



universität
wien

Einführung in Anwendungsgebiete der Informatik – Wintersemester 2012/13

Scientific Computing
5. Oktober 2012

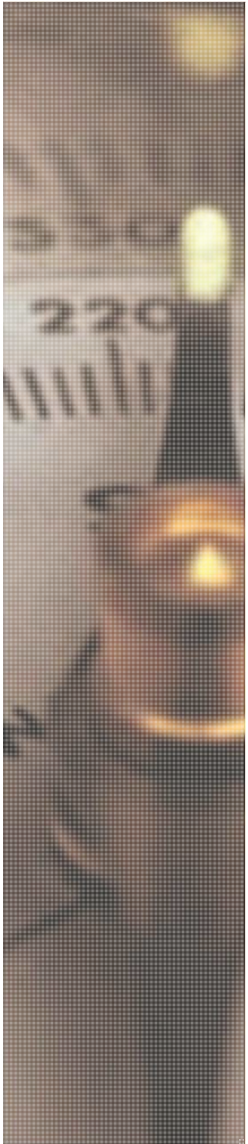
Univ.-Prof. Dipl.-Ing. Dr. Siegfried Benkner
Assoc. Prof. Dipl.-Ing. Dr. Wilfried Gansterer
Fakultät für Informatik
Universität Wien

www.cs.univie.ac.at



Scientific Computing – Orientierung zum Studium

- Inhalte des Studiums
- Anwendungsbereiche des Scientific Computing
- Technologien und Infrastrukturen
- Berufsbilder
- Organisatorisches, Fragen





Scientific Computing – Inhalte

Computing (Computermodellierung, -simulation, etc.) ist die dritte Säule des Erkenntnisgewinns (neben Theorie und Experiment) in verschiedensten Wissenschaftsbereichen (Chemie, Physik, Biologie, Astronomie, etc)

- *Algorithmen*
 - Komplexität, effiziente Implementierung, Datenorganisation, etc.
- *Software*
 - effizientes Abbilden von Algorithmen auf Hardware
- *Hardware*
 - Konzepte und Eigenschaften aktueller und zukünftiger Hardwareplattformen (der „Formel 1“ unter den Computersystemen)
- *Infrastrukturen, Technologien*

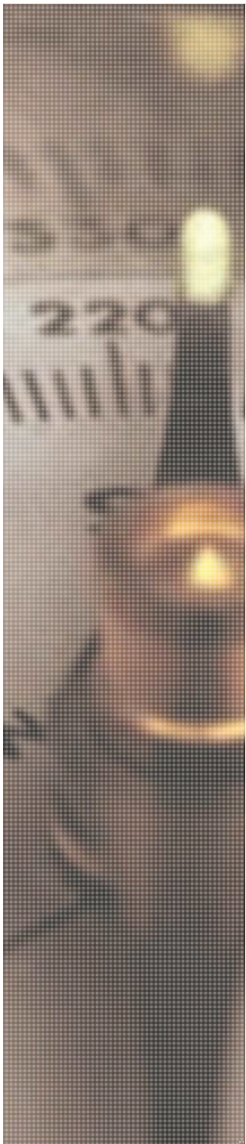


Aktuelle Herausforderungen

„Big Data“ (Schätzungen)

- 224 Giga(10^9)byte ... neue Wetterdaten pro Tag im National Climatic Data Center (USA) [2012]
- 20 Tera(10^{12})byte ... Foto upload auf Facebook/Monat [2008]
- 120 Terabyte ... Datensammlung des Hubble Teleskops [2008]
- > 1.2 Peta(10^{15})byte ... gesamte Wetterdaten im NCDC [2008]
- 10 Petabyte ... alle Videos auf YouTube [2011]
- 24 Petabyte ... von Google pro Tag verarbeitete Datenmenge [2009]
- 200 Petabyte ... am CERN gesammelte Kollisionsdaten [2012]

Permanent wachsend: Seit 2007 generieren wir pro Jahr mehr neue Daten als alle bis dahin existierenden Daten!

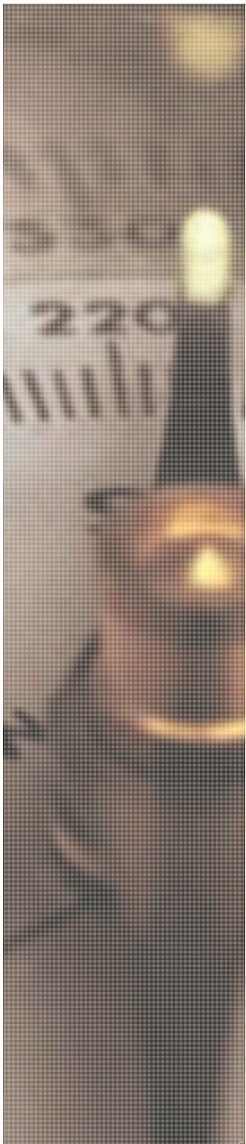




Aktuelle Herausforderungen

Hardwareentwicklungen

- *Multi-/Manycore*: massiver Parallelismus auf mehreren Ebenen (in einem Chip, zwischen Chips, zwischen Knoten, etc.)
- Verstärkte Integration von spezialisierten Chips, teilweise als „Beschleuniger“ (GPUs, FPGAs, etc.)
- „Exascale“
- Immer stärkeres Auseinanderklaffen zwischen theoretischer Maximalleistung der Hardware (die ungebrochen rasch anwächst) und praktisch erzielter Leistung

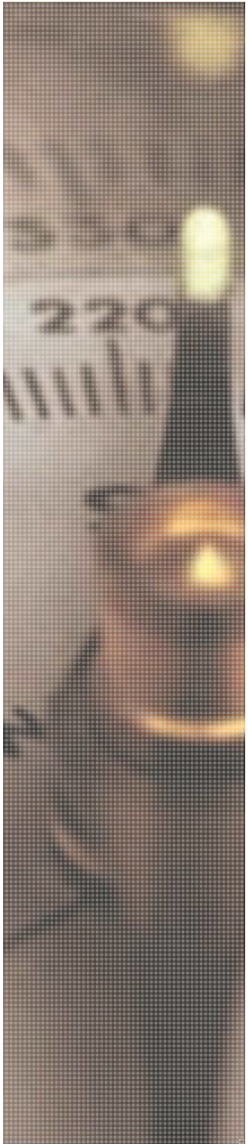




Aktuelle Herausforderungen

Algorithmen

- Massiv parallel
- Wenig Synchronisation
- Nur lokale Interaktion
- Fehlertoleranz

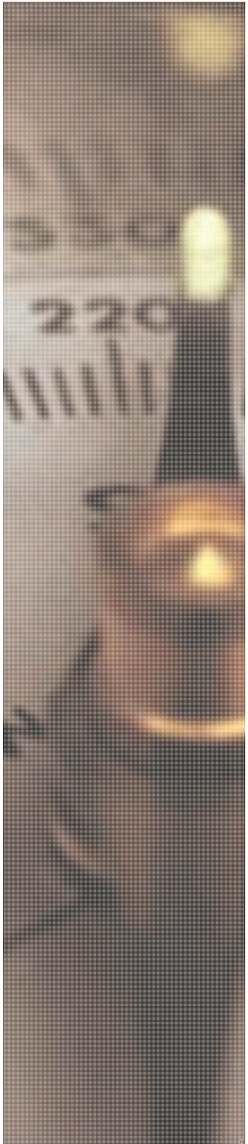




Aktuelle Herausforderungen

Software

Wie können wir Software und Toolkits für die effiziente Verwendung der immer komplexer werdenden Hardwareinfrastruktur entwickeln?

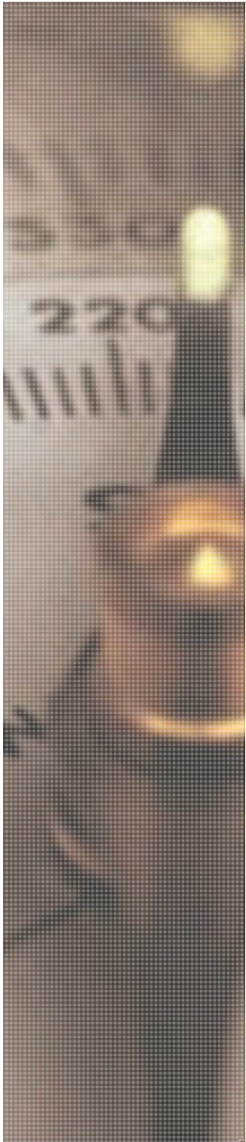




Aktuelle Herausforderungen

Multi-/Interdisziplinarität

- Problemstellungen, die im Scientific Computing auftreten, betreffen üblicherweise mehrere Wissenschaftsdisziplinen (*Multidisziplinarität*)
 - Unterschiedliche Kulturen
 - Unterschiedliche Terminologien
 - Unterschiedliche Organisationsstrukturen
- ➔ Kommunikation und Interaktion besonders wichtig !





Anwendungsbereiche des Scientific Computing

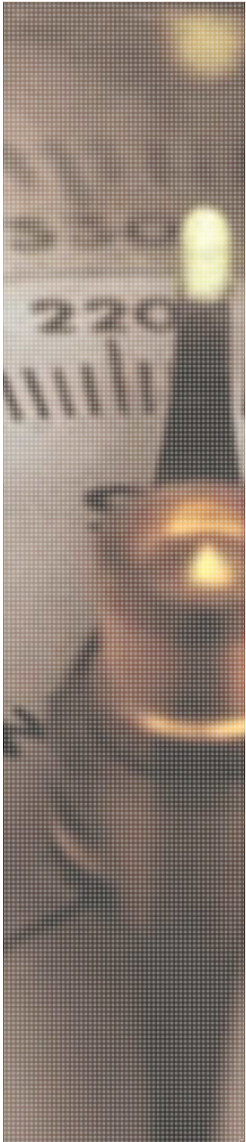
Es gibt viele Bereiche, wo Methoden des Scientific Computing eine (oft sehr wichtige) Rolle spielen

→ Tendenz steigend!

Unterstützung und Ergänzung von Theorie und Experiment

Grundlegende Arten von Anwendungsbereichen:

- Modellbasiert – datenbasiert
- Makroskopisch – mikroskopisch





Anwendungsbeispiele

Die folgende Sammlung von Anwendungsbeispielen erhebt **keinen Anspruch auf Vollständigkeit**

Gewählte Reihenfolge:

- zuerst allgemeine Beispiele
- gegen das Ende diejenigen Anwendungsbereiche, die (derzeit) im Studium vertreten sind bzw. in denen es Forschungsaktivitäten an der Fakultät für Informatik gibt

Im Rahmen des Studiums vertiefen Sie Ihr Wissen in einem Anwendungsfach Ihrer Wahl (z.B. Computational Life Sciences, Molecular Modeling, etc.)

Die unterrichtete *Methodik* ist aber in vielen Bereichen anwendbar!



Anwendungen Scientific Computing (Beispiele)

Wetter-, Klimaforschung

Wettervorhersage, Tornadovorhersage, Klimaveränderungen, ...

