralgebra-Systemen (CAS) realisiert. In Präzisionsrechnungen höherer Ordnung der Störungstheorie, wie sie für den Vergleich gegenwärtiger Experimente mit den theoretischen Vorhersagen oft erforderlich sind, treten häufig Tausende von Dia-

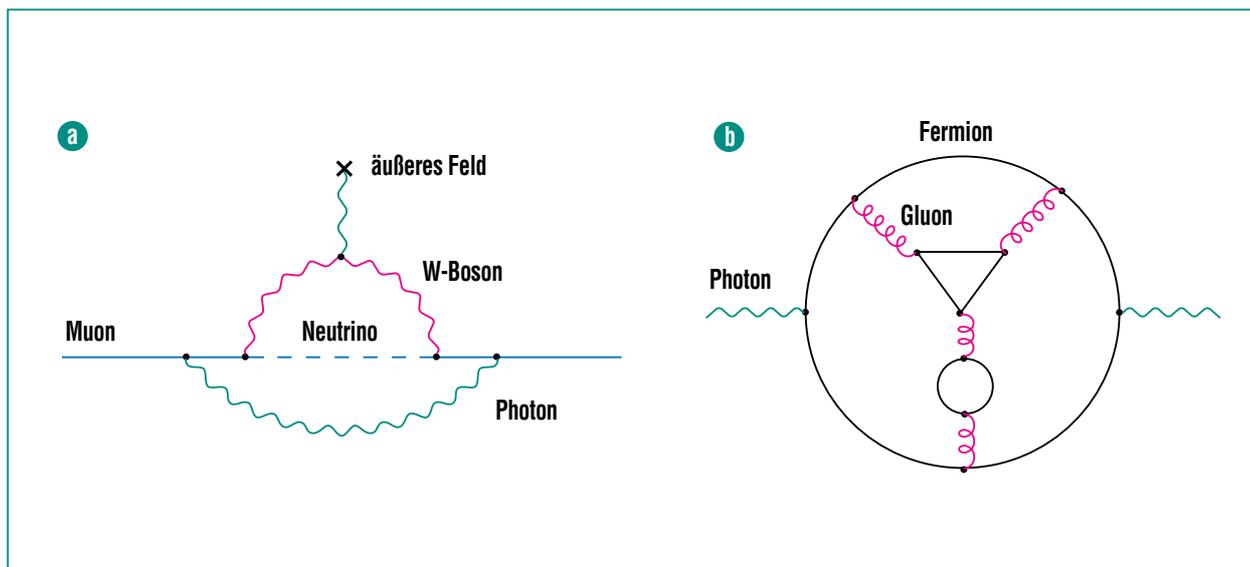


Abb. 1: Beispiele für kurzzeitige Teilchen-Antiteilchen-Erzeugung und -Vernichtung in der Teilchenphysik: (a) Feynmandiagramm, das zum anomalen magnetischen Moment des Muons in Zweischleifen-Näherung beiträgt; (b) Beitrag zur Vakuumpolarisation in Fünfschleifen-Näherung der Quantenchromodynamik. Gluonen sind die Vermittler der starken Wechselwirkung.

grammen auf, von welchen jedes faktisch eine komplizierte Summe multidimensionaler, häufig hochgradig singularer Integrale repräsentiert. Bei deren Auswertung zeigt sich nun die besondere Stärke der computeralgebraischen Behandlung gegenüber numerischen Methoden. Während die direkte numerische Integration bei der großen Zahl von Termen häufig ähnlicher Größe und entgegengesetzten Vorzeichens zwangsläufig scheitert, gelingt es mittels geeigneter in CAS implementierter Algorithmen, die riesige Zahl von Integralen auf wenige Basisintegrale zu reduzieren, welche im Idealfall sogar analytisch oder jedenfalls numerisch berechnet werden können. Die konkreten Probleme und die derzeit verwendeten Methoden bei der Auswertung werden weiter unten noch genauer diskutiert.

Das erste, von Nobelpreisträger Martinus Veltman vor etwa dreißig Jahren für die Teilchenphysik programmierte CAS, genannt SCHOONSCHIP [1], enthält bereits alle auch für neuere Systeme wesentlichen Elemente: es erlaubt die Substitution algebraischer Terme durch andere Ausdrücke und vereinfacht das Resultat selbständig auf eine vorgegebene Normalform. Darüber hinaus sind bereits eine Reihe für die Teilchenphysik wichtiger Algorithmen, wie die Auswertung von Spuren der Gamma-Matrizen, die Manipulation von Lorentz-Indizes und anderes vom Autor implementiert. Mittels vom Anwender vorgegebener Ersetzungsregeln können damit Integration, Differentiation und andere, auch re-

kursiv definierte Algorithmen implementiert werden.

FORM [2], ein neueres CAS, beruht auf ähnlichen Konzepten, erlaubt aber deutlich umfangreichere Rechnungen und ist das derzeit effizienteste CAS. Es wurde speziell für die schnelle Behandlung extrem umfangreicher Ausdrücke konzipiert und kann, da in der Programmiersprache C codiert, in einfacher Weise auf verschiedene Hardware portiert werden.

Mathematica [3], ein kommerziell vertriebenes und das vermutlich am weitesten verbreitete CAS hat seinen Ursprung ebenfalls in der Teilchenphysik, nämlich im Programm SMP, das von S. Wolfram für Anwendungen aus diesem Bereich entwickelt wurde. Im Vergleich zu FORM bietet es einen erheblich größeren Befehlssatz, eine Vielzahl mathematischer Algorithmen – beispielsweise zur Lösung von Gleichungssystemen oder zur analytischen oder numerischen Integration – und insgesamt eine komfortable Nutzeroberfläche. Mathematica wird daher ebenfalls für Rechnungen in der Teilchenphysik eingesetzt. Auf diese, auf Mathematica beruhenden Programmpakete, die sich vorwiegend auf Einschleifenrechnungen beschränken, wird später ebenfalls kurz eingegangen.

Die Berechnung von Feynmandiagrammen in der Quantenchromodynamik, der Theorie der starken Wechselwirkung, ist eines der zentralen Arbeitsgebiete des Instituts für Theoretische Teilchenphysik der Universität Karlsruhe. Derzeit lassen sich gewisse Klassen von Vier- und teilweise sogar

Fünfschleifendiagrammen exakt behandeln. Das in Abb. 1 (b) abgebildete Diagramm repräsentiert die Fluktuation eines virtuellen Photons in masselose Quarks und Gluonen und wurde kürzlich in analytischer Form berechnet. Es liefert einen Beitrag zum Wirkungsquerschnitt bei der Elektron-Positron-Vernichtung in Hadronen. In Verbindung mit der Messung dieses Wirkungsquerschnitts führt die vollständige Berechnung aller Diagramme in dieser Ordnung zur gegenwärtig genauesten und theoretisch bestfundierten Messung der starken Kopplungskonstanten. Die Berechnung auch nur einer Teilklasse von Diagrammen (etwa 60 von insgesamt 3800) erforderte mehrere Wochen Rechenzeit auf einer 8-Prozessor ALPHA-SMP-Maschine [4].

Im Folgenden sollen exemplarisch zwei ausgewählte Programmpakete kurz beschrieben und Gemeinsamkeiten und Unterschiede erläutert werden. Eine ausführlichere Darstellung findet man in [5] sowie unter der URL <http://www-ttp.physik.uni-karlsruhe.de/Links/algprog.html>. Das erste System besteht aus den drei Paketen FeynArts, FormCalc und LoopTools auf der Grundlage von Mathematica und wurde an den Universitäten Würzburg sowie Karlsruhe entwickelt [6]. Das primäre Einsatzgebiet sind Einschleifen-Korrekturen im Standardmodell der Elementarteilchenphysik und dessen Erweiterungen. Komplementär dazu soll das ebenfalls an der Universität Karlsruhe entwickelte Paket GEFICOM [7] vorgestellt werden. Dieses erlaubt unter Benutzung

der asymptotischen Entwicklung die vollautomatische Berechnung physikalischer Prozesse mit ausgeprägter Hierarchie verschiedener Massenskalen.

FeynArts erzeugt für eine vom Nutzer vorgegebene Reaktion bei vorgegebener Ordnung der Störungstheorie in einem ersten Schritt alle möglichen Topologien von Feynmandiagrammen. In einem zweiten Schritt wird jeder Linie ein physikalisches Feld zugeordnet. Dies geschieht auf der Grundlage des vorher ausgewählten Modells, durch das die Feynman-Regeln festgelegt sind. Das Resultat, eine symbolische Darstellung der einzelnen Diagramme, kann einerseits in grafischer Form dargestellt, andererseits mit FormCalc weiterverarbeitet werden.

Durch algebraische Umformungen mittels FormCalc wird nun das Resultat von FeynArts so verarbeitet, dass es von einem numerischen Programm ausgewertet werden kann. Dies geschieht in zwei Schritten: Zunächst werden diverse algebraische Vereinfachungen vorgenommen und das Resultat als Linearkombination von Basisintegralen dargestellt. Dabei benutzt FormCalc in Zwischenschritten auch das Algebraprogramm FORM, da Mathematica große Ausdrücke nur sehr langsam bearbeitet. In einem zweiten Schritt wird das Resultat in ein Fortran-Programm übergeführt, welches dann mittels des Pakets LoopTools numerisch ausgewertet wird.

