

Steinbuch Centre for Computing

NEWS

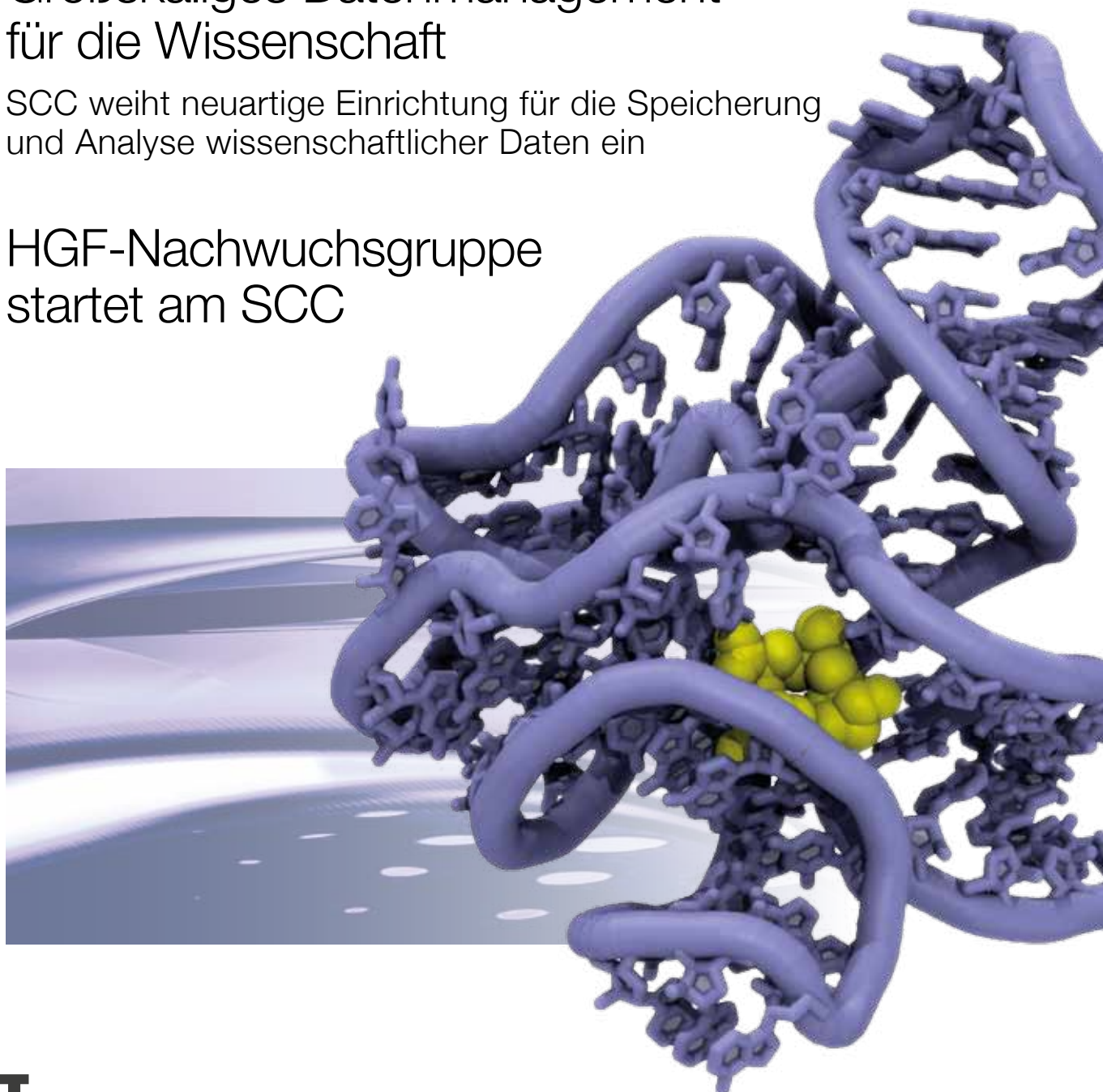
SCC

Prof. Dr. Wilfried Juling seit Oktober
Chief Science and Information Officer

Großskaliges Datenmanagement für die Wissenschaft

SCC weiht neuartige Einrichtung für die Speicherung
und Analyse wissenschaftlicher Daten ein

HGF-Nachwuchsgruppe startet am SCC



INHALT

4
Prof. Dr. Wilfried Juling seit Oktober
Chief Science and Information Officer

4
Wissenschaft und IT – mit Herzblut und
strategischem Weitblick
Zur Berufung von Prof. Dr. Wilfried Juling auf
die Positionen des Chief Science Officer und
des Chief Information Officer in das Erweiterte
Präsidium des KIT

5
„Wenn man eine Tätigkeit mit Leiden-
schaft ausgeübt hat, lässt man nicht
gerne davon ab“
Im Gespräch: Prof. Dr. Wilfried Juling

7
HGF-Nachwuchsgruppe startet am SCC

8
Mastering big data in science –
großskaliges Datenmanagement
für die Wissenschaft
SCC weiht neuartige Einrichtung für die
Speicherung und Analyse wissenschaftlicher
Daten ein

10
Ausbildungsoffensive im Software-Cluster
9.500 Aus- und Weiterbildungsplätze und 5.000
spezialisierte Ausbildungs- und Studienplätze bis
2015 angestrebt

11
MMM@HPC – Mehrskalige Materialmodell-
rechnungen auf Hochleistungsrechner-
architekturen

13
Wie Peer-to-Peer Netzwerke
sich wirklich verhalten
Messungen am BitTorrent-Netzwerk
ermöglichen neue Einblicke

16
Job Scheduling Karlsruhe (JoSchKa)
Jobverteilung in heterogenen unzuverlässigen
Umgebungen und auf On-Demand-Knoten
in der Cloud

20
Das SCC stellt sich vor
In dieser Ausgabe: Die Abteilung Scientific
Computing Services (SCS)

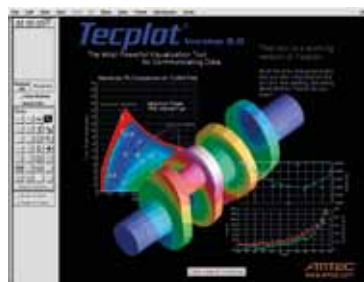
27
Softwareshop des KIT
Professionelle Anwendungen für
Forschung und Lehre



8



20



27



34

30
100 Gigabit/s Ethernet erfolgreich
zwischen FZJ und KIT getestet

31
Grid-Nutzung am KIT

33
TRANSITS – internationaler Workshop
für Computernotfallteams am KIT

34
SimLab Workshop für HPC-Nutzer

34
xchange - Auslandserfahrung
für Auszubildende

35
OE-Betreuer – Ansprechpartner für
IT-Beauftragte

36
SCC auf internationaler
Supercomputing Conference SC10
Rekordwerte und heiße Trends

38
Einführungskurs LaTeX

EDITORIAL

Liebe Leserinnen und Leser,

wenn Sie diese Ausgabe der SCC-News in Händen halten, werden Sie feststellen, dass etwas auf der Editorial-Seite fehlt: das Foto von Prof. Dr. Wilfried Juling. Der langjährige Direktor des ehemaligen Rechenzentrums der Universität Karlsruhe und zuletzt Geschäftsführender Direktor des SCC hat am 1. Oktober 2010 sein Amt als Chief Science and Information Officer im Erweiterten Präsidium des KIT angetreten.

Das Amt des Geschäftsführenden Direktors des SCC ist auf Prof. Dr. Hannes Hartenstein übergegangen, der die Bereiche IT-Management und Prozessintegration, IT-Sicherheit, Service Management und innovative Netztechnologien vertritt. Prof. Dr. Bernhard Neumair ist nach wie vor stellvertretender Geschäftsführender Direktor und für das Management der IT-Dienste verantwortlich. Prof. Dr. Achim Streit vertritt den Bereich Verteilte Systeme, Grid und Cloud Computing sowie HPC und Scientific Computing.

Prof. Juling leitete das Universitätsrechenzentrum über viele Jahre mit großem strategischen Geschick und herausragendem Engagement. Das SCC avancierte unter seiner Führung zu einem der leistungsfähigsten und größten wissenschaftlichen IT-Zentren in Europa. Dafür gilt ihm unser besonderer Dank. Im Bereich der Informationstechnologie – vor allem was ihr Management anbetrifft - hat er nicht nur am KIT deutliche Akzente gesetzt, beispielsweise mit dem Karlsruher Integrierten InformationsManagement (KIM), sondern auch die Informationslandschaft in ganz Deutschland mitgeprägt. Darüber hinaus engagierte er sich neben seiner Lehrstuhlätigkeit und seinen Forschungsaktivitäten zusätzlich in vorbildlicher Weise ehrenamtlich in der Wissenschafts-Community, so zum Beispiel als Vorsitzender des Vorstands des Vereins zur Förderung eines Deutschen Forschungsnetzes (DFN).

Mit dem Ausscheiden von Prof. Juling ist am SCC eine Ära zu Ende gegangen. Doch sehen wir dies nicht nur mit einem weinenden, sondern auch mit einem lachenden Auge, da wir nun auf der Präsidialebene des KIT mit Prof. Juling einen weiteren Wissenschaftler und Manager erster Güte dazugewonnen haben.

Im Namen des gesamten SCC gratulieren wir Prof. Juling an dieser Stelle nochmals ganz herzlich zu seinem neuen Amt und wünschen ihm viel Erfolg!

Prof. Dr. Hannes Hartenstein
Prof. Dr. Bernhard Neumair
Prof. Dr. Achim Streit



Prof. Dr. Hannes Hartenstein
Foto: Privat



Prof. Dr. Bernhard Neumair
Foto: Privat



Prof. Dr. Achim Streit
Foto: Privat

IMPRESSUM

März 2011

Herausgegeben im Auftrag des Direktoriums des Steinbuch Centre for Computing (SCC) von der Stabsstelle Öffentlichkeitsarbeit und Kommunikation

Anschrift: Steinbuch Centre for Computing (SCC)

Redaktion SCC-News

Zirkel 2

76128 Karlsruhe bzw.

Hermann-von-Helmholtz-Platz 1

76344 Eggenstein-Leopoldshafen

Fax: 0721/32550

<http://www.scc.kit.edu/publikationen/scc-news.php>

Redaktion:

Ursula Scheller (verantwortlich)

Telefon: 0721/608-44865

E-Mail: ursula.scheller@kit.edu

Layout und Bildredaktion: John Atkinson

Redaktionell bearbeitete Texte werden mit (red) gekennzeichnet. Nachdruck und elektronische Weiterverwendung von Texten und Bildern nur mit ausdrücklicher Genehmigung der Redaktion.

Prof. Dr. Wilfried Juling seit Oktober Chief Science and Information Officer

Prof. Dr. Wilfried Juling, zuletzt Geschäftsführender Direktor des SCC, hat am 1. Oktober 2010 sein Amt als Chief Science and Information Officer (CSO/CIO) im Erweiterten Präsidium des KIT angetreten.

Als CIO ist er KIT-weit für die technische, organisatorische und nutzungsrechtliche Integration und Koordination aller Aktivitäten in den Bereichen Information und Kommunikation sowie für den Einsatz neuer Informationstechnologien zuständig.

Als CSO für den Bereich 3 ist er dem Vizepräsidenten für Forschung und Information, Prof. Dr. Detlef Löhe, zugeordnet und betreut die Fakultäten für Mathematik und Informatik. Neben den KIT-Schwerpunkten COMMMputation und Anthropomatik & Robotik, die ebenfalls zu seinem Verantwortungsbereich gehören, obliegt ihm die Weisungsbefugnis für den KIT-Teil des HGF-Programms Supercomputing (SuCo), zu dem auch das Grid Computing Centre Karlsruhe (GridKa) zählt. Für das Institut für Angewandte Informatik (IAI) und das

Institut für Prozessdatenverarbeitung und Elektronik (IPE) am Campus Nord und die zentralen Einrichtungen KIT-Archiv, KIT-Bibliothek, Steinbuch Centre for Computing sowie die KIT-Dienstleistungseinheit Informationsmanagement (IMAG) und das Studienzentrum für Sehgeschädigte (SZS) nimmt Prof. Juling die Linienverantwortung des Präsidiums wahr.

(red)

Wissenschaft und IT – mit Herzblut und strategischem Weitblick

Zur Berufung von Prof. Dr. Wilfried Juling auf die Positionen des Chief Science Officer und des Chief Information Officer in das Erweiterte Präsidium des KIT

Ein Zeitraum von über 12 Jahren entspricht in der Moore'schen Sichtweise sechs bis acht Perioden, in denen sich – vereinfacht ausgedrückt – die Rechenleistung verdoppelt hat. Im Jahre 1998, als Wilfried Juling zum Direktor des Universitätsrechenzentrums ernannt worden war, begann der Siegeszug des Mobilfunks, Wireless LANs wurden aufgebaut und die Firma Google gegründet. Betrachtet man diese 12 Jahre von 1998 bis 2010, ist es offensichtlich, dass die Amtszeit von Prof. Wilfried Juling als Direktor des Universitätsrechenzentrums und später als Geschäftsführender Direktor des Steinbuch Centre for Computing von erheblichen technologischen Veränderungen geprägt war, die es zu gestalten galt. Genau dieser Gestaltungsaufgabe widmete sich Prof. Juling als RZ- bzw. SCC-Direktor mit größter Leidenschaft. Dabei ging es ihm nicht alleine um die Technik, sondern vielmehr um deren Einsatz in der Wissenschaft und damit verbunden um das Begeistern von Menschen für die sich bietenden Möglichkeiten.

Der „Dreiklang“ von Strukturen, Integration/Integrations-technologien und technischen Ressourcen erscheint am besten geeignet, die wesentlichen Handlungsstränge von Wilfried Juling als RZ- und SCC-Direktor zu beschreiben:

- Schon frühzeitig entwickelte Prof. Juling ein klar strukturiertes Steuerungs- und Regelungssystem für die Organisation von IT an der Universität Karlsruhe. Diese Vorschläge, welche auch außerhalb der Universität Karlsruhe große Beachtung fanden, stellen auch konsequenterweise die Grundlage der „IT-Governance“ des KIT dar. Das Zusammenführen

von Universitätsrechenzentrum und IWR (Institut für Wissenschaftliches Rechnen) des Forschungszentrums Karlsruhe zum heutigen Steinbuch Centre for Computing wurde von Herrn Juling in gewohnt umfassender und umsichtiger Weise angeführt; das SCC kann deshalb schon jetzt als etablierte Größe sowohl am KIT wie auch im Land und im nationalen und internationalen Kontext gelten.

- Eine Reihe von Projekten und deren nachfolgende Umsetzung in einen Regelbetrieb widmeten sich dem Aspekt der Innovation und Integration von IT: während „DUKATH“ (Drahtlose Universität Karlsruhe (TH)) das Wireless LAN der Universität Karlsruhe aufbaute, widmete sich „NUKATH“ (Notebook Universität Karlsruhe (TH)) der Nutzung von Anwendungen mobiler Systeme im Umfeld universitären Lebens und Arbeitens. Mit dem „Karlsruher Integrierten InformationsManagement“ (KIM) wurden seit dem Jahre 2005 die Prozesse für Lehrveranstaltungsmanagement, Prüfungsverwaltung, Studienassistenten sowie für ein integriertes Identitätsmanagement neugestaltet. Die Ergebnisse dieser ersten Teilprojekte von KIM sowie die derzeit laufenden Projekte haben auch insbesondere das Zusammenwachsen des KIT unterstützt.

- Im Hinblick auf den Aufbau von dedizierten IT-Ressourcen für die Wissenschaft sind insbesondere der Landeshöchstleistungsrechner Baden-Württemberg, die Simulation Labs, das Baden-Württemberg Grid (bwGRid) und die Schaffung zur Nachhaltigkeit der Grid-Infrastruktur in Deutschland

und Europa als Prof. Julings „Themen mit Variationen“ zu nennen. Die vergangenen Monate waren vor allem geprägt durch seinen Einsatz zur Verankerung des SCC als Zentrum für „Large Scale Data Management and Analysis“ in der Helmholtz-Gemeinschaft und in Baden-Württemberg.

Die vorangestellten Punkte können und sollen nicht den Anspruch einer vollständigen Auflistung der Aktivitäten und Verdienste von Wilfried Juling darstellen: sie können es nicht, da Prof. Juling bei seinem unbändigen, ja extremen Einsatz noch deutlich mehr vorangetrieben hat als das Genannte; sie

sollen es nicht, da Prof. Julings Wirken in Sachen „Wissenschaft und IT“ ja in keiner Weise beendet, sondern im Gegenteil nun auf präsidialer Ebene fortgeführt wird. Wir freuen uns auf das Zusammenwirken in dieser neuen Konfiguration, wissen um unsere Verantwortung in der Fortsetzung der oben skizzierten Entwicklungslinien und bedanken uns bei Prof. Juling zusammen mit allen Mitarbeiterinnen und Mitarbeitern des SCC für sein Herzblut und seinen strategischen Weitblick in Sachen Wissenschaft und IT.

Prof. Dr. Hannes Hartenstein



Wechsel ins Erweiterte Präsidium: Prof. Dr. Wilfried Juling wurde zum 1. Oktober zum CSO und CIO des KIT ernannt.
Foto: SCC

„Wenn man eine Tätigkeit mit Leidenschaft ausgeübt hat, lässt man nicht gerne davon ab“

Im Gespräch: Prof. Dr. Wilfried Juling
anlässlich seines Wechsels ins erweiterte Präsidium des KIT

Sie nehmen nun bereits seit einigen Monaten das Amt des Chief Science and Information Officer wahr. Welche Aufgaben und Kompetenzen sind mit dieser neu geschaffenen Führungsposition verbunden?

Prof. Juling: Generell gesagt, erfüllt man als CSO gegenüber den zugeordneten Instituten und Einrichtungen eine so genannte Linienfunktion, das heißt, man ist der unmittelbare Vorgesetzte, so wie das früher beim Forschungszentrum der Vorstand war. Dies gilt auch im Hinblick auf die Verantwortung für die zugeordneten HGF-Programme. Bei den Fakultäten sieht das etwas anders aus, hier würde ich eher von einer Betreuungsfunktion sprechen.

Zu den Aufgaben und Kompetenzen gehört insbesondere auch die strategische Entwicklung der einzelnen Teile des Bereichs, dazu zählen beispielsweise auch die zugehörigen Personalentwicklungen.

Ihre Ernennung zum CSO/CIO erfolgte für viele Mitarbeiter und Mitarbeiterinnen relativ überraschend. Wie kam es dazu?

Prof. Juling: Man redet ja bereits seit über einem Jahr über die Einrichtung von Bereichsvorständen als erweiterte Ebene des Präsidiums, da es insbesondere angesichts der Doppelmission nahezu unmöglich ist, ein solch großes wissenschaftliches Gebilde wie das KIT mit zwei Präsidenten und einer Vizepräsidentin sowie drei Vizepräsidenten zu führen. Die Zustimmung durch den Aufsichtsrat und die Zuwendungsgeber, sprich Bund und Land, zur Einrichtung der so genannten Bereichsvorstandsebene erfolgte Ende Juli. Im August hat mich dann das Präsidium in persönlichen Gesprächen intensiv gebeten, gegebenenfalls für die Übernahme einer solchen Aufgabe zur Verfügung zu stehen. Prof. Hippler und Prof. Umbach haben mich daraufhin der Findungskommission vorgeschlagen, die mich nach einer

Anhörung am 13. September benannt hat. Der Senat hat dies am 20. September bestätigt, und anschließend konnte die Ernennung zum CSO und CIO durch die Präsidenten erfolgen. Natürlich ist es selbstverständlich, dass man als Betroffener während eines solchen Verfahrens nicht über den Hergang spricht. Daher sind solche Ernennungen wohl immer ein wenig überraschend.

Welche Auswirkungen hat Ihre neue Tätigkeit insbesondere für das SCC?

Prof. Juling: Ich hoffe, nur gute Auswirkungen, da ich nun die Möglichkeit habe, mich auf Präsidiumsebene für das SCC weiter zu engagieren und die positive Entwicklung voranzutreiben. Dies gilt natürlich nicht nur für das SCC, sondern für den gesamten Bereich, der mir verantwortlich zugewiesen wurde. U. a. wird eine meiner Aufgaben darin bestehen, auf eine gute strategische Ausrichtung hinzuwirken, um

weitere Synergien und möglichst effiziente Strukturen insbesondere im Hinblick auf die KIT-Entwicklung selbst sowie die Beteiligung an großen Projekten und eine noch bessere Positionierung innerhalb der HGF-Programmatik und des Landes zu erreichen.

Wenn wir zum Beispiel jetzt die nächste programmorientierte Förderphase innerhalb der HGF vorzubereiten haben, geht es vor allem darum, gleich zu Beginn die Kompetenzen besser zu bündeln. Auch wird es uns ohne eine konsequente Forschungskohärenz nur schwer gelingen, die nächsten Hoch- und Höchstleistungsrechner mit Finanzierung des Bundes gemäß Artikel 91b zu beschaffen. Dies bedeutet, dass das SCC als HPC-Service-Zentrum eine solche Mitfinanzierung nicht allein begründen kann, da hier vor allem überregional anerkannte Forschungsaspekte zu berücksichtigen sind, die auch sogenannte wissenschaftliche Querschnittsaspekte beinhalten müssen.

Und natürlich dürfen die bereits gestarteten Aktivitäten zur Exzellenzinitiative II nicht vergessen werden.

Eine weitere direkte Auswirkung ist natürlich mein Ausscheiden aus dem Direktorium. Dies ist mir, ehrlich gesagt, sehr schwer gefallen. Wenn man eine Tätigkeit mit Leidenschaft ausgeübt hat, lässt man nicht gerne davon ab. Ich denke, niemand kennt das SCC so gut wie ich, und selbstverständlich werde ich weiterhin versuchen, dass SCC nach Kräften zu fördern. Ich trage jetzt zwar nicht mehr die unmittelbare operative Verantwortung, werde aber auf der darüber angesiedelten Aufsichts- und Steuerungsebene die weitere Entwicklung des SCC bestmöglich unterstützen. Von daher gehe ich dem SCC nicht ganz verloren.

In welche Richtung werden sich die strategischen Entwicklungen in Ihrem Bereich bewegen, welche ersten Ansatzpunkte werden Sie verfolgen?

Prof. Juling: Aus CIO-Sicht sind zunächst eine bessere Integration von Anwendungssystemen und die Konsolidierung von gemeinsamen Dienstleistungen auf den Weg zu bringen. Dies bedeutet, dass die großen Service Provider, allen voran das SCC, die Bibliothek, das Archiv und die Dienstleistungseinheiten der Verwaltung auf der Grundlage optimierter Prozessab-

läufe effizienter zusammenwirken müssen. Das vorrangige Ziel ist dabei, deutliche Serviceverbesserungen zu erreichen.