→ Experiment nicht möglich oder zu kostspielig

Beispiele: Earth Simulator, Simulation Supertwister

Geowissenschaften

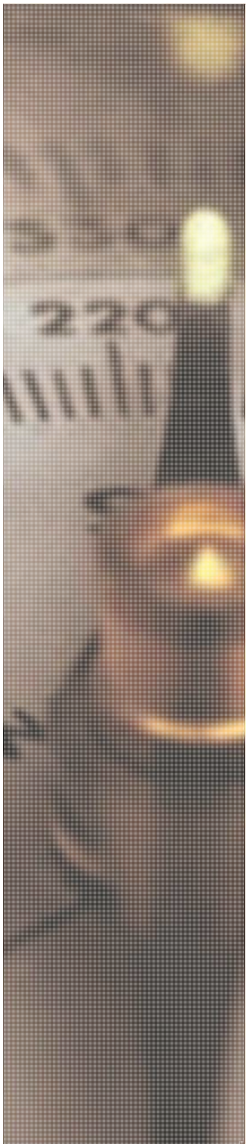
Erdbebenmonitoring und –vorhersage, ...

Strömungsmechanik

Luftwiderstand (Ergänzung/Ersatz Windkanal), Turbulenzen, Schockwellen, Schadstoffausbreitung, Simulation von Verbrennungsmotoren, ...

Physik

Materialwissenschaften (z.B. Entwicklung von supraleitenden Materialien), Struktureigenschaften, magnetische Eigenschaften, etc. von Materialien, ...





Anwendungen Scientific Computing (Beispiele)

Struktur- und Designberechnungen

Belastungen, Deformationen, Bruchstellen, ...

→ Experiment zu kostspielig oder sehr aufwändig

Beispiele: Flugzeugbau, Simulationen von Crashtests

Ingenieurwissenschaften

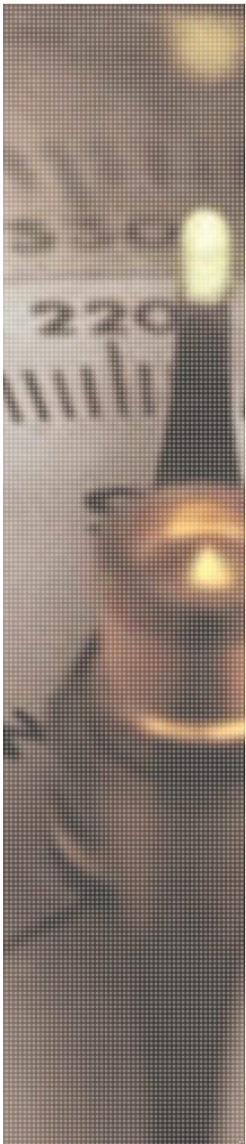
Verkehrsinfrastrukturplanung, Verkehrsmodellierung, erneuerbare Energien, Simulationen zur Lokalisierung von Ölreserven, ...

Wirtschaft, Unternehmungen

Vorhersage Wirtschaftsentwicklung, Management komplexer Informationsflüsse, Materialien, Finanzen, Logistik, supply chains, ...

Finanzwirtschaft

Makro- und mikroskopische Analyse und Vorhersage für Finanzmärkte, ...





Anwendungen Scientific Computing (Beispiele)

Astronomie

Entwicklung von Galaxien, Sternen, ...

→ nicht experimentell reproduzierbar, schwer meßbar

Biologie, Medizin

Entschlüsselung des menschlichen Genoms, Seuchenausbreitung, Suche nach Impfstoffen, Modellierung von Atemwegserkrankungen durch Umwelteinflüsse, ...

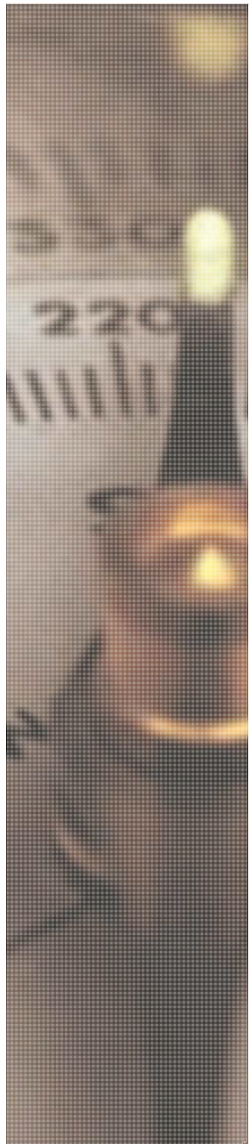
→ Experiment nicht möglich, zu gefährlich oder zu kostspielig

[Anknüpfungspunkte zur Medizininformatik, Bioinformatik]

Pharmakoinformatik

Informationstechnologie für Medikamentendesign und –entwicklung, Proteindynamik, ...

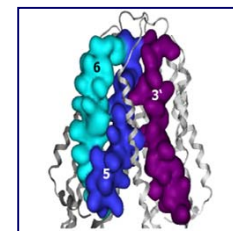
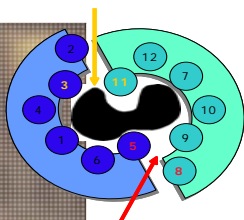
→ Experiment zu gefährlich oder zu kostspielig





Pharmakoinformatik

Verwendung von Informationstechnologien für pharmazeutische Probleme

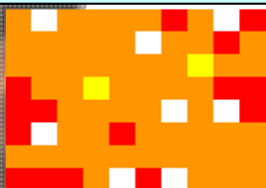


Target Identification and Validation

For human ABC-transporter, the dataset of Gottesman is the main source for substrate patterns. For infectious diseases targets are identified via screening of bacterial and fungal genomes for ABC-pumps. Considerations on drugability includes analysis of the regulatory networks

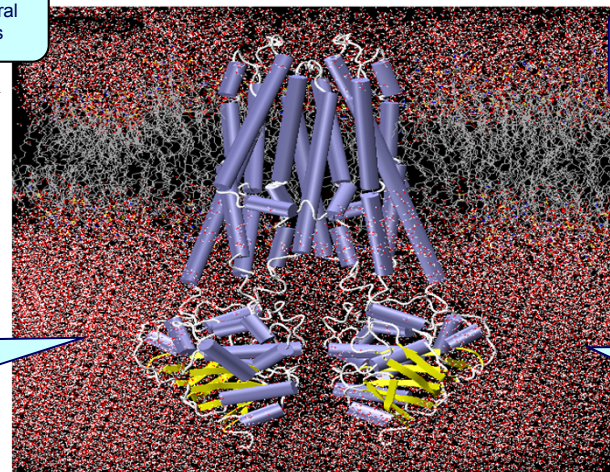
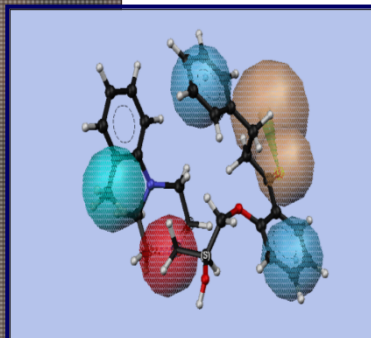
Hit Identification

Hit identification comprises use of artificial neural networks, autocorrelation vectors and VolSurf descriptors



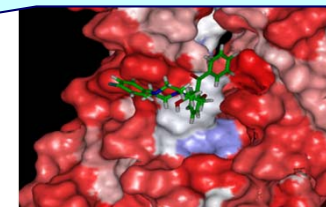
Virtual Screening

Virtual screening of large compound libraries is performed on basis of our recently developed SIBAR-technology, which utilises similarity based algorithms.



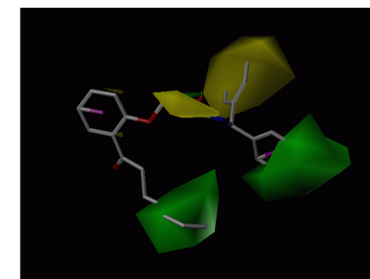
Protein-Protein Interactions

Selectivity profiling is performed on a systemic level and includes modeling of protein-protein interactions.



Structure based Design

Up to now no high resolution structure of a human ABC-transporter is available. Photoaffinity labeling and protein homology modeling pave the way for structure-based design and molecular basis of function.



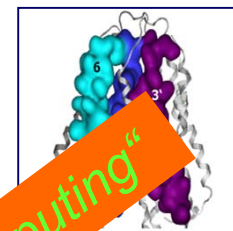
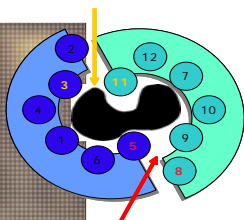
Lead Optimisation

For lead optimisation, we use conventional 2D- and 3D-QSAR methods as well as pharmacophore modeling and molecular holograms



Pharmakoinformatik

Verwendung von Informationstechnologien für pharmazeutische Probleme

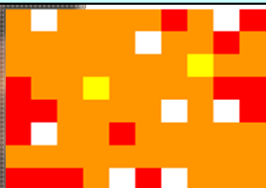


Target Identification and Validation

For human ABC-transporter, the dataset of Gottesman is the main source for substrate patterns. For infectious diseases targets are identified via screening of bacterial and fungal genomes for ABC pumps. Considerations on drugability includes analysis of the regulatory networks

Hit Identification

Hit identification comprises use of artificial neural networks, autocorrelation vectors and VolSurf descriptors



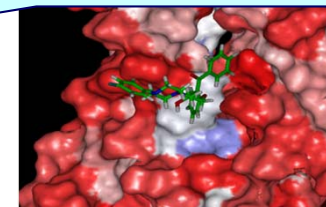
Virtual Screening

Virtual screening of large compound libraries is performed on basis of our recently developed SIBAR-technology, which utilises similarity based algorithms.



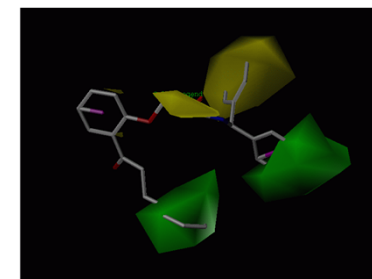
Protein-Protein Interactions

Selectivity profiling is performed on a systemic level and includes modeling of protein-protein interactions.



Structure based Design

Up to now no high resolution structure of a human ABC-transporter is available. Photoaffinity labeling and protein homology modeling pave the way for structure-based design and molecular basis of function.



Lead Optimisation

For lead optimisation, we use conventional 2D- and 3D-QSAR methods as well as pharmacophore modeling and molecular holograms

Mehr Information in der VU „Einführung in Scientific Computing“



Anwendungen Scientific Computing (Beispiele)

Web Science

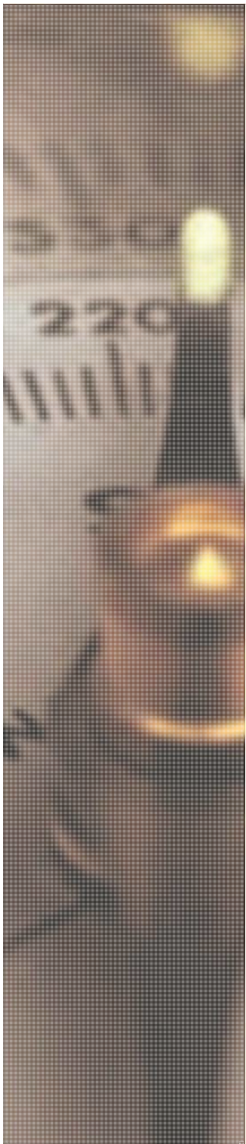
Informationsgewinn aus grossen Datenmengen, Internet-Suchmaschinen (z.B. Google), ...