Das zweite System, GEFICOM, greift seinerseits auf die Pakete QGRAF, MATAD/MINCER und

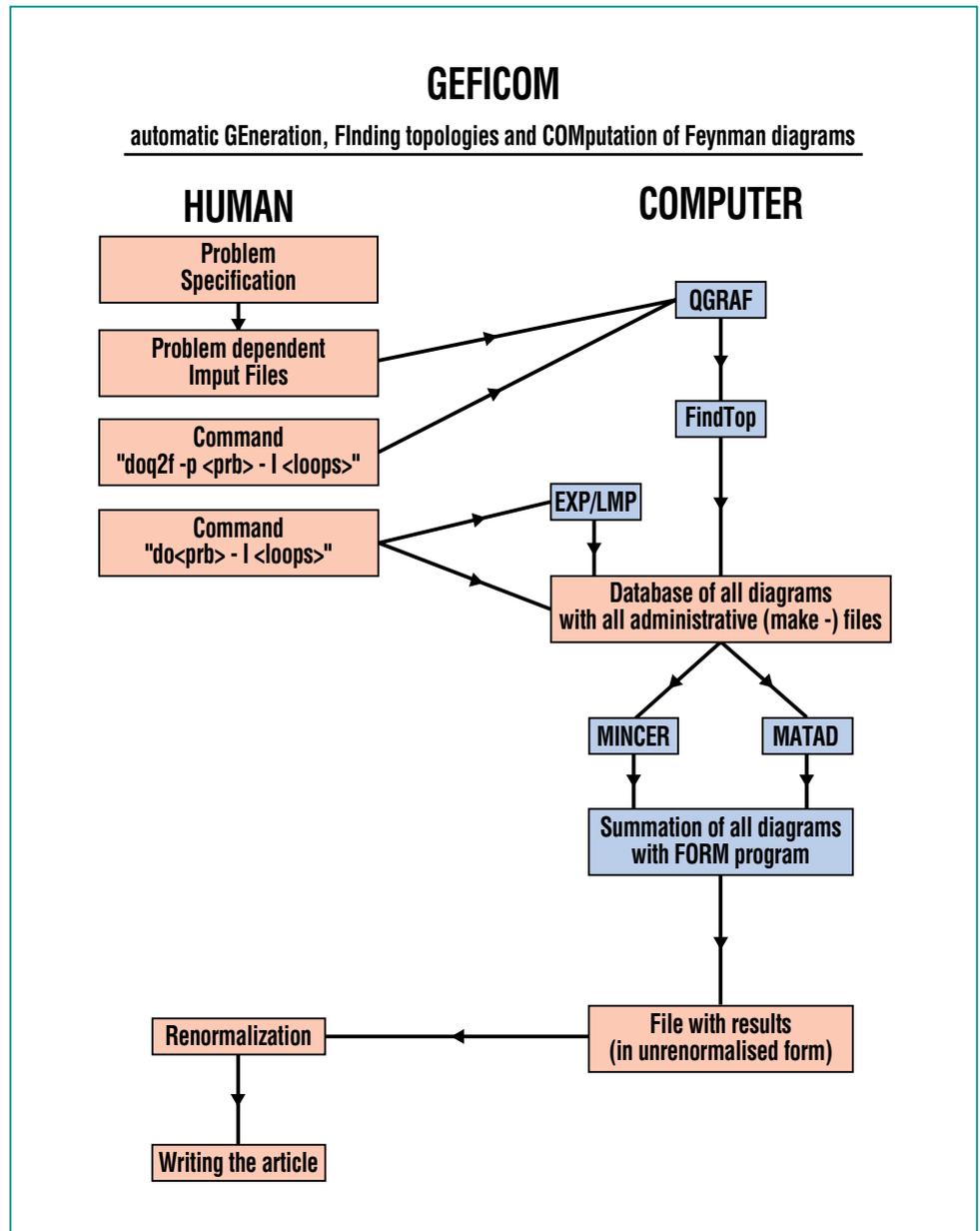


Abb. 2: Flussdiagramm von GEFICOM.

EXP zurück (Abb. 2). Bei vorgegebenem Anfangs- und Endzustand erzeugt und berechnet GEFICOM alle relevanten Diagramme bis zur Dreischleifenordnung in analytischer Form unter der Voraussetzung, dass die verschiedenen Massenskalen hierarchisch geordnet sind. Das Ergebnis ist,

wie oben beschrieben, eine Mehrfachentwicklung in den Verhältnissen der Massenskalen.

Der erste Modul, QGRAF [8], ist in FORTRAN 77 geschrieben und höchst effizient bei der Erzeugung von Feynman-Diagrammen mit drei und mehr Schleifen. Mittels GEFICOM wird anschließend das

Resultat von QGRAF in von MATAD bzw. MINCER lesbare Ausdrücke umgeschrieben.

MINCER [9] berechnet Ein-, Zwei- und Dreischleifenintegrale, wobei alle Linien masselos sein müssen und nur ein äußerer Impuls erlaubt ist. MATAD [10] erfüllt die komplementäre Aufgabe und berechnet massive Integrale ohne äußere Impulse, also ebenfalls Einskalenintegrale. Das Herzstück bei-

der Pakete bildet die Implementierung eines sog. „integration-by-parts“-Algorithmus [11], der jedes auftretende Integral auf die Linearkombination von wenigen, tabellierten Integralen zurückführt, so dass die komplette Rechnung analytisch durchgeführt werden kann.

Mittels der bisher beschriebenen Kombination aus QGRAF und MATAD/MINCER lassen sich Pro-

bleme behandeln, die nur eine Massenskala enthalten. Falls eine Hierarchie verschiedener Massenskalen vorliegt, so kann, nachdem QGRAF die Arbeit beendet hat, das in C++ geschriebene Programm EXP [12] benutzt werden, welches die (evtl. geschachtelte) Anwendung der asymptotischen Entwicklung bei vorgegebener Hierarchie der Massenparameter erlaubt.

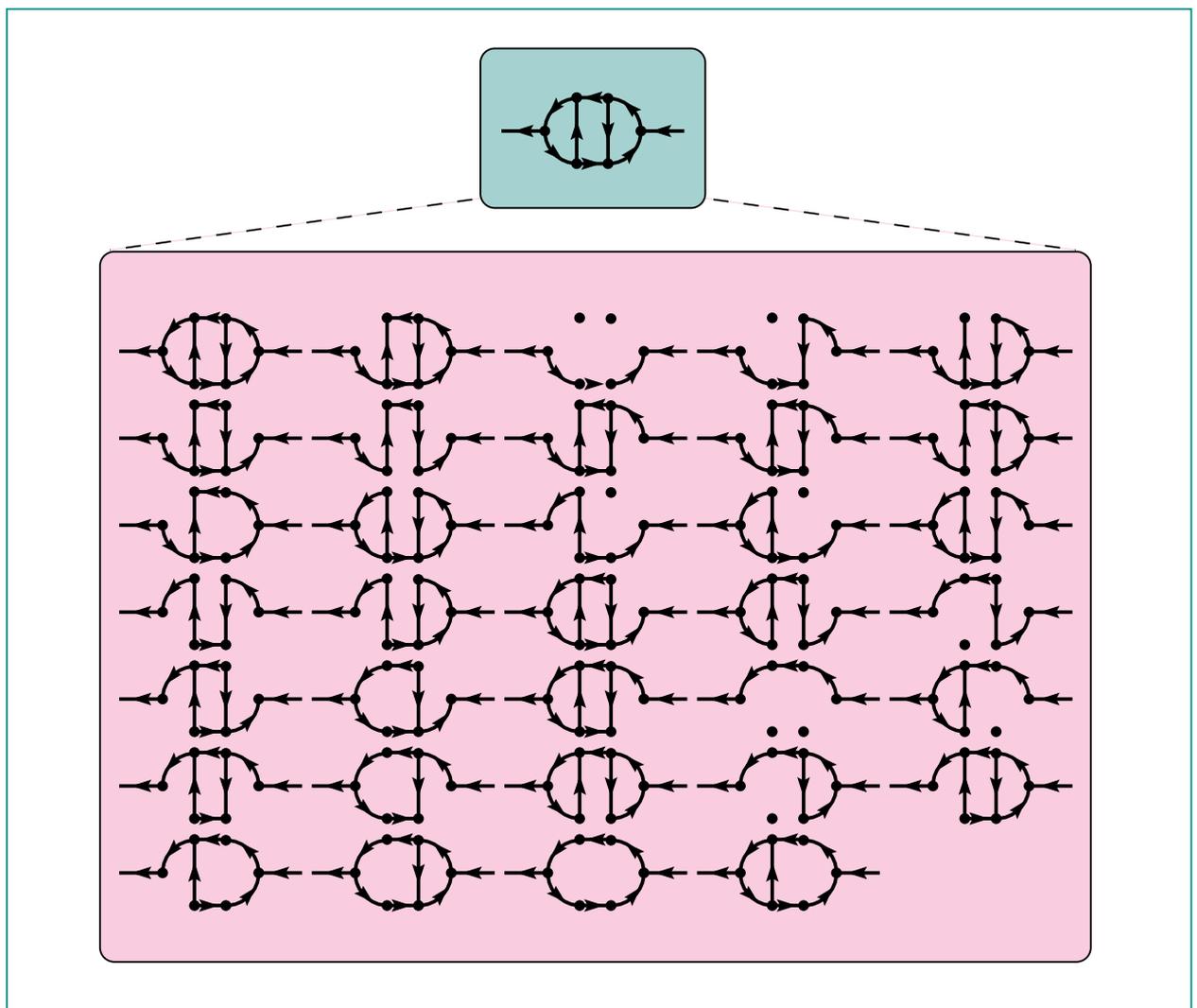


Abb. 3: Diagrammatische Darstellung der asymptotischen Entwicklung des in der ersten Zeile dargestellten Diagramms für den Fall, dass der äußere Impuls viel größer ist als die Masse der virtuellen Teilchen.

Unabhängig von den verwendeten Tools ist in erster Linie die Zahl der Schleifen des Diagramms wesentlich für die Komplexität der jeweiligen Rechnung. Daneben beeinflusst aber auch die Zahl der verschiedenen Massenskalen, von denen der Integrand letztlich abhängt, in erheblichem Maß den Aufwand der Rechnung. Die Resultate werden daher stark vereinfacht, wenn die Massen gegenüber den Impulsen vernachlässigt werden können. Beispielsweise symbolisiert Abb. 1 (b) dann ein sog. Einkalen-Propagatorintegral, das sich als einfacher numerischer Zahlenwert berechnen lässt. In anderen Fällen, in denen man die Effekte relativ kleiner Massen approximativ behandeln will, lassen sich asymptotische Entwicklungen formulieren, die beispielsweise ein Zweiskalenintegral als Summe von Produkten von Einkalenintegralen darstellt (Abb. 3). Solche Entwicklungen lassen sich für Mehrskalenskalenintegrale auch iterativ durchführen.