Aus CSO-Sicht ist derzeit die Portfolio-Diskussion in der HGF ein bedeutendes Thema. Ziel ist hier, die Programmatik im Hinblick auf die Themen des Bereichs frühzeitig zu beeinflussen und die nächste Phase der programmorientierten Förderung vorzubereiten. Dabei spielen insbesondere die jüngsten Entwicklungen im Programm SuCo und im Bereich der Speicher- und Grid-Themen eine große Rolle. Bei der Neuausprägung der nächsten HGF-Förderphase besteht eines der Ziele darin, die am Campus Süd vorhandene Kompetenz noch mehr mit einzubeziehen, dies gilt vor allem für meinen Bereich.

Inwieweit unterscheidet sich Ihre neue Tätigkeit von der vorherigen?

Prof. Juling: Natürlich ist meine unmittelbare Einflussnahme auf die strategischen Entwicklungen im Bereich der Informationstechnologie am KIT gewachsen. Man trägt nicht nur Verantwortung, sondern hat auch entsprechende Gestaltungsmöglichkeiten. Dabei zeigt sich in der jetzigen Umstellungsphase, dass die operativ strategische Erfahrung als ehemaliger Geschäftsführender Direktor des SCC doch von großem Vorteil ist.

Wird Ihre Position eines SCC-Direktors nachbesetzt und falls ja, wird es sich dabei um eine Persönlichkeit aus dem KIT oder von extern handeln?

Prof. Juling: Ich gehe davon aus, dass es eine Nachfolge geben wird, da mir beide Präsidenten dies so bestätigt haben. Es wird daher auch zu meinen Aufgaben gehören, die Position inhaltlich und funktional in den KIT-Gremien zu begründen, um eine Freigabe für eine entsprechende Professur zu erwirken. Dabei werde ich natürlich auf eine zeitgerechte und inhaltlich entsprechend ausgeprägte Ausschreibung Wert legen, die von der Widmung her den Zukunftsentwicklungen des SCC und des KIT Rechnung trägt.

Wie lang wird dieser Prozess in etwa dauern?

Prof. Juling: Das ist schwer zu sagen, aber man kann davon ausgehen, dass die Besetzung einer Professur mindestens ein Jahr in Anspruch nimmt, vermutlich wird es

aber aufgrund der hohen Anforderungen, die mit dieser Position verbunden sind, sogar länger dauern.

Was wünschen Sie dem SCC für die Zukunft?

Prof. Juling: Weiterhin eine positive Entwicklung und eine erfolgreiche Zukunft! Insbesondere wünsche ich den Mitarbeitern und Mitarbeiterinnen, dass sie immer besser zueinanderfinden; denn in Gemeinsamkeiten liegt besondere Kraft.

Vielen Dank für das Interview und viel Glück und Erfolg in Ihrem neuen Amt!

Interview: Ursula Scheller

Zur Person

Prof. Dr. Wilfried Juling, Jahrgang 1949, studierte von 1970 bis 1976 Mathematik an der Rheinisch-Westfälischen Technischen Hochschule Aachen (RWTH), wurde dort 1982 promoviert mit einer Dissertation auf dem Gebiet der Strömungsmechanik und Gasdynamik und arbeitete danach als wissenschaftlicher Mitarbeiter am Rechenzentrum der RWTH Aachen. 1992 folgte er einem Ruf auf eine Universitätsprofessur für „Super- und Parallelcomputing“ am Fachbereich Informatik der Universität Rostock und übernahm gleichzeitig die Leitung des dortigen Universitätsrechenzentrums. Zum 1. März 1998 berief ihn die Universität Karlsruhe (TH) als Direktor des Rechenzentrums auf den Lehrstuhl für „Rechnersysteme und Infrastruktur der Informationsverarbeitung“ an der Fakultät für Informatik. Ab 2008 war er Geschäftsführender Direktor des Steinbuch Centre for Computing (SCC). Seine Forschungsgebiete umfassen IT-Infrastrukturen, IT-Management, Hochleistungsrechnen und Hochleistungskommunikation, Telematik, Web Engineering und Service-orientierte Architekturen. Mit Wirkung zum 1. Oktober 2010 trat Prof. Juling sein Amt als Chief Science and Information Officer (CSO/CIO) im Erweiterten Präsidium des KIT an.

Seit vielen Jahren engagiert sich Prof. Juling ehrenamtlich für die Wissenschafts-Community, u.a. ist er seit 2005 Vorsitzender des Vorstandes des Vereins zur Förderung eines Deutschen Forschungsnetzes (DFN), seit 2004 Mitglied des Auswahlausschusses der Alexander-von-Humboldt-Stiftung, und von 1999 bis 2005 war er Mitglied der Kommission für Rechenanlagen der Deutschen Forschungsgemeinschaft.

HGF-Nachwuchsgruppe startet am SCC

Zum 1. April 2011 startet am SCC eine Helmholtz-Hochschul-Nachwuchsgruppe im Forschungsgebiet der rechnergestützten Lebenswissenschaften. Der Leiter der Gruppe, Dr. Alexander Schug, konnte sich im diesjährigen mehrstufigen Wettbewerbsverfahren um Fördermittel der Helmholtz-Gemeinschaft (HGF) durchsetzen. Wie zwei weitere Antragsteller am KIT gehört er zu den deutschlandweit 18 Antragstellern, die nun als Helmholtz-Hochschul-Nachwuchsgruppen ihre eigene Forschungsgruppe an einer HGF-Einrichtung aufbauen und leiten können.

Ziel der Nachwuchsgruppe ist es, die strukturellen und dynamischen molekularen Mechanismen genetischer Regulation in Computersimulationen quantitativ zu verstehen, wobei die Methodenentwicklung mit der Anwendung direkt verknüpft ist. Dazu werden verschiedene genetische Regulationssysteme untersucht, unter diesen Zwei-Komponenten-Signal-Transduktions-Systeme und regulatorische nicht-kodierende RNS wie Riboswitches oder das Gruppe-II-Intron. Die Herausforderung in diesen Simulationen besteht darin, ausreichend lange Zeitskalen unter Beibehaltung einer realistischen Beschreibung des Systems zu erreichen. Ein Fokus ist die Entwicklung und Anwendung neuartiger Rechenwerkzeuge in einem mehrskaligen Ansatz, in dem effiziente Suchtechniken auf einer vereinfachenden Beschreibungsebene mit rechentechnisch aufwändigeren detaillierten Modellen kombiniert werden. Letztendlich könnte das gesammelte Wissen dazu dienen, die genetische Regulation der untersuchten biomolekularen Systeme zu kontrollieren, und Anwendung in der synthetischen Biologie oder der Pharmazie finden.

Die Nachwuchsgruppe wird am SCC angesiedelt und erhält ein Budget von 1,25 Mio. Euro für fünf Jahre, wodurch die Anstellung eines PostDocs und zwei Doktoranden sichergestellt ist. Im Rahmen ihrer Forschung wird die Gruppe eng mit dem SimLab NanoMikro am SCC zusammenarbeiten. Durch diese enge Kooperation wird die Nachwuchsgruppe eine größtmögliche Kompetenzbündelung in Bezug auf den wissenschaftlichen Auftrag, die Algorithmenentwicklung und das Hochleistungsrechnen erreichen. Zusätzlich wird ein Wissenspool für molekulare Simulationsmethoden aufgebaut, der Forschern bei der Lösung von Aufgabenstellungen ihres Gebietes auf Hochleistungsrechnern hilft.

Dr. Ivan Kondov



Foto: Privat

Dr. Alexander Schug studierte Physik an der Universität Dortmund und promovierte 2005 am Institut für Nanotechnologie (INT). Anschließend hatte er mehrere Aufenthalte im Ausland: 2005 in der Gruppe von Prof. S. Tanaka an der Universität Kobe in Japan und 2006 bis 2010 bei Prof. J. Onuchic an der University of California, San Diego. Zurzeit hat er eine Assistenz-Professur an der Universität Umeå in Schweden inne.

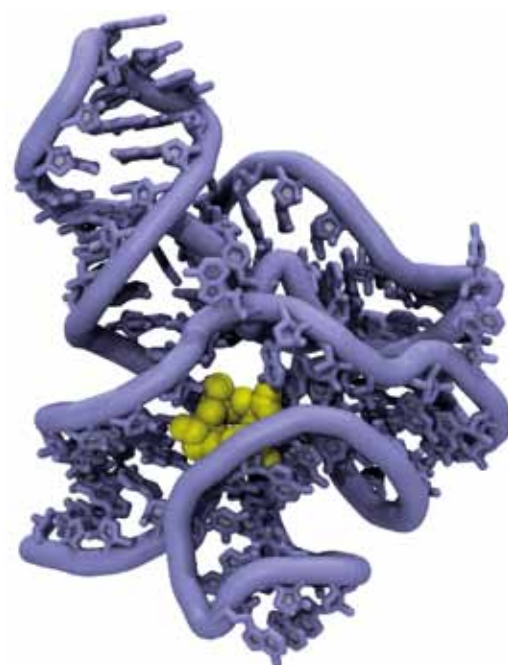


Abbildung 1: Um in DNA gespeicherte genetische Information zu verarbeiten, gibt es in Zellen eine komplexe molekulare Maschinerie von Kontrollmechanismen. Ein Beispiel dafür sind die aus RNA bestehenden Riboswitches (blau), die sich an kleine Moleküle (gelb) binden können und abhängig von dieser Bindung die Verarbeitung genetischer Information regulieren.

Mastering big data in science – großskaliges Datenmanagement für die Wissenschaft

SCC weiht neuartige Einrichtung für die Speicherung und Analyse wissenschaftlicher Daten ein

Daten gewinnen in der Wissenschaft und für die Allgemeinheit eine immer größere Bedeutung. Durch Experimente, Beobachtungen, Instrumentierungen und Messungen werden Tag für Tag rund um die Uhr gewaltige Datenmengen produziert. Um diese Daten zu analysieren und daraus neue wissenschaftliche Erkenntnisse zu extrahieren, sind Services mit entsprechender Infrastruktur erforderlich, die eine Verarbeitung und Auswertung solcher Daten ermöglichen. Das Steinbuch Centre for Computing (SCC) hat mit der Large Scale Data Facility (LSDF) ein neuartiges Konzept für die Speicherung, Verwaltung, Archivierung und Analyse von wissenschaftlichen Daten entwickelt und in Betrieb genommen. Anlässlich der Einweihung der LSDF fand Anfang Februar am KIT ein zweitägiges Kolloquium mit hochkarätigen Vorträgen statt.

Prof. Dr. Wilfried Juling, Mitglied des erweiterten Präsidiums und Chief Science and Information Officer (CSO/CIO) des KIT, begrüßte die rund 100 Teilnehmer und betonte, dass es bei LSDF keineswegs nur um eine reine Datenspeicherung gehe, sondern um adäquate Infrastrukturen und Services zur Unterstützung des kompletten Data Life Cycle – von der Generierung der Rohdaten über erste Analysen und Zwischenspeicherung bis hin zur Langzeitarchivierung. Datenintensive Anwendungen beispielsweise würden schon bald Infrastrukturen erfordern, die über eine Peta-Skalierung weit hinausgingen.



Vortrag beim LSDF-Kolloquium: Prof. Dr. Achim Streit erläuterte u. a. die Gründe für den Aufbau der LSDF.



Innovative Lösungen beim Management großskaliger Daten seien daher von hoher wissenschaftlicher Relevanz

Die LSDF-Infrastruktur des SCC umfasst die Bereitstellung von Speicher- und Rechenressourcen und liefert auch neue Forschungsaspekte. Im Fokus stehen dabei neben der engen Benutzerunterstützung, der Analyse der Daten und der Bereitstellung verlässlicher Services Forschungsthemen wie der schnelle und sichere Zugriff auf Speicher und Rechner, automatisierte Workflows zur Verschiebung der Daten in unterschiedliche Stufen einer Speicherhierarchie, die Langzeitarchivierung sowie die Entwicklung von Daten-Schnittstellen zur LSDF und deren tiefe Integration in den wissenschaftlichen Alltag bei Benutzern.

trug insbesondere auf die Hintergründe zur Realisierung der LSDF ein. Es existiere ein großes Spektrum an Wissenschaftsbereichen und Anwendergruppen, für die die effiziente und sichere Speicherung und Verwaltung großer Datenmengen, ihre semantische Auswertung und Langzeitarchivierung einen enormen technologischen Fortschritt bedeute, der zu neuem Erkenntnisgewinn führen werde. Zwar gebe es in Deutschland schon große Datenspeicher, diese dienten aber entweder dazu, die I/O-Daten von Simulationen für kurze

Prof. Dr. Achim Streit, Direktor am SCC, ging in seinem Vor-



Zeit zu speichern, oder seien themenspezifisch ausgerichtet und nicht für breite Wissenschaftsbereiche konzipiert. Im Gegensatz dazu stehe die LSDF am SCC allen Wissenschaftsdisziplinen zur Verfügung.

Das LSDF-Projekt, an dem auch das Rechenzentrum und das Institut BIOQUANT der Universität Heidelberg beteiligt sind, hat von der Firma IBM einen „Shared University Research Grant“ in Höhe von 75.000 US-Dollar erhalten, der symbolisch von Dieter Münk, Vice President IBM STG Storage, im Rahmen der Veranstaltung überreicht wurde. Dabei erklärte Münk, IBM sei sehr froh über die Partnerschaft in diesem Projekt, da das LSDF-Konzept großes Potenzial habe und völlig neue Möglichkeiten im Bereich der Datenspeicherung eröffne.



Im Anschluss startete der fachspezifische Teil des Kolloquiums. Das Spektrum der Beiträge reichte von der Simulation digitaler Embryos über Aspekte der Langzeitarchivierung bis hin zu Softwareentwicklungen für das LSDF-Datenmanagement. Wesentliche Erkenntnisse dabei waren, dass in Zukunft das Daten-Management den größten Kostenfaktor darstellt und nicht mehr die Speicher-Hardware an sich. Daher sind Methoden zur Datenreduktion – schon an der Datenquelle – notwendig; gleichzeitig stellen die Daten jedoch ein sehr wertvolles Gut dar. Ein Beispiel aus der Linguistik zeigte, dass die Gewinnung eines Terabyte an Daten rund 1,5 Millionen Euro kostet. Daher ist der Langzeitarchivierung – physikalisch, logisch-technisch wie auch intellektuell – ein besonderes Augenmerk zu widmen. Gleichfalls müssen Technologien zur Repräsentation der Daten (grafisch, in Form von Wikis für den freien Zugriff über das World Wide Web, usw.) mindestens einen Technologiesprung durchlaufen, um fit für großskalige Datenmengen zu sein.



Insgesamt bestätigten die Vorträge, dass die Zielsetzung der LSDF die Anforderungen der breiten Wissenschaft nach Speicherung



und Management großskaliger Datenmengen, Datenauswertung mittels Data Intensive Computing (DIC), Langzeitspeicherung sowie einer engen Benutzerunterstützung erfüllt. Die LSDF ist damit ein wichtiger Baustein für den zukünftigen Erkenntnisgewinn aus wissenschaftlichen Daten.

(red)
Fotos: Rolf Mayer



Eine Gruppe von Kolloquiumsteilnehmern am KIT-Campus Nord.

Ausbildungsoffensive im Software-Cluster

9.500 Aus- und Weiterbildungsplätze und 5.000 spezialisierte Ausbildungs- und Studienplätze bis 2015 angestrebt

In der Software- und IT-Dienstleistungsindustrie in Deutschland sollen nach Angaben des Fraunhofer ISI bis 2030 über 450.000 neue Arbeitsplätze entstehen. Der Software-Cluster „Softwareinnovationen für das digitale Unternehmen“, an dem auch das KIT beteiligt ist, wird daher in den kommenden Jahren seine Ausbildungsprogramme gezielt weiterentwickeln. Ergänzend dazu sollen spezifische Ausbildungsschwerpunkte Fachwissen vermitteln, das zukünftig im „Europäischen Silicon Valley“, wie die Clusterregion auch genannt wird, dringend erforderlich sein wird. Schon heute ist der „Software-Cluster“ - ein Zusammenschluss aus Unternehmen, Universitäten und Forschungseinrichtungen um Darmstadt, Kaiserslautern, Karlsruhe, Saarbrücken und Walldorf - Europas größter Softwarecluster.

Bei der Suche nach einer freien Ausbildungsstelle oder einem passenden Studienplatz spielt neben persönlichen Interessen und Begabungen auch die Frage nach den Zukunftschancen



im angestrebten Beruf eine immer größere Rolle. In diesem Zusammenhang ist die IT-Industrie als Wachstumssektor besonders interessant. Der Software-Cluster plant daher für die nächsten Jahre die Ausbildungsmöglichkeiten signifikant zu steigern. Das bezieht sich sowohl auf Ausbildungsstellen für Fachinformatiker als auch auf duale Studiengänge mit hohem Spezialisierungsgrad. Die Jobperspektiven gelten schon heute als hervorragend. Laut einer Studie des Fraunhofer Instituts ISI sollen die hohen Wachstumsraten der Softwareindustrie auch in den nächsten Jahren stabil bleiben. ISI prognostiziert bis zum Jahr 2030 über 450.000 neue Arbeitsplätze und damit einen gesteigerten Bedarf an Fachkräften in der Software- und IT-Dienstleistungsbranche in Deutschland. Schon heute klagt der ITK-Branchenverband BITKOM über Fachkräftemangel und 20.000 offene Stellen im ITK-Bereich.

Der Software-Cluster wird daher in den kommenden Jahren seine Ausbildungsprogramme gezielt weiterentwickeln. Im Januar 2010 gewann der Cluster den Spitzencluster-Wettbewerb des Bundesministeriums für Bildung und Forschung und erhält nun als „Spitzencluster“ öffentliche Mittel in Höhe von rund 40 Millionen Euro. Industrie und Wissenschaft investieren noch einmal die gleiche Summe. Diese Mittel sollen nicht zuletzt der Ausbildung von Nachwuchskräften zu Gute kommen: Schon jetzt werden die Cluster-Partner bis 2015 rund 9.500 Personen im Bereich Informationstechnologie ausbilden. Weiterhin ist die Ausbildung von 5.000 spezialisierten Fachkräften geplant. Dabei bietet der Software-Cluster Plätze für jährlich 400 Auszubildende, 400 Bachelorstudenten, 200 Masterstudenten, 800 Diplomanden und 100 Doktoranden. Spezifische Ausbildungsschwer-

punkte garantieren den Anwärtern Fachwissen, das künftig in der Clusterregion von entscheidender Bedeutung sein wird. Dazu zählt beispielsweise der Studiengang „International Master Software Engineering for Embedded Systems“ oder die Promotion im Bereich „Emergente Software“.

Die Weiterbildungsmaßnahmen werden in Zusammenarbeit mit der Exzellenz-Graduiertenschule des Clusters von den unterschiedlichsten Trägern angeboten: Nachwuchstalente können hier sowohl von Universitäten – beispielsweise dem KIT – als auch von Unternehmen wie etwa der SAP AG aus Walldorf, der Software AG aus Darmstadt und der zum Software AG-Konzern gehörenden saarländischen IDS Scheer AG profitieren. Auch Einrichtungen wie das Fraunhofer Institut und das Deutsche Forschungszentrum für Künstliche Intelligenz (DFKI) sind Teil des Spitzenclusters. Insgesamt sind über 350 Institutionen, darunter auch zahlreiche kleine und mittelständische Betriebe, in den Software-Cluster eingebunden.

(red)

Truffle Capital: Software-Cluster hat weiterhin die Nase vorn

Die Region des Software-Cluster „Softwareinnovationen für das digitale Unternehmen“ hat in der europäischen Softwarebranche weiterhin die Nase vorn. Truffle Capital, das führende europäische Private-Equity-Unternehmen, veröffentlichte im November die zweite Ausgabe seiner Studie „Truffle 100 European Clusters“, eine Rangliste der Top 40-Regionen der europäischen Softwareindustrie. Laut dieser Bewertungsliste erwirtschafteten Unternehmen des deutschen Software-Clusters 40 Prozent des weltweiten Umsatzes der 100 führenden europäischen Softwareanbieter. Ausdrücklich hebt die Studie die Rolle der Software-Cluster-Region hervor, die ihre Position als „Silicon Valley“ Europas behaupten konnte. „Wir freuen uns sehr, dass die Software-Cluster-Region um Darmstadt, Kaiserslautern, Karlsruhe, Saarbrücken und Walldorf in diesem international anerkannten Ranking erneut so hervorragend abgeschnitten hat“, so Prof. Dr. Wilfried Juling, Vertreter des KIT im Strategieboard des Software-Cluster, „die Studie bestätigt nicht nur das enorme Innovationspotenzial der Region, sondern auch die internationale Wettbewerbsfähigkeit des Cluster“. Auf Platz 2 nennt die Rangliste Südostengland, auf Platz 3 die Region Paris.

(red)

MMM@HPC – Mehrskalige Materialmodellrechnungen auf Hochleistungsrechnerarchitekturen

Mit der Beschleunigung der Entwicklungszyklen von neuen Materialien erscheint der effektive Einsatz von Simulationsverfahren für die prädiktive Charakterisierung und Optimierung von Materialien und Device-Eigenschaften als eine große Herausforderung in der europäischen F&E-Agenda. Ein vereinheitlichter multidisziplinärer Ansatz für die systematische Integration von Modellen, Algorithmen, Simulations- und Visualisierungstechniken ist erforderlich, um bestehende isolierte Lösungen für spezielle Probleme in eine umfassende und produktionsreife Infrastruktur umzusetzen und damit Eigenschaften komplexer Materialien auf der Basis ihrer mehrskaligen Bestandteile zu erschließen.

Während viele Methoden existieren, welche nur spezifische Fragestellungen oder Skalen behandeln (siehe Abbildung 1), begrenzt die mangelnde Integration dieser Methoden in eine einsatzbereite Infrastruktur die Anwendung von Simulationsverfahren in der Materialentwicklung mit nanoskaligen Komponenten.

In einem neuen FP7-Projekt bündeln Expertengruppen aus dem Materialforschungsbereich (Code-Entwickler), HPC-Ressourcenanbieter und Anwender aus der Industrie ihre Kompetenzen, um eine integrierte e-Infrastruktur für mehrskalige Modellrechnungen von Materialien (MMM@HPC) zu entwickeln und zur Verfügung zu stellen. Die offene MMM@HPC-Infrastruktur wird die Integration existierender Softwaremodule in adaptierbare Hochleistungsprotokolle und Workflows und gleichzeitig den Einsatz auf verteilten Rechenressourcen ermöglichen. MMM@HPC bringt die Wissenschaftler aus dem Bereich „Computational Science“ in einer Community zusammen und trägt den Bedürfnissen dieser Community Rechnung, indem das Projekt auch mit anderen europäischen Projekten, wie PRACE (Partnership for Advanced Computing in Europe), und Organisationen zusammenarbeitet. Das Projekt wird insbesondere einige sichtbare F&E-Herausforderungen in den Bereichen Energie-

speicherung (Li-Ion-Batterien) und Umwandlung (OLEDs), sowie für Materialien für Kohlenstoff- und Polymer-basierte Elektronik adressieren.