Chemie

Berechnung und Simulation von Teilchenstruktur und -dynamik, Modellierung von Molekülen, ...

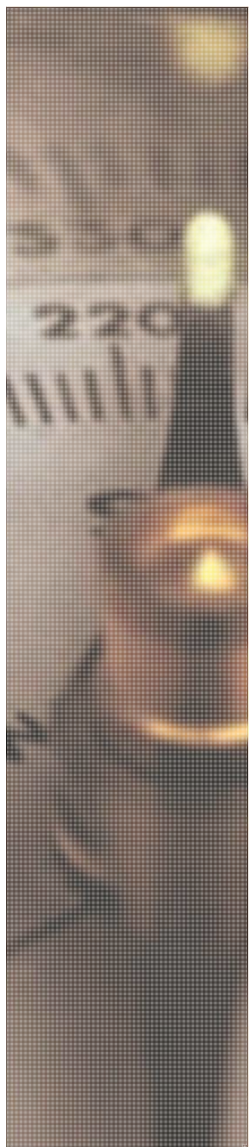
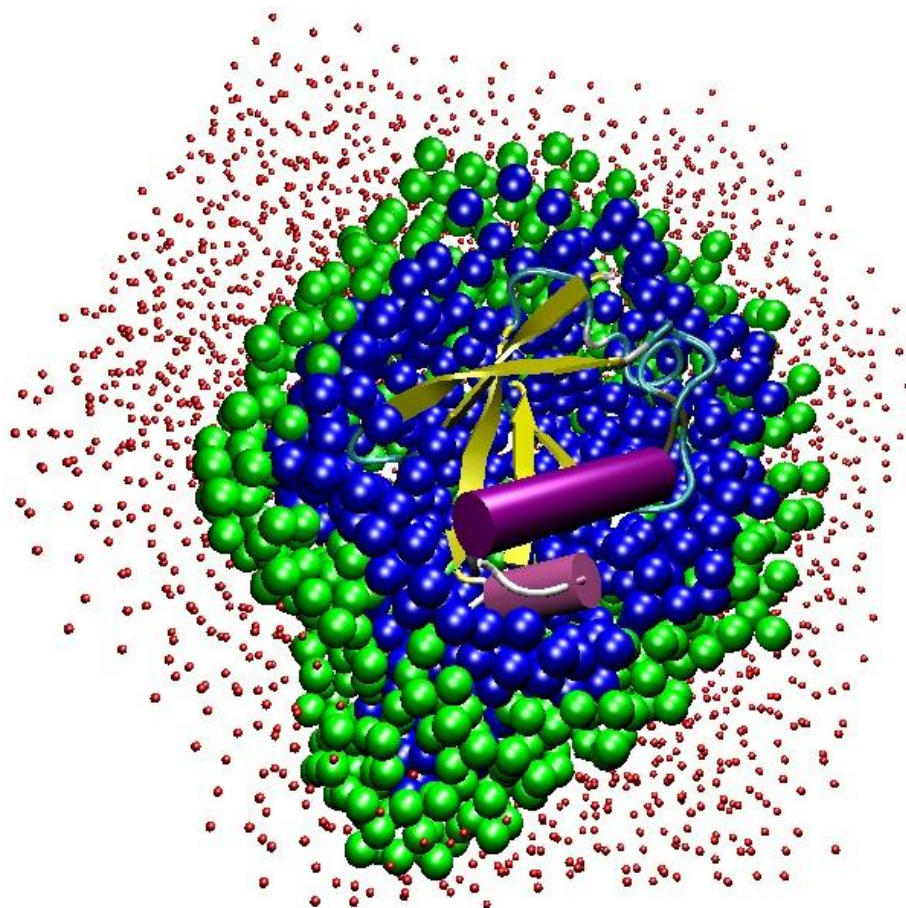
→ Phänomene, die nicht oder nur sehr schlecht messbar sind

Beispiele: Quantenchemie, Molekulardynamik und Biomolekulare Simulation





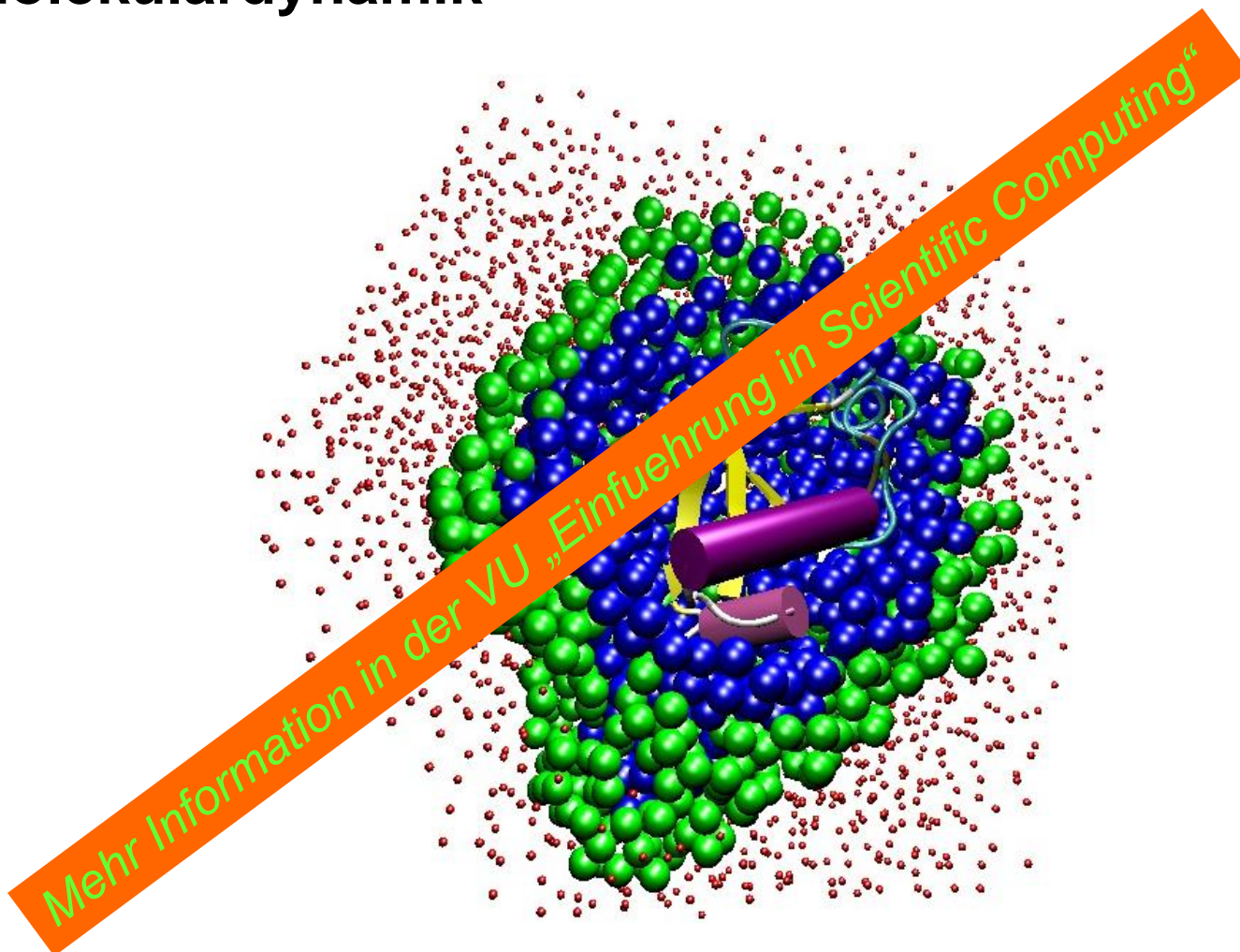
Molekulardynamik (in Loesungen)





universität
wien

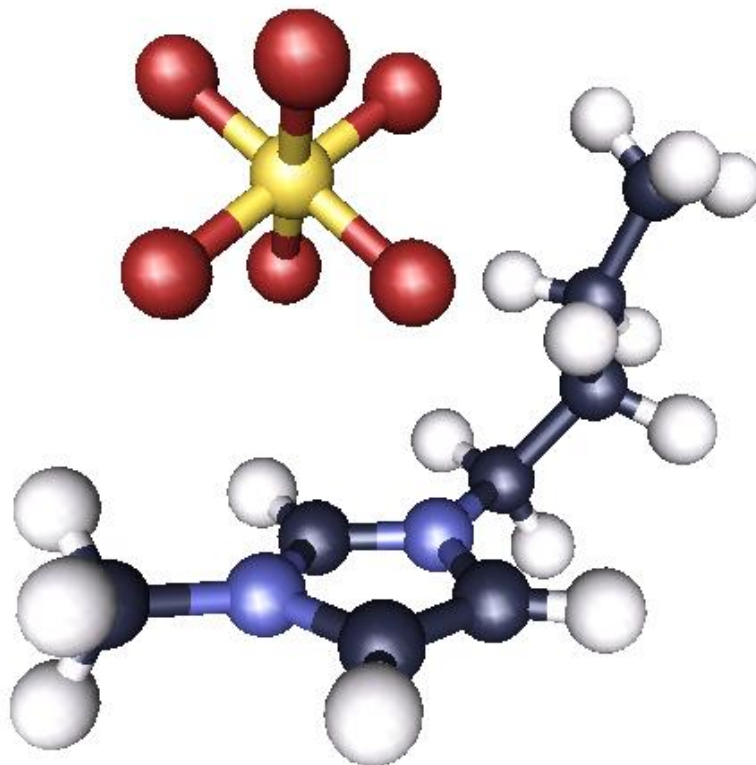
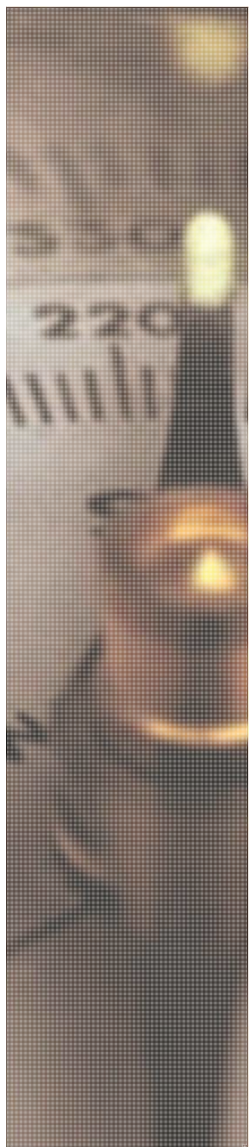
Molekulardynamik





