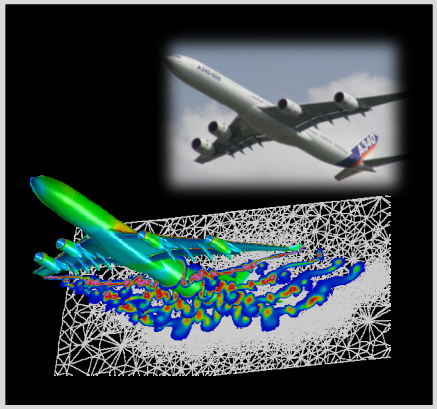


High Performance Computing

ADVANCED SCIENTIFIC COMPUTING

Prof. Dr. – Ing. Morris Riedel
Adjunct Associated Professor
School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland
Research Group Leader, Juelich Supercomputing Centre, Forschungszentrum Juelich, Germany

SHORT LECTURE 14

@Morris Riedel @MorrisRiedel @MorrisRiedel

Molecular Systems & Libraries

November 21, 2019
Webinar

UNIVERSITY OF ICELAND
SCHOOL OF ENGINEERING AND NATURAL SCIENCES
FACULTY OF INDUSTRIAL ENGINEERING, MECHANICAL ENGINEERING AND COMPUTER SCIENCE

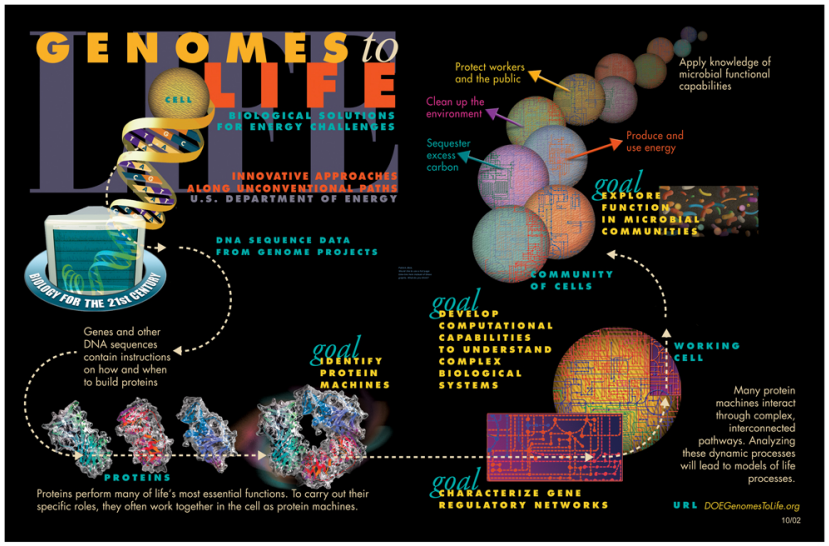
JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

DEEP Projects

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

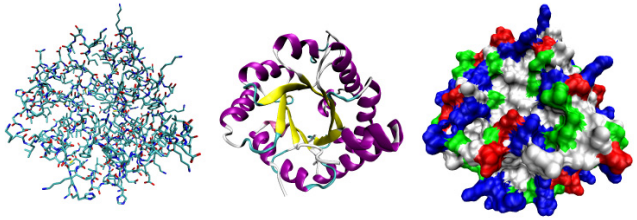
HAICU HELMHOLTZ ARTIFICIAL INTELLIGENCE COOPERATION UNIT

Review of Short Lecture 13 – Systems Biology & Bioinformatics



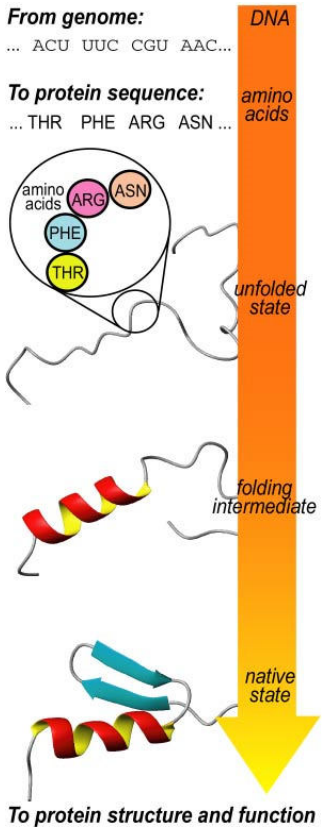
(all atom representation colored by atom type)