MMM@HPC wird vom KIT (Institut für Nanotechnologie und Steinbuch Centre for Computing) koordiniert. Das Konsortium des Projekts besteht aus mehreren Partnern: Commissariat à l'Énergie Atomique (CEA), Frankreich; Consorzio Interuniversitario (CINECA), Italien; IT Center for Science (CSC), Finnland; Korea Institute of Science and Technology (KIST), Korea; Nokia, Finnland; Sony Deutschland GmbH; Science and Technology Facilities Council (STFC), Großbritannien; Universität Mons, Belgien; und Universität Patras, Griechenland. Das Projekt startete am 1. Oktober 2010 und wird zunächst drei Jahre laufen. Zu den Aufgaben des SCC im Projekt MMM@HPC gehören

- Koordination des Projekts, insbesondere im Bereich HPC; Anbindung an das SimLab NanoMikro; Kooperationen mit anderen Projekten im SCC-Umfeld (EUFORIA, EGI, D-Grid¹) und darüber hinaus, z.B. PRACE und CECAM² Ansprechpartner: Dr. Ivan Kondov, SCC.

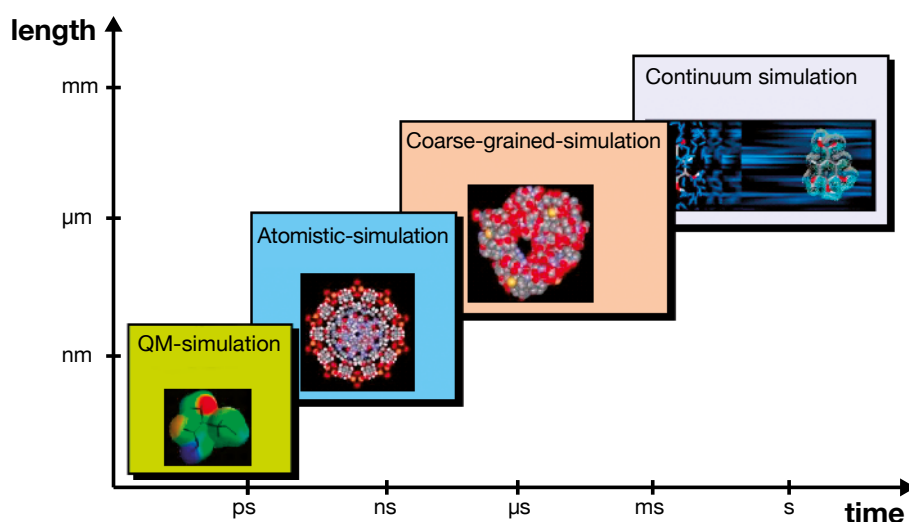


Abbildung 1: Für die Berechnung von Eigenschaften nano-strukturierter Materialien sind verschiedene Zeit- und Längenskalen relevant. Jeder Modell-Partner hat Expertise und Codes in vorwiegend einem dieser vier Bereiche. Diagramm: Dr. Robert Maul, SCC.

¹ EUFORIA (EU Fusion FOR Iter Applications)
EGI (European Grid Initiative)
D-Grid (Deutsche Grid-Initiative)

² PRACE (Partnership for Advanced Computing in Europe)
CECAM (Centre Européen de Calcul Atomique et Moléculaire)

- Entwicklung und Verifizierung von Anwendungsschnittstellen (siehe Abbildung 2) in Zusammenarbeit mit der SCC-Abteilung VSG. Ansprechpartner: Stefan Bozic, SCC.
- Entwicklung von neuen Simulationsprotokollen und Einsatz in Referenzsimulationen in Kooperation mit Sony Deutschland und mit CEA Frankreich. Ansprechpartner: Dr. Velimir Meded, INT.

Erstes Projekttreffen

Bereits vor dem offiziellen Projektstart kamen Vertreter aller Partner beim Kick-off-Meeting am 13. und 14. September 2010 am SCC zusammen. Das Treffen diente der Vorstellung der beteiligten Organisationen und Arbeitsgruppen, der Ausarbeitung von detaillierten Arbeitsplänen und der Umsetzung letzter Schritte zur Unterzeichnung des Konsortialvertrages.

Nach den Begrüßungsworten des Projektsprechers, PD Dr. Wolfgang Wenzel (INT), stellten alle Partner ihre Institutionen und Pläne vor. Am zweiten Tag präsentierten sich die Industriepartner, darüber hinaus fanden die Break-Out-Sitzungen statt, bei denen die Aufgabenplanung im Detail ausgearbeitet wurde.

Sehr positiv aufgenommen wurde die Demonstration von UNICORE GridBeans und Workflows von Stefan Bozic (SCC). Diese Elemente werden als Basis für die Implementierung der integrierten Simulationsprotokolle dienen (siehe Abbildung 2). Nach der Demonstration und bis zum Ende der Veranstaltung gab es mehrere Nachfragen und Diskussionen zu Grid Computing und der Nutzung von UNICORE.

Dr. Ivan Kondov

Weitere Informationen

<http://www.multiscale-modelling.eu>

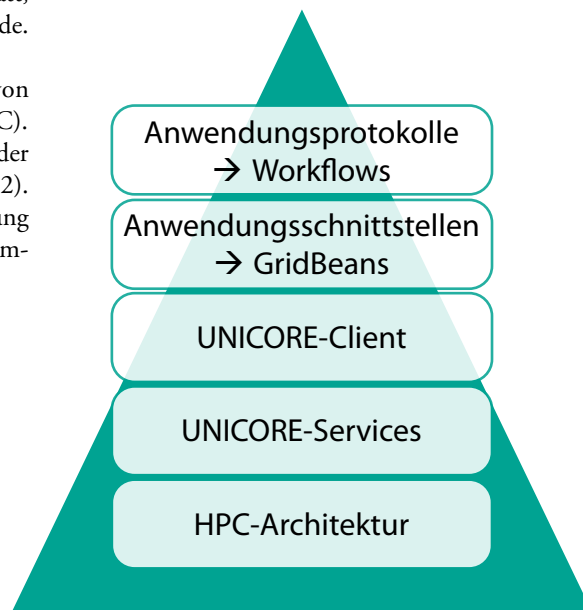
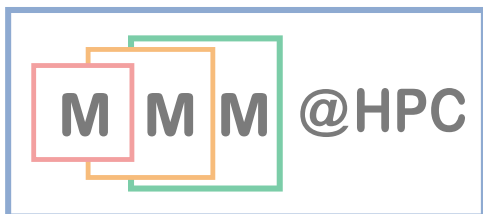


Abbildung 2: Das Konzept basiert auf bestehenden HPC-Grid-Infrastrukturen, die Rechen- und Speicherressourcen über UNICORE-Services bereitstellen. Die Anwendungsschnittstellen werden in so genannten GridBeans im UNICORE Client Layer implementiert und in mehreren Workflows (Simulationsprotokollen) wiederverwendet. GridBean Deployment und Workflow Engine werden ebenso als UNICORE Services bereitgestellt.

Wie Peer-to-Peer Netzwerke sich wirklich verhalten

Messungen am BitTorrent-Netzwerk ermöglichen neue Einblicke

BitTorrent¹ kann heute als einer der am weitesten verbreiteten und bekanntesten Vertreter der Peer-to-Peer (P2P)-Netzwerke gelten. P2P-Netzwerke sind logische, meist auf dem Internet aufsetzende Rechnernetze, deren Teilnehmer (engl. Peers) untereinander gleichberechtigt einen bestimmten Dienst sowohl anbieten als auch nachfragen – im Gegensatz zum Client-Server-Prinzip, das beispielsweise beim World-Wide-Web eingesetzt wird. Neben der Nutzung als Tauschbörse bilden P2P-Netzwerke heute die Basis für eine Vielzahl anderer Systeme aus Wirtschaft und Forschung, beispielsweise der verbreitete Sprach- und Videotelefoniedienst „Skype“. Da Peers zu jeder Zeit einem P2P-Netzwerk beitreten oder es wieder verlassen können, unterliegt Größe und Zusammensetzung des P2P-Netzes meist starken Schwankungen. Die Forschungsgruppe für Dezentrale Systeme und Netzdienste (DSN) führt derzeit umfangreiche Messungen am BitTorrent-Netzwerk durch, das auf dem Kademia-Protokoll basiert. Die durch die Messungen gewonnenen Erkenntnisse ermöglichen die Optimierung von P2P-Protokollen und deren Einsatz als „IT-Basisdienst“.

Das Erwachen der Peer-to-Peer Netze

P2P-Netzwerke erlangten erstmalig Ende der 90er Jahre mit der Verbreitung der Software Napster globale Aufmerksamkeit aber auch zweifelhaften Ruhm als Tauschbörsen für zum großen Teil lizenzrechtlich geschützte Musikstücke und Filme. Die wichtigste Eigenschaft von P2P-Systemen ist ihre Dezentralität, durch die Skalierbarkeit und Robustheit des Systems verbessert werden kann. So haben die Anwender im Fall von Napster Musikstücke nicht etwa von einem zentralen Server herunter geladen, sondern von anderen Napster-Benutzern. Dadurch, dass jeder Teilnehmer Musikstücke nicht nur konsumierte sondern gleichzeitig auch anbot, wuchs das Angebot an Download-Kapazität mit der Nachfrage, das System skalierte also mit der Anzahl der Teilnehmer. Allerdings benötigte Napster für die interne Koordination ein zentrales Datenverzeichnis, insbesondere zur Beantwortung der Frage, welche Teilnehmer eine bestimmte Datei anzubieten hatten. P2P-Netze, die solche zentralen Elemente aufweisen, werden auch als hybride P2P-Netze bezeichnet. Napsters zentrales Verzeichnis war nicht nur Ziel zahlreicher juristischer Attacken der Musikindustrie, es stellte als Single-Point-of-Failure auch eine technische Schwachstelle des Systems hinsichtlich Skalierbarkeit und Robustheit dar.

Im modernen BitTorrent-Netzwerk erfüllt die Rolle dieses Verzeichnisses der so genannte Tracker. Ein Tracker speichert also, welche Teilnehmer gerade eine bestimmte Datei anbieten. Anders als bei Napster kann es bei BitTorrent allerdings für jede Datei einen anderen Tracker geben, was sowohl die Robustheit als auch Skalierbarkeit des Gesamtsystems erhöht. Um auch dieses zentrale Element überflüssig zu machen, wurde 2008 der so genannte Trackerless-Modus spezifiziert², in dem die Funktion des Trackers von einem weiteren P2P-Netz übernommen wird, einer so genannten verteilten Hash-Tabelle (engl. Distributed Hash Table, DHT). DHTs ermöglichen es, Werte effizient unter einem bestimmten Schlüssel abzulegen oder zu speichern. Dabei werden die Schlüssel/Wert-Paare von jeweils wechselnden Peers gespei-

chert, zentrale Elemente sind hierfür nicht erforderlich. Im Falle von BitTorrent spielt die Prüfsumme einer Datei die Rolle des Schlüssels und die IP-Adressen des Peers, der diese Datei anbietet, die Rolle des Wertes. P2P-Netze, die wie die BitTorrent-DHT ohne zentrale Elemente auskommen und eine DHT-Funktionalität implementieren, werden auch der Klasse der strukturierten P2P-Netze zugeordnet.

Die Dynamik hochverteilter Systeme

Global eingesetzte P2P-Netze wie die BitTorrent-DHT bestehen aus einem losen Verbund von Peers, den fortwährend Peers verlassen und betreten. Aufgrund dieses laufenden Peer-Wechsels (engl. Churn) können P2P-Netze nicht als statisch angesehen werden, sondern unterliegen einer starken Dynamik. So schwankt beispielsweise die Anzahl an Peers innerhalb des P2P-Netzes nicht nur kurzfristig über die Tageszeit und Woche, sondern auch langfristig über Monate und Jahre. Auch die prozentuale Verteilung der Peers nach Kriterien wie Herkunftsland oder Güte der Erreichbarkeit ändert sich laufend. Insbesondere langfristige Entwicklungen hängen zudem nicht zuletzt von gesellschaftlichen Trends und Entwicklungen ab und sind daher schwer vorauszusagen. Das Fehlen zentraler Elemente in strukturierten P2P-Netzen wie der BitTorrent-DHT erschwert zudem die Erhebung solcher Kenngrößen, da sich Peers im P2P-Netz an keiner zentralen Stelle an- oder abmelden. Daher können nur durch Messungen am lebenden Objekt und nicht etwa allein durch Simulationen verlässliche Aussagen getroffen werden.

Die Forschungsgruppe für Dezentrale Systeme und Netzdienste (DSN) führt langfristig angelegte Messungen durch, die insbesondere Größe und Zusammensetzung der BitTorrent-DHT erfassen sollen. Hierzu werden von einem Programm regelmäßig und automatisiert repräsentative Stichproben von Peers erhoben und hinsichtlich unterschiedlicher Metriken ausgewertet. Die erhobenen Messdaten sollen ne-

¹BitTorrent Website: <http://www.bittorrent.com/>. 2010.

²BitTorrent Protocol Specification: BEP 5: DHT Protocol (Draft, V. 11031). http://www.bittorrent.org/beps/bep_0005.html. 2010

ben dem Ziel der Optimierung des verwendeten P2P-Protokolls insbesondere auch als Grundlage für die Konzeption ähnlicher hochverteilter Systeme dienen können. So basiert etwa der verbreitete Sprach- und Video-Kommunikationsdienst „Skype“ auf einem unterliegenden P2P-Netzwerk, während der US-Spielehersteller „Blizzard Entertainment“ das BitTorrent-Protokoll zum effizienten Verteilen von Softwareupdates unter Millionen von Kunden verwendet. Analog existieren im Forschungsbereich Ideen, den Energiemarkt von morgen dezentral auf Basis von P2P-Systemen zu organisieren. Für solche Projekte ist es von wesentlicher Bedeutung, möglichst genaue Kenntnisse über Struktur und Zusammensetzung des jeweiligen P2P-Netzes zu besitzen.

Entwicklung der BitTorrent-DHT im letzten Jahr

Die seit mehr als einem Jahr erhobenen Messdaten zeichnen ein genaues Bild sowohl kurz- als auch langfristiger Größenveränderungen der BitTorrent-DHT: Abbildung 1 illustriert die Entwicklung der Anzahl der sich gleichzeitig innerhalb der BitTorrent-DHT befindenden Peers über die Woche vom 29. Oktober bis zum 05. November 2010. Die Kurve stellt eine Extrapolation einer repräsentativ gewählten Stichprobe dar. Es sind mehrere Effekte zu erkennen:

- Die Größe der DHT schwankt stark zwischen etwa 8,5 Mio. und 4,7 Mio. Peers. Insbesondere nehmen also zu jeder Tageszeit mehr als 4,5 Mio. Peers an der BitTorrent-DHT teil.
- Die DHT erreicht ihre maximale Größe in den Abendstunden mitteleuropäischer Zeit, ihr Minimum meist um etwa 4 Uhr morgens. Dies ist insofern überraschend, als die DHT ein weltweit zugängliches P2P-Netz ist und diese Zeit insbesondere für amerikanische Peers die am stärksten genutzte Zeitspanne darstellen sollte.
- Die Amplitude der dargestellten Kurve bleibt nicht konstant, sondern wächst zum Wochenende und nimmt danach wie-

der deutlich ab. Die DHT ist am Wochenende also stärker frequentiert als unter der Woche. Der Zuwachs liegt bei etwa 7 Prozent.

Besonders die starke Korrelation der Schwankungen mit dem mitteleuropäischen Lebensrhythmus kann auf den ersten Blick überraschen, können doch weltweit Nutzer der DHT beitreten. Daher wurde die Zusammensetzung des P2P-Netzes in Bezug auf die Herkunft der Peers genauer untersucht. Abbildung 2 zeigt die Ergebnisse der bereits in Abbildung 1 gezeigten Messung, unterteilt nach dem Herkunftskontinent der Peers. Insbesondere kann dem Diagramm Folgendes entnommen werden:

- Die einzelnen Kurven vollziehen die durchschnittliche Zeitverschiebung des jeweiligen Kontinents nach. Dabei sind die Schwankungen umso ausgeprägter, je weniger unterschiedliche Zeitzonen auf dem Kontinent liegen.
- Die BitTorrent-DHT ist entgegen eventueller Erwartungen vor allem europäisch und asiatisch dominiert. So stammen selbst während der „Rush Hour“ auf dem amerikanischen Kontinent weniger Peers aus Nord- und Südamerika als zur gleichen Zeit aus Europa. Die Ursachen, die zu diesem Phänomen führten, sind zurzeit noch nicht abschließend aufgeklärt.

Die gezeigten Messungen werden nun bereits mit einigen Entwicklungspausen seit mehr als einem Jahr durchgeführt. Die Betrachtung dieser langfristigen Messergebnisse zeigt, dass die BitTorrent-DHT wächst (vgl. Tabelle 1): Bestand die DHT im November 2009 in Spitzenzeiten noch aus maximal lediglich 5,3 Mio. Peers, so waren im August 2010 gleichzeitig bereits bis zu 7 Mio. Peers online. Messungen vom November 2010 zeigen bereits über 8,5 Mio. Peers. Das schnelle Wachstum von fast 60 Prozent innerhalb eines Jahres lässt sich sowohl durch die zunehmende Verbreitung von BitTorrent-Clients, die den Trackerless-Mode unterstüt-

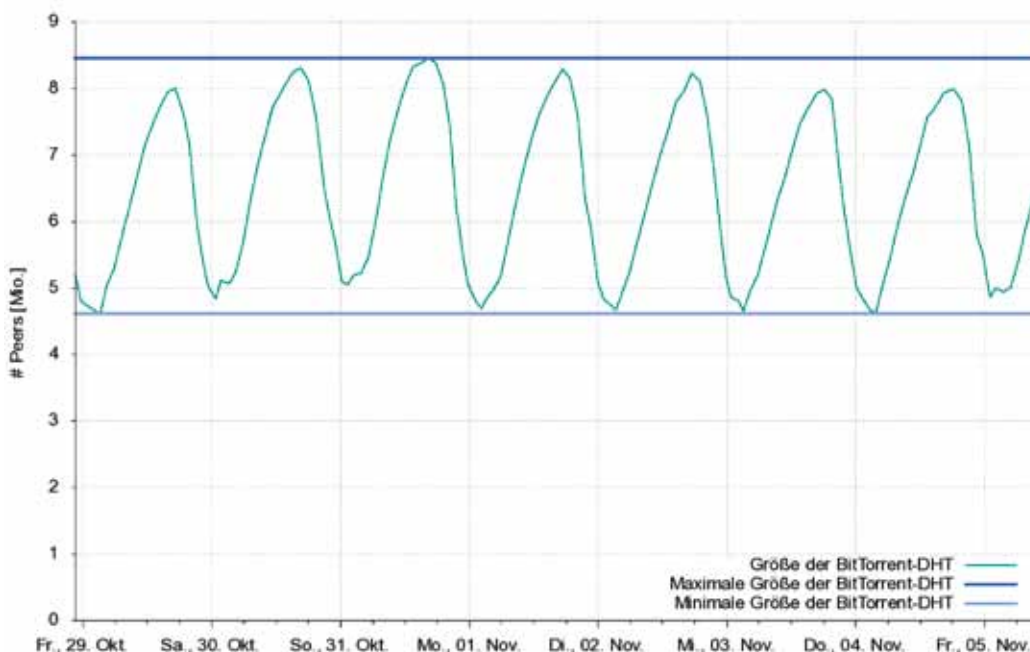


Abbildung 1: Entwicklung der Größe der BitTorrent-DHT über Allerheiligen 2010.

zen als auch durch eine allgemein zunehmende Beliebtheit von BitTorrent erklären. Es bleibt abzuwarten, wie lange sich dieser Trend fortsetzt.

Datum	Minimale DHT-Größe	Maximale DHT-Größe
November 2009	3,3	5,3
August 2010	4,1	7,2
September 2010	4,2	8,0
Oktober 2010	4,4	8,5

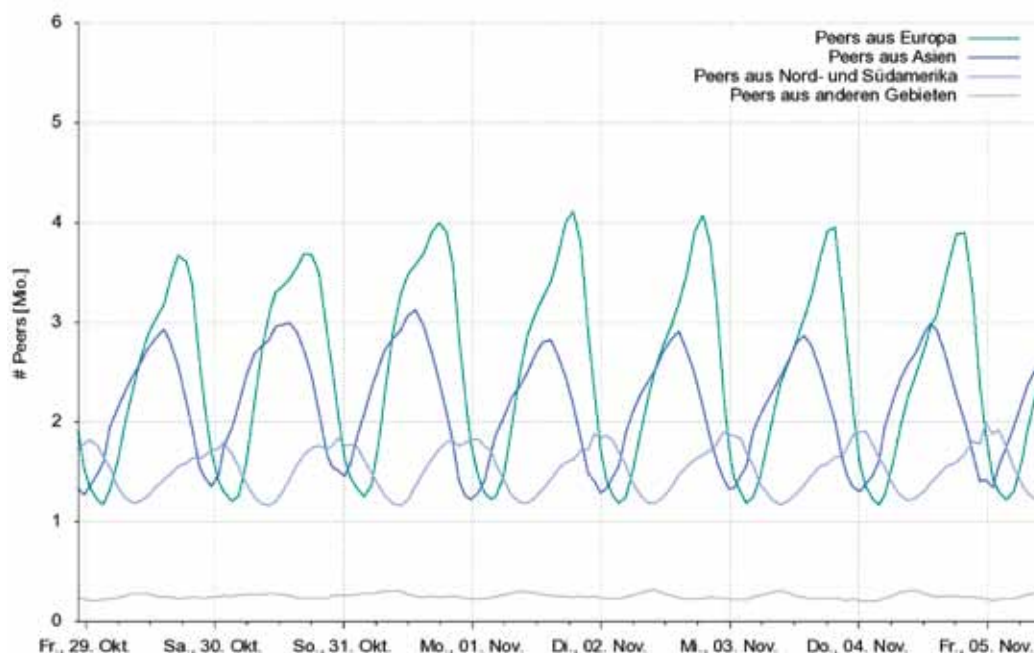
Tabelle 1:
Langfristiges Wachstum der BITTorrent-DHT [in Millionen Peers]

Fazit

Langfristig durchgeführte Messungen von P2P-Netzen wie der BitTorrent-DHT ermöglichen neue Einblicke in kurz- oder langfristige Veränderungen innerhalb von P2P-Netzen und hochverteilten Systemen. So lassen sich Effekte wie eine international schwankende Verbreitung oder die zeitliche Entwicklung besser verstehen und abschätzen. Diese Informationen sind sowohl für die Entwicklung vergleichbarer Systeme als auch für die Entwicklung von Anwendungen, die auf P2P-Systemen beruhen, von besonderer Bedeutung. Ergebnisse aktueller Live-Messungen werden automatisiert³ veröffentlicht. Weiter sollen die Auswirkungen des Ausgangsortes der Messläufe auf die Messergebnisse untersucht werden. Hierdurch soll geklärt werden, ob es zur geographischen Fraktionierung strukturierter P2P-Netze kommen kann.