universität
wien

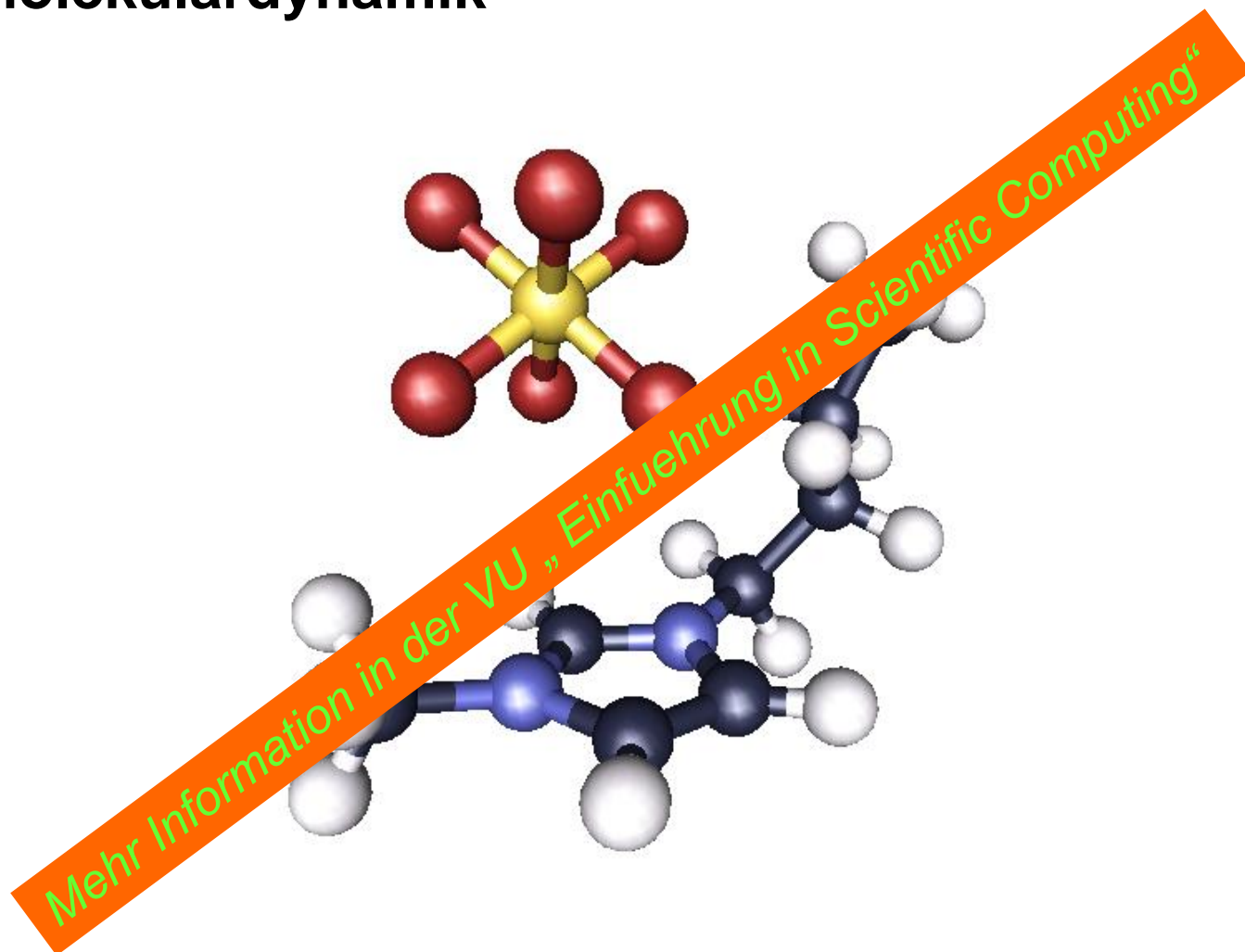
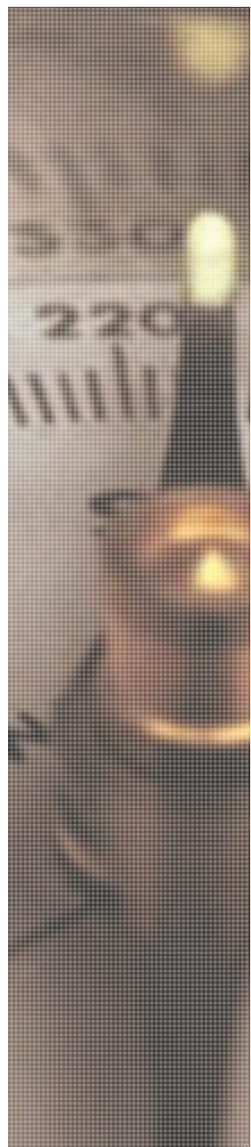
Molekulardynamik (mit Wechselwirkungen)





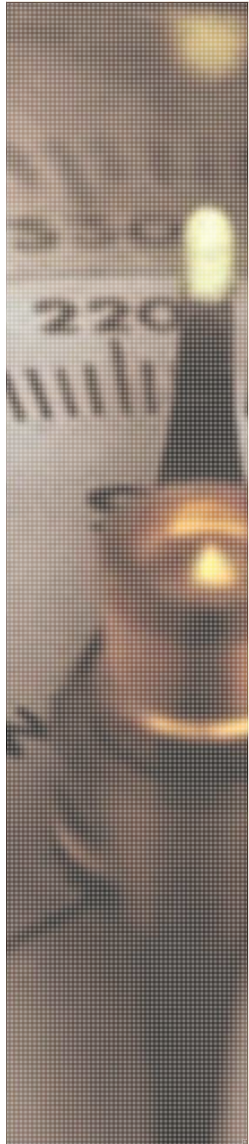
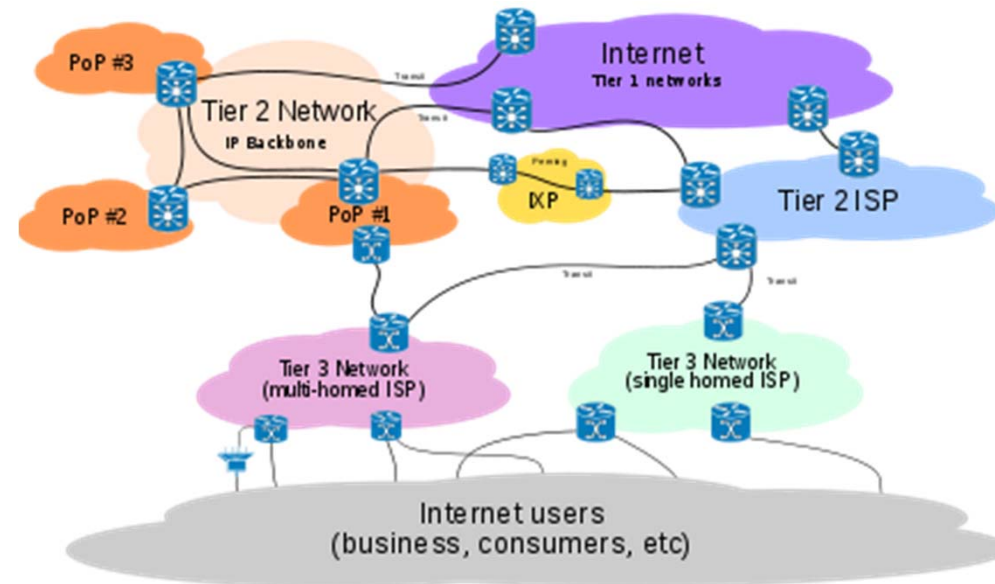
universität
wien

Molekulardynamik



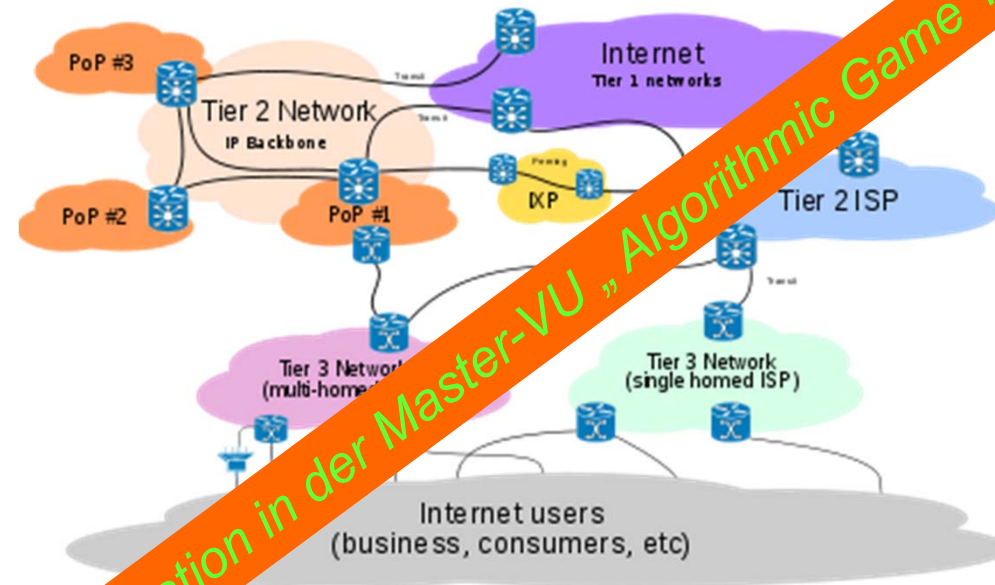


Algorithmische Spieltheorie

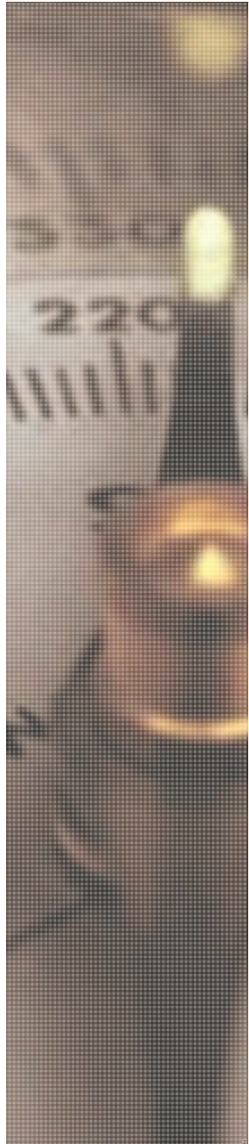




Algorithmische Spieltheorie



Mehr Information in der Master-VU „Algorithmic Game Theory“





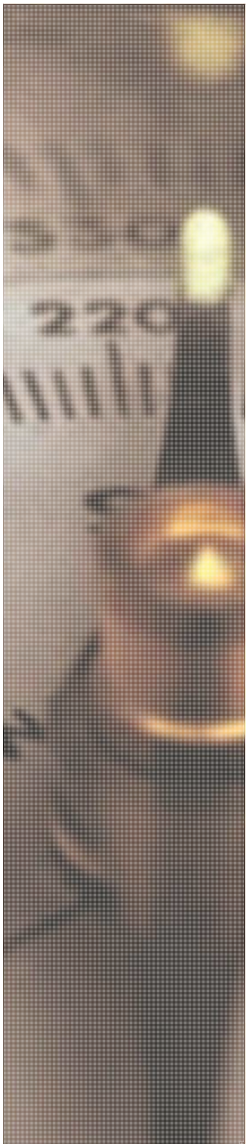
Weitere Anwendungsbeispiele

<http://www.distributedcomputing.info/projects.html>

LVA-Hinweis:

VU Einführung in Scientific Computing – Anwendungen und Algorithmen (empfohlen im 3. Semester)