(once linked in the protein chain, an individual amino acid is called a residue)

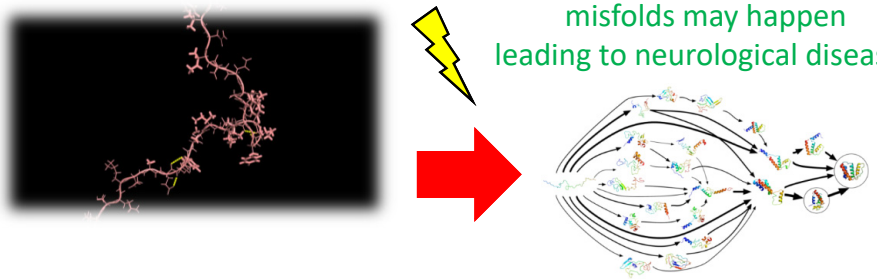


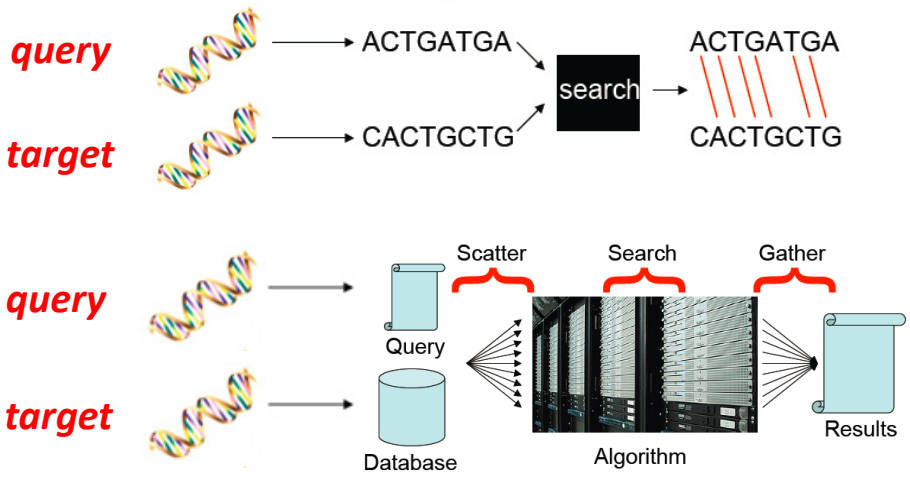
(simplified representation illustrating backbone conformation, colored by secondary structure)