Konrad Jünemann

Abbildung 2:
Europäische Dominanz
in der BitTorrent-DHT



³BitMON – Ein Werkzeug zum automatisierten Monitoring der BitTorrent-DHT. Website der DSN-Forschungsgruppe. <http://dsn.tm.kit.edu/bitmon.php>. 2010.

Job Scheduling Karlsruhe (JoSchKa)

Jobverteilung in heterogenen unzuverlässigen Umgebungen und auf On-Demand-Knoten in der Cloud

Seit Mitte der 1990er Jahre verbreiten sich PCs in rasantem Tempo. Aus dem universitären und geschäftlichen Umfeld sind sie nicht mehr wegzudenken und auch an den heimischen Arbeitsplätzen setzen sie sich immer mehr durch. Heutzutage besitzt ein Standardrechner eine moderne Mehrkern-CPU mit 4 bis 8 GB Hauptspeicher. Diese Leistung wird allerdings nur selten abgerufen, denn bei der Standardnutzung – Office und Web – wartet der Rechner meist auf Benutzereingaben. Gleichzeitig existiert seit rund zwei Jahren der Trend zur Auslagerung rechenintensiver Aufgaben in die so genannte Cloud. Das sind professionell und meist kommerziell betriebene Großrechenzentren, die es ermöglichen, bei Bedarf per Selbstbedienung Rechenressourcen zu beziehen und diese nach Nutzung zu bezahlen. Das JoSchKa (Job Scheduling Karlsruhe)-System erlaubt einerseits brachliegende Rechenressourcen für rechenintensive Tasks optimal zu nutzen, andererseits eröffnet es den Nutzern zusätzlich die Möglichkeit, bei Bedarf dynamisch neue Rechenressourcen in der Cloud zu erstellen.

Das JoSchKa-System

Im Zuge der immer weiteren Verbreitung des Internets seit Ende der 1990er Jahre sind einige Systeme entstanden, die es ermöglichen, die weltweit brachliegende Rechenleistung von Heim-, Pool- oder Büro-PCs zu nutzen, wie etwa BOINC [1] oder zeta-Grid [2]. Um ein lokales Desktop-Grid aufzubauen, kann man zum Beispiel auf Systeme wie Condor [3] oder Alchemi [4] zurückgreifen. Sie sind jedoch nicht notwendigerweise dazu geeignet, die Leerlaufzeit beliebiger Internet-PCs zu nutzen und im Allgemeinen auch nicht programmiersprachenunabhängig.

Typisch für JoSchKa sind autonom auf den Rechenknoten laufende Agenten sowie die Unterstützung vieler Programmier- und Skriptsprachen (sofern auf dem Zielsystem lauffähig). Die Kommunikation von Agenten und Servern basiert vollständig auf HTTP, ein gemeinsames Dateisystem ist nicht notwendig. Die Rechenknoten können also problemlos hinter Firewalls und NAT-Routern eingesetzt werden. Nutzer können auf grafische Tools und ein modernes objektorientiertes Software Development Kit zurückgreifen um Batch-Jobs zu verteilen oder interaktive verteilte Anwendungen zu entwickeln.

Server

Beim Server handelt es sich um die zentrale Komponente des Systems, er ist für den Up- bzw. Download der Dateien sowie für die Verwaltung der Jobdaten zuständig. Hier wird auch in Echtzeit die Entscheidung getroffen, welcher Job einem anfragenden Agenten zugeteilt wird.

Die Dateitransferschnittstellen für den Benutzer sind SMB-basiert, um bequemes Dateimanagement zu ermöglichen. Alle von den Agenten und sonstige automatisiert genutzten Schnittstellen basieren jedoch ausschließlich auf HTTP als Anwendungsprotokoll, meist werden SOAP-Webservices eingesetzt.

Auf Anfrage eines Agenten nach einem Job erzeugt der Server zunächst eine Kandidatenliste mit allen Jobs, die auf dem betreffenden Knoten ausgeführt werden können. Dabei werden unter anderem Plattformanforderungen des Jobs (zum Beispiel Betriebssystem, Hauptspeicherbedarf), Daten-, und Reihenfolgeabhängigkeiten geprüft. Sind alle Tests positiv ausgewertet, wird der Job in eine Kandidatenliste eingefügt, aus der der Scheduler dann einen einzelnen auszuwählen hat. Doch warum ist es überhaupt ein Problem, den am besten geeigneten Job auszuwählen, warum nimmt man nicht den ersten besten? Es gibt eine Menge Literatur, die sich mit intelligenten Verteilverfahren beschäftigt, aber sie alle werden für zuverlässige Cluster oder HPC-Großrechner entwickelt bzw. betrachten das Problem nur theoretisch oder in Simulationen. Üblicherweise besitzen diese Verfahren auch vollständige Informationen über die Knoten und die Laufzeitanforderungen



Abbildung 1: JoSchKa-Arbeitsprinzip

Das Kernprinzip des JoSchKa-Systems (Job Scheduling Karlsruhe) [5, 6] liegt darin, Jobs nach dem Anfrage/Antwort-Prinzip zu verteilen: Ein zentraler Server verwaltet alle Jobs, während autonome Agentenprogramme auf den Rechenknoten nach Jobs anfragen und diese ausführen. Erreicht eine Anfrage den Server, wird mit geschickten (und im Bereich des Public Resource Computing einzigartigen) Verfahren ein Job ausgewählt, der bestmöglich zu den Eigenschaften der anfragenden Maschine passt, und dessen Beschreibung zum Agenten übertragen. Der Agent lädt dann alle angegebenen Dateien herunter, führt den Job aus und überträgt die Ergebnisse zurück zum Server (siehe Abbildung 1).

der Jobs¹. Im heterogenen Desktopbereich hat man es jedoch mit verschiedensten, nicht vorhersehbar ausfallenden Knoten zu tun. Es sind also grundlegend andere Methoden notwendig, insbesondere ein geeignetes Monitoring ist essentiell.

Fehlertoleranz

Setzt man einen dedizierten Rechencluster ein, kann man sich sicher sein, dass jeder Knoten seine Jobs zuverlässig ohne auszufallen abarbeitet – von Programmier- oder Bedienfehlern der Nutzer einmal abgesehen. Sind die Knoten jedoch Teil eines mehr oder weniger öffentlichen Pools oder handelt es sich gar um Arbeitsplatzrechner, sind Ausfälle und nicht vollständig durchgeführte Berechnungen an der Tagesordnung, denn derartig eingesetzte Rechner werden immer wieder einmal abgeschaltet oder neu gestartet. Auch so genannte Sicherheitssoftware wirkt sich in diesem Umfeld immer wieder als störend aus. In JoSchKa sind einige Mechanismen integriert, die diesem Problem entgegenwirken: Der regelmäßige Upload von Zwischenergebnissen zum Server und ein „Heartbeat“-Signal vom Agenten zum Server.

Falls der Upload der Zwischenergebnisse vom Nutzer aktiviert wurde, überträgt der Agent in regelmäßigen Abständen periodisch alle neu entstandenen bzw. geänderten Dateien zum Server. So kann man als Nutzer entscheiden, ob ein Job gestoppt oder im Falle eines Rechnerausfalls neu gestartet werden soll. Fällt ein Rechenknoten aus bzw. sendet er eine Zeitlang das Heartbeat-Signal nicht zum Server, wird der betreffende Job dort automatisch neu zur Berechnung für andere Agenten freigegeben, ohne dass sich der Nutzer darum kümmern muss. Um jedoch Programmier- oder Konfigurationsfehlern entgegenzuwirken, ist die Anzahl der Neustarts für jeden Job begrenzt. Jobs, die zu oft nicht korrekt vollständig durchgeführt werden, werden vom Server gesperrt bis der Benutzer diese erneut explizit für einen neuen Lauf freigibt.

Monitoring der Rechenknoten

Eine grundsätzliche Annahme ist also die Heterogenität der Rechenknoten. Neben unterschiedlichster Hard- und Softwareausstattung sowie diversen Betriebssystemen betrifft das vor allem die Zuverlässigkeitscharakteristik. Manche Knoten führen jeden Job erfolgreich aus, andere wiederum verhalten sich völlig unvorhersehbar. Ebenso kann sich das Verhalten mit der Zeit ändern: Zuverlässige Knoten fallen plötzlich aus und unzuverlässige werden zuverlässig so wie Rechenressourcen denkbar wären, die am Wochenende und nachts durchgängig verfügbar sind, während des Poolbetriebs jedoch dauernd von Studierenden neu gestartet werden. Büroarbeitsplatzrechner laufen 8 bis 10 Stunden täglich, zum Feierabend werden sie jedoch heruntergefahren oder in den Stand-by geschickt.

Um dieses Verhalten abschätzen zu können, wurde eine Komponente entwickelt, die es ermöglichen soll, das Verhalten eines Rechenknotens in der näheren Zukunft basierend auf seinem bisherigen Verhalten vorherzusagen [5, 6]. Unter anderem werden vom Server charakterisierende Werte erfasst, so dass man für jeden Rechenknoten weiß, wie schnell er im Vergleich zu allen anderen Knoten arbeiten kann, wie lange er im Schnitt verfügbar ist bevor er ausfällt, wie lange er seit dem letzten Start bereits verfügbar ist und wie zuverlässig er die letzten Jobs bearbeitet hat. Dieses Wissen wird bei verschiedenen Verteilstrategien eingesetzt.

Heuristiken zur Job-Selektion

Wie bereits beschrieben hat man es also nicht mit einem spezialisierten Rechencluster zu tun, sondern mit einem heterogenen Pool aus unterschiedlichsten Rechnern, die unterschiedliche Rechenleistung und vor allem unterschiedlichste Zuverlässigkeit besitzen. Ein Knoten kann tagelang ununterbrochen zuverlässig laufen, um dann plötzlich wiederholt unvorher



Abbildung 2: Möglichkeiten der Nutzung von JoSchKa (Server-Netzwerkdateisystem, grafisches Frontend, Modellierung von Jobabhängigkeiten und/oder Programmierschnittstelle mit automatischer Programmcodeübertragung) und das Zusammenspiel von Nutzer, Server und den Agenten auf den Rechenknoten im Überblick. Für weitergehende Informationen sei auf [5] verwiesen.

¹Trotz dieses kompletten Wissens über alle Systemparameter sind die meisten Scheduling-Probleme NP-vollständig.

sehbar neu gestartet zu werden. Jobs mit einer Laufzeit von mehreren Stunden können dann dort nicht mehr sinnvoll ausgeführt werden. Andererseits wäre es eine Verschwendung von wertvoller Rechenleistung, wenn ein zuverlässiger Knoten Jobs berechnen würde, die nach wenigen Minuten erledigt sind. Wenn nun also das Verteilsystem diverse Jobs von verschiedenen Benutzern zur Verfügung hat, ist es naheliegend, sie den anfragenden Knoten derart zuzuteilen, dass die ausfallbedingte Rechenzeitverschwendung minimiert wird² und die Jobs zum frühestmöglichen Zeitpunkt fertig berechnet sind. Gleichzeitig müssen Fairness-Randbedingungen zwingend eingehalten werden, denn kein Nutzer soll sich benachteiligt fühlen.

Unter den vielen (auch miteinander kombinierten) Heuristiken sollen hier nur zwei erwähnt werden: Das einfachste Verfahren und das, welches in der Praxis überwiegend eingesetzt wird.

- (a) Ausbalancierte (faire) Verteilung: Für jeden Nutzer arbeitet die gleiche Anzahl an Knoten.
- (b) Uptime-basierte Verteilung: Die Entscheidung, welcher Job einem anfragenden Agenten zugeteilt wird, basiert auf dem Uptime-Verhalten des jeweiligen Knotens.

Im praktischen Einsatz wird eine Mischung dieser Verfahren verwendet, um einen Mindestgrad an Balance bzw. Fairness zu garantieren. Für weiterführende Informationen zu den Verteilverfahren und deren Evaluation durch Simulation sei auf [5, 6] verwiesen.

Agent

Beim JoSchKa-Agenten handelt es sich um die Komponente, die auf den verteilten Rechenknoten läuft. Er erfordert im Betrieb keinerlei Benutzerinteraktion sondern wird über eine einzelne Konfigurationsdatei gesteuert, welche einmalig dort eingerichtet werden muss. Danach arbeitet der Agent völlig autonom, kann als Hintergrunddienst ge-

startet werden (optional mit einem beim Start durchgeführten Auto-Update) und erledigt die folgenden Aufgaben:

- (1) Zunächst wird die lokale Systemumgebung auf einige für die Jobanfragen relevanten Parameter untersucht. Dies betrifft den installierten Hauptspeicher, die Prozessorarchitektur (x86/x64), die Verfügbarkeit verschiedener Laufzeit- und Simulationsumgebungen (.NET, Mono, Java, Python, Perl, GAMS) und natürlich das Betriebssystem (Windows- oder Unix-artig).
- (2) Der Agent fragt dann beim Server nach einem Job. Sollte der Server mit einer Aufgabe antworten, werden die darin spezifizierten Quelldateien heruntergeladen.
- (3) Der Job (bzw. die dort angegebene Befehlszeile) wird dann mit der niedrigsten verfügbaren Prozesspriorität gestartet und auf dessen Ende gewartet. Während der Ausführung kontaktiert der Agent periodisch den Server, um das Heartbeat-Signal zu übermitteln. Falls vom Benutzer gewünscht, wird die Konsolenausgabe des Jobs und des Agenten mitprotokolliert.
- (4) Beendet sich der Job ohne Fehler, werden alle neu erstellten Dateien sowie (auch im Fehlerfall) die Konsolenausgaben zum Server übertragen.
- (5) Sind alle Dateiübertragungen fehlerfrei durchgeführt worden, wird ein letztes Commit zum Server geschickt und mit Schritt (2) fortgefahren, indem eine neue Aufgabe beim Server angefragt wird.

Nutzung

Für einen Nutzer/Entwickler gestaltet es sich vergleichsweise einfach, nebenläufige Jobs mit JoSchKa zu verteilen. Kommt das Batchsystem zum Einsatz, muss sich die parallele Anwendung in einzelne Teile auftrennen lassen, von denen sich jeder einzelne formal wie folgt beschreiben lassen muss:

- Eine Menge Eingabedateien $I = \{i_1, \dots, i_n\}$,

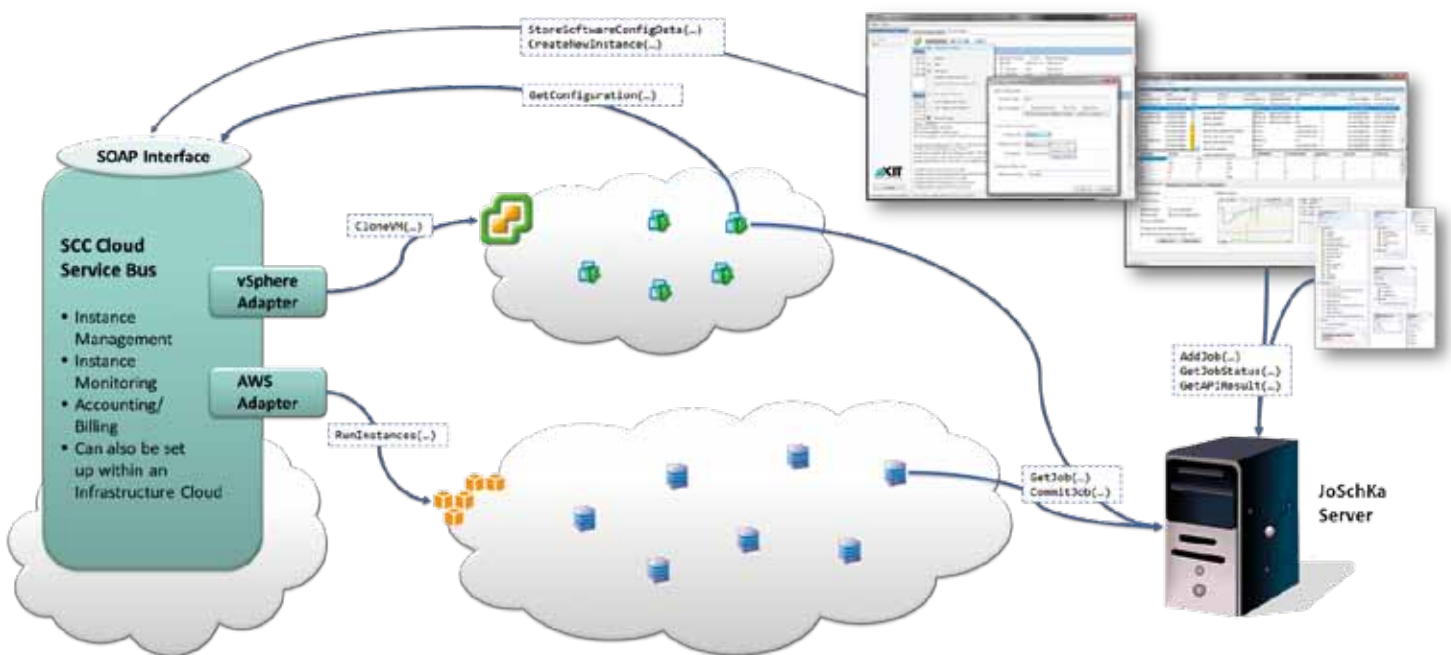


Abbildung 3: Rechenknoten in der Cloud

²Die Ausfälle selbst sind natürlich nicht zu verhindern.

- ein Kommandozeilenbefehl (oder Skript) C , der auf den Rechenknoten ausgeführt wird und
- eine Menge von Ausgabedateien $O = \{o_1, \dots, o_m\}$, die von dem oben genannten Kommando erzeugt werden.

Sollte das nebenläufig zu bearbeitende Problem in derartige Einheiten $J = \{I, C, O\}$ unterteilt werden können und ist ausführbar auf den Rechenknoten, kann eine Verteilung mit dem Batchsystem vorgenommen werden. Hierzu stehen grafische Tools zur Verfügung (siehe Abbildung 2).

Möchte man jedoch eine parallele Anwendung mit mehr Interaktivität entwickeln, programmgesteuerte Threads erzeugen, diese remote ausführen lassen und direkt auf deren Ergebnis reagieren können, kann das objektorientierte JoSchKa SDK mit seiner Thread API für .NET verwendet werden. Hiermit wird es möglich, dynamisch kompilierten Programmcode zum Server zu übertragen, Threads, die diesen parallelen Code ausführen, zu starten und sich ereignisbasiert über deren Ergebnisse benachrichtigen zu lassen. Die Entwicklung eines verteilten Programms ähnelt damit der Entwicklung einer lokalen (Master/Slave) Multithread-Anwendung. Testweise wurde mit diesem SDK eine Worthäufigkeitssuche in Goethes „Faust“ unter Verwendung des Map/Reduce-Paradigmas implementiert.

Rechenknoten on demand in der Cloud

Im Zuge der immer weiteren Verbreitung und Nutzung von Cloudcomputing-Angeboten (wie Amazon AWS, Google App-Engine oder Microsoft Azure), wurde ein Dienst entwickelt, der es ermöglicht, bei Bedarf dynamisch JoSchKa-Rechenknoten auf Cloudressourcen zu erzeugen. Sollte ein Nutzer nicht genügend Rechenleistung erhalten oder Anforderungen an die Rechenknoten haben, die nicht erfüllt werden können, besteht die Möglichkeit, sich bei einem Infrastrukturprovider eben diese erstellen zu lassen, sie nach Gebrauch wieder zu entfernen und gegebenenfalls – je nach Infrastrukturprovider – für die verwendeten Ressourcen zu zahlen.

Der am SCC entwickelte Servicebus [7] ermöglicht es, per SOAP-Schnittstelle derartige Rechenknoten aus Vorlagen (VMWare-Templates oder Amazon Machine Images) zu klonen und die jeweilige Ausstattung der neuen Instanz (CPU-Anzahl, RAM, Service Level) entsprechend den Nutzerwünschen zu konfigurieren. Im Weiteren wird der Ressourcenverbrauch der virtuellen Rechenknoten protokolliert und entsprechend abgerechnet. Auf den Templates sind die JoSchKa-Agenten vorinstalliert, sie verbinden sich nach dem Start mit dem Servicebus, erfragen ihre nutzerspezifische Konfiguration und arbeiten dann wie normale Agenten weiter, jedoch ausschließlich für die vom Cloud-Nutzer konfigurierbaren JoSchKa-Nutzer. Für den Cloud-Nutzer wurde ein grafisches Frontend erstellt, mit dem sich diese virtuellen Rechenknoten bequem in Selbstbedienung verwalten lassen. Neben Funktionen zum Erzeugen, Starten, Stoppen, Konfigurieren und Löschen von virtuellen Rechenknoten kann sich der Nutzer über Statistiken zum Ressourcenverbrauch und die anfallenden Kosten informieren. Zurzeit werden Rechenknoten in VMware vSphere und Amazon EC2-Umgebungen unterstützt. Die anfallenden Kosten können mit flexiblen Preismodellen berechnet und dem Cloud-Nutzer in Rechnung gestellt werden. Die Zusammenhänge sind in Abbildung 3 dargestellt.

Praxiserfahrungen

Das Batchsystem sowie die Verteilverfahren wurden in der Arbeitsgruppe „Effiziente Algorithmen“ am Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB) des KIT implementiert, wo der Server auch betrieben wird. Die Weiterentwicklung – vor allem des Client-Thread-SDKs und des Cloudsystems – wird am SCC in der Forschungsgruppe Cloud durchgeführt. Seit 2006 läuft JoSchKa durchgehend auf den Rechnern des CPI-Pools der Fakultät für Wirtschaftswissenschaften am KIT und auf weiteren kleineren Pools sowie einigen Servern. Die Poolrechner sind Standard-PCs und mit 2 bis 3GHz-Prozessoren sowie 2 GB Hauptspeicher ausgestattet. Bisher wurden rund 300.000 (hauptsächlich Java-basierte) Jobs im Bereich der naturalen Optimierungverfahren und Simulationen für das Organic Computing durchgeführt. Die interaktiv an den Rechnern arbeitenden Studierenden merken praktisch nichts von der zusätzlichen Hintergrundlast. Wie in [5] gezeigt wurde, sinkt die Ausführungsgeschwindigkeit einer Standard-Desktop-anwendung nicht, wenn der Rechner eine CPU-intensive Task im Hintergrund mit Leerlaufpriorität ausführt. Es besteht jedoch die Möglichkeit, die Agenten so zu konfigurieren, dass keine Jobs ausgeführt werden, sobald am Rechner interaktiv gearbeitet wird.