050111 VU, DO 12:00-13:30, SR 7





Forschungsaktivitäten in der Fakultät

Forschungsgruppe Scientific Computing

➔ Prof. Benkner

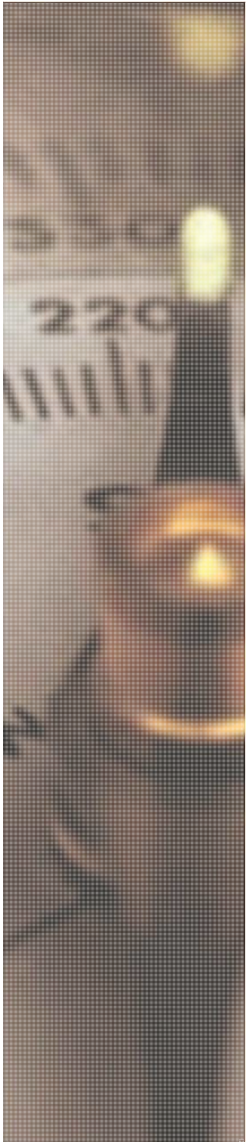
Forschungsgruppe Theory & Applications of Algorithms (TAA)

➔ Prof. Henzinger, Prof. Gansterer

Forschungsplattform „Computational Science Center“ an der Universität Wien

- Gemeinsam mit der Fakultät für Mathematik
- Inverse Probleme, Bildverarbeitung, Modellierung

➔ Prof. Scherzer



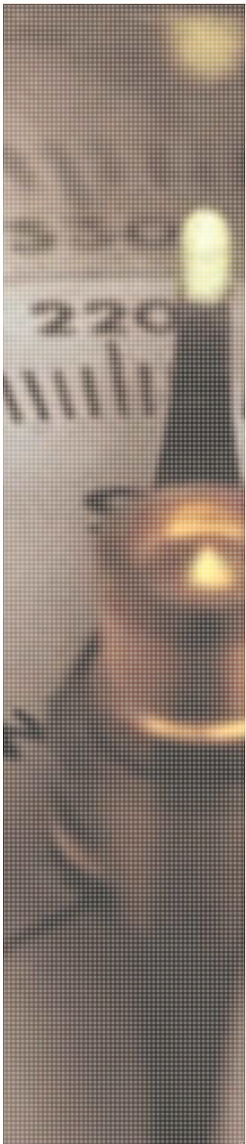


Forschungsaktivitäten in der Fakultät

Beispiele Forschungsprojekte TAA

- **DICIP**: Distributed Computing and Information Processing
 - Verteilte Algorithmen, Signalverarbeitung, Sensornetzwerke, Mobilnetzwerke
- **MixSVM**: Mixed-Precision SVM Classification on FPGAs
 - Bildklassifikation in Echtzeit für automatische Materialsortierung, Mülltrennung, etc.
- **CPAMMS**: Computing Paradigms and Algorithms for Molecular Modeling and Simulation: Applications in Chemistry, Molecular Biology, and Pharmacy
 - Computational Drug Design, in-silico drug screening
 - Computational Biology, Bioinformatik

... siehe <http://rlcta.univie.ac.at>





Forschungsaktivitäten in der Fakultät

Institut für Scientific Computing

➔ Prof. Benkner

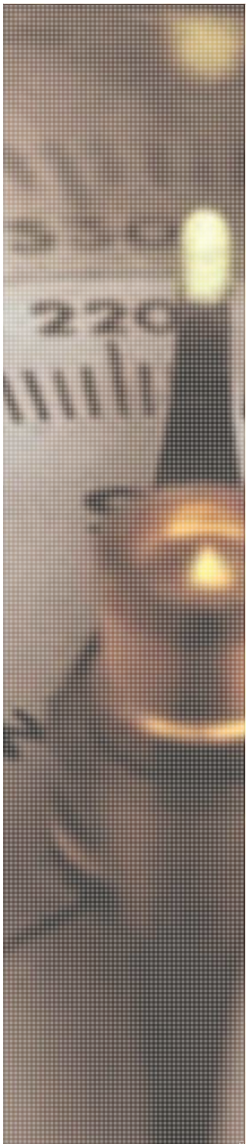
Forschungsgruppe Theory and Applications of Algorithms

➔ Prof. Henzinger, Prof. Gansterer

Forschungsplattform „Computational Science Center“ an der Universität Wien

- Gemeinsam mit der Fakultät für Mathematik
- Inverse Probleme, Bildverarbeitung, Modellierung

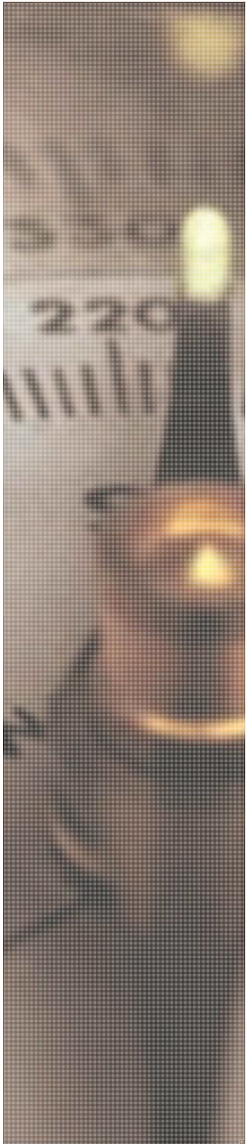
➔ Prof. Scherzer





Scientific Computing Technologien und Infrastrukturen

- High Performance Computing - Aktuelle Hochleistungsrechner
- Software und Programmierung
- Grid & Cloud Computing

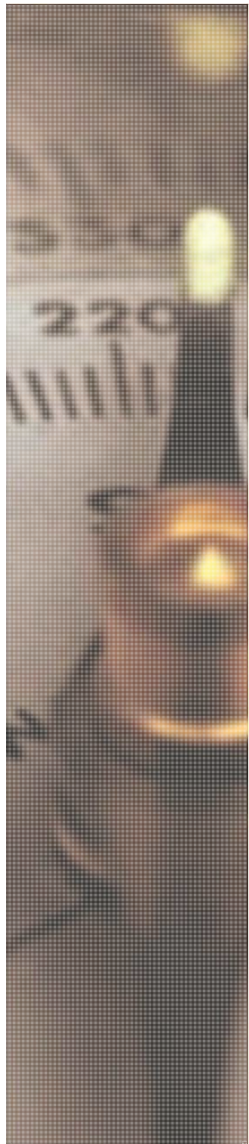
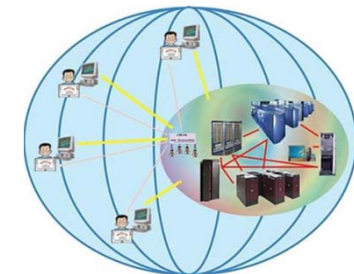




Scientific Computing – Infrastrukturen

Komplexe wissenschaftliche/technische Simulationen erfordern den Einsatz von Hochleistungsrechnern.

- **Supercomputer**
spezielle Hardware; eng gekoppelt
- **Cluster**
handelsübliche Hardware; lose gekoppelt
- **Grids**
globale Vernetzung von IT-Ressourcen;
„virtueller Supercomputer“
- **Clouds**
Virtualisierung von IT Ressourcen;
Everything-as-a-Service (XaaS)





Neue Herausforderungen – Hardware

- Steigende Komplexität neuer Prozessorarchitekturen



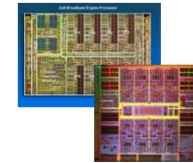
FPGAs



Accelerators



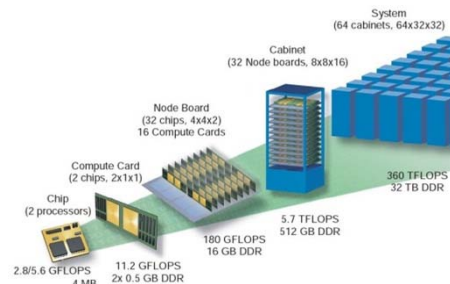
GPUs



Multicore

- Steigende Komplexität von High-End Systemen

Supercomputer mit >1000000 Cores → Global-scale Grids/Clouds



Weitere Steigerung der Rechenleistung → Erhöhung der **Parallelität!**



Top 500 Supercomputer - Juni 2012

Rank	Site	Computer/Year Vendor	Cores	R _{max}	R _{peak}	Power
1	DOE/NNSA/LLNL United States	Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom / 2011 IBM	1572864	16324.75	20132.66	7890.0
2	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer , SPARC64 VIIIfx 2.0GHz, Tofu interconnect / 2011 Fujitsu	705024	10510.00	11280.38	12659.9
3	DOE/SC/Argonne National Laboratory United States	Mira - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	786432	8162.38	10066.33	3945.0
4	Leibniz Rechenzentrum Germany	SuperMUC - iDataPlex DX360M4, Xeon E5-2680 8C 2.70GHz, Infiniband FDR / 2012 IBM	147456	2897.00	3185.05	3422.7
5	National Supercomputing Center in Tianjin China	Tianhe-1A - NUDT YH MPP, Xeon X5670 6C 2.93 GHz, NVIDIA 2050 / 2010 NUDT	186368	2566.00	4701.00	4040.0
6	DOE/SC/Oak Ridge National Laboratory United States	Jaguar - Cray XK6, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA 2090 / 2009 Cray Inc.	298592	1941.00	2627.61	5142.0
7	CINECA Italy	Fermi - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	163840	1725.49	2097.15	821.9
8	Forschungszentrum Juelich (FZJ) Germany	JuQUEEN - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	131072	1380.39	1677.72	657.5
9	CEA/TGCC-GENCI France	Curie thin nodes - Bullx B510, Xeon E5-2680 8C 2.700GHz, Infiniband QDR / 2012 Bull	77184	1359.00	1667.17	2251.0
10	National Supercomputing Centre in Shenzhen (NSCS) China	Nebulae - Dawning TC3600 Blade System, Xeon X5650 6C 2.66GHz, Infiniband QDR, NVIDIA 2050 / 2010 Dawning	120640	1271.00	2984.30	2580.0





universität
wien

Sequoia – BlueGene/Q



~20 Petaflops (peak)

Cores: 1572864

Power: ~8 MWatts



'K' Supercomputer



~11 Petaflops (peak)
Cores: 705024
Power: ~12 MWatts

Protein Folding

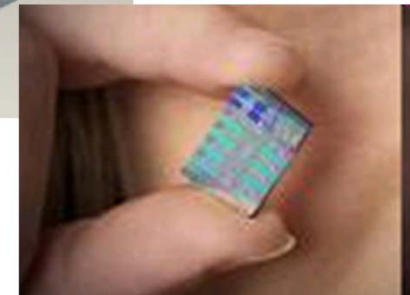
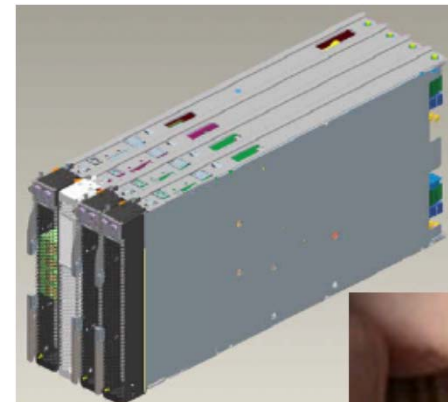
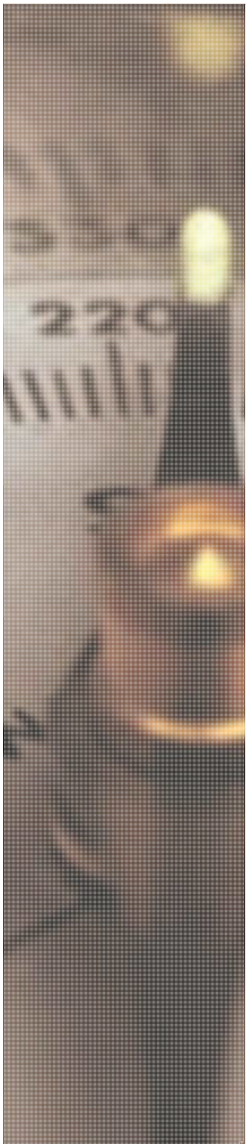