Der Vorteil liegt in der Darstellung des Integrals durch Reihen mit einfachen, in analytischer Form bekannten Koeffizienten, der Nachteil im exponentiell wachsenden Umfang der Rechnung,

der nur mit optimierter Software und modernster Hardware zu bewältigen ist.

Die automatisierte Berechnung einer derartig großen Zahl komplizierter Integrale, insbesondere die rekursive Reduktion auf Basisintegrale, führt in vielen Fällen auf Problemstellungen an der Grenze gegenwärtiger Hard- und Software. Laufzeiten der Programme von mehreren Wochen auf den schnellsten derzeit verfügbaren Workstations und Speicherplatzbedarf von einigen Hundert Gigabyte bei der Bearbeitung von Zwischenresultaten sind häufig die begrenzenden Faktoren bei der Behandlung aktueller Fragestellungen.

Im Rahmen der an der Universität Karlsruhe angesiedelten DFG-Forschergruppe „Quantenfeldtheorie, Computeralgebra und Monte-Carlo-Simulation“ wurde deshalb PARFORM [13], eine parallelisierte Version des CASFORM entwickelt und erfolgreich für physikalische Anwendungen eingesetzt. Die auf einer 8-Prozessor SMP-Architektur implementierte Version hat sich dabei als besonders nutzerfreundlich herausgestellt.

Zusammenfassung

Die computeralgebraische Behandlung störungstheoretischer Rechnungen hat sich während der letzten beiden Dekaden zum Standardwerkzeug der theoretischen Teilchenphysik entwickelt. Sie entlastet von mühsamen, letztlich langweiligen Rechnungen und erlaubt es, das Schwergewicht auf die Entwicklung effizienter und innovativer Algorithmen und auf die Formulierung physikalischer Fragestellungen zu legen. In Verbindung mit der rapide wachsenden Rechengeschwindigkeit konnten dabei qualitativ neue Probleme aufgegriffen werden, deren Bearbeitung noch vor einer Dekade undenkbar erschien.

Literatur

- [1] M. Veltman, *SCHOONSCHIP, CERN preprint, 1967; H. Strubbe, Comp. Phys. Commun. 8 (1974) 1*
- [2] J.A.M. Vermaseren, *Symbolic Manipulation with FORM, (Computer Algebra Netherlands, Amsterdam, 1991)*
- [3] S. Wolfram, *Mathematica - a system for doing mathematics by computer (Addison-Wesley, New York, 1988)*
- [4] P.A. Baikov, K.G. Chetyrkin, J.H. Kühn, *Phys. Rev. Lett. 88 (2002) 01200.*
- [5] R. Harlander, M. Steinhauser, *Prog. Part. Nucl. Phys. 43 (1999) 167*
- [6] T. Hahn, *Comp. Phys. Commun. 140 (2001) 418, eine zusammenfassende Darstellung findet sich in Nucl. Phys. Proc. Suppl. 89 (2000) 231*
- [7] K.G. Chetyrkin, M. Steinhauser, *unpublished. A description of GEFICOM can also be found in [5]*
- [8] P. Nogueira, *J. Comp. Phys. 105 (1993) 279.*
- [9] S.A. Larin, F.V. Tkachov, J.A.M. Vermaseren, *Rep. No. NIKHEF-/91-18 (Amsterdam, 1991)*
- [10] M. Steinhauser, *Comp. Phys. Commun. 134 (2001) 335*
- [11] F.V. Tkachov, *Phys. Lett. B 100 (1981) 65; K.G. Chetyrkin and F.V. Tkachov, Nucl. Phys. B 192 (1981) 159*
- [12] T. Seidensticker, *„Automatic application of successive asymptotic expansions of Feynman diagrams“ published in New computing techniques in physics research VI - Proceedings of the Sixth International Workshop on Software Engineering, Artificial Intelligence and Expert Systems, 12. - 16. Apr. 1999, Heraklion, Crete*
- [13] D. Fliegner, A. Retey, J.A.M. Vermaseren, *„The Parallelization of the Symbolic Manipulation Program FORM“, published in: Electronic Proceedings of the International Conference on Computing in High-Energy Physics (CHEP98), 31. Aug. - 4. Sep. 1998, Chicago, Illinois, <http://www.hep.net/chep98/PDF/57.pdf>*

Modellierung vom Feinsten

N. Riemer, H. Vogel, IMK

Beschreibung der Gesamtproblematik

In den letzten Jahren wurden große Anstrengungen unternommen, die Ausbreitung und die Umwandlung von gasförmigen Luftbeimengungen zu erforschen. Zur Beschreibung der in der Atmosphäre ablaufenden Prozesse wurden Chemie-Transport-Modelle, wie das am Institut für Meteorologie und Klimaforschung (IMK) betriebene Modellsystem KAMM/DRAIS, entwickelt. Diese haben mittlerweile einen Stand erreicht, der – analog zur täglichen Wettervorhersage – unter anderem eine Prognose der alljährlich wiederkehrenden Sommersmog-episoden ermöglicht.

Neben den gasförmigen Luftbeimengungen sind in der Atmosphäre jedoch auch partikelförmige Luftbeimengungen, sogenannte Aerosolpartikel, vorhanden. Der Wissensstand bezüglich der Aerosolphase weist, verglichen mit dem der Gasphase, wesentlich größere Defizite auf. Dabei stehen gerade die Aerosolpartikel im Verdacht, erhebliche Gesundheitsbeeinträchtigungen zu verursachen. Unklar ist im Moment noch, ob der entscheidende Parameter, der zur Schädigung führt, die Konzentration der Partikel, die Größe der Partikel oder die Zusammensetzung der Partikel ist. Aufgrund der möglichen Gesundheitsgefährdung wurden von der Europäischen Kommission neue Grenzwerte für die Partikelmasse verabschiedet, die deutlich unter denen bislang in Deutschland geltenden Werten liegen. Tatsächlich werden aber auch die letzteren in

vielen Gebieten Deutschlands nicht eingehalten.

Neben ihrer Wirkung auf die Gesundheit des Menschen greifen Aerosole aber auch in vielfältiger Weise in das System Atmosphäre ein und stellen einen entscheidenden Klimafaktor dar. Ihr Einfluss auf den Strahlungshaushalt kann von jedermann als reduzierte Sichtweite wahrgenommen werden. Strahlung wird absorbiert, kurzwellige Sonnenstrahlung wird zurückgestreut. Diese Vorgänge werden als direkter Klimaantrieb bezeichnet. Weiterhin sind Aerosolpartikel bei den Prozessen der Wolkenbildung wichtig, wodurch sie das Klima indirekt beeinflussen.

Während Spurengase durch einen Parameter, nämlich die Konzentration, charakterisiert werden können, benötigt die Beschreibung des atmosphärischen Aero-

sols mehrere Parameter. So spielen, außer der chemischen Zusammensetzung, die Größenverteilung, die Morphologie, der Phasenstatus und der Mischungszustand des Aerosols eine entscheidende Rolle für sein physikalisches und chemisches Verhalten. In die Atmosphäre gelangen Aerosole auf zwei Wegen. Die sogenannten Primärpartikel werden direkt in die Atmosphäre emittiert. Zu dieser Kategorie gehören Mineralstaub, Seesalz, aber auch Partikel, die infolge von Verbrennungsprozessen entstehen. Sekundärpartikel dagegen werden aus Vorläufersubstanzen in der Gasphase gebildet. Diese unterschiedlichen Bildungsprozesse haben zur Folge, dass sich die Durchmesser der atmosphärischen Aerosolpartikel über mehrere Größenordnungen erstrecken (Abb. 1). Die Nukleation aus der Gasphase führt zu sehr

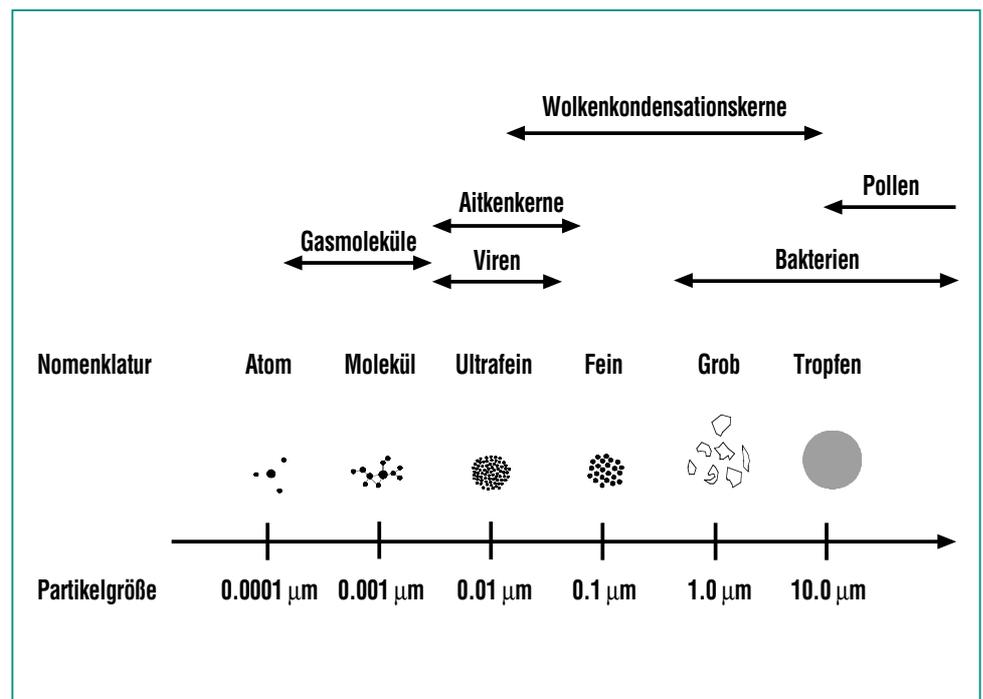


Abb. 1: Größeneinteilung des atmosphärischen Aerosols.

kleinen Partikeln, deren Durchmesser im Nanometerbereich liegt, während Dispersionspartikel eher im Bereich von 0.1 - 1 μm zu finden sind.