Die Studierenden und das wissenschaftliche Personal der Fakultät profitierten erheblich von JoSchKa: Die Studierenden entwickeln ihre Optimierungsprogramme normalerweise in Java auf Windows, wie sie es in Programmierkursen gelernt haben. Aber die üblichen HPC- oder Gridplattformen benötigen C/C++ oder gar Fortran und laufen unter speziellen Linux-Varianten, so dass deren Verwendung aus technischer Sicht schwierig ist. Denn die Notwendigkeit, Messläufe vielfach ausführen lassen zu müssen, ergibt sich oft erst gegen Ende einer Studien- oder Diplomarbeit, so dass eine Umstellung der Simulationsprogramme auf Supercomputer- oder Gridtechnik schon aus Zeitgründen nicht mehr möglich ist. Die nutzerfreundliche Handhabung von JoSchKa ermöglicht in diesem Falle eine nebenläufige Ausführung auf gewöhnlichen Rechenknoten (ohne Architektur- oder Sprachumstellung), selbst wenn eine verteilte Ausführung vorab nicht angedacht war.

Dr. Matthias Bonn, Dr. Marcel Kunze

Referenzen

- [1] Anderson, D. (2004). BOINC: A System for Public Resource Computing and Storage. 5th IEEE/ACM International Workshop on Grid Computing, (pp. 365–372). Pittsburg, PA.
- [2] Wedeniwski, S. (2008). ZetaGrid – Computations connected with the Verification of the Riemann Hypothesis, Foundations of Computational Mathematics Conference, Minnesota, USA, August 2002.
- [3] Litzkow, M., Livny, M., & Mutka, M. (1988). Condor – a Hunter of idle Workstations. 8th International Conference on Distributed Computing Systems, (pp. 104–111).
- [4] Luther, A., Buyya, R., Ranjan, R., & Venugopal, S. (2005). Peer-to-Peer Grid Computing and a .NET-based Alchemi Framework. In L. Y. Guo, High Performance Computing: Paradigm and Infrastructure. New Jersey, USA: Wiley Press.
- [5] Bonn, M. (2008). JoSchKa: Jobverteilung in heterogenen und unzuverlässigen Umgebungen. Dissertation an der Universität Karlsruhe (TH), Universitätsverlag Karlsruhe, 2008.
- [6] Matthias Bonn, Hartmut Schmeck. The JoSchKa System: Organic Job Distribution in Heterogeneous and Unreliable Environments. In Christian Müller-Schloer, Wolfgang Karl, and Sami Yehia. Proceedings of the 23rd International Conference on Architecture of Computing Systems (ARCS 2010), Seiten: 73-86, Springer, LNCS, 5974, Berlin Heidelberg, Februar, 2010
- [7] Matthias Bonn, Marcel Kunze. Cloud Computing – SCC entwickelt Servicebus, SCC News, Seiten 12–14, September, 2010

Das SCC stellt sich vor

In dieser Ausgabe: Die Abteilung Scientific Computing Services (SCS)



Foto: Rolf Mayer

Prof. Dr. Rudolf Lohner studierte Diplom-Mathematik an der Universität Karlsruhe, wo er nach einem Auslandsjahr in den USA auch promovierte und habilitierte. Seine wissenschaftliche Tätigkeit liegt in den Bereichen der angewandten und numerischen Mathematik mit besonderem Schwerpunkt auf den Gebieten der Rechnerarithmetik, Intervallrechnung und selbstverifizierenden Algorithmen. Prof. Lohner ist Leiter der Abteilung Scientific Computing Services (SCS) am SCC und damit zuständig für die Bereitstellung von Hochleistungsrechnerressourcen und technisch-wissenschaftliche Anwendungen.



Foto: Rolf Mayer

Horst Gernert ist stellvertretender Leiter der Abteilung SCS. Nach dem Abschluss seines Studiums als Diplom-Wirtschaftsingenieur nahm er 1980 seine Tätigkeit als Wissenschaftliche Mitarbeiter am ehemaligen Rechenzentrum der Universität Karlsruhe auf. Im Verlauf seiner langjährigen Tätigkeit beschäftigte er sich mit verschiedensten Aspekten der Datenverarbeitung, wie Entwicklung und Betreuung systemnaher Software, Datenhaltung und Netzwerken. Zurzeit liegt sein Hauptaufgabengebiet im Bereich Beschaffung und Betrieb von HPC-Systemen.

(red)

Die Abteilung SCS von oben nach unten: Dr. Elisabeth Syrjakow und Mehmet Soysal; Dr. Holger Obermaier und Christoph-Erdmann Pfeiler; Richard Walter, Frauke Bösert und Roland Laifer; Dr. Klaus Braune, Rolf Mayer und Dr. Paul Weber. Fotos: Rolf Mayer, Holger Obermaier



Die Abteilung Scientific Computing Services (SCS) bietet den wissenschaftlichen Mitarbeiterinnen und Mitarbeitern am KIT eine Vielzahl von Diensten rund um das Wissenschaftliche Rechnen und technisch-wissenschaftliche Anwendungen. Abteilungsleiter Prof. Dr. Rudolf Lohner, sein Stellvertreter Horst Gernert sowie alle Mitglieder der Abteilung betrachten diese Aufgaben als ganzheitlichen Service, der alle Aspekte des Wissenschaftlichen Rechnens und der Nutzung technisch-wissenschaftlicher Anwendungen berücksichtigt. Das Dienstespektrum reicht vom Betrieb verschiedener Hochleistungsrechner und Hochleistungs-File-Systeme, über die Bereitstellung einer umfangreichen Basissoftware, wie Compiler, Bibliotheken, Debugger und Batchsysteme bis hin zu einem breitgefächerten Angebot an Anwendungssoftware für die unterschiedlichsten technisch-wissenschaftlichen Bereiche.

Das Dienstangebot endet nicht mit dem bloßen Bereitstellen der Anwendungssoftware, sondern beinhaltet auch Beratung beim Prä- und Post-Processing der Daten bis hin zur Präsentation wissenschaftlicher Ergebnisse. SCS bietet Unterstützung bei der Visualisierung von Daten, der Erstellung komplexer wissenschaftlicher Dokumente oder beim wissenschaftlichen Austausch von Forschergruppen über moderne Medien- und Teledienste.

Für alle Aspekte dieser Dienste offerieren die Mitarbeiterinnen und Mitarbeiter den Nutzern eine kompetente und weitreichende Betreuung: Schulungen, allgemeine und individuelle Beratungen, Hilfestellung bei der Fehlersuche und -beseitigung in Programmen oder bei der Parallelisierung und Optimierung von Anwendungen für konkrete Plattformen, um diese bestmöglich auszunutzen.

Hochleistungsrechner und Cluster

Derzeit betreibt SCS die folgenden Hochleistungsrechner und Cluster:

- HP XC3000 („HC3“): Die neueste Anlage, basierend auf Intel Xeon-Prozessoren der Nehalem-Generation mit 334 Rechenknoten mit je 8 CPU-Cores pro Knoten. Dieser Rechner steht allen Wissenschaftlerinnen und Wissenschaftlern am KIT zur Verfügung.
- HP XC4000 („XC2“): Diese auf AMD-Opteron basierende Maschine mit 750 Rechenknoten mit je 4 CPU-Cores ist der Landeshöchstleistungsrechner für Baden-Württemberg. Die Nutzung der Anlage muss über ein Webportal bei dem zuständigen Lenkungsausschuss beantragt und von diesem genehmigt werden. Im Rahmen einer Kooperation mit den Universitäten Stuttgart und Heidelberg, dem Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg und der Wirtschaft steht diese Maschine auch industriellen Nutzern zur Verfügung.
- Institutscluster („IC1“): Dieser auf Intel Xeon-Prozessoren basierende Cluster mit 200 Rechenknoten wurde gemeinsam von verschiedenen Instituten des KIT zusammen mit dem SCC beschafft und steht diesen Einrichtungen auch zur Verfügung. Das dahinter stehende Konzept, anstelle von einzelnen, kleineren, institutseigenen Clustern lieber in eine gemeinsame große Anlage zu investieren, hat sich als so erfolgreich erwiesen, dass bereits die Beschaffung eines weiteren Clusters in Vorbereitung ist.

- BW-Grid-Cluster („bwGRiD“): Seit etwas über zwei Jahren steht eine über acht Standorte in Baden-Württemberg verteilte Grid-Infrastruktur aus größeren Compute-Clustern zur Verfügung. Einer dieser Cluster mit 140 IBM Blade-Center-Knoten steht auch am KIT und wird von SCS betrieben. Diese bwGRiD-Infrastruktur hat zu einer guten Zusammenarbeit der Rechenzentren an den Landesuniversitäten geführt. Durch geeignete Grid-Middleware auf allen Clustern können die Nutzer auch alle Systeme landesweit verwenden.

Nähere Informationen zu den Anlagen unter <http://www.scc.kit.edu/dienste/4665.php>.

Technisch-Wissenschaftliche Anwendungen

Das SCC stellt den Wissenschaftlern und Studierenden des KIT eine Fülle von Programmsystemen zur Verfügung, um technisch-wissenschaftliche Fragestellungen auf den verschiedensten Rechnern, seien es Arbeitsplatzrechner in den Instituten oder die HPC-Cluster des SCC, zu untersuchen. Zu dem ganzen Komplex gehören neben den CAE-Anwendungen, mit denen physikalische Modelle numerisch simuliert werden, auch die Werkzeuge zur Konstruktion und Modellierung solcher Systeme sowie die grafische Darstellung und Präsentation der Simulationsergebnisse. Darüber hinaus werden Computer-Algebra-Systeme eingesetzt, die mathematische Probleme mit algebraischen Methoden behandeln.

Ein weiterer Bereich der technisch-wissenschaftlichen Anwendungen umfasst die Datenanalyse und Statistik sowie die wissenschaftliche Textverarbeitung.

Bei den Anwendungen in diesen Bereichen leistet die Abteilung SCS eine durchgängige Unterstützung. Hierzu gehören u.a.:

- eine erste Support-Anlaufstelle bei Problemen mit der Anwendung (Beratung, Schulung, Fehlersuche, etc.)
- die Installation der Anwendungen auf den HPC- und Clustersystemen sowie die Anpassung der Anwendungsumgebungen
- die Installationshilfe auf lokalen Rechnern der Nutzer
- das Lizenzmanagement der Produkte
- Verhandlungen mit Herstellern und Lieferanten.

CAE-Anwendungen

Unter CAE-Anwendungen versteht man im weitesten Sinne Software, mit der ingenieurwissenschaftliche Probleme simuliert werden. Dazu gehören beispielsweise strukturmechanische und Festigkeitsprobleme aus dem Bereich des Ingenieurbaus, der Automobil- und Luftfahrtindustrie sowie der Materialforschung. Das dabei zugrundeliegende numerische Verfahren ist die so genannte Finite-Elemente-Methode. Zu den wichtigen Programmen, die das SCC in diesem Bereich bereit stellt, zählen:

- ABAQUS
- ANSYS Mechanical
- ADINA
- MD Nastran
- MSC.MARC
- LS-DYNA.

Die Simulation von Strömungen, Verbrennungen, chemische Reaktionen, die Ausbreitung von Partikeln sowie Mehrphasensysteme gehören ebenso zu den wichtigen Aufgaben, die sich Ingenieuren und Wissenschaftlern stellen. Unter ANSYS Academic Research ist ein ganzes Bündel von Softwaresystemen zusammengefasst, welches das oben schon erwähnte ANSYS Mechanical enthält, aber auch Programme zur Berechnung von Strömungen, wie

- ANSYS Fluent
- ANSYS CFX

und Werkzeuge zur Erzeugung von Geometrien bzw. zum Import und zur Bearbeitung von CAD-Modellen und zur Vernetzung:

- ANSYS ICEM_CFD
- ANSYS Workbench.

Weitere wichtige Programme sind

- STAR-CD
- STAR CCM
- COMSOL Multiphysics
- PowerFLOW.

Im Bereich des Computer Aided Design im 2D- und 3D-Bereich setzt das SCC auf die führenden Produktfamilien von Autodesk, Dassault und PTC. Mit diesen Produktfamilien kann praktisch die gesamte Spanne der Anwendungsprofile vom Maschinenbaubereich über den Architektur- und Bauingenieur-Bereich bis hin zu den Geowissenschaften abgedeckt werden. Als Produkte sind hier u.a. zu nennen:

- AutoCAD
- Autodesk Inventor
- AutoCAD Architecture und Revit
- CATIA
- Pro/ENGINEER.

Die Ausbreitung elektromagnetischer Wellen (Wellenleiter, Antennen, Resonatoren) und die Wechselwirkung mit geladenen Partikeln werden ebenfalls mit numerischen Programmen simuliert. Hier offeriert das SCC folgende Anwendungen:

- ANSOFT EDA System
- CST Microwave Studio
- COMSOL Multiphysics.

Die Simulation von elektronischen Schaltungen, der Entwurf von Leiterplatten und die Messdatenerfassung und -auswertung sowie die Gerätesteuerung kann mit

- PSPICE
- Electronics Workbench
- LabVIEW

sowie mit Modulen aus dem ANSOFT EDA-System und MATLAB durchgeführt werden. Weitere EDA-Systeme werden über die EUROPRACTICE-Initiative bereitgestellt.

Bei der den meisten CAE-Systemen zugrunde liegenden Methode der Finiten Elemente oder Finiten Volumen, muss die Geometrie des Lösungsraumes modelliert und in Zellen zerlegt werden. Dazu gibt es programmspezifische und produktneutrale Präprozessoren wie die Systeme

- Patran
- HyperWorks,

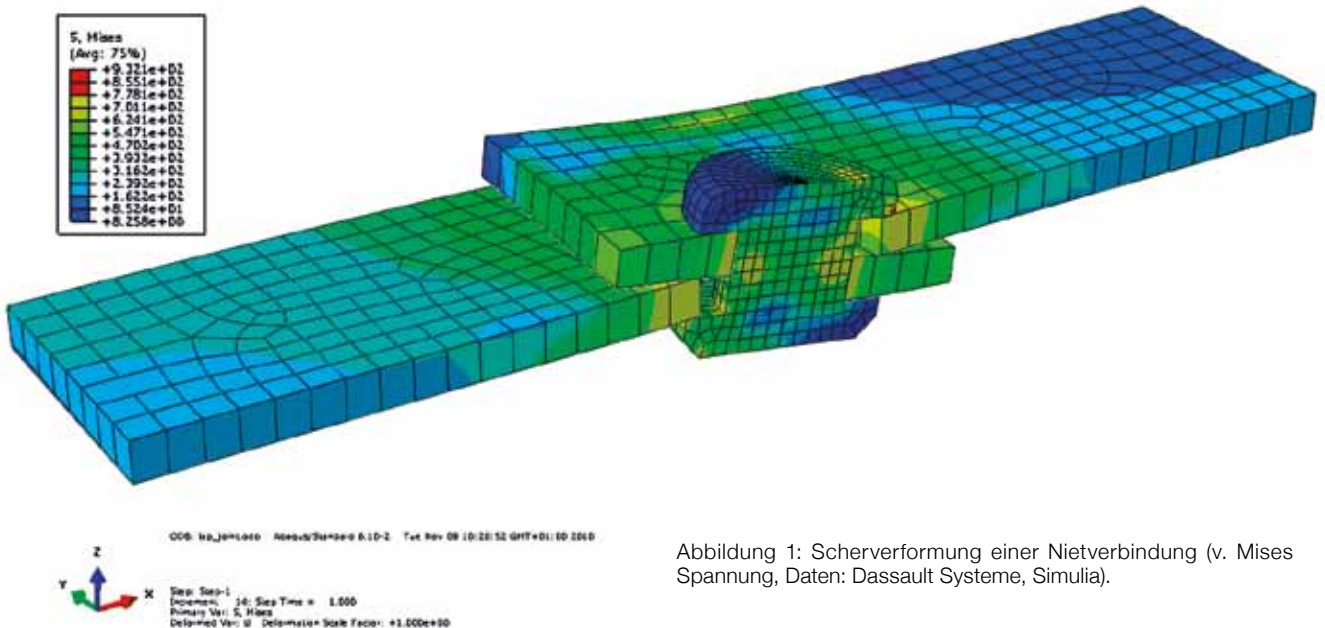


Abbildung 1: Scherverformung einer Nietverbindung (v. Mises Spannung, Daten: Dassault Systeme, Simulia).

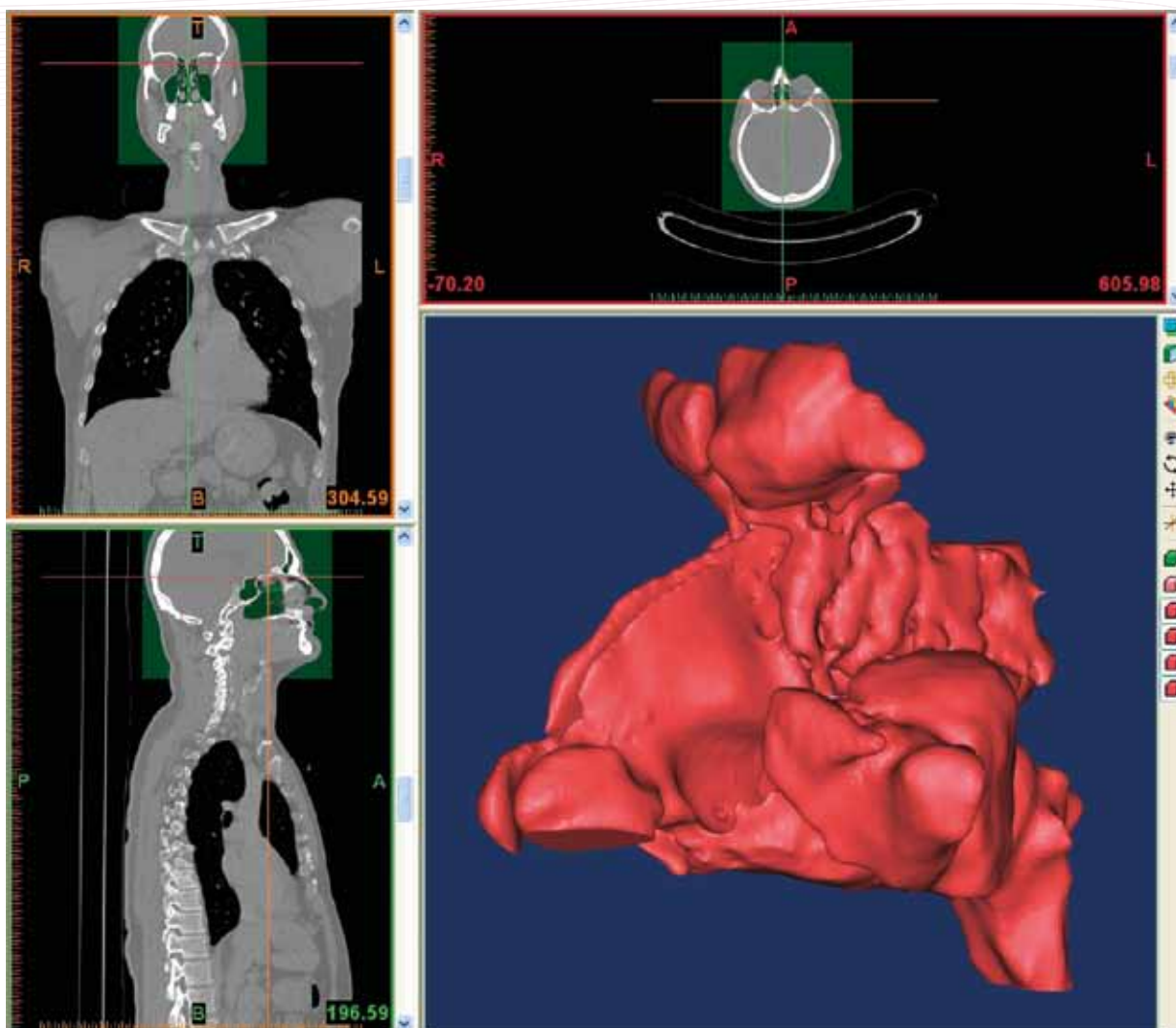


Abbildung 2: Ermittlung der Atemwege im Kopfbereich aus CT-Daten.

die noch zusätzliche Funktionalitäten anbieten, insbesondere auch die Auswertung und grafische Darstellung und Animation der Berechnungsergebnisse.

Bei der Erstellung und Nachbearbeitung von 3D-Modellen aus Daten der medizinischen und technischen Bildgebung (CT, μ CT, MRI) werden spezielle Anforderungen an die Software gestellt. Hier setzt das SCC auf die Produktfamilie der Firma Materialise. Das Produkt Mimics erlaubt es, 2D-Schichtbilder unterschiedlicher Formate mit hoher Genauigkeit in 3D-Daten umzuwandeln. Das Programm 3-matic ermöglicht das komfortable Editieren von Scan-Daten und das Arbeiten mit unterschiedlichen CAD-Formaten bis hin zum CAE-Post-Prozess.

Eine zunehmend wichtige Rolle im CAE-Bereich spielen so genannte Multiphysik-Anwendungen, bei denen verschiedene physikalische Phänomene miteinander gekoppelt werden, zum Beispiel die Fluid-Struktur-Kopplung. Viele der bisher genannten Produkte lassen solche Kopplungen zu, insbesondere sind hierbei

- ADINA
- COMSOL Multiphysics

zu nennen. Es gibt aber auch spezielle Kopplungssoftware, die unterschiedliche CAE-Programme, beispielsweise ABAQUS mit ANSYS Fluent, koppeln kann. Das SCC stellt dafür das Programm MpCCI zur Verfügung.

Ein Softwareprodukt, das in den verschiedensten Bereichen wie Mathematik, Grafik und Visualisierung, Mess- und Regelungstechnik, Statistik, Wirtschaftswissenschaft und Operations Research, Messdatenerfassung und -auswertung, Signalverarbeitung und neuronale Netze etc. eingesetzt wird, ist MATLAB/Simulink. Das SCC hat dafür eine TAH-Lizenz erworben, die es erlaubt, dass KIT-weit alle Mitarbeiter und Studierenden das Produkt auf ihren Arbeitsplatzrechnern und auch auf ihren für Forschung und Lehre genutzten Computern installieren dürfen. Mittlerweile hat sich MATLAB/Simulink nicht nur am KIT zur Basissoftware etabliert.

Im Bereich der technisch-wissenschaftlichen Datenanalyse und -visualisierung (2D und 3D) bietet das SCC das Produkt Origin Pro an. Das Programmpaket vereinigt leistungsfähige Datenanalysewerkzeuge, die freie Programmierbarkeit mit C, die Nutzung der NAG-Bibliothek und eine Vielzahl von manipulierbaren Diagrammtypen unter einer intuitiven und leicht erlernbaren Benutzeroberfläche.