Simulating 100 microseconds
of protein folding requires
 $\sim 10^{23}$ operations.
→ >> 4 months on 'K'



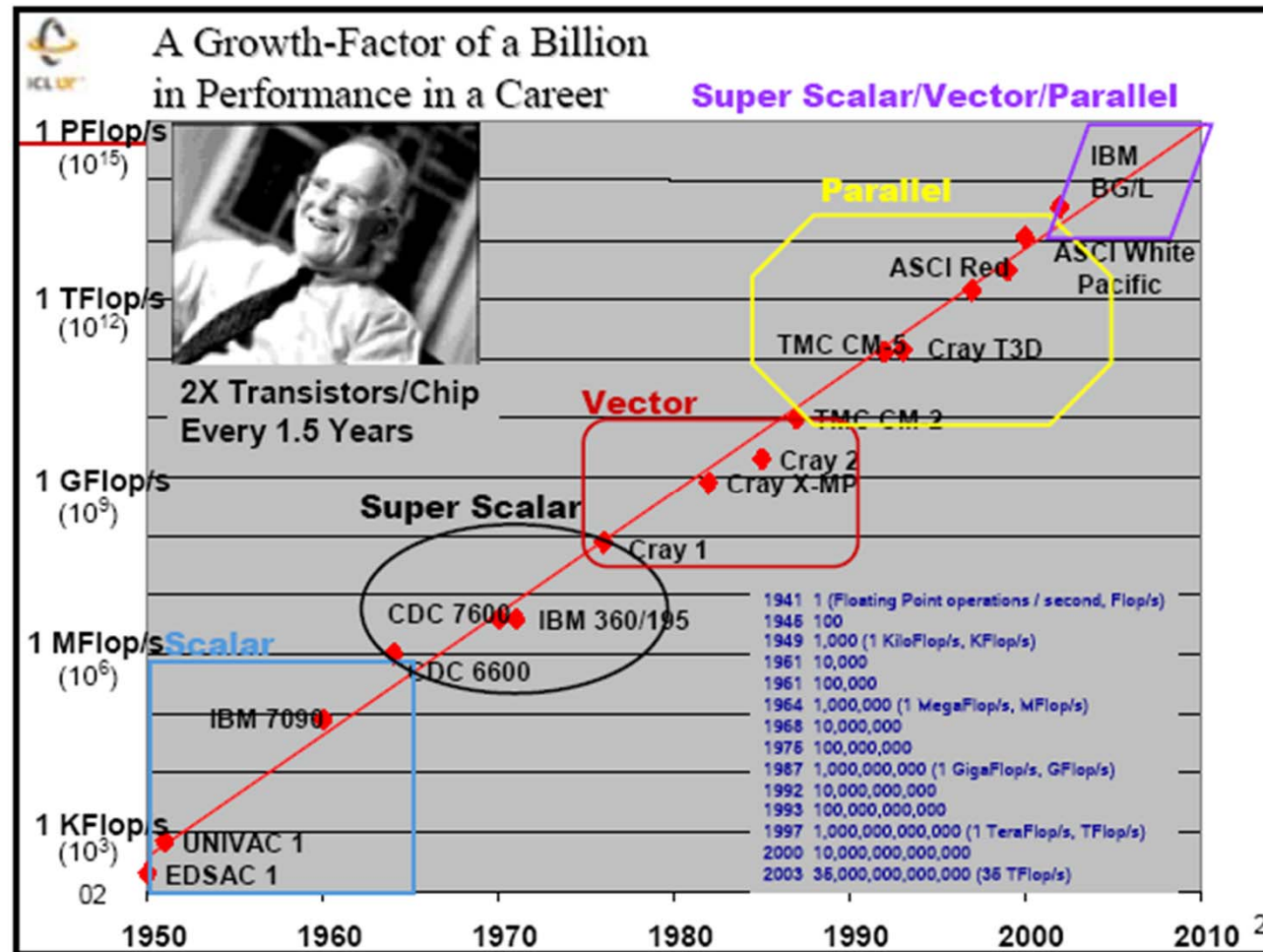
IBM Roadrunner

- Erster Petaflops-Rechner
- Los Alamos National Lab
- Hybrider Supercomputer
- 129600 Cores
 - 6480 Dual-Core Opteron
 - 12960 Cell eDP Prozessoren
- 104 TB Memory
- 1.3 PFlop/s
(1.3 Milliarden Gleitkommaberechnungen / sec)



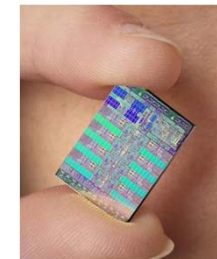
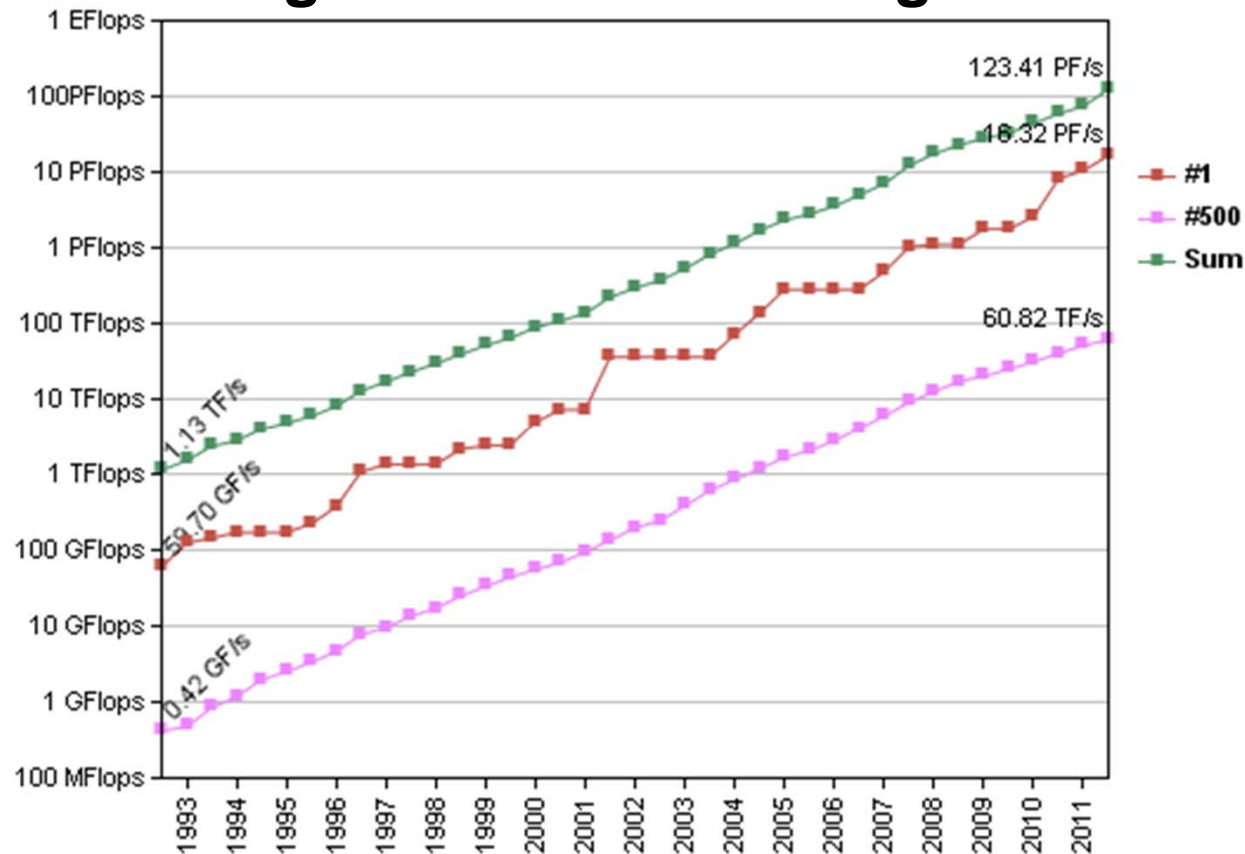


Entwicklung der Rechenleistung





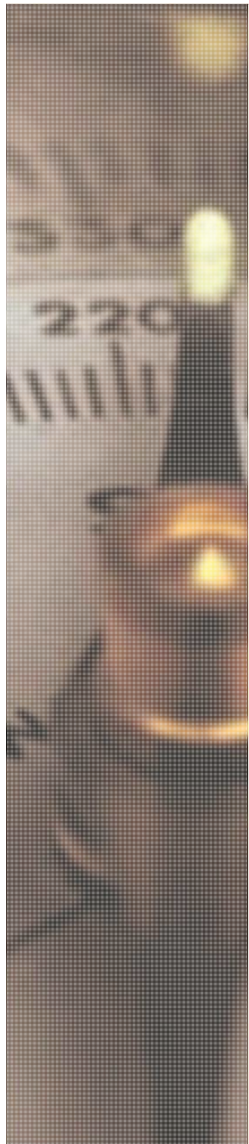
Entwicklung der Rechenleistung: 1993-2012