Der Zustand und die Dynamik der Atmosphäre beeinflussen ganz entscheidend den Eintrag und die Ausbreitung von primären Aerosolpartikeln und Aerosol-Vorläufergasen einschließlich deren Umwandlung in sekundäre Aerosole. Beeinflusst werden auch die Größe und Zusammensetzung der Partikel durch physikalische

und chemische Alterungsprozesse, sowie schließlich das Entfernen der gealterten Partikel aus der Atmosphäre durch trockene und nasse Deposition. So haben z.B. die Windstärke und der Turbulenzzustand der Atmosphäre in Bodennähe einen erheblichen Einfluss auf die Quellstärke und die Ausbreitung von Primärpartikeln aus Reifenabrieb oder aus der Bau- und Landwirtschaft. Gleiches gilt für biogene Primärpartikel wie Bakterien, Sporen und Pollen. Der Eintrag von natürlichem Wüstenstaub wird in Mitteleuropa

durch bestimmte großräumige Transportprozesse ermöglicht. Ob bei der Oxidation von Aerosol-Vorläufergasen neue Partikel entstehen oder lediglich die Masse bereits vorhandener Partikel durch Kondensation zunimmt, hängt von den Konzentrationen der Vorläufergase und der spezifischen Oberfläche des bereits vorhandenen Aerosols und damit ebenfalls vom Zustand der Atmosphäre und seiner Vorgeschichte ab. Abbildung 2 fasst alle wichtigen Prozesse schematisch zusammen, die zusätzlich zu

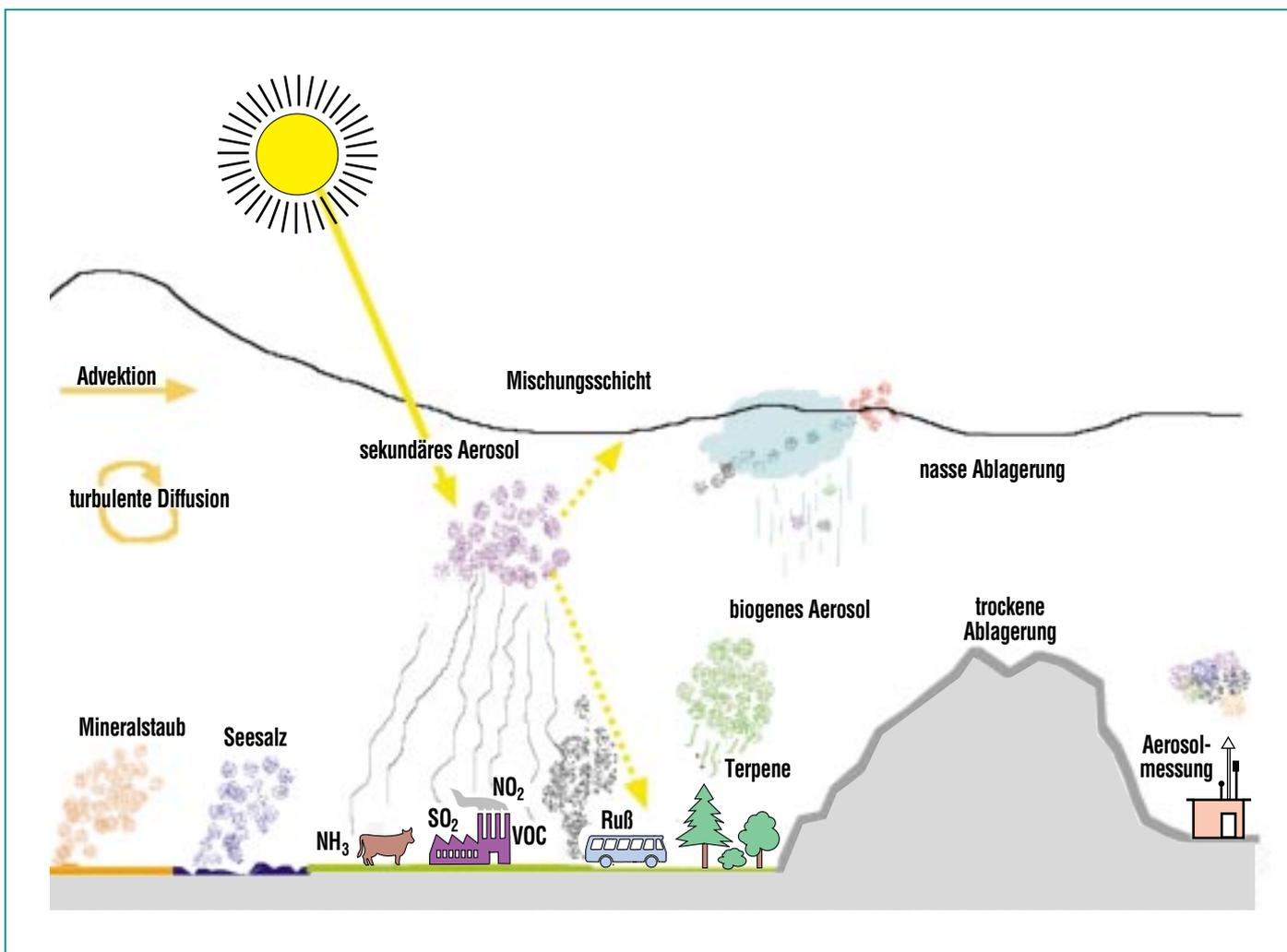


Abb. 2: Aerosoltypen und ablaufende Prozesse in der Atmosphäre.

den direkt kontrollierbaren anthropogenen Emissionen die Menge, die Größenverteilung und die chemische Zusammensetzung des Aerosols bestimmen und zwar dort, wo es auf Mensch und Umwelt einwirkt.

Die dargestellten Prozesse verlaufen räumlich sehr inhomogen und auf unterschiedlichen Zeitskalen. Sie sind zudem stark witterungsabhängig. Ein Problem ergibt sich aus der Forderung, die tatsächliche Belastung durch menschliche Aktivitäten zu erfassen. Allein durch Messungen lässt sich diese Frage nur unzureichend beantworten. Noch schwieriger zu beurteilen ist die Wirksamkeit von emissionsmindernden Maßnahmen zum Schutze der Bevölkerung vor Grenzwertüberschreitungen.

Zur Klärung dieser Fragen muss auf komplexe dreidimensionale Aerosol-Chemie-Transport-Modelle zurückgegriffen werden. Für gegebene Emissionen lassen sich durch numerische Simulation die in Abb. 2 zusammengefassten Transport- und Transformationsprozesse einschließlich ihrer zeitlichen Änderung detailliert darstellen. Damit wird eine wichtige Grundlage geschaffen, um die Wirkung von Aerosolemissionen und die Auswirkung emissionsmindernder Maßnahmen auf die Bevölkerung zu bewerten. Gleiches gilt für andere unerwünschte Wirkungen des Aerosols wie z.B. die Minderung der Sichtweite oder die Modifikation des Klimas. Die wesentlichen Ziele, die mit der Modellierung verfolgt werden, sind deshalb:

- die Quantifizierung der beteiligten Prozesse,
- die Quantifizierung des Einflusses der partikelförmigen Luftbeimengungen auf den physikalischen und chemischen Zustand der Atmosphäre,
- die Quantifizierung des anthropogenen Beitrages zur Gesamtbelastung,
- die Prognose des räumlichen und zeitlichen Verhaltens partikelförmiger Luftbeimengungen analog zur täglichen Wettervorhersage.

Beschreibung eines erweiterten Chemie-Transport-Modells

Das am IMK vorhandene mesoskalige Chemie-Transport Modellsystem KAMM/DRAIS [1] wurde erweitert, um die aerosoldynamischen Prozesse zu berechnen. Das Modell kommt hauptsächlich in der regionalen Skala zum Einsatz. Typische Gebietsgrößen haben eine horizontale Ausdehnung von 200 x 200 km². Die horizontalen Maschenweiten liegen zwi-

schen 500 m und 4 km. In der Vertikalen arbeitet das Modell mit 25 Schichten. In Bodennähe beträgt die vertikale Auflösung 20 m, am Oberrand des Modells in 8 km Höhe beträgt die vertikale Maschenweite 400 m. Die verwendeten Zeitschritte bewegen sich im Sekundenbereich.

Die Behandlung der Aerosole erfolgt mit dem Modell MADEsoot [2, 3, 4]. Es hat modalen Charakter, das bedeutet, die Partikelverteilungen werden durch überlagerte Log-Normalverteilungen beschrieben. Die Partikelneubildung durch Nukleation von Schwefelsäuredampf wird in stark vereinfachter und parametrisierter Form behandelt. Bis auf den Dieselruß liegen die Bestandteile der Partikel (Nitrat, Sulfat, Ammonium, organische Substanzen) stets als interne Mischung vor. Dies bedeutet, dass ein Partikel aus mehreren Substanzen besteht und die Zusammensetzung nicht von der Größe abhängt. Im Gegensatz dazu bestehen bei einer externen Mischung die einzelnen Partikel jeweils aus einer reinen Substanz. Abbildung 3 illustriert die beiden Mischungszustände. Frisch emit-

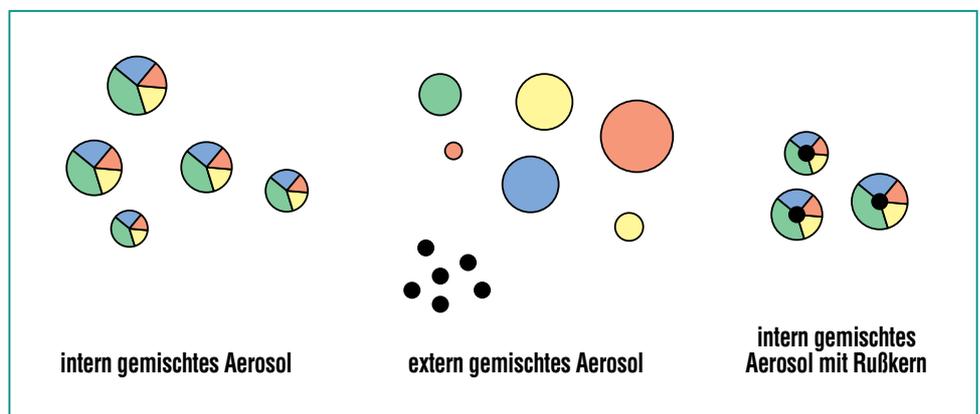


Abb. 3: Verschiedene Aerosolmischungstypen. Die Farben stehen für unterschiedliche Aerosolinhaltsstoffe.

tierter Dieselruß ist zunächst extern gemischt, kann jedoch durch Koagulation und Aufkondensieren von Schwefelsäuredampf in die interne Mischung übergeführt werden.