Function of a protein is related to its 3D shape



misfolds may happen leading to neurological diseases





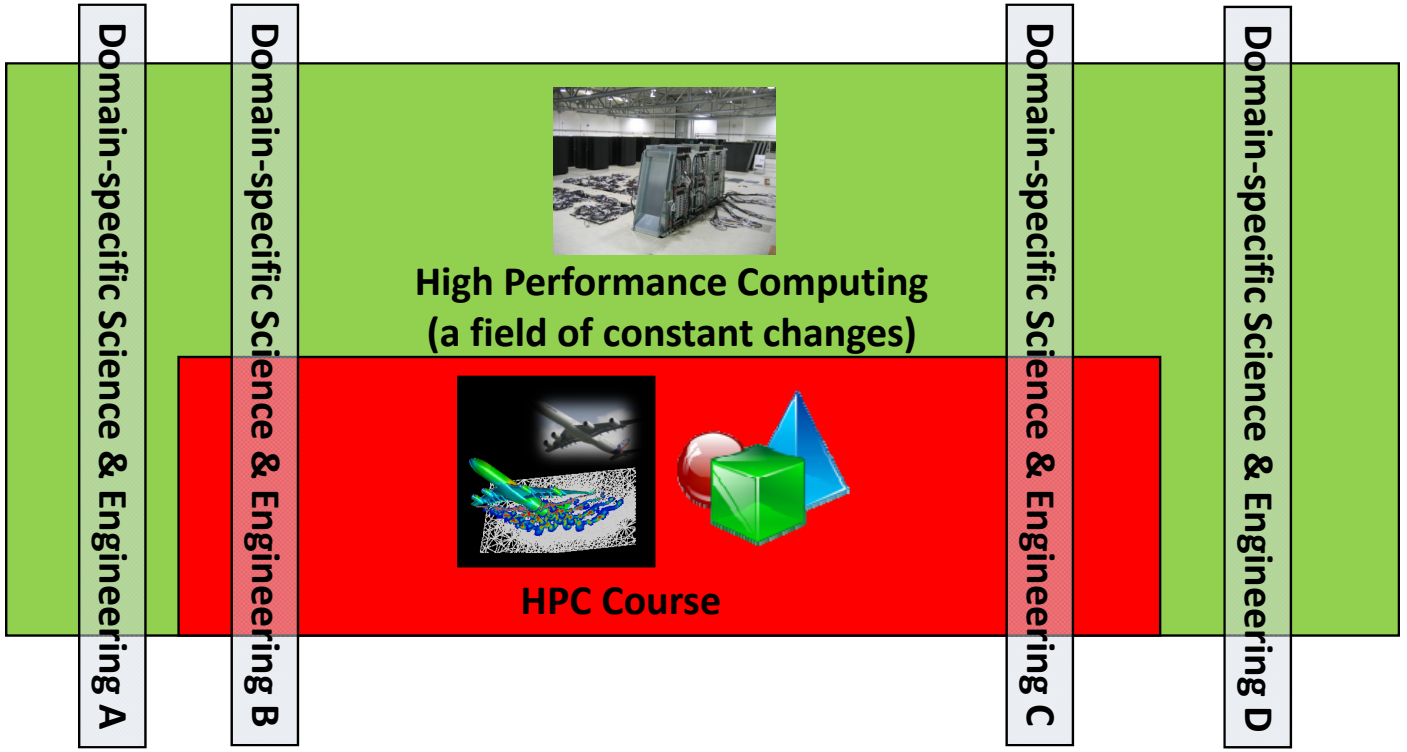
modified from [5] Sequence Searching

[4] Folding@Home *[3] S. Mohanty et al, JSC Simlab Biology*

modified from [1] Wikipedia on Systems Biology *[2] Wikipedia on Protein*

HPC-A[dvanced] Scientific Computing (cf. Prologue) – Second Part

- Consists of techniques for programming & using large-scale HPC Systems
  - Approach: Get a broad understanding what HPC is and what can be done
  - Goal: Train general HPC techniques and systems and selected details of domain-specific applications



Outline of the Course

1. High Performance Computing
2. Parallel Programming with MPI
3. Parallelization Fundamentals
4. Advanced MPI Techniques
5. Parallel Algorithms & Data Structures
6. Parallel Programming with OpenMP
7. Graphical Processing Units (GPUs)
8. Parallel & Scalable Machine & Deep Learning
9. Debugging & Profiling & Performance Toolsets
10. Hybrid Programming & Patterns
11. Scientific Visualization & Scalable Infrastructures
12. Terrestrial Systems & Climate
13. Systems Biology & Bioinformatics
14. Molecular Systems & Libraries
15. Computational Fluid Dynamics & Finite Elements
16. Epilogue

+ additional practical lectures & Webinars for our hands-on assignments in context

- Practical Topics
- Theoretical / Conceptual Topics