Moore's Law: #transistors on a chip doubles every 18-24 month

#1: '93: 1024 processors, 60 GFlops; '12: 1572864 cores, 20,132 PFlops

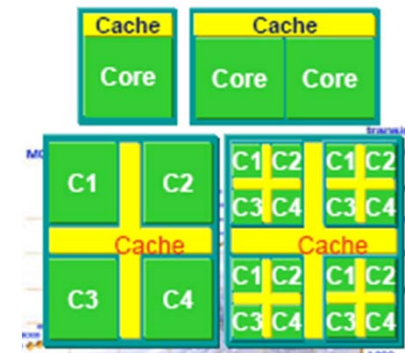
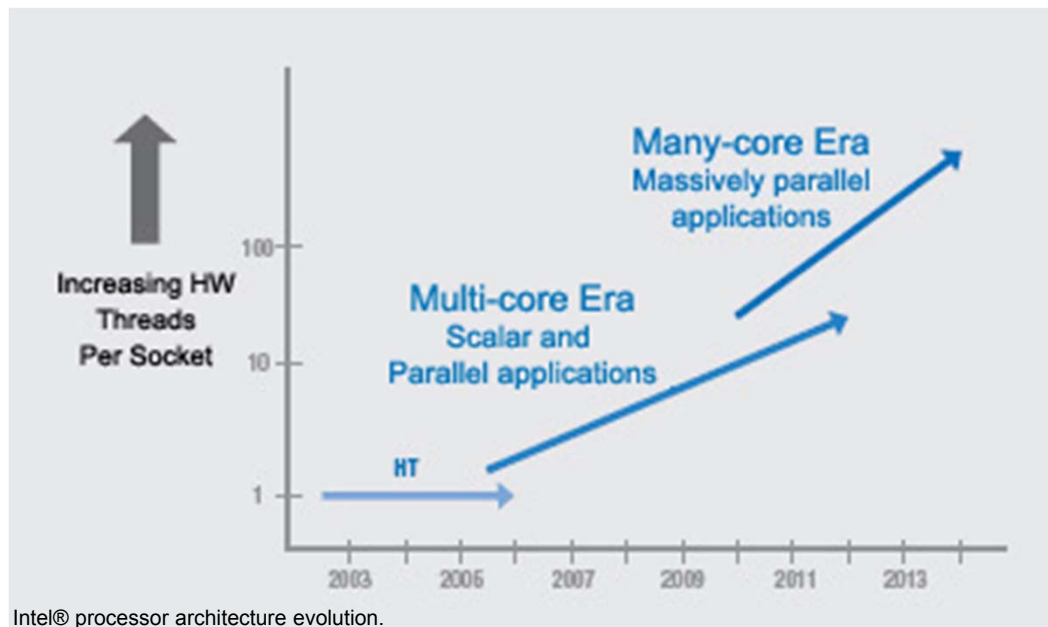
→ **335000X**





Prozessoren – Zukünftige Entwicklung

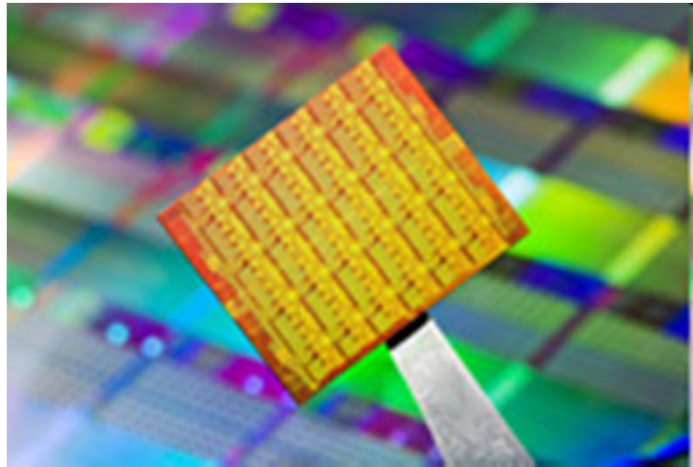
- Steigerung der Taktfrequenz ist an Grenzen gestoßen
 - Multi-core Technologie: mehrere Prozessorkerne/Chip
 - Konvergenz von CPU und GPU Technologien
- „Parallel computing on everyone ‘s desktop“
- Große Herausforderung an Anwendungssoftware!



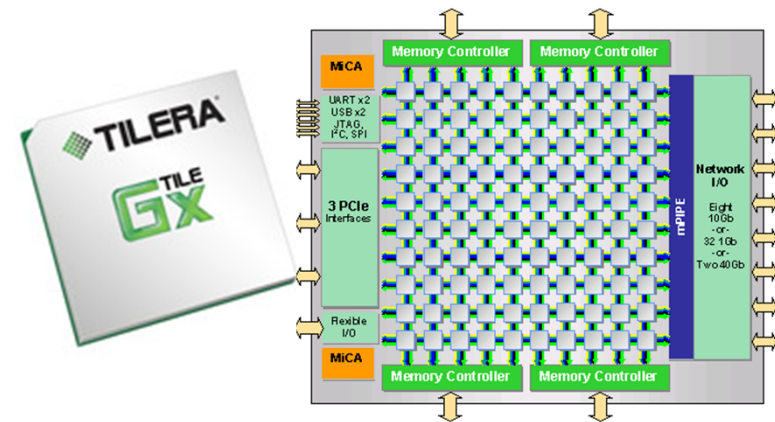


Prozessoren – Neuere Entwicklungen

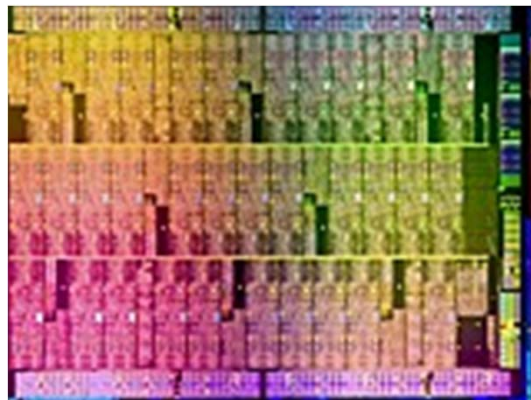
Intel SCC (2010)



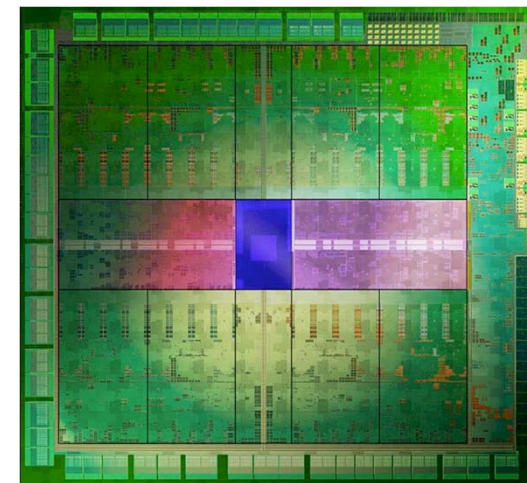
Tilera Tile-Gx (2010)



Intel Xeon Phi (2012)



NVIDIA Kepler (2012)





Neue Herausforderungen – Software

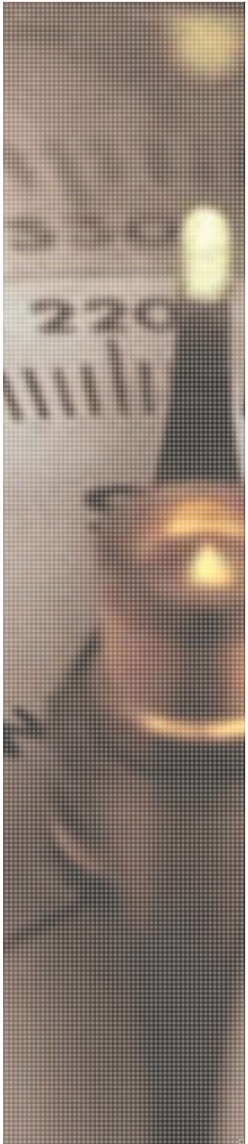
Wie entwickeln wir Software und Toolkits für die effiziente Verwendung der immer komplexer werdenden Hardware?

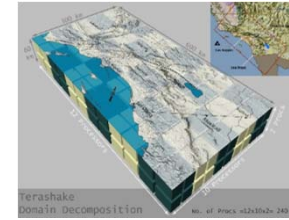
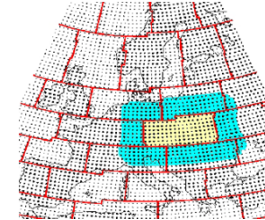
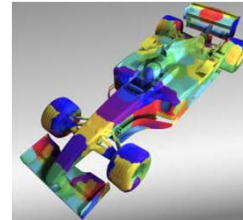
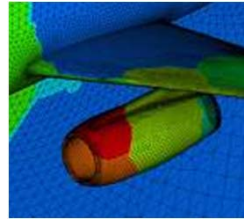
Parallele Programmiermodelle.

Parallele Algorithmen.

Wie kann bestehende Software auf neue Hardware portiert werden?

Anwendungsentwicklung für Grids/Clouds?



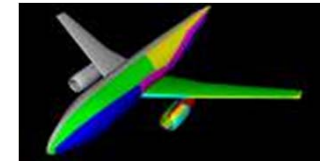


Software für Hochleistungsrechner

Um die Rechenleistung von parallelen Computern nutzen zu können, müssen Anwendungsprogramme parallelisiert werden.

Parallelisierung

- Aufteilung der Daten
- Aufteilung der Arbeit
- Kommunikation/Synchronisierung



Parallele Programmiermodelle/Sprachen

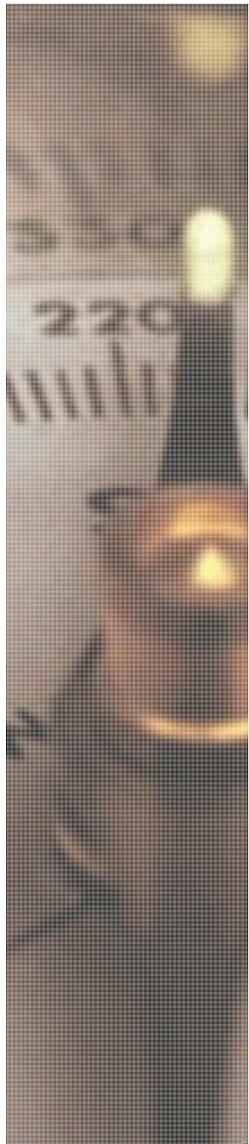
- MPI, OpenMP, HPF, UPC, ...



Parallele Programmbibliotheken

- LAPACK, ScaLAPACK, ...

Programmentwicklung/adaptierung für Parallelrechner ist komplex, zeitaufwendig und kostenintensiv! (***“no free lunch for software”***)





Grid Computing

- globale Infrastruktur zur **sicheren, flexiblen, koordinierten Nutzung verteilter Ressourcen** (Rechner, Netzwerke, DBs, Instrumente, ...)
- nächste Generation von Internet-basierten Anwendungen

EGEE Grid Infrastructure

250 Sites

48 Countries

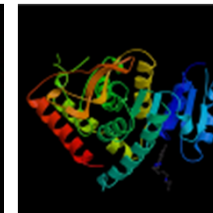
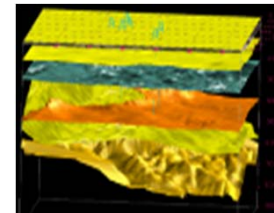
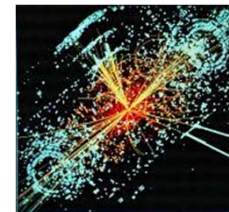
50,000 CPUs