Zusätzlich zu den Differentialgleichungen, die zur Beschreibung des atmosphärischen Zustandes erforderlich sind und den Gleichungen für die gasförmigen Substanzen müssen nun weitere Differentialgleichungen für die verschiedenen Aerosolkomponenten gelöst werden, was den Rechenaufwand weiter deutlich erhöht. Um mit der Rechenzeit unter der Realzeit zu liegen, was für Prognosesimulationen unabdingbar ist, wurde das Programm in Zusammenarbeit mit HIK hinsichtlich der Vektorisierbarkeit deutlich optimiert. Im Folgenden soll exemplarisch dargestellt werden, für welche Fragestellungen

das Modellsystem eingesetzt werden kann.

Ergebnisse

Als Ausgangsfall für die Untersuchungen dient ein typischer Sommertag, bei dem das Modellgebiet, welches große Teile von Baden-Württemberg, das Elsass, Teile der Nordschweiz, sowie von Reinland-Pfalz und dem Saarland abdeckt, großräumig aus Osten angeströmt wird. An anthropogenen Emissionen werden dem Modell flächendeckende Werte für SO_2 , NO_x , VOC, NH_3 und Dieselruß vorgegeben. Biogene VOC- und NO -Emissionen werden simultan vom Modell in Abhängigkeit von Landnutzung, Temperatur und Strahlung berechnet.

Abbildung 4 zeigt simulierte Horizontalverteilungen des extern gemischten Rußes und der Gesamt-

partikelmasse in Bodennähe. Während der extern gemischte Ruß noch die Emissionsstruktur erkennen lässt, zeigt die Verteilung der Gesamtpartikelmasse keinen Zusammenhang mit der Lage der Quellen. Dieses Ergebnis unterstreicht die Notwendigkeit des Einsatzes numerischer Simulationsmodelle, wenn es darum geht, den Erfolg von Emissionsminderungsmaßnahmen zur Reduktion der Belastung der Atmosphäre mit Aerosolpartikeln abzuwägen. Des weiteren zeigt die Abbildung, dass es im Rheintal infolge der Überlagerung der Abluftfahnen der verschiedenen städtischen Gebiete zu einer Anreicherung der Partikelmasse kommt. Neben der Massenkonzentration ist auch die Zusammensetzung der Aerosole von Bedeutung. Abbildung 5 zeigt für einen Punkt in einem Stadtgebiet den berechneten zeitlichen Ver-

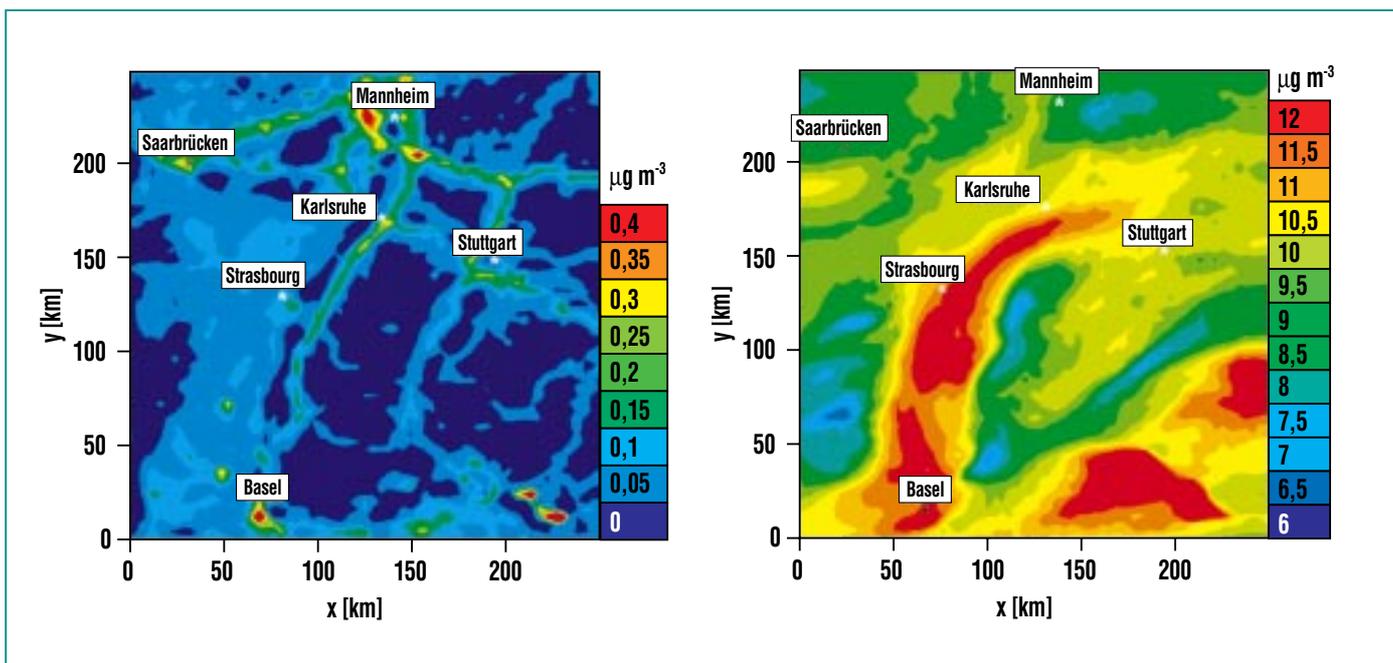


Abb. 4: Simulierte Horizontalverteilung des extern gemischten Rußes (links) und der Gesamtpartikelmasse (rechts) in 20 m über Grund um 14 Uhr.

lauf verschiedener Aerosolkomponenten. Hierbei wird deutlich, dass es aufgrund der schon angesprochenen Variabilität sowohl des atmosphärischen Zustands als auch der Emissionen zu deutlichen Unterschieden im Tagesverlauf der einzelnen Komponenten kommt. Der hohen zeitlichen Auflösung, die mit dem Modell erreicht wird, stehen auf der Messseite noch recht lange Probenahmezeiten von mehreren Stunden gegenüber. Hier besteht daher auch messtechnisch gesehen ein erheblicher Forschungsbedarf.

Basierend auf den simulierten dreidimensionalen Aerosolverteilungen können die Auswirkungen auf die atmosphärische Strahlung berechnet werden. Exemplarisch zeigt Abb. 6 die berechnete Sichtweite für die Station Hornisgrinde in den Morgen- und Mittagsstunden. In die Berechnung fließt die Zusammensetzung der Partikel und ihre Größenverteilung ein. Es zeigt sich, dass die hohen Aerosolmassenkonzentrationen im Rheingraben, wie in Abb. 3 dargestellt, in Kombination mit der tageszeitlichen Grenzschichtentwicklung zu einer deutlichen Reduktion der Sichtweite in den Mittagsstunden führen. Neben diesem unmittelbar beobachtbaren Einfluss der Aerosole auf die Strahlung wird durch Absorption und Streuung auch der Strahlungshaushalt modifiziert. Sensitivitätsstudien haben ergeben, dass für die untersuchte Sommersituation die an der Erdoberfläche ankommende solare Strahlung aufgrund der vorhandenen Aerosole um 25 W m^{-2} verringert wird. Dabei sind 30 % des Effektes auf die Wirkung der Rußparti-

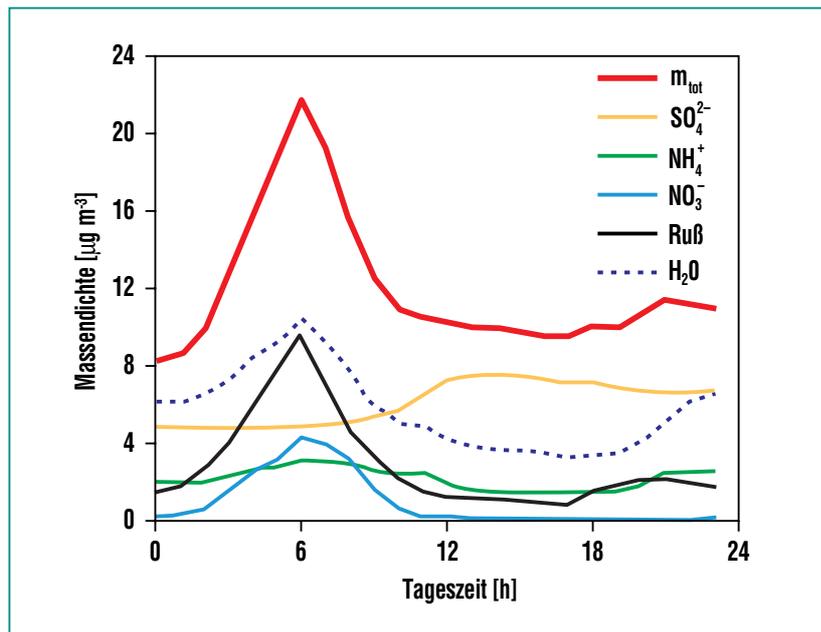


Abb. 5: Tagesgänge der Massendichten ausgewählter Substanzen und der trockenen Gesamtmassendichte für einen Gitterpunkt bei Mannheim, 20 m über Grund.