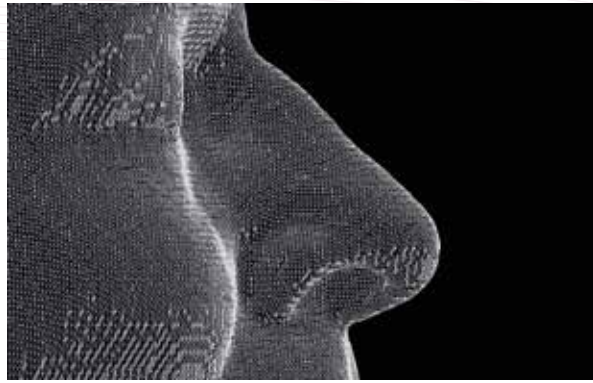
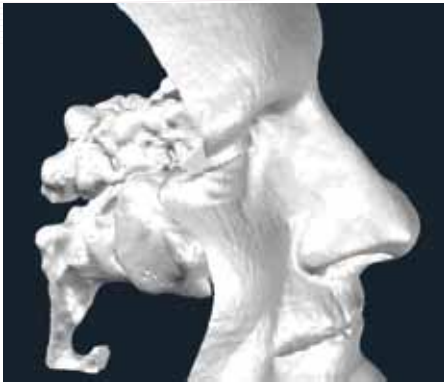


Abbildung 3: Flächenerzeugung aus CT-Daten.

Datenanalyse und Statistik

In diesen Bereich gehören die eigentlichen Statistikpakete sowie ein Programm zur Optimierung. Als Software zur Datenanalyse und für statistische Auswertungen stellt das SCC die folgenden Softwarepakete zur Verfügung:

- SAS
- SPSS
- Origin Pro.

Für die beiden Pakete SAS und SPSS gibt es eine Campuslizenz zur Installation der Software auf Instituts- und eigenen Rechnern in Verbindung mit einem vom SCC betriebenen Lizenzserver. Die Verteilung der Software erfolgt über den Softwareshop des KIT CS bei der Firma asknet.

Zur Arbeit mit Strukturgleichungen mit sehr vielen Variablen und für Optimierungen kann das folgende Programm genutzt werden:

- CPLEX.

Das SCC verfügt über eine Floating-Lizenz für 10 gleichzeitige Nutzer, die an Rechnern auf dem Campus Süd genutzt werden kann. Zusätzlich ist CPLEX an der HC3 des SCC installiert.

Wissenschaftliche Textverarbeitung

Zur Erstellung wissenschaftlicher Texte, vor allem im Zusammenhang mit mathematischen Formeln, bietet das SCC die frei verfügbare Software TeX auf den Windows-Rechnern und unter Linux an. Zum Arbeiten mit TeX stehen unter Windows und Linux Oberflächen zur Verfügung.

Computeralgebrasysteme

Neben den an numerischen Verfahren orientierten Softwaresystemen spielen die vorwiegend auf algebraischen und symbolischen Methoden basierenden Computeralgebrasysteme eine wichtige Rolle zur Lösung von Problemstellungen in der Mathematik, den Naturwissenschaften und den Ingenieurwissenschaften. Ein wesentlicher Vorteil gegenüber numerischen Verfahren ist beispielsweise, dass Lösungen für ganze Problemklassen ermittelt werden können und dass die Ergebnisse

exakt sind, das heißt, sie sind nicht mit Approximations- oder Rundungsfehlern behaftet. Im Laufe der Jahre haben sich viele Computeralgebrasysteme darüber hinaus zu mächtigen Problemlösungsumgebungen entwickelt, die neben den symbolischen auch numerische Verfahren implementieren. Auch bieten sie eine Fülle von Visualisierungsmöglichkeiten bis hin zu der Möglichkeit, vollständige und publikationsfähige wissenschaftliche Dokumente erstellen zu können. Aus der Vielzahl von professionellen Computeralgebrasystemen bietet das SCC die folgenden weit verbreiteten Produkte an:

- Maple
- Mathematica
- MathCAD.

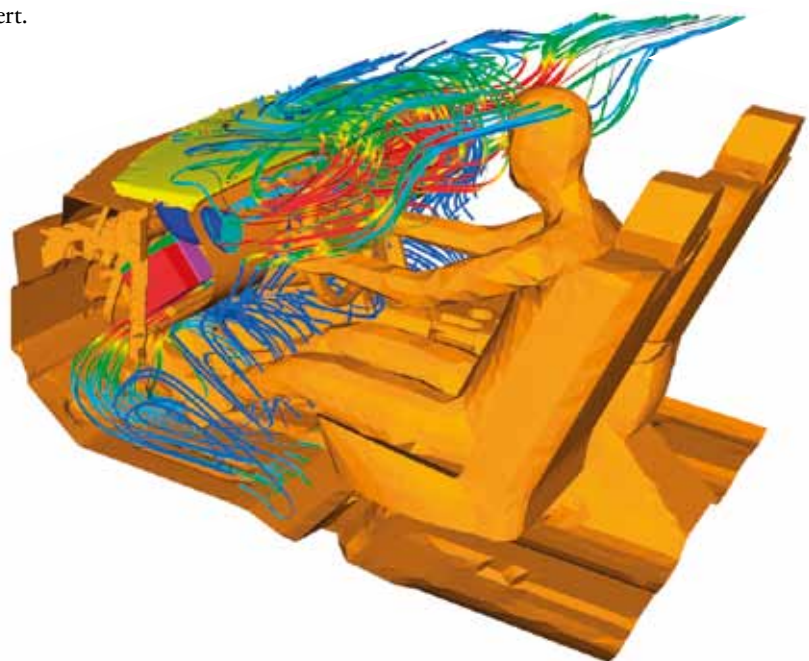


Abbildung 4: Klimatisierung in einem PKW (Daten: CD-adapco).

Visualisierung

Die Welt ist voll von Informationen, Simulationen, Experimenten und Datensammlungen mit ständig wachsendem Umfang. Daher stellt sich zunehmend das Problem, dass für teilweise riesige und unübersichtliche Datenmengen Möglichkeiten geschaffen werden müssen, um nicht direkt zugängliche Informationen finden zu können. Eine gute Möglichkeit hierfür bietet die Visualisierung. Sie bildet ein Werkzeug zur verdichteten Darstellung komplexer Informationen, unter Ausnutzung des sensorischen Apparates und der hoch entwickelten Aufnahmefähigkeit des Menschen.

Sie wird in vielen Szenarien benutzt, etwa für Design und Evaluierungszwecke in Industrie und Architektur und zur Präsentation digitaler Prototypen. Die meisten dieser Visualisierungen basieren auf Modellen von Industrieprozessen, Wirtschaftsmodellen oder mathematischen Modellen. Andere Anwendungen liegen im Bereich der Informations- oder Unterhaltungsmedien. Obwohl die Anwendungen und Anforderungen in diesen Bereichen unterschiedlich sind, werden in allen Szenarien zunehmend anspruchsvolle Visualisierungstechniken genutzt.

Für die Visualisierung von Daten stehen verschiedenste

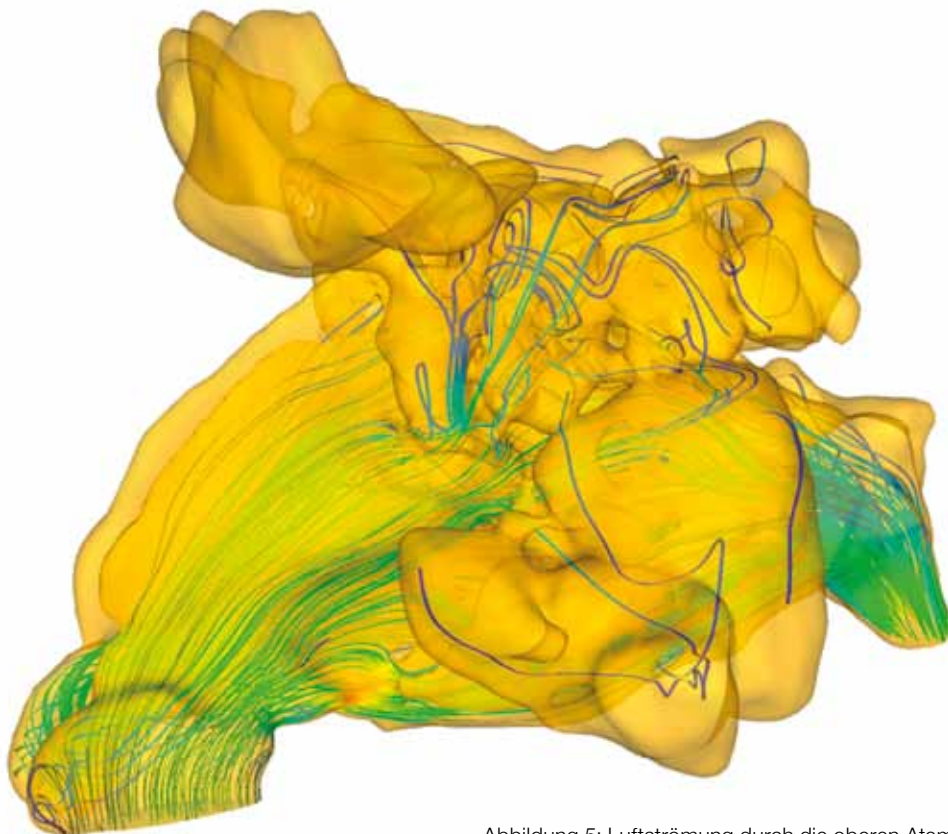


Abbildung 5: Luftströmung durch die oberen Atemwege (Daten: Dr. Paul Weber, Rolf Mayer).

Softwareprodukte zur Verfügung:

ArcGIS, ArcInfo, ArcView (ESRI)
EnSight
Origin Pro
Tecplot.

Das SCC betreibt u.a. in Kooperation mit dem Institut für Informationsmanagement im Ingenieurwesen (Prof. Dr. Dr.-Ing. Jivka Ovtcharova) die Rechnerinfrastruktur der stationären Großprojektion für hochauflösende, immersive Visualisierung des Lifecycle Engineering Solutions Center, kurz LESC.

Das LESC ist eine neuartige Forschungseinrichtung am KIT für die interdisziplinäre Zusammenarbeit zwischen internen und externen Forschungsinstituten sowie für den Wissensaustausch und Technologietransfer in Wirtschaft und Gesellschaft. Im LESC stehen skalierbare und flexible Entwicklungsumgebungen vom Einzelarbeitsplatz bis hin zur



Diskussion an einem studentischen Lehr-CAD-Modell.
Foto: SCC



Abbildung 6: Vorlesung gehalten von Prof. Dr. Wolfgang Reichel direkt aus Mexiko.

Großprojektion für innovative Ingenieurösungen im Rahmen des Product Lifecycle Management und Virtual Engineering zur Verfügung. Nähere Informationen unter <http://www.lesc.kit.edu/>.

Ansprechpartner: Rolf Mayer, Tel. 0721/608-46435, E-Mail: rolf.mayer@kit.edu; Dr. Paul Weber, Tel. 0721/608-44035, E-Mail: paul.weber@kit.edu.

Videokonferenzen, Teleteaching und Mediendienste

Eine Videokonferenz ist häufig der einfachste und preiswerteste Weg, um mit Partnern nicht nur audiovisuell zu kommunizieren, sondern auch um digital vorliegende Unterlagen zu präsentieren und gegebenenfalls gemeinsam zu bearbeiten.

Das SCC bietet eine Reihe von Lösungen für die unterschiedlichen Anforderungen von Diskussionen und Vorträgen. Je nachdem, wie viele Teilnehmer erreicht werden sollen und wie viel Zeit für die Vorbereitung bleibt, kann das SCC spontan ISDN- oder IP-Konferenzen in hoher Qualität aufsetzen. Dabei erfolgt auf Wunsch eine Moderation und Betreuung durch das Personal.

Neben den H.323-basierten Systemen bietet das SCC auch den Zugang zu einem Web-Conferencing System (Adobe Connect) an. Mit diesem Web-Browser-basierten System können Vorträge, Seminare und Projektsitzungen weltweit einfach und komfortabel durchgeführt werden.

Es können auch vorhandenen Filme, Animationen, Visualisierungen und Einzelbilder zu Filmen oder Animationen zusammgebaut werden. Diese Filme kann das SCC in digitaler Form (MPEG1, MPEG2, Quicktime, AVI, Real Media, Windows Media, etc.), in Standard Definition (SD) und in High Definition-Auflösung (HD) zur Verfügung stellen, aber auch Videobänder (VHS, DV), Video-CDs oder DVDs erstellen. Zuvor digitalisiertes Film- und Bildmaterial kann ohne sichtbaren Qualitätsverlust integriert werden.

Ansprechpartner: Rolf Mayer, Tel. 0721/608-46435, E-Mail: rolf.mayer@kit.edu

Prof. Dr. Rudolf Lohner, Rolf Mayer, Dr. Paul Weber

Softwareshop des KIT

Professionelle Anwendungen für Forschung und Lehre

Das KIT bietet Studierenden und Mitarbeitern ein breites Angebot an professionellen Softwarepaketen für Forschung und Lehre in Form von Campus- beziehungsweise Landeslizenzen. Dazu betreibt das KIT einen eigenen Online-Softwareshop. Dort können sich Studierende und Mitarbeiter mit ihrer Benutzerkennung und ihrem Passwort registrieren und so auf einfachste Weise Software im Rahmen der Campuslizenzen beziehen.

Eine Nutzung der Lizenzen für andere Zwecke, wie beispielsweise für kommerzielle Auftragsforschung oder Einsatzgebiete außerhalb von Forschung und Lehre ist in allen Lizenzverträgen dieser Art ausgeschlossen. Dafür müssen in der Regel gesonderte Lizenzvereinbarungen mit den Herstellern abgeschlossen werden.

Aus meiner Sicht, der eines Elektrotechnik-Studenten im zweiten Semester, weiß ich schon jetzt die gegebenen Möglichkeiten zu schätzen. Mein Studium hat mich bereits früh zu Beginn mit dem Standardtool für Ingenieure, MATLAB, vertraut gemacht. In einer Projektarbeit war es unsere Aufgabe, mithilfe von MATLAB und einem Spice-Programm eine Mikrofonverstärkerschaltung zu entwickeln, die Übertragungsfunktionen zu berechnen und die gesamten Ergebnisse mit MATLAB zu dokumentieren. Durch ein unterstützendes MATLAB-Tutorium wurden so schon am Anfang des ersten Semesters meines Studiums Grundsteine für die Bedienung und Programmierung dieser Software gelegt. Um aber gerade eine Standardsoftware wie MATLAB für aktuelle wissenschaftliche Probleme nutzbar zu machen, ist die MATLAB-Programmumgebung durch viele Toolboxen ergänzbar und passt sich den Problemstellungen flexibel an. Im Softwareshop des KIT steht die neueste Version von MATLAB im Release 2010b zur Verfügung, die weiterhin von Studierenden als auch von Mitarbeitern des KIT kostenfrei genutzt werden kann.

Die neue MATLAB-Version bietet unter anderem noch mehr Möglichkeiten zum automatisierten Image Processing und wurde hinsichtlich der Ausnutzung von Mehrkernprozessoren optimiert.

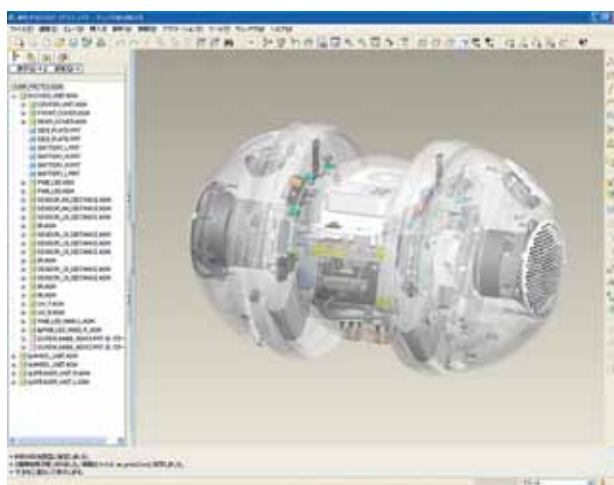


Abbildung 2: Konstruktion eines Produkts in Creo Elements/Pro.

Creo Elements/Pro: Flexibel entwickeln

Creo Elements/Pro (neuer Produktname von Pro/ENGINEER) ist eine volumenbasierte CAD-Software, die die komfortable Erstellung von 2D- und 3D-Konstruktionen erlaubt. Sie ist das zentrale Element einer kompletten Produktfamilie zur Produktentwicklung der Firma Parametric Technology Corporation (PTC).

Als Windows-Plattformen werden XP, Vista und Windows 7 unterstützt. Die Unterstützung von UNIX-Plattformen wurde eingestellt. Besonders hervorzuheben sind:

- eindeutiges 3D-Volumenmodell
- einheitliches Datenmodell von der Schnittdarstellung über die 2D-Zeichnung bis zur NC-Programmierung
- Änderungen in 3D werden sofort in 2D übernommen und umgekehrt
- Erstellung der Modelle durch fertigungsorientierte Funktionen, wie Bohren, Fräsen, Fasen usw.
- Kollisionsprüfungen
- Berechnungen von physikalischen Eigenschaften wie Masse, Schwerpunkt, Widerstandsmomente, Oberfläche

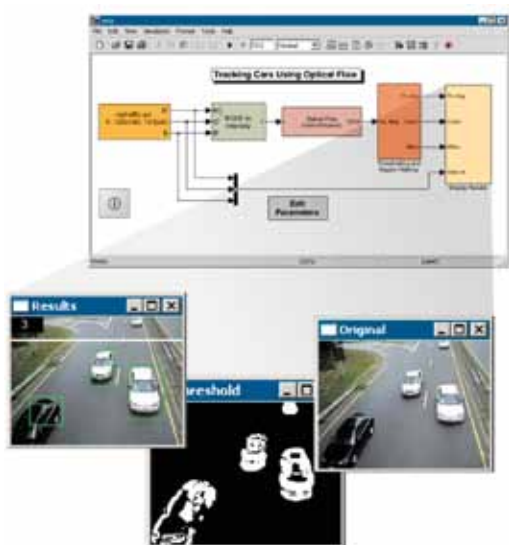


Abbildung 1: Verwendung der MATLAB Image Processing Toolbox zur Fahrzeugerkennung.

usw. Komfortable Konstruktionsverwaltung und -historie
· realitätsnahe, farbschattierte Modelldarstellung.

Creo Elements/Pro ist ein in der Industrie weit verbreitetes CAD-System mit besonderen Stärken im 3D-Entwurf, aber auch in der Produktion. Es wird bereits seit mehreren Jahren im Rahmen einer Campuslizenz eingesetzt. Die Software steht damit allen Einrichtungen am Campus Süd kostenlos zur Verfügung.

Nähere Informationen unter
<http://www.scc.kit.edu/produkte/3833.php>.

TecPlot: Daten anschaulich darstellen

Schon die ersten Aufgabenstellungen in einem Elektrotechnikstudium behandeln das Lösen von Problemstellungen, die durch viele Einflüsse und Faktoren geprägt sind. Oft ist es daher notwendig, gewisse Bedingungen oder aber auch Messdaten visuell aufzuarbeiten, um die enthaltenen Informationen sicht- und nutzbar zu machen. Genau für solche Aufgaben eignet sich die Profisoftware TecPlot. Weitere Informationen



Abbildung 3: Visualisierung an einem Getriebeteil durch TecPlot.

zur Nutzung beispielsweise in den Pools des SCC, finden sich unter <http://www.scc.kit.edu/produkte/3855.php>. Die Stärken dieser Software liegen vor allem in der visuellen Aufarbeitung großer Datenmengen, die während einer Simulation oder einer Messreihe entstanden sind. Mit TecPlot gelingt es auch, komplexe Strömungsdaten von Objekten zu simulieren und daraus Rückschlüsse für die Konstruktion zu ziehen. Besonders für spätere Ingenieure ist es nicht nur sehr wichtig, komplexe Problemstellungen verstehen und lösen zu können, sondern diese auch vereinfacht darzustellen, um beispielsweise Angestellten im Vertrieb einer Firma deutlich zu machen, an welchen Punkten die eigene Ingenieursarbeit anknüpft und welche Ziele verfolgt werden.

ArcGIS: Daten örtlich strukturieren

Eine weitere, sehr interessante Lizenz für die wissenschaftliche Arbeit an den Hochschulen des Landes Baden-Württemberg und somit auch an den Instituten des KIT sowie für die Ausbildung von Studierenden aus allen Fachbereichen und

insbesondere der Geowissenschaften und Geodäsie ist die Software ArcGIS von ESRI. ArcGIS ist eine ganze Sammlung von Programmen zur Zusammenstellung, Verarbeitung und Auswertung von ortsbezogenen Daten aller Art. Das Softwaremodul ArcInfo bietet für dieses Einsatzgebiet viele Werkzeuge für die Automatisierung von Prozessen zur räumlichen Modellierung und zur kartografischen Aufbereitung von Daten. Die Software erlaubt es, im Mehrbenutzerbetrieb eine gemeinsame Datenbank zur modellbasierten Datenerfassung zu erstellen und zu administrieren. Als typisches Anwendungsbeispiel lassen sich bestimmte Informationen über Gesteinseigenschaften an unterschiedlichen Orten eines Gebiets nennen, aber auch infrastrukturelle Daten, wie die Einwohnerdichte und Wirtschaftskraft können so bestimmten Orten zugewiesen werden. Mit ArcView bietet sich dann die Möglichkeit, die gesammelten Daten einfach zu erfassen und diese visuell darzustellen. Das Ergebnis sind interessante Modelle, die die Verteilung der Informationen darstellen.

Das ArcGIS-Softwarepaket steht als Landeslizenz allen Mitarbeitern des KIT sowie allen Studierenden seit dem 01. Januar 2011 kostenfrei für die Nutzung in Forschung und Lehre zur



Abbildung 4: Verwendung der Software ArcGIS zur Betrachtung infrastruktureller Daten.

Verfügung. Für Studierende gibt es derzeit noch eine spezielle Version, die über DVDs abgegeben wird, jedoch den vollen Funktionsumfang bietet. Künftig wird der Zugriff dann auch über den Softwareshop möglich sein.

Weitere Informationen unter
<http://www.scc.kit.edu/produkte/3864.php>.

Borland Together: Softwareentwicklung optimieren

Nicht nur im Elektrotechnikstudium ist die Softwareprogrammierung eine entscheidende Fertigkeit. Heutzutage sind fast alle technischen Geräte über mehr oder weniger komplexe Programmierung geregelt. Gerade weil Programmierarbeit ein so wichtiges Kapitel in der Ingenieurstätigkeit darstellt, sollte jedes mögliche Mittel zu Steigerung der Effizienz dieses Prozesses genutzt werden. Borland Together ist ein mächtiges UML-Tool und kann kostenfrei genutzt werden. Together übernimmt und automatisiert viele Teile der Softwareerstellung, erzeugt aus grafisch erstellten Designs im Hintergrund

Programmcode und erlaubt so eine noch effizientere und fehlerfreiere Programmierung. Gerade bei großen Projekten helfen die vielen Analysefunktionen bei der Erkennung und Vermeidung sich wiederholender Fehler.