13 PetaBytes

>5000 Users

>200 VO's

>140,000 Jobs/Day





Global Distributed Computing

Seti@Home

- 500 000 CPUs, 1500 CPU hours/day



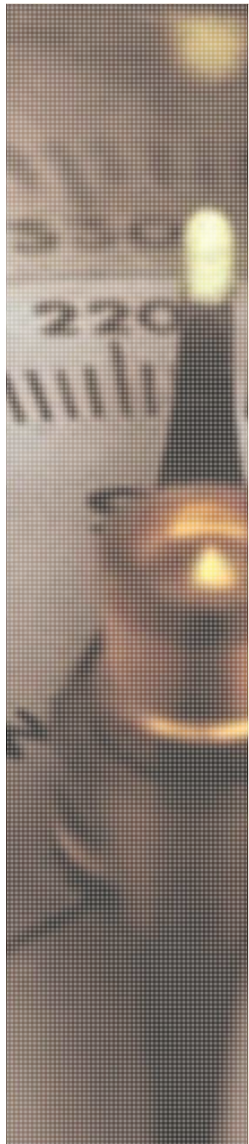
Folding@Home

OS Type	Native TFLOPS*	x86 TFLOPS*	Active CPUs	Total CPUs
Windows	181	181	173928	4310585
Mac OS X/PowerPC	2	2	1893	152528
Mac OS X/Intel	47	47	11352	174996
Linux	80	80	29473	716395
ATI GPU	1951	2058	13737	342200
NVIDIA GPU	991	2091	6232	279709
PLAYSTATION®3	467	985	16555	1262752
Total	3725	5450	257543	8370211



Google

- > 900 000 servers, > 2000 queries/sec, > 150 000 000 queries/day

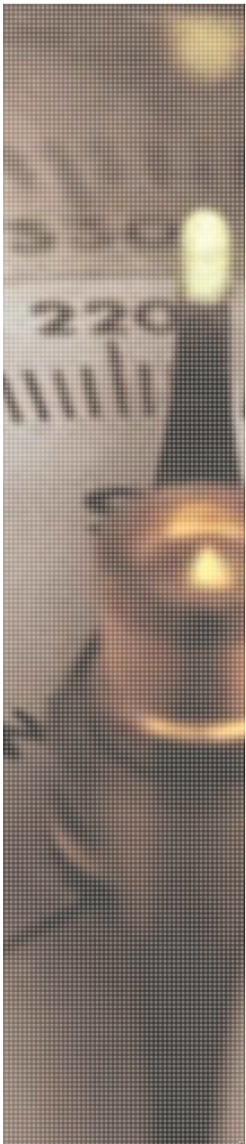




Cloud Computing

- New model for service provisioning based upon: Grid Computing, On-Demand Computing, Virtualization, Web Services Technologies
- Clouds provide the illusion of unlimited compute power available on demand within a pay-per-use model
- Different facets of Cloud computing: *aaS
 - **SaaS – Software as a Service**
 - **IaaS – Platform as a Service**
 - **IaaS – Infrastructure as a Service**

Commercial Cloud offers: Amazon EC2, Microsoft Azure, Google AppEngine, ...

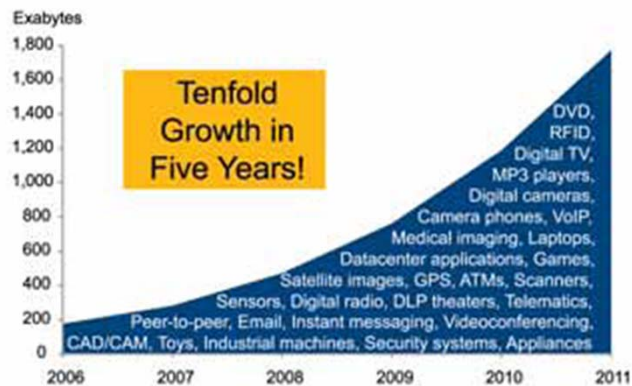




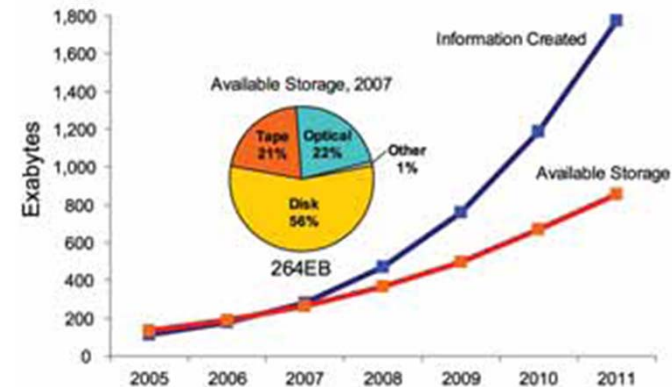
Neue Herausforderungen – Daten

1 Peta(10^{15})bytes ... von Google pro Stunde verarbeitet

Digital Information Created, Captured, Replicated Worldwide



Information Creation and Available Storage



Digitales Datenvolumen steigt um Faktor 10 alle 5 Jahre

Seit 2007 mehr neue Daten als verfügbarer Speicher

Digitales Datenvolumen 2011: ~1700 Exa(10^{18})bytes

→ Neue Methoden um Information aus Daten zu gewinnen

→ Neue Methoden um große Datenmengen zu verwalten



HPC @ Uni Wien

- **Luna** (Inst. f. Scientific Computing)

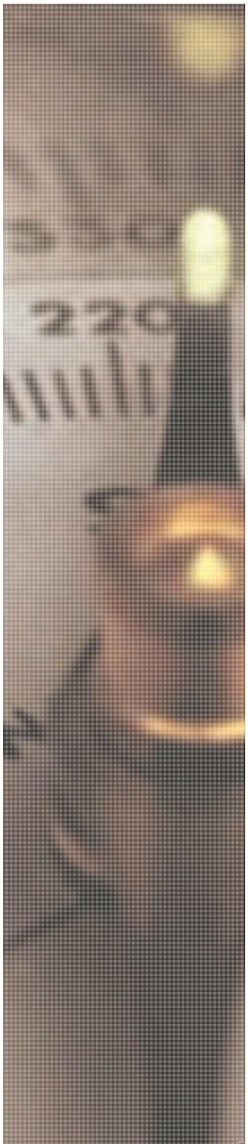
288 processor cores, Opteron, Infiniband

- **Cora** (Inst. f. Scientific Computing)

72 CPU cores, ~6000 GPU cores, Infiniband

- **Vienna Scientific Cluster 2** (TU, Uni, Boku)

21024 processor cores, AMD Opteron 6132 HE





Related Research Efforts @ Uni Wien

- **EU Project Aneurist**

Interated Biomedical Informatics, www.aneurist.org

- **EU Project ADMIRE**

Advanced Data Mining and Integration Research for Europe

- **EU Project PEPPER**

Performance Portability and Programmability for Heterogeneous Many-core Architectures, www.peppher.eu

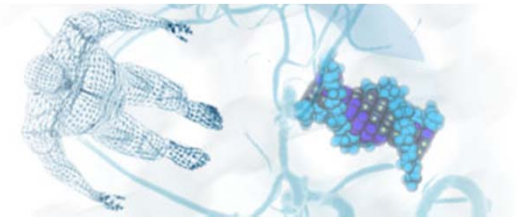
- **EU Project VPH-Share**

Virtual Physiological Human: Sharing for Healthcare

- **EU Project AutoTune**

Automatic Online Tuning

See: <http://www.par.univie.ac.at/>





Berufsbild

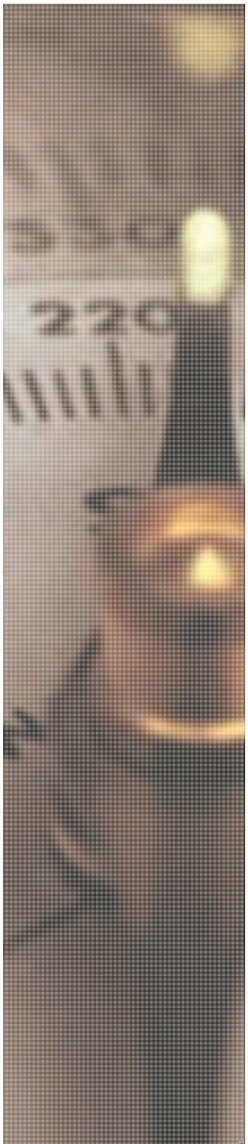
Computergestützte Forschung und Entwicklung

- Einsatz neuester Technologien
- Entwicklung neuer Methoden und Technologien

Tätigkeiten in Wissenschaft oder Industrie

Methodenorientiert, offen für verschiedenste Anwendungsfächer

- Über diejenigen hinaus, die im Curriculum vertreten sind





Master Scientific Computing @ Universität Wien

Fundament: Bachelor Informatik

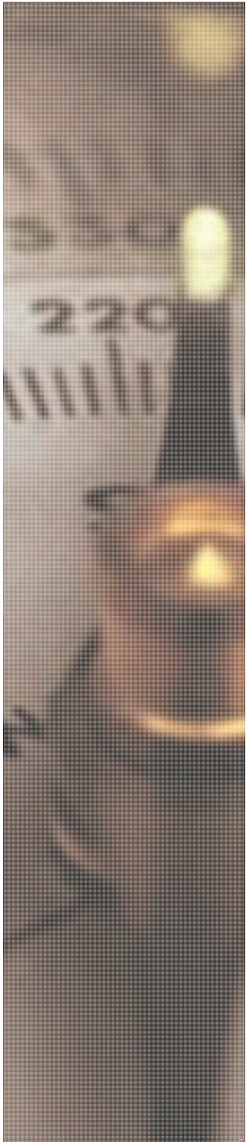
Ausprägungsfach „Scientific Computing“

Zulassung:

Informatiker, Naturwissenschaftler, Mathematiker mit
Universitätsabschluss (z.B. Bachelor) bzw. verwandtem
Fachhochschulabschluss

Dauer: 2 Jahre (4 Semester)

Abschluss: akademischer Grad Dipl.-Ing.





Master Scientific Computing @ Universität Wien

Ziel: Heranführen an

State-of-the-art in der Methodik des Scientific Computing

Aktuelle Technologien und Infrastrukturen

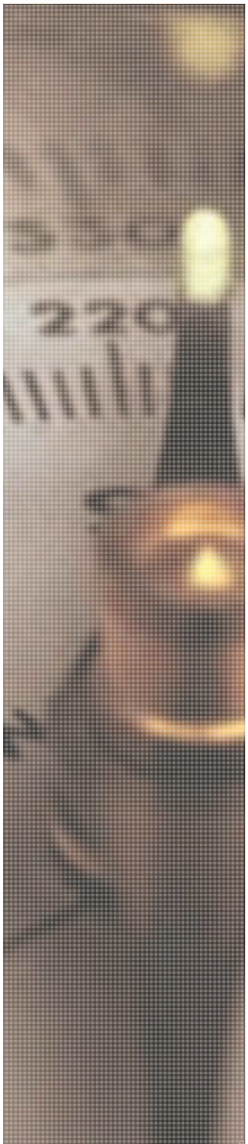
Aktuelle Forschungsthemen

Vertiefende Beschäftigung mit exemplarischen
Problemstellungen aus Anwendungsdisziplinen

Abstraktionsprozess:

Motivation aus Anwendungswissenschaft

→ allgemeine Fragestellungen der Informatik





Master Scientific Computing @ Universität Wien

Grundlagenfächer & Interdisziplinäre Informatik

Parallele Architekturen (z.B. Multicore, GPUs, Supercomputer)

Parallele Programmiermodelle und Programmiersprachen

Zentrale Algorithmen und deren effiziente Implementierung

Ein Anwendungsfach aus:

- Pharmakoinformatik
- Computational Chemistry
- Algorithmen und Optimierung

Kernfachkombination (2 aus 3 Modulen)

- High-End Computing
- Algorithmen und Optimierung
- Praktikum Scientific Computing

Masterarbeit