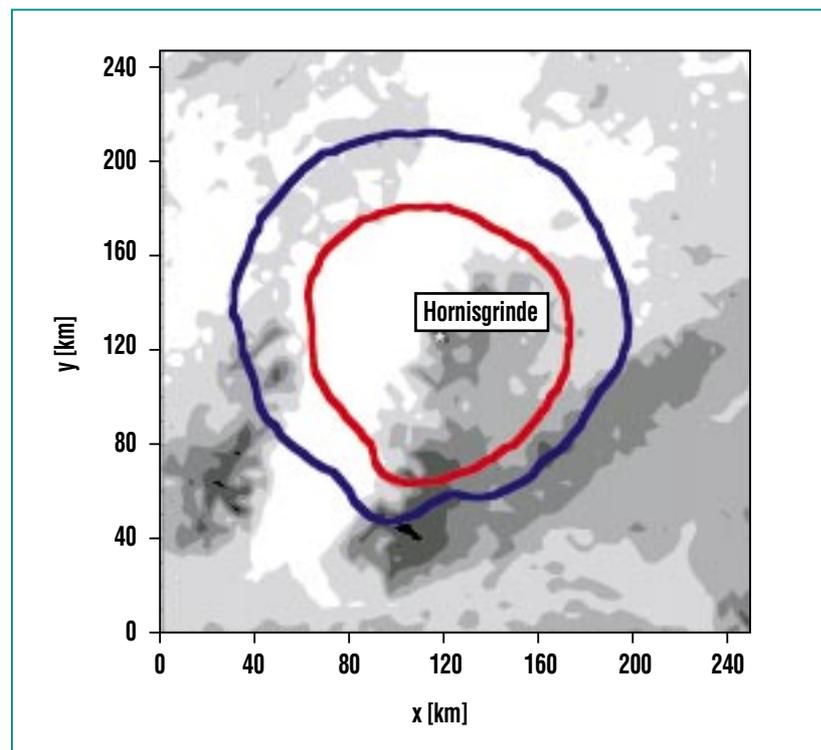


Abb. 6: Berechnete Sichtrosen unter Berücksichtigung der Aerosolverteilung für die Station Hornisgrinde um 6:00 MEZ (blau) und 14:00 MEZ (rot).

kel zurückzuführen, obgleich diese nur eben 10 % der Gesamtmasse ausmachen.

Zusammenfassung und Ausblick

Dreidimensionale Aerosol-Chemie-Transportmodelle wie das Modellsystem KAMM/DRAIS sind ein wichtiges Instrument zur Analyse der Aerosolbefrachtung der Atmosphäre und ihrer Wirkung.

Neben der Analyse des Istzustands kann durch die Betrachtung unterschiedlicher Emissionsszenarien der Beitrag des natürlichen Aerosols an der Belastung der Atmosphäre und der voraussichtliche Erfolg von Minderungsstrategien untersucht werden. Daneben ist auch die Bestimmung von regionalen Beiträgen zur Belastung der globalen Atmosphäre und deren Folgen von Bedeutung.

Infolge der Entwicklung immer leistungsfähigerer Computer und verbesserter Algorithmen wird es zukünftig möglich sein, längere Zeiträume und größere Regionen zu simulieren und somit diese Vorgänge noch besser zu prognostizieren.

Literatur

- [1] B. Vogel, F. Fiedler, H. Vogel, *J. Geophys. Res.*, 100, 22907-22928, 1995
- [2] I.J. Ackermann, H. Hass, M. Memmesheimer, A. Ebel, F.S. Binkowski, U. Shankar, *Atmos. Environ.* 32, 2981-2999, 1998
- [3] B. Schell, I.J. Ackermann, H. Hass, F.S. Binkowski, A. Ebel, *J. Geophys. Res.*, 106, 28275-28293, 2001
- [4] N. Riemer, *Wiss. Ber. Inst. Meteorol. Klimaforsch. der Univ. Karlsruhe*, 29, 2002

Das Karlsruher Simulationsmodell der mittleren Atmosphäre – Modellbeschreibung und Anwendungen

W. Kouker, T. Reddmann, R. Ruhnke, IMK

Einführung

Trotz internationaler Anstrengungen zur Begrenzung von Ozon zerstörenden Substanzen muss befürchtet werden, dass sich die Erholung der Ozonschicht in der Stratosphäre wegen der Kopplung an die globale Klimaänderung erheblich verzögern wird. Als eine langfristige Aufgabe des Klimaschutzes muss daher das wesentliche Ziel der angewandten Stratosphärenforschung sein, die in der Stratosphäre ablaufenden Prozesse soweit zu verstehen, dass auch unter geänderten Bedingungen, wie sie der Treibhauseffekt in der Atmosphäre hervorrufen wird, eine zuverlässige Vorhersage des Zustandes der Stratosphäre möglich ist. Wegen der außerordentlichen Komplexität des Systems Atmosphäre kann sich dieser Aufgabe nur im Zusammenspiel von detaillierten Beobachtungen und aufwendigen, möglichst alle relevanten Prozesse beinhaltenden Modellsimulationen genähert werden. Die Ergebnisse solcher Studien müssen dann in parametrisierter Form in aufwendige Klimamodelle integriert werden.

Im IMK steht mit dem Karlsruher Simulationsmodell der Mittleren Atmosphäre (KASIMA) eine solche Modellumgebung zur Verfügung. KASIMA wird für die Interpretation von bei unterschiedlichen Messkampagnen erhaltenen Beobachtungsdaten eingesetzt, wobei im Vordergrund Messungen des IMK stehen, die mittels verschiedener Fernerkundungsmethoden gewonnen wurden (siehe nächste Ausgabe der Nachrichten).

Modellbeschreibung

KASIMA ist ein globales atmosphärisches Modell, dessen Anwendbarkeit sich auf die mittlere Atmosphäre, d.h. die Stratosphäre, Mesosphäre und untere Thermosphäre, zwischen etwa 10 und 100 km begrenzt. Zur Beschreibung troposphärischer Prozesse (< 10 km) fehlt die Behandlung der Feuchte und die damit verbundenen Umsetzungen latenter Wärme, sowie wichtige Parametrisierungen, z.B. der planetarischen Grenzschicht und der Konvektion [1].

Die horizontale Auflösung des Modells beträgt etwa 3°, vertikal 1-5 km. Da horizontale Schall- und Schwerwellen Lösung des meteorologischen Gleichungssystems sind, ist ein Zeitschritt von etwa 10 Minuten erforderlich.

Die Meteorologie

Die zentrale Komponente von KASIMA ist ein numerischer Löser des meteorologischen Grundgleichungssystems (prognostisches Modell). Das sind Bilanzgleichungen für die Horizontalkomponen-

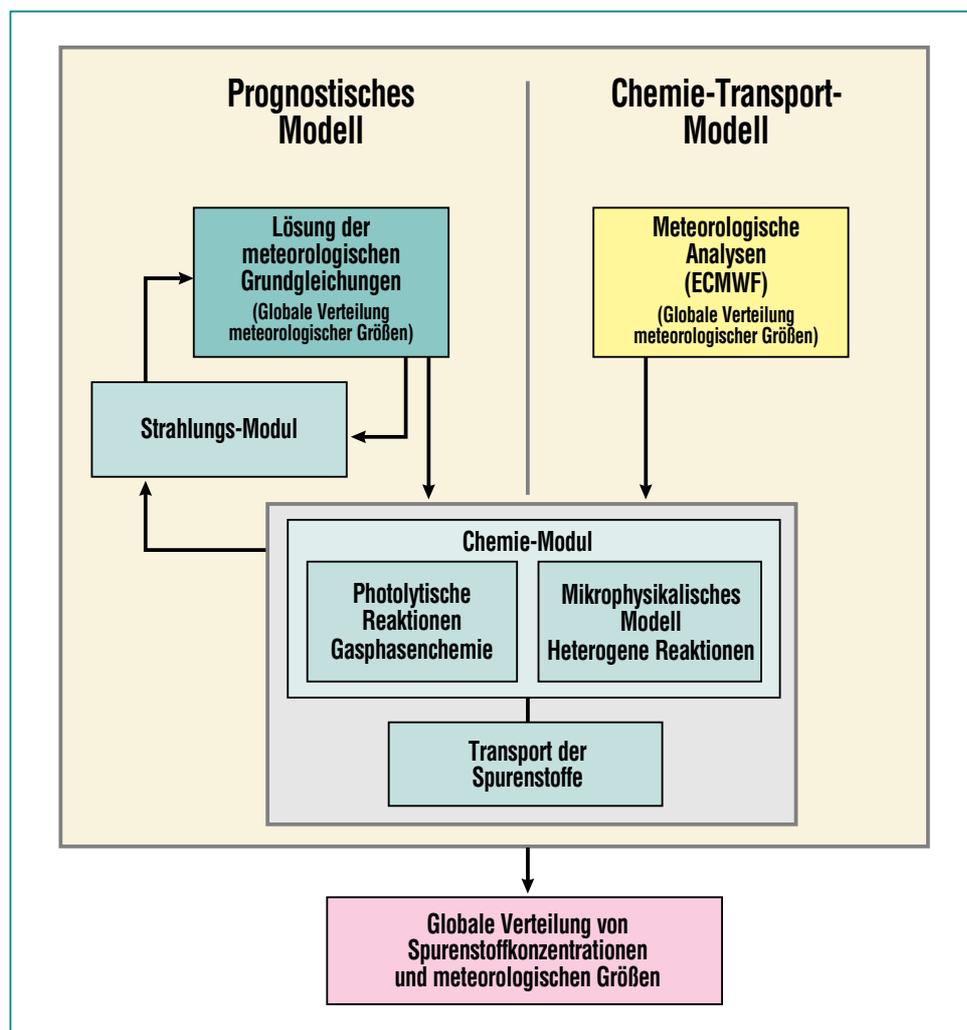


Abb. 1: Schematische Darstellung der in KASIMA enthaltenen Module.

ten des Impulses, der Luftmasse, und Wärme (1. Hauptsatz). Die Vereinfachungen gegenüber dem Navier-Stokeschen Gleichungssystem sind im Wesentlichen die Approximation der Vertikalkomponente der Impulsbilanz durch die Hydrostatik, das ist die Balance von Auftriebs- und Schwerkraft, sowie die Nichtberücksichtigung der molekularen Viskosität.