Weitere Informationen unter
<http://www.scc.kit.edu/produkte/3868.php>.

Corel DRAW Graphics Suite X5: Designen und Bilder bearbeiten

Eine gute Bildbearbeitungs- und Designumgebung ist besonders für Studierende als Privatlizenz kaum erschwinglich. Studierende und Mitarbeiter des KIT können jedoch über eine Campuslizenz die CorelDRAW Graphics Suite nutzen. In der Lizenz sind alle wichtigen Tools zur Bearbeitung von Fotos und zur Erstellung von Digital Art und Animationen enthalten. Seit kurzer Zeit steht die neuste Version X5 im Softwareshop des KIT zur Verfügung und ist weiterhin in vollem Umfang kostenfrei nutzbar. Die Änderungen und Verbesserungen sind immens, sodass die Corel Suite durchaus mit den Adobe-Konkurrenzprodukten mithalten kann. Zum einen wurde die Geschwindigkeit der Software besonders in Hinblick auf Mehrkernprozessoren optimiert. Auf der anderen Seite vervollständigen neue Werkzeuge die große Palette von Anwendungen in CorelDRAW und PhotoPaint. So können mit der neuen Version X5 auch Flashanimationen erstellt werden.

Die CorelDRAW Graphics Suite bietet eine vollständige Grafik-, Design- sowie Layout-Umgebung und lässt mit Corel CONNECT auch den Wunsch nach einer Medienverwaltung zwischen den einzelnen Corel Programmen nicht offen.

Seit Wintersemester erweitertes Angebot

Seit Beginn des Wintersemesters 2010/2011 wurde das Softwareangebot erneut erweitert. Mit der Zitat- und Literaturverwaltungssoftware Endnote der Firma Thomson Reuters können Studierende und Mitarbeiter ein umfangreiches Tool

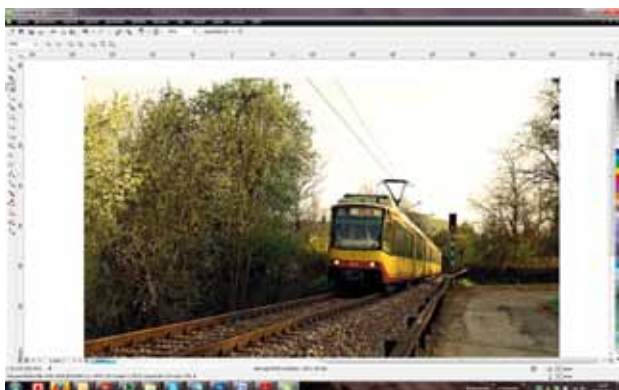


Abbildung 5: Bildbearbeitung mit der Corel Graphics Suite X5.

nutzen, das unter der Verwendung vieler verschiedener Literaturdatenbanken und Abbildungsverzeichnisse das Schreiben einer wissenschaftlichen Arbeit enorm vereinfacht. Doch nicht nur die Recherche in knapp 4.000 Fachdatenbanken zu verschiedenen Fachrichtungen und Bibliothekskatalogen steigert die Produktivität. Die Programmsuite erlaubt es, alle Daten zu Literatur, Zitaten und Belegen bequem zu verwal-

ten. Literaturverzeichnisse für die Bachelor-, Diplom- oder Masterarbeiten werden mit der Software in ihrer Struktur und Organisation sehr vereinfacht. Doch auch schon während der Recherche bietet Endnote mit seinen zahlreichen Datenbank-schnittstellen (beispielsweise zu Datenbanken wie PubMed oder Z39.50) eine wunderbare Informationsquelle und Ergänzung zur Literaturrecherche in der Bibliothek und im Internet.

Da viele Studien- und Forschungsarbeiten mit Microsoft Word geschrieben werden, ist es besonders praktisch, dass Endnote perfekt mit aktuellen Versionen von Microsoft Word zusammenarbeitet und sich die Vorteile der Software in den Workflow des Verfassers einbinden lassen. Anmerkungen zu den verwendeten Quellen müssen daher nicht nach, sondern können während des Schreibens direkt und förmlich richtig eingebunden werden. Die Arbeitsweise mit Endnote besteht darin, verschiedene Stapeldateien anzulegen, in der alle verwendeten Informationen einer Arbeit enthalten sind. Diese Dateien lassen sich durch Endnote sortieren, ändern, löschen, ergänzen.

Erwähnenswert ist zusätzlich die Kompatibilität der erstellten Endnote-Daten: Diese können von vielen anderen Literaturverwaltungsprogrammen importiert werden (zum Beispiel JabRef oder Zotero), sodass der Informationsaustausch zwischen Studierenden und Wissenschaftlern problemlos möglich ist. Die Software ist sowohl für Mac als auch für Windows in der momentan aktuellen Version X4 erhältlich und für Studierende und Mitarbeiter kostenfrei nutzbar.

Neben den Campuslizenzen stehen den Studierenden und Mitarbeitern des KIT auch verschiedene andere Softwarepakete zur Verfügung. Besonders erwähnenswert ist hier die MSDN Academic Alliance (MSDN AA). Dabei handelt es sich um ein Förderprogramm, das Studierenden und Mitarbeitern der teilnehmenden Einrichtungen eine kostenfreie Nutzung hochwertiger Microsoft-Software ermöglicht. Die aus dem MSDN AA-Programm bezogene Software ist sowohl auf den Rechnern des KIT als auch auf privaten Rechnern nutzbar.

Die Softwarepalette der MSDN Academic Alliance ist enorm: Microsoft Betriebssysteme wie Windows 7 Professional, Entwicklungsumgebungen wie Visual Studio 2010, Microsoft Viso, Microsoft Virtual PC und diverse Serveranwendungen sind nur wenige Beispiele. Um die Software zu beziehen, ist lediglich eine Anmeldung auf einer für diesen Zweck eingerichteten Webseite erforderlich. Nach der Authentifizierung erhalten Sie als Studierender oder Mitarbeiter des KIT einen persönlichen Zugang zur MSDN AA-Weboberfläche und können die gewünschte Software herunterladen.

Die Anmeldung zur Nutzung des MSDN AA-Software ist unter folgendem Link oder unter der Kategorie „Microsoft Academic Alliance“ im Softwareshop zu finden: <http://www.stud.uni-karlsruhe.de/~usath/elms/index.html>.

Patrick Schmich

Weitere Informationen

<https://rzunika.asknet.de/cgi-bin/home>.

100 Gigabit/s Ethernet erfolgreich zwischen FZJ und KIT getestet

Das deutsche Forschungsnetz (DFN), das Forschungszentrum Jülich (FZJ) und das SCC haben zusammen mit den Firmen Cisco Systems und Huawei Technologies neu entwickelte 100 Gigabit/s Ethernet-Komponenten unter realen Bedingungen auf einer Glasfaserverbindung zwischen dem FZJ und dem KIT erfolgreich getestet. Die Strecke zwischen Jülich und Karlsruhe ist weltweit eine der ersten Langstrecken-Verbindungen, die den kürzlich verabschiedeten IEEE 802.3ab-Standard für 100 Gigabit/s Ethernet verwendet.

Die Projektpartner evaluierten insbesondere die Kompatibilität der 100 Gigabit/s-Komponenten in einer heterogenen Umgebung. Die Firma GasLine stellte die Glasfaser-Infrastruktur mit einer Länge von 447 km bereit, um die 100GE-Fähigkeit der GasLine-Infrastruktur zu verifizieren. Huawei lieferte das DWDM-Equipment mit einem 100G-Transponder, der die Netzwerkstrecke zwischen dem FZJ und dem KIT mit zwei 56 Gigabit-Farbmodulationen beleuchtet, sowie den 100GE CFP Client Transceiver (CFP=100 Gigabit Small Form Factor Pluggable). Der CSR-3 Router von Cisco, der über eine „Backplane-Kapazität“ von 140Gbps verfügt, wurde über ein CFP gleicher Bauweise (LR4 (4x25G)) angeschlossen. Die qualitativ überraschend guten Testergebnisse zeigten, dass

die 100GE-Komponenten in hohem Maße geeignet sind, eine nachhaltige und verlässliche 100 Gigabit/s-Bandbreite für IP-basierte Anwendungen zur Verfügung zu stellen. Dies wurde durch UDP-Datenströme, die eine Bandbreite von bis zu 99,6Gbps erlauben, sowie TCP-Datenströme, die bis zu 94,6Gbps der 100Gbps nutzen, bestätigt. Die Ergebnisse dieser Tests sind insbesondere für die weitere Entwicklung von Netzwerkinfrastrukturen und speziellen Netzwerken im Bereich des High Performance Computing (HPC) von Bedeutung.

Bruno Hoefl

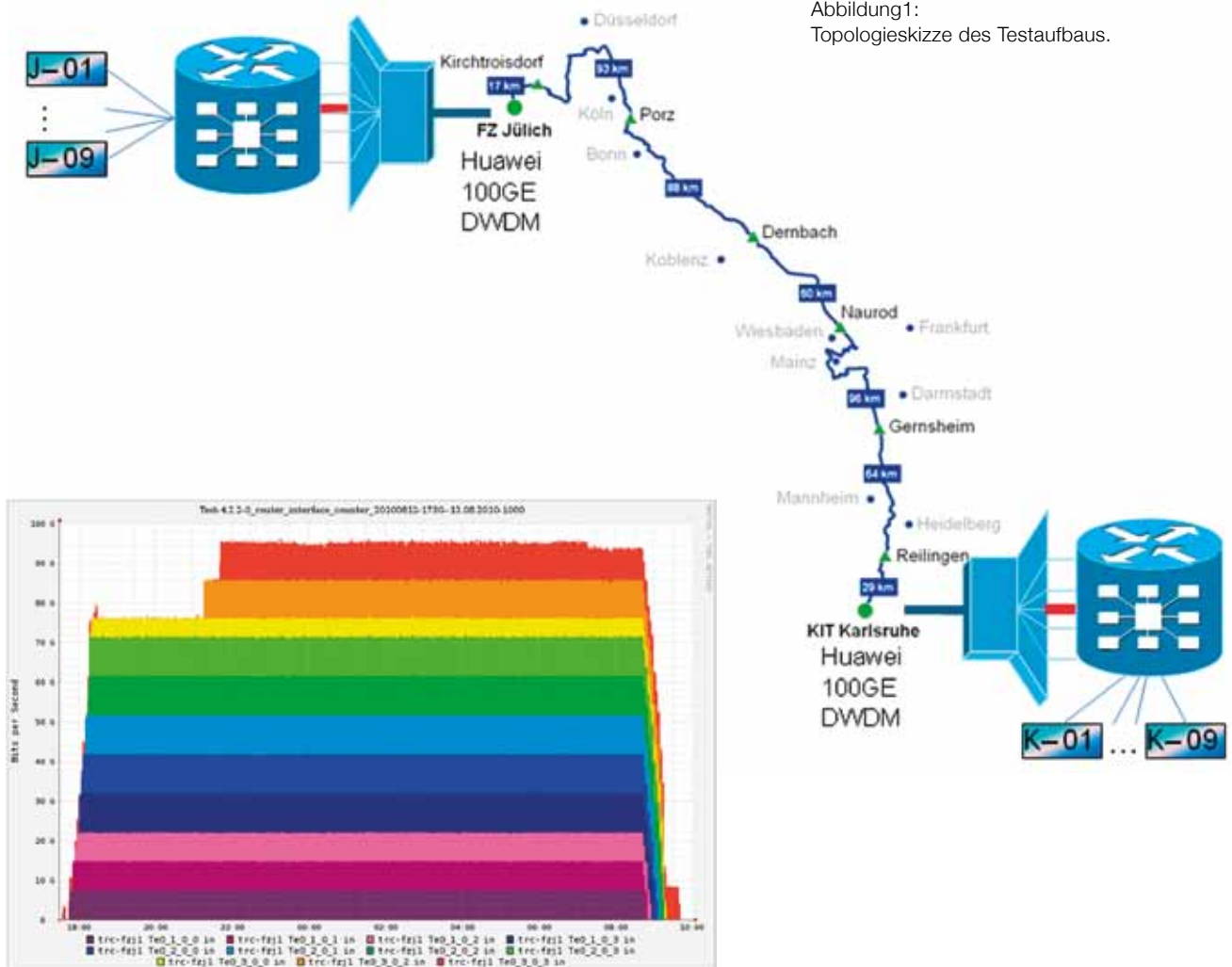


Abbildung1: Topologieskizze des Testaufbaus.

Abbildung 2: TCP-Durchsatz FZJ → KIT, 11 TCP-Datenströme.

Grid-Nutzung am KIT

Grid-Computing ermöglicht eine entfernte Nutzung verteilter heterogener Rechnersysteme und Speicherressourcen. Für den Zugriff auf Grid Services ist spezielle Software (die so genannte Middleware) erforderlich. Der Dienst „Grid-Zugang am KIT“ bietet eine vorkonfigurierte Arbeitsumgebung zur unmittelbaren Grid-Nutzung ohne aufwändige Software-Installation durch den Nutzer.

Wozu Grid?

Beginnen wir mit einem Beispiel: Bis zur Konferenz in sechs Wochen müssen sehr viele Simulationen gerechnet werden. Der Cluster im Institut ist nicht groß genug und wird auch von anderen Kollegen benötigt. Zum Glück gibt es noch die zentralen Systeme am SCC. Aber selbst mit dieser zusätzlichen Rechenleistung wird die Zeit knapp werden. Grid Computing bietet einen Ausweg aus solchen und ähnlichen Situationen. Statt auf einige wenige lokale Rechenressourcen angewiesen zu sein, stehen dem Grid-Benutzer nahezu beliebig viele Ressourcen verteilt in Deutschland und weltweit zur Verfügung. Um diese Infrastruktur zu verwenden, müssen Rechen-Jobs mit einer speziellen Software, der so genannten Grid-Middleware gestartet werden. Die Installation und Konfiguration der Middleware setzt allerdings spezielle Kenntnisse und etwas Erfahrung voraus. Um diese Hürde zu umgehen, stellt das SCC allen KIT-Angehörigen eine fertig konfigurierte Arbeitsumgebung (Grid User Interface) zur Verfügung.

Wie funktioniert das?

Praktisch alle gebräuchlichen Middleware-Systeme stellen dem Benutzer oder Anwendungsentwickler eine möglichst einheitliche Schnittstelle zum Ausführen von Berechnungen zur Verfügung. Die Besonderheiten der unterschiedlichen Hardware, Software und Batch-Job-Systeme werden verborgen. Das Starten von Jobs erfolgt über die Kommandozeile ganz ähnlich wie bei klassischen Batchsystemen oder durch grafische Benutzeroberflächen.

Für den Zugang zu Grid-Infrastrukturen benötigt man meist ein Grid-Benutzer-Zertifikat. Dabei handelt es sich um einen kryptographischen Ausweis, den eine spezielle Zertifizierungsstelle, die so genannte Certificate Authority (CA), ausstellt. Der Ausweis (das Zertifikat) dient dann als Identitätsnachweis des Benutzers im Grid. Grid-Zertifikate sind Bestandteil einer Public Key Infrastructure (PKI), das Zertifikat ist ein öffentlicher Schlüssel (public key) und nur der Besitzer des zugehörigen privaten Schlüssels (private key) kann seine Identität mit Hilfe des Zertifikats belegen. Je nach verwendeter Software werden Zertifikat und privater Schlüssel als einzelne Dateien gespeichert oder zusammen in einem so genannten Keystore. Um ein Grid-Zertifikat zu erhalten, wenden Sie sich an die GridKa-CA „GermanGrid“ (siehe weitere Informationen).

Um in einem Grid arbeiten zu können, müssen Sie zusätzlich Mitglied einer passenden Virtuellen Organisation (VO) sein. Für KIT-Angehörige kommt insbesondere eine Mitgliedschaft in der D-Grid-VO „bwGRiD“ infrage.

Für die Nutzung des SCC-Dienstes „Grid-Zugang am KIT“ melden Sie sich einfach per SSH auf „gridui.scc.kit.edu“ mit Ihrem normalen SCC-Benutzerkonto an. Dort können Sie die Middleware-Systeme Globus Toolkit und UNICORE nutzen. Bei den ersten Schritten im Grid hilft Ihnen die Online-Dokumentation auf dem System und auf den SCC-Webseiten.

Um selbst geschriebene Programme im Grid ausführen zu können, müssen diese natürlich auf die Zielplattform portiert werden. Häufig ist es ausreichend, das Programm auf dem Zielsystem neu zu übersetzen. Auf weit verbreiteten Plattformen wie Linux 64bit kann gegebenenfalls auch ein anderswo übersetztes Binär-Programm ausgeführt werden. Bei parallelen Programmen (MPI) oder speziellen Bibliotheken kann ein größerer Aufwand nötig sein. Für Beratung und Unterstützung bei solchen Vorhaben wenden Sie sich bitte an den SCC-Servicedesk mit dem Stichwort „Wissenschaftliches Rechnen in der Forschung“ oder direkt an die Abteilung Scientific Computing Labs (SCL).

Welche Grid-Ressourcen kann ich nutzen?

D-Grid

Das D-Grid wird seit 2005 vom Bundesministerium für Bildung und Forschung (BMBF) gefördert. Es besteht aus deutschlandweit verteilten Rechen- und Speicher-Ressourcen und wird von über 20 Virtuellen Organisationen genutzt. Wer D-Grid einmal ausprobieren möchte, kann in einer Test-VO (dgtest) Mitglied werden.

bwGRiD

Das bwGRiD ist eine dezentrale Ansammlung von Rechenknoten, die von acht Universitäten des Landes Baden-Württemberg betrieben werden und den Wissenschaftlern aller beteiligten Hochschulen zur Verfügung stehen. Die Parallelrechencluster werden im Rahmen der D-Grid Initiative zu einem bwGRiD-Verbund gekoppelt. Auf den Sites des bwGRiD ist eine große Anzahl von Anwendungssoftware-Paketen bereits vorinstalliert und kann mittels einheitlicher Module genutzt werden.

EGEE und EGI

EGEE (2004-2010) war ein in drei Phasen von der Europäischen Kommission gefördertes Projekt, in dem basierend auf Grid-Technologien Wissenschaftlern rund um die Uhr eine weltweit verteilte, produktive Infrastruktur zur Datenspeicherung und Auswertung zur Verfügung gestellt wird. Im Rahmen des EGEE-Betriebs wurde ein Produktionsgrid aufgebaut, das heute mit etwa 300 Rechenzentren, verteilt über 50 Länder, betrieben wird. Es bietet Ressourcen mit über 140.000 CPU-Kernen, die bis zu 330.000 Jobs pro Tag

verarbeiten. Das SCC trägt als regionales Operation Centre für die Region Deutschland maßgeblich zum Betrieb dieser Grid-Infrastruktur bei. Seit April 2010 werden die Aktivitäten zum Betrieb der Europäischen Grid Infrastruktur in die neu gegründete European Grid Initiative (EGI) überführt.

WLCG und GridKa

Am Europäischen Forschungszentrum CERN wird der Large Hadron Collider (LHC) betrieben. Die gigantischen Datenmengen, die die vier LHC-Experimente produzieren, werden im Worldwide LHC Computing Grid (WLCG) gespeichert und analysiert. Eines von 11 Herzen des WLCG ist das Grid Computing Centre Karlsruhe (GridKa), der Hauptknoten (Tier-1 Centre) für Deutschland und Mitteleuropa. Mit mehr als 10.000 CPU-Kernen, über 8.000 Terabyte an Festplattenspeicher und mehr als 10.000 Terabyte an assoziiertem Bandspeicherplatz ist GridKa zudem eine der größten Installationen seiner Art. Neben den vier LHC-Experimenten unterstützt GridKa Teilchenphysik-Experimente sowie weitere Communities innerhalb von EGI und D-Grid.

DEISA und PRACE

Distributed European Infrastructure for Supercomputing Applications (DEISA) ist ein von der EU gefördertes Konsortium von elf Supercomputing-Zentren in Europa. Die Zentren sind durch das GÉANT2-Netzwerk verbunden und stellen Grid-Dienste wie UNICORE zum Zugriff auf die Ressourcen zur Verfügung. Partnership for Advanced Computing in Europe (PRACE) ist ein Projekt zur Etablierung europaweiter HPC-Dienste im Exascale-Bereich.

Grid User Support

Probleme bei der Nutzung von bwGRiD, D-Grid oder europäischen Grid-Infrastrukturen (EGEE, EGI) sollten über den NGI-DE (nationale Grid Initiative für Deutschland)-Helpdesk gemeldet werden.

Foued Jrad, Dr. Olaf Schneider

Weitere Informationen

- SCC-Dienst „Grid-Zugang am KIT“: www.scc.kit.edu/dienste/5750.php
- GridKa-CA „GermanGrid“: gridka-ca-sec.fzk.de
- D-Grid User Portal: www.d-grid.de/userportal
- bwGRiD: www.bwgrid.de
- European Grid Initiative (EGI): www.egi.eu
- NGI-DE Helpdesk: helpdesk.ngi-de.eu

Was ist UNICORE?

UNICORE (Uniform Interface to Computing Resources) ist eine in Europa unter der Koordination des Forschungszentrums Jülich entwickelte Grid-Middleware auf Basis von Web-Services und offenen Standards. Durch die grafische Benutzeroberfläche und die enge Verzahnung von Client- und Serverkomponenten bietet es einen einfachen, sicheren und intuitiven Zugang zu verteilten Grid-Ressourcen wie Supercomputern und Clustern (www.unicore.eu).

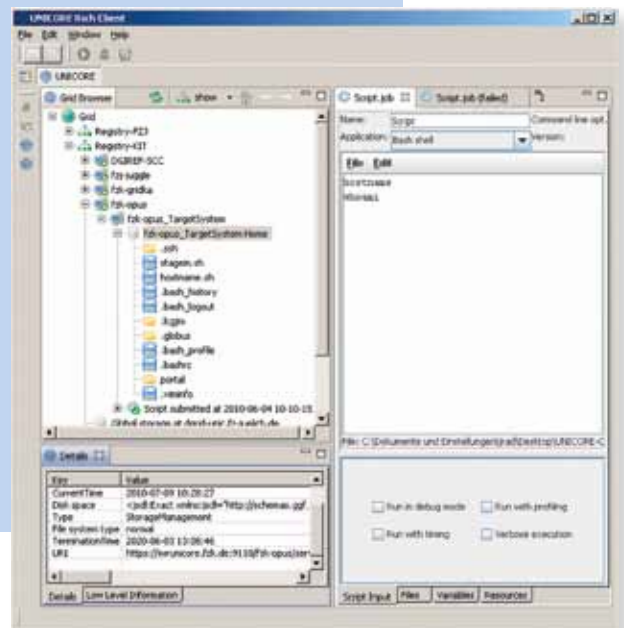


Abbildung 1: Unicore Client

TRANSITS – internationaler Workshop für Computernotfallteams am KIT

Kernel-Exploits, Identitätsdiebstahl, Einbruchserkennung - dies sind nur einige Schlagworte aus der täglichen Arbeit des Computernotfallteams des KIT. Die Reaktion auf Sicherheitsverletzungen und drohende Gefahren ist anspruchsvoll, denn diese muss nicht nur verschiedene technische und organisatorische Aspekte berücksichtigen, sondern auch in den meisten Fällen in kürzester Zeit erfolgen. Um solchen Anforderungen gerecht zu werden, müssen Mitarbeiter eines Notfallteams systematisch aus- und weitergebildet werden, um in kritischen Situationen adäquat reagieren zu können.

Die europäische Kommission hatte das Projekt TRANSITS (Training of Network Security Incidents Teams Staff) ins Leben gerufen, um die Ausbildung in diesem Bereich im europäischen Raum voranzutreiben und weiter zu entwickeln. Das Projekt wurde von 2002 bis 2005 in Zusammenarbeit mit TERENA (Trans-European Research and Education Networking Association) durchgeführt. Seit Beendigung der Projektphase wird das Programm von TERENA allein im Rahmen der Taskforce „TF-CSIRT“ dauerhaft fortgeführt. Die aktuellen Entwicklungen im Rahmen dieses Aus- und Weiterbildungskonzeptes sind so erfolgreich, dass voraussichtlich ab Herbst 2010 das Programm der TRANSITS2-Workshops starten kann. TRANSITS2 wird auf dem bisherigen TRANSITS-Workshop aufbauen und technische Aspekte vertiefen.

Damit alle Teilnehmer eines TRANSITS-Workshops die Möglichkeit haben, sämtliche Kursmodule zu besuchen, ist der Workshop modular aufgebaut. Auf diese Weise soll sichergestellt werden, dass jeder Teilnehmer die für die Arbeit in einem Notfallteam erforderlichen Basiskenntnisse erlernt hat. Die Module, die innerhalb des TRANSITS-Kurses gelehrt werden, sind:

- Organisatorische Aspekte eines CERT
- Warnungen und Warnmeldungen
- Technische Aspekte eines CERT
- Juristische Fragestellungen

Im Notfallteam des KIT haben alle Mitarbeiter mit Abschluss dieses Workshops, der vom 7. bis 8. September stattfand, eine TRANSITS-Zertifizierung erworben. Nach Evaluation der Bewertungsbögen war der Workshop am KIT mit 32 externen Teilnehmern aus insgesamt 14 verschiedenen Ländern außerordentlich erfolgreich.

Weitere Informationen

- TERENA: <http://www.terena.org>
- TRANSITS: <http://www.terena.org/activities/csirt-training/>
- TF-CSIRT: <http://www.terena.org/activities/tf-csirt/>
- KIT-CERT: <http://www.kit.edu/cert>

Adrian Wiedemann



Internationales Interesse: Die Teilnehmer und Teilnehmerinnen des Transit-Workshops.
Foto: SCC

SimLab Workshop für HPC-Nutzer



Dr. Paul Gibbon vom Forschungszentrum Jülich informierte über die Ausprägungen der SimLabs in Jülich.
Foto: Frank Schmitz

Um die Kompetenzen im Wissenschaftlichen Rechnen besser zu bündeln, wurden am KIT Simulation Laboratories (SimLabs) als neuartige Schnittstelle zwischen Anwendern und Betreibern von Hoch- und Höchstleistungsrechnern etabliert. Thematisch orientieren sich die SimLabs an den vier KIT-Zentren Energie, Klima und Umwelt, NanoMikro sowie Elementarteilchen- und Astroteilchenphysik und sind am SCC angesiedelt. Aufgabe der SimLabs ist die interdisziplinäre Forschung und Entwicklung, insbesondere die Ertüchtigung wissenschaftlicher Software zur effizienten Nutzung von Supercomputern und verteilten Systemen. Ende November letzten Jahres veranstaltete das SCC den ersten SimLab Workshop „SimLab@KIT“.

Der Workshop richtete sich hauptsächlich an die Nutzer von High Performance Computing (HPC). Das neuartige Konzept der SimLabs wird im Rahmen des Helmholtz-Programms „Supercomputing“ vom KIT gemeinsam mit dem Forschungszentrum Jülich umgesetzt. Prof. Dr. Wilfried Juling, Mitglied des erweiterten Präsidiums und Chief Science and Information Officer (CSO/CIO) des KIT, begrüßte die rund 60 Teilnehmer und betonte insbesondere die Bedeutung der interdisziplinären Zusammenarbeit im Bereich des wissenschaftlichen Rechnens. Nach Vorstellung der einzelnen SimLabs und ihres Leistungsportfolios startete der fachspezifische Teil. Die Beiträge bezogen sich unter anderem auf hochmoderne Rechnersysteme und Infrastrukturen in den wissenschaftlichen Rechenzentren wie auch auf Anwendungen und Forschungsarbeiten in verschiedenen Fachdisziplinen. So berichtete der Präsident des Deutschen Wetterdienstes, Prof. Dr. Gerhard Adrian, über die Herausforderungen bei der kurz- und langfristigen Wettervorhersage. Prof. Claus-Dieter Munz von der Universität Stuttgart hielt einen Vortrag zur Simulation von turbulenten Strömungen. Den Abschluss bildeten Vorträge zum KIT-Hochleistungsrechner HP XC3000 (HC3) sowie eine Führung durch die Rechnerräume des SCC. Ein zweiter Workshop zum Thema SimLabs soll noch in diesem Jahr angeboten werden.

Frank Schmitz

Weitere Informationen

www.scc.kit.edu/forschung/5960.php.

xchange - Auslandserfahrung für Auszubildende

Das von der EU geförderte Austauschprogramm xchange bietet Auszubildenden die Möglichkeit, ein vierwöchiges Praktikum in der Firma eines anderen Landes zu absolvieren. An dem Programm beteiligen sich sechs europäische Länder: Deutschland, Schweiz, Österreich, Italien, Liechtenstein und Frankreich.

Im Rahmen von xchange hatte die Hochschule Liechtenstein bereits 2008 eine Auszubildende des SCC aufgenommen, im Gegenzug konnte in diesem Jahr Andreas Risch von der Hochschule Liechtenstein ein Praktikum am SCC durchführen. Auch im nächsten Jahr sollen Auszubildende am SCC die Möglichkeit erhalten, an diesem Programm teilzunehmen. Mit über 200 mitwirkenden Auszubildenden konnte in diesem Jahr ein neuer Rekord aufgestellt werden.

Petra Spanger



Freudige Gesichter: Im Rahmen der Abschlussfeier im November 2010 erhielten die Absolventinnen und Absolventen ein Zertifikat der Projektträger.

Foto: Landespressestelle Vorarlberg

Weitere Informationen

www.xchange-info.net

OE-Betreuer – Ansprechpartner für IT-Beauftragte

Um die Zusammenarbeit, den Informationsfluss und das Verständnis zwischen den Organisationseinheiten des KIT und dem SCC zu verbessern und auszubauen, stehen insbesondere den IT-Beauftragten persönliche Ansprechpartner als so genannte OE-Betreuer zur Verfügung. Diese bieten Hilfestellung u. a. in folgenden Fällen an:

- wenn das SCC-Serviceangebot die Kundenanforderungen nicht abdeckt
- bei Unzufriedenheit mit der Dienstqualität
- bei Fragen im Hinblick auf Zuständigkeiten und Ansprechpartner innerhalb des SCC
- bei Projekten oder sonstigen gemeinsamen Vorhaben mit dem SCC.

In der Vergangenheit nutzten bereits viele Organisationseinheiten persönliche Ansprechpartner des SCC im Sinne einer umfassenden Betreuung. Da der Zusammenschluss von Forschungszentrum und Universität Karlsruhe zum KIT aber auch in neuen Strukturen und Organisationseinheiten resultiert, die optimal von den Dienstleistungen des SCC profitieren sollen, hat das SCC diese zusätzliche Kontaktmöglichkeit etabliert. Hiermit möchte das SCC seine "Kunden" ganz individuell unterstützen und einbinden.

Daher sind die OE-Betreuer insbesondere an den Erfahrungen der Benutzer mit dem Serviceangebot des SCC interessiert: Wie werden die Dienste beurteilt, wo sind Verbesserungen, Veränderungen oder Ergänzungen notwendig? Auf Grund dieser Anregungen werden die OE-Betreuer gemeinsam mit SCC-Kolleginnen und -Kollegen an der nachhaltigen Verbesserung von Problemsituationen arbeiten, die IT-Betreuer in die SCC-Prozesse mit einbeziehen und partnerschaftlich Lösungen erarbeiten.

Seit Juni 2010 stehen den OEs des KIT zusätzlich folgende Personen als OE-Betreuer zur Verfügung:

- Horst Gernert, CS 46422
- Achim Grindler, CN 24506
- Alexander Helget, CN 24509
- Rolf Mayer, CS 46435
- Ulrich Weiss, CS 44868

Die Kommunikation mit dem SCC kann und soll natürlich auch weiterhin über bestehende und gut funktionierende Kontakte geführt werden. Für Anfragen zu einzelnen Diensten stehen die Verantwortlichen jederzeit zur Verfügung.

Für alle Nutzer des SCC-Serviceangebots nimmt der zentrale Service Desk Anfragen und Probleme unter der Telefonnummer - 8000 bzw. servicedesk@scc.kit.edu entgegen.

Achim Grindler

Weitere Informationen

<http://www.scc.kit.edu/dienste/>



Horst Gernert
Fotos: Privat



Achim Grindler



Alexander Helget



Rolf Mayer



Ulrich Weiss

SCC auf internationaler Supercomputing Conference SC10

Rekordwerte und heiße Trends

Seit 1988 findet alljährlich im November die Supercomputing Conference (SC) in den USA statt. Sie ist zusammen mit der in Deutschland bereits seit 1986 jeweils im Juni veranstalteten International Supercomputing Conference (ISC) die weltweit wichtigste Konferenz im Bereich des High Performance Computing. Fachleute aus anerkannten Supercomputing-Zentren sowie führende Hard- und Softwareanbieter präsentieren dort modernste Anwendungen, Architekturen und Trends.

Mit über 10.000 Teilnehmern und ca. 35.000 m² Ausstellungsfläche wurden wieder Rekordwerte erreicht. Das SCC war auf der Konferenz in New Orleans mit einem eigenen Stand vertreten, um Kontakte zu knüpfen und eigene Entwicklungen zu präsentieren.

Gemeinsam mit dem Forschungszentrum Jülich (FZJ) wurde auf dem Stand des SCC eine Vortragsreihe über eigene Entwicklungen veranstaltet. Themen waren dabei u.a. die Simulation Laboratories (SimLabs), die Large Scale Data Facility (LSDF), High Performance Computing as a Service, das Grid Computing Centre Karlsruhe (GridKa) sowie der erfolgreiche Test der 100GE-Verbindung zwischen dem SCC und dem Forschungszentrum Jülich.

Auf beiden Konferenzen wird regelmäßig die Liste der 500 leistungsfähigsten Supercomputercomputer bekanntgegeben (<http://www.top500.org>), die seit vielen Jahren die unangefochtene Messlatte darstellt, welche Institutionen und Hersteller im Höchstleistungsrechnen die Spitzenpositionen einnehmen. Gerade anhand dieser Liste lassen sich viele Trends und Technologien ablesen und vorhersagen, die in der Vergangenheit und möglicherweise in der nahen Zukunft die Schwerpunkte des Höchstleistungsrechnens ausmachen.

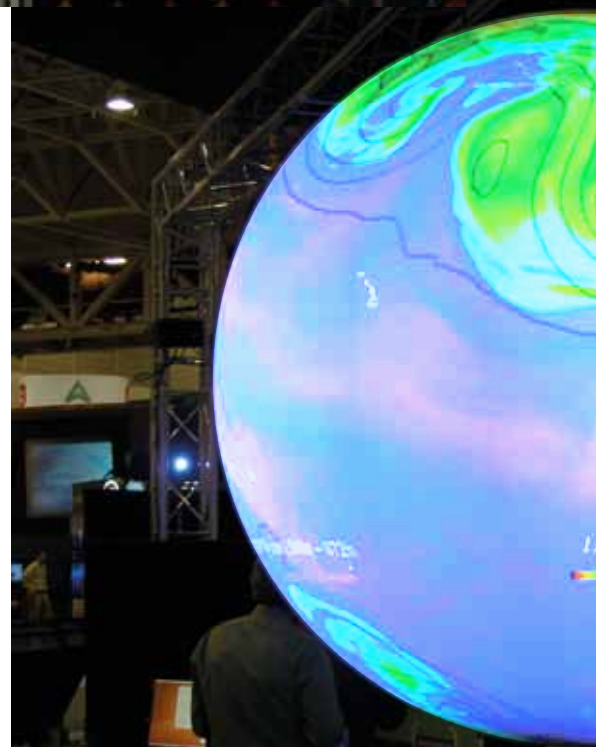
Hohe Linpack-Leistung durch Arithmetik-Acceleratoren

Bei dem Linpack-Benchmark, der dieser Top500-Liste zugrunde liegt, ist ein lineares Gleichungssystem mit voll besetzter Matrix zu lösen. Ein heißer Trend, der sich seit etwa zwei Jahren dabei abzeichnet, ist die Tatsache, dass eine besonders hohe Linpack-Leistung durch den Einsatz von Arithmetik-Acceleratoren erreicht wird. Diese Acceleratoren sind meist Grafikkarten (GPUs) der Firmen Nvidia und AMD/ATI, die durch einen extrem hohen Parallelisierungsgrad der Gleitkommaberechnungen eine sehr hohe Rechenleistung erzielen. Diese wird allerdings nur bei dafür geeigneten Anwendungen erreicht (wie etwa beim Linpack).

Paradigmenwechsel bei der Chip-Programmierung

Auch wenn es sich in den Jahren seit ca. 2000 so abgezeichnet hat, dass die Supercomputerarchitektur im Wesentlichen einen Cluster aus quasi Standard-PCs mit einem adäquaten schnellen Netzwerk als Interconnect darstellt, bei der zudem eine regelmäßige Leistungssteigerung durch die ständig steigende Taktfrequenzen der CPUs erwartet werden kann, zeigt sich die Situation heute wesentlich differenzierter.

Die durch „Moore’s Law“, d.h. die Verdoppelung der Anzahl von Transistoren pro Flächeneinheit auf einem Chip alle 18 Monate, mögliche Geschwindigkeitssteigerung der CPUs, die bisher meist durch Erhöhung der Taktfrequenzen stattfand, stößt mit dem dadurch enorm steigenden Energieverbrauch an ihre Grenzen. Stattdessen muss die zusätzlich mögliche Leistung nun in weitere Cores mit leicht reduzierter Taktfrequenz



Stand des SCC

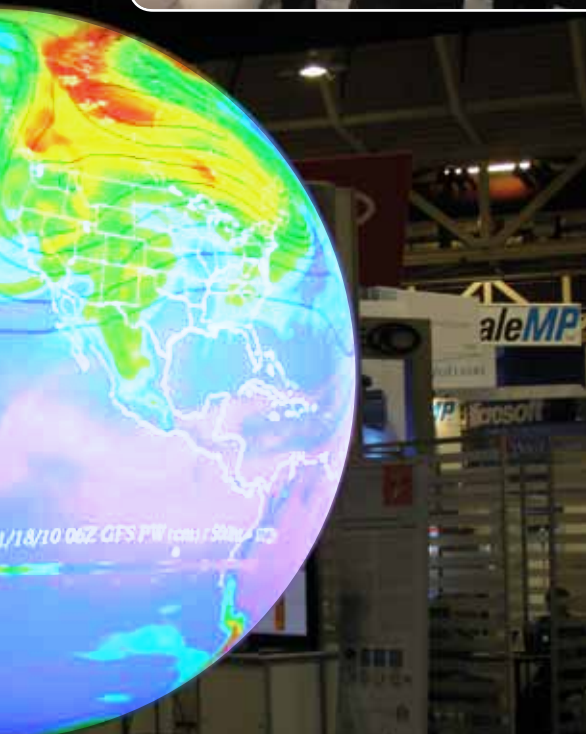


auf einem Chip realisiert werden. Diese Erhöhung der Anzahl von Cores pro CPU bedingt auch einen Paradigmenwechsel beim Programmieren dieser Chips: Eine Steigerung der Geschwindigkeit ist nicht mehr durch steigende Taktfrequenzen erreichbar, sondern nur durch Parallelisierung der Anwendungen für diese neuen Multicore-Chips.



Für das High Performance Computing bedeutet dies insgesamt, dass das bisher vorherrschende Message Passing-Modell mit Hilfe von MPI einem komplexeren Programmiermodell weichen muss (oder zumindest damit ergänzt werden muss), welches sowohl Multicore-Chips als auch potenziell vorhandene Acceleratoren wie GPUs möglichst optimal berücksichtigt muss.

Nicht unerwähnt bleiben soll an dieser Stelle, dass auch weitere Formen von Acceleratoren interessant sind, wie FPGAs (Field Programmable Gate Arrays), die es erlauben, Algorithmen dynamisch und zur Laufzeit in Hardware zu „gießen“ und damit bei speziellen Anwendungen vielfache Geschwindigkeitssteigerungen im Vergleich zu herkömmlichen CPUs zu erzielen. Ein herausragender Vertreter dieses Ansatzes ist die Firma Convey. In den nächsten Jahren werden sich somit umfangreiche Änderungen ergeben, um immer leistungsfähigere Simulationen realisieren zu können. Insgesamt müssen viele Anwendungen umgeschrieben werden, neue Algorithmen sind erforderlich und auch die Systemsoftware muss weiterentwickelt werden. Dass dieser Prozess bereits begonnen hat, zeigt sehr konkret die aktuelle Top500-Liste, in der sowohl Systeme mit mehreren hunderttausend klassischen CPU Cores wie auch immer mehr heterogene Systeme mit GPUs führend vertreten sind.



Neue Konsortien zur Weiterentwicklung des Open Source Lustre

Ein anderes wichtiges Thema betraf das parallele Dateisystem Lustre, das derzeit im HPC-Bereich am weitesten verbreitet ist und auch am SCC seit Jahren auf verschiedenen Systemen eingesetzt wird. Lustre steht unter einer Open Source-Lizenz und wurde bisher im Wesentlichen von der Firma Sun weiterentwickelt. Nach der Übernahme von Sun durch die Firma Oracle wurde ein Strategiewechsel angekündigt, der u.a. beinhaltete, Teile der Weiterentwicklung nicht mehr unter Open Source zur Verfügung zu stellen. Aus diesem Grund sahen einige Lustre-Lieferanten und -Nutzer Handlungsbedarf, um weiterhin ein offenes System mit Unterstützung durch verschiedene Hersteller zu gewährleisten und nicht von einem Hersteller abhängig zu sein. Im Rahmen der Supercomputing Conference haben sich nicht weniger als drei neue Konsortien vorgestellt, die sich gebildet hatten, um die Weiterentwicklung des Open Source Lustre zu fördern. Es bleibt derzeit spannend, wie sich das Verhältnis zwischen der Firma Oracle und den verschiedenen Lustre-Lieferanten auf lange Sicht ausprägen wird. Jedenfalls haben viele Organisationen bereits erkannt, dass parallele Dateisysteme eine wichtige und kritische Komponente von HPC-Systemen sind und dass die Verfügbarkeit eines offenen Produkts gefördert werden muss.

Prof. Dr. Rudolf Lohner, Roland Laifer



Fotos: SCC



Einführungskurs LaTeX

LaTeX ist ein Makropaket zu TeX, mit dessen Hilfe auf relativ einfache Weise Dokumente mit umfangreichen mathematischen Formeln, Abbildungen und Querverweisen erstellt werden können.

In der Zeit vom 14. bis 18.3.2011 findet ein Einführungskurs in LaTeX statt. Ziel des Kurses ist es, LaTeX kennenzulernen und das Erstellen von Texten mit Hilfe von LaTeX zu erlernen. Die Übungen zum Kurs finden wahlweise unter Windows oder Linux unter Verwendung einer für LaTeX geeigneten Entwicklungsumgebung statt. Für die Teilnahme am Kurs sind Vorkenntnisse nicht erforderlich. Die im Kurs erworbenen Kenntnisse können bei der Textverarbeitung an PCs ebenso angewendet werden wie an anderen Rechnern.

Kursbeginn: 14.3.2011, 9.00 Uhr
Kursende: Freitag, 18.3.2011, 17.00 Uhr
Ort: SCC-Süd, Raum 217, Übungen im Raum 114.2
Anmeldung: BIT8000

Behandelte Themen:

- Allgemeine Informationen über TeX und LaTeX
- Genereller Aufbau und Gliederung eines Dokumentes
- Die vordefinierten Dokumenttypen und Änderungen des Layouts
- Standardschriften und die Verwendung zusätzlicher Schriften
- Silbentrennung, Umlaute und scharfes S
- Listen, Tabellen, Zitate, Fußnoten
- Einfache Grafiken
- Inhaltsverzeichnis, Literaturverzeichnis und weitere Verzeichnisse
- Setzen mathematischer Formeln.

Der Kurs findet täglich in den Zeiten 9.00 bis 12.00 Uhr und 14.00 bis 17.00 Uhr im Raum 217 (Vorträge) bzw. 114.2 (L-Pool, Übungen) des SCC-Süd statt. Die Teilnehmerzahl ist auf 25 begrenzt. Zur Anmeldung liegen vorbereitete Listen beim BIT 8000 im Foyer des SCC-Süd aus (Tel. -8000, E-Mail: bit8000@scc.kit.edu).

Dr. Klaus Braune

Weitere Informationen

zu TeX und zum LaTeX-Kurs unter
<http://www.scc.kit.edu/produkte/3967.php>

Literatur

F. Mittelbach, M. Goossens, J. Braams, D. Carlisle, C. Rowley:
Der LaTeX-Begleiter, 2. überarbeitete und erweiterte Auflage, Pearson Studium, 2005, ISBN 3-8273-7166-X

L. Lamport: The LaTeX Document Preparation System,
2. Auflage, AddisonWesley USA, 1994, ISBN 0-201-52983-1

H. Partl, E. Schlegl, I. Hyna: LaTeX-Kurzbeschreibung, im Rahmen
der verschiedenen TeX-Installationen als LaTeX-Datei verfügbar.



Steinbuch Centre
for Computing

Steinbuch Centre for Computing (SCC)
76128 Karlsruhe
Tel: 0721/608-43754 oder 07247/82-25601
E-Mail: contact@scc.kit.edu

www.scc.kit.edu