Die horizontale Abhängigkeit der Modellvariablen wird durch eine endliche Reihe von Koeffizienten der Kugelflächenfunktionen, die assoziierten Légendre Polynome, dargestellt. Die Vertikalordinate ist eine variable Funktion des Luftdruckes. Die nichtlinearen Terme zur Berechnung der Impuls- und Wärmeflüsse werden über die spektrale Transformationsmethode berechnet. Probleme numerischer nichtlinearer Instabilität (Aliasing) durch das Auftreten von Koeffizienten von kleineren räumlichen Skalen werden durch Nichtberücksichtigung verhindert [2].

Das Modell wird angetrieben durch die Nettoerwärmungsrate (siehe Strahlung) sowie durch vorgegebene Randbedingungen. Die untere Randbedingung sind aus Analysen des europäischen Zentrums für mittelfristige Wettervorhersage (ECMWF) entnommene Felder für das Geopotential (die Höhe der Druckfläche) zur Integration der hydrostatischen Gleichung, sowie Temperatur und Horizontalwindkomponente für die Vertikalflüsse.

Die Wechselwirkung von kurzwelligen (ca. 100 km Wellenlänge) Schwerewellen, die explizit nicht mehr aufgelöst werden, mit der

großräumigen Strömung werden in parametrisierter Form berücksichtigt.

Die Strahlung

Der primäre Antrieb atmosphärischer Zirkulation ist die Absorption von Sonnenstrahlung in den Luftschichten und am Erdboden. Da die Erde als Ganzes und im zeitlichen Mittel gesehen im thermischen Gleichgewicht ist, muss die als im wesentlichen durch sichtbares Licht zugeführte Energie wieder in den Weltraum abgegeben werden. Dies kann ebenfalls nur in Form von Strahlung erfolgen, und zwar entsprechend der Temperatur der Atmosphäre bzw. des Erdbodens im infraroten Spektralbereich, also durch Wärmestrahlung.

In der Stratosphäre erfolgt die Absorption ultra-violetten Lichtes (Heizung) durch Elektronenübergänge hauptsächlich am Ozon, in der Mesosphäre am molekularen Sauerstoff. Die Emission (Kühlung) erfolgt im wesentlichen in den Schwingungs- und Rotationsbanden der Moleküle CO_2 , O_3 und H_2O . Der rechnerische Aufwand für den Strahlungstransport in der Atmosphäre ist im allgemeinen so hoch, dass Parametrisierungen für die Berechnung von Heiz- und Kühlraten benutzt werden müssen. Über Ozon (und in geringerem Maße auch über den Wasserdampf) sind der chemische und der dynamische Zustand der Atmosphäre über die Strahlung gekoppelt [3].

Die Chemie

Die in KASIMA integrierte Chemie berücksichtigt die für die Simula-

tion der Stratosphäre wichtigen chemischen Prozesse. Dazu gehören Reaktionen in der Gasphase und auf flüssigen oder festen Aerosolen sowie die Photodissoziationen (die durch Sonnenlicht induzierte Spaltung) der Spurenstoffe in der Stratosphäre. Neben diesen insgesamt 165 Reaktionen, an denen insgesamt 49 Substanzen teilnehmen, werden auch mikrophysikalische Prozesse in parametrisierter Form berücksichtigt. Diese Parametrisierungen enthalten die Aufnahme von Spurenstoffen (z.B. HNO_3) in flüssige Aerosole oder die Bildung von polaren stratosphärischen Wolken, die ab einer gewissen Größe sedimentieren.

Anwendungen

Die KASIMA-Modellumgebung kann für eine Reihe von Problemen eingesetzt werden. Im Folgenden werden einige Beispiele aufgezeigt.

Die chemischen Lebenszeiten von Spurenstoffen in der Stratosphäre reichen bis hin zu Jahren. Eine wesentliche Voraussetzung für die erfolgreiche Modellierung der Spurengaskonzentrationen in der Stratosphäre ist daher eine möglichst realistische Wiedergabe des Transportes von Luftmassen. Die notwendige Verifikation der Transporteigenschaften des numerischen Modells kann durch detaillierte Vergleiche von Modelldaten mit Beobachtungen langlebiger Spurengase (Tracer) untersucht werden. Zwei Beispiele solcher Untersuchungen werden im Folgenden herausgegriffen (Streamer und Mittleres Alter).

Streamer

Der Eintrag von Spurengasen von der Troposphäre in die Stratosphäre erfolgt im wesentlichen in den Tropen. Hierdurch erhält man für die Quellgase in der Stratosphäre eine typische Verteilung mit maximaler Konzentration in den Tropen und Abnahme zu den Polen.

Durch großskalige planetarische Wellen, besonders im Winter, werden Scherströmungen hervorgerufen, durch die tropische Luft innerhalb weniger Tage in Form langgezogener Zungen, sogenannte Streamer, in mittlere und höhere Breiten transportiert wird.

Ein solcher Prozess wurde im November 1994 mit dem räumlich hochauflösenden gekühlten Spektrometer CRISTA der Universität Wuppertal beobachtet und mit KASIMA anhand der Ertelschen Potenziellen Vorticity (EPV), einem Analogon der Hydrodynamik zum Drehimpuls starrer Körper, analysiert [6].

Abb. 2 zeigt eine Zeitreihe der EPV, mit der Streamerstruktur über dem Nordatlantik und Sibirien (Linien). Deutlich erkennbar sind die starken Änderungen des Gradienten der EPV durch Advektion (farbig) im Bereich der Streamerentstehung von Südwesten nach Nordosten.

Mittleres Alter stratosphärischer Luft

Auf Zeitskalen von Jahren werden die Substanzen in der gesamten Stratosphäre verteilt. Dies kann durch das sogenannte (mittlere) Alter stratosphäri-

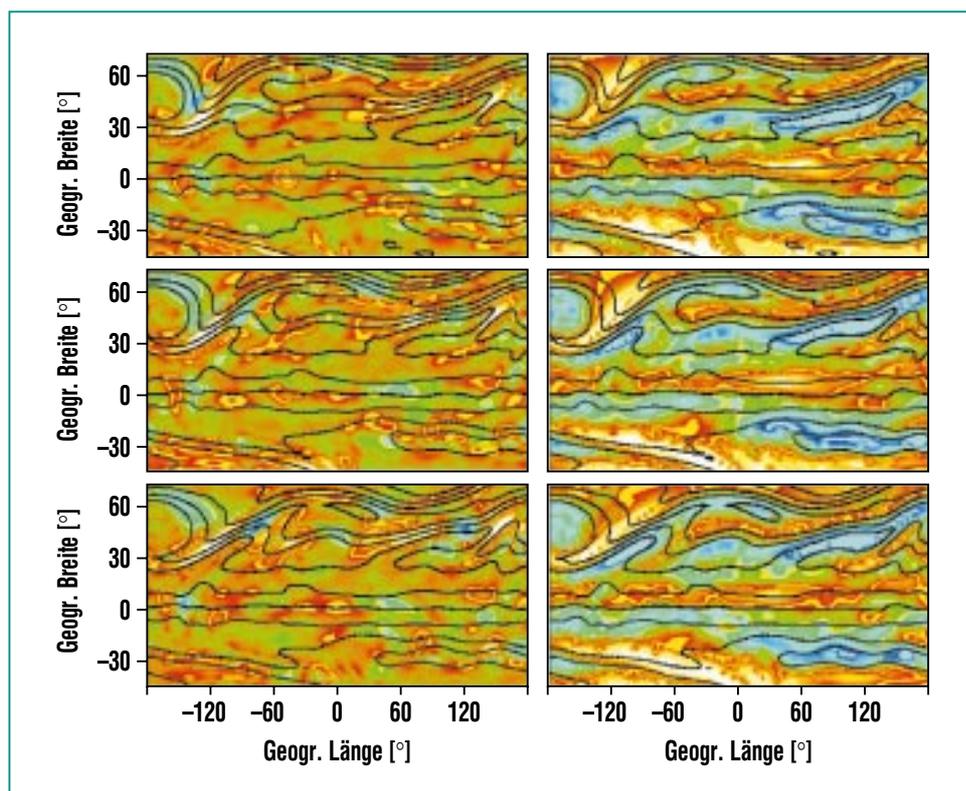


Abb. 2: Ertels Potenzielle Vorticity (EPV, Linien), die Änderung des Betrages (farbig links) und die Winkelgeschwindigkeit (rechts) ihres Gradienten für den 5.-7. November 1994 (von oben nach unten). Die Streamer sind an den Linien konstanter EPV erkennbar.

scher Luft charakterisiert werden, die die Zeit angibt, wie lange im Mittel ein Luftpaket von der tropischen Tropopause zu einem Punkt der Stratosphäre benötigt hat. Mittels eines stabilen Gases, wie z.B. dem Gas SF_6 , dessen Konzentration in der Troposphäre beständig wächst, kann diese zunächst theoretische Größe experimentell über die Differenz der Konzentration am Ort der Messung und an der tropischen Tropopause bestimmt und mit Modellsimulationen verglichen werden. Da aber auch sehr inerte Gase in den oberen Luftschichten (größer etwa 60 km) von der harten UV-Strahlung und Ionenreaktionen angegriffen werden, wur-

den für die teilweise unsichere Chemie dieses Gases verschiedene Szenarien gerechnet, siehe Abb. 3, und mit den Beobachtungen verglichen. Die Vergleiche deuten auf eine wesentlich längere chemische Lebenszeit hin, als bisher angenommen wurde. Dies ist auch deshalb wichtig, weil SF_6 eine um Größenordnungen stärkere spezifische Treibhauswirkung als CO_2 besitzt und damit trotz viel geringerer Konzentration langfristig zum anthropogenen Treibhauseffekt beiträgt [7].

Antarktisches Ozonloch

Ein weiteres Beispiel gibt einen Einblick, inwieweit mit KASIMA

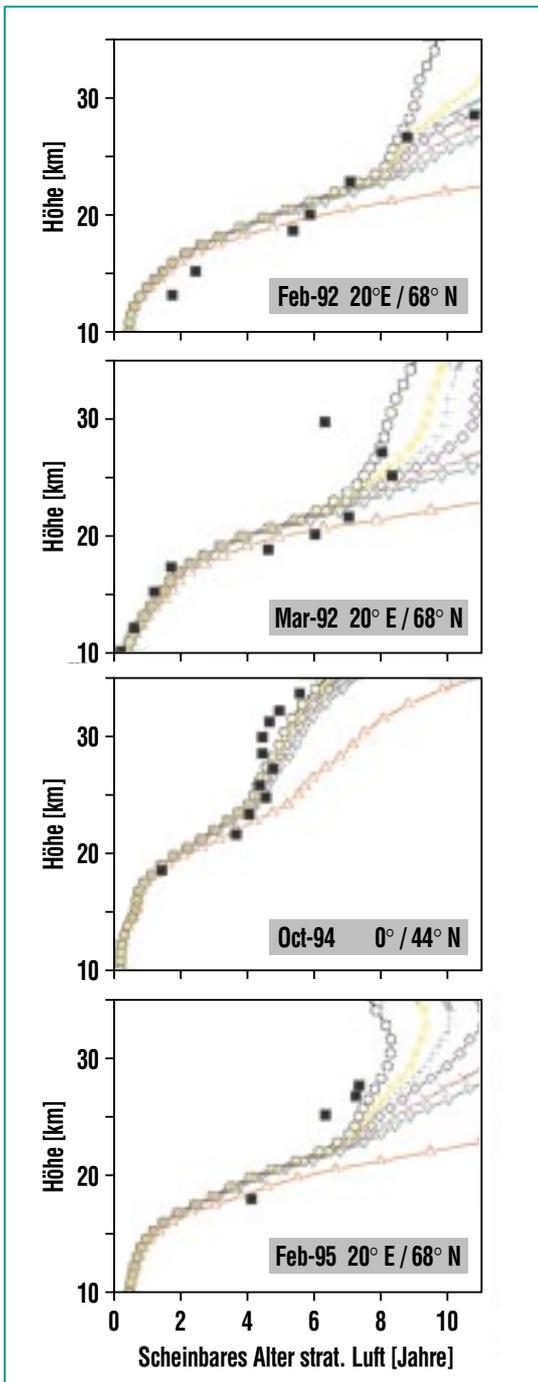


Abb. 3: Aus SF₆-Messungen bestimmtes Alter der Luft (Quadrate) und Vergleich mit verschiedenen chemischen Szenarien für SF₆ (Linien mit Symbolen). Die Szenarien mit größter Lebenszeit ergeben die beste Übereinstimmung zwischen Modell und Beobachtung.

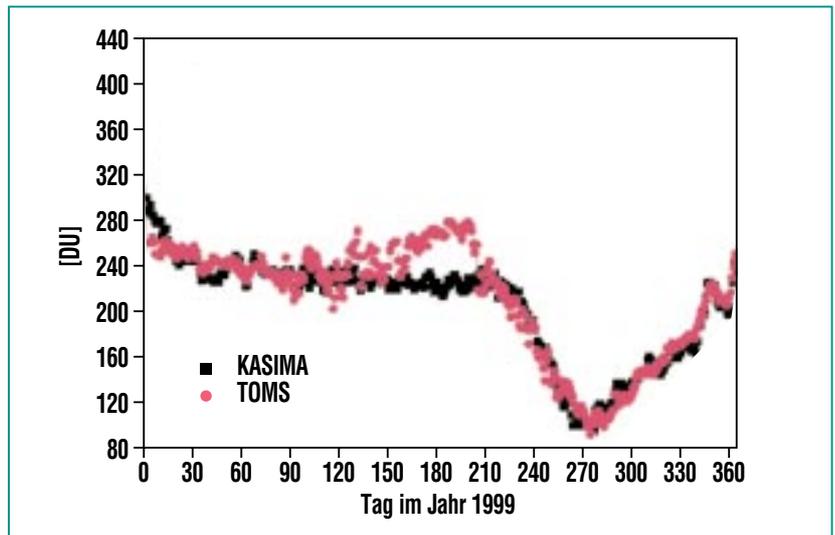


Abb. 4: Minimaler Ozonsäulengehalt in Dobson-Einheiten (1 Dobson-Einheit entspricht einer auf Normalbedingungen komprimierten Ozonschicht von 1 mm Dicke) südlich von 60° Süd im Laufe des Jahres.

die chemischen Prozesse in der Stratosphäre reproduziert werden können. Das Ozonloch ist ein jährlich über dem Südpol auftretendes Phänomen, bei dem durch die aufgehende Sonne im Frühjahr Chlor- und Bromatome in der Stratosphäre freigesetzt werden, die dann in katalytischen Zyklen das Ozon abbauen. Gasphasenchemie, Transporte und physikalische und chemische Prozesse an polaren stratosphärischen Wolken spielen hierbei zusammen.

Ein Beispiel für einen Vergleich zwischen KASIMA-Modelldaten und Satellitenmessungen ist in Abb. 4 zu sehen. Die mit KASIMA berechneten südhemisphärischen Minimalwerte stimmen in ihrem Jahresgang gut mit den Messungen des TOMS Satellitenexperimentes überein. Die Ozonverteilung über der Südhemisphäre am 30. September 1999, dem Tag des ausgeprägtesten Ozonlochs, ist in Abb. 5 darge-

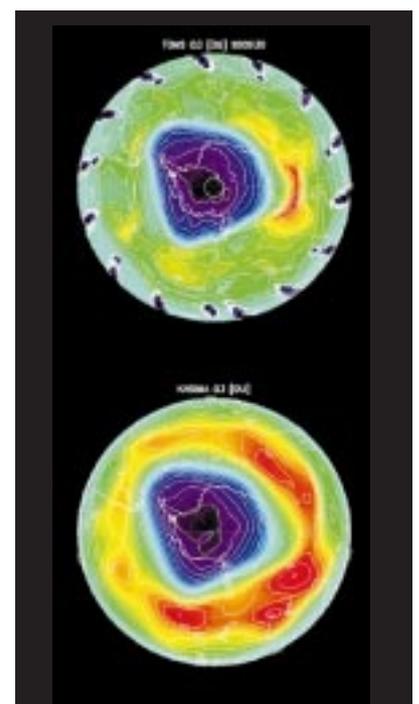


Abb. 5: Vergleich des Ozonsäulengehaltes über der Südhemisphäre am 30. September 1999 (Oben: TOMS-Messdaten, unten: KASIMA-Modellergbnisse)

stellt. Die Simulation des Ozonloches entspricht der tatsächlichen Größe und Stärke in vielen Details, die Ozonverteilung in mittleren Breiten wird jedoch vom Modell überschätzt.

Die Ursache dieser Überschätzung des Ozon, die sowohl dynamische (Ozon wird in der Stratosphäre über den Tropen gebildet und in höhere Breiten transportiert) als auch chemische Ursachen haben kann, sind momentan Bestandteil der Untersu-

chungen am IMK und werden unter anderem im Rahmen von Projekten, wie sie von der EU und vom BMBF innerhalb des Atmosphärenforschungsprogramms (AFO 2000) gefördert werden, durchgeführt.

Schlussbemerkung

Nach dem erfolgreichen Start des europäischen Umweltsatelliten ENVISAT werden im IMK globale Datensätze vieler wichtiger strato-

sphärischer Spurengase ausgewertet und mit dem Modell KASIMA interpretiert. Dadurch wird ein erheblicher Fortschritt beim Verständnis stratosphärischer Prozesse und ihrer Modellierung erwartet.

Literatur

Detaillierte Informationen zu KASIMA gibt es im Internet unter <http://www.fzk.de/imk/imk2/kasima>

- [1] W. Kouker, I. Langbein, Th. Reddmann, R. Ruhnke, *Wissenschaftliche Berichte No. 6278, Forschungszentrum Karlsruhe, 1999*
- [2] W. Kouker, A. Beck, H. Fischer, K. Petzoldt, *J. Geophys. Res.*, 100, 11069-11084, 1995
- [3] Th. Reddmann, R. Ruhnke, W. Kouker, *Ann. Geophysicae*, 17, 415-429, 1999.
- [4] R. Ruhnke, W. Kouker, Th. Reddmann, *J. Geophys. Res.*, 104, 3755-3772, 1999.
- [5] R. Ruhnke, W. Kouker, Th. Reddmann, H. Berg, G. Hochschild, G. Kopp, R. Krupa, M. Kuntz, *Geophys. Res. Letters*, 26, 839-842, 1999.
- [6] W. Kouker, D. Offermann, V. Küll, R. Ruhnke, Th. Reddmann, A. Franzen, *J. Geophys. Res.*, 104, 16405-16418, 1999.
- [7] Th. Reddmann, R. Ruhnke, W. Kouker, *J. Geophys. Res.*, 106, 14525-14537, 2001.

Berichtigung

In Heft 2-3/2002 der „Nachrichten“ ist auf Seite 187 die Unterschrift der unteren Abbildung zu ersetzen durch:

Abb. 7: Mikrotiterplatte aus PMMA mit 96 identischen mikrofluidischen Kanalstrukturen für die Kapillarelektrophorese (mit freundlicher Genehmigung der Firma Greiner Bio-one, Frickenhausen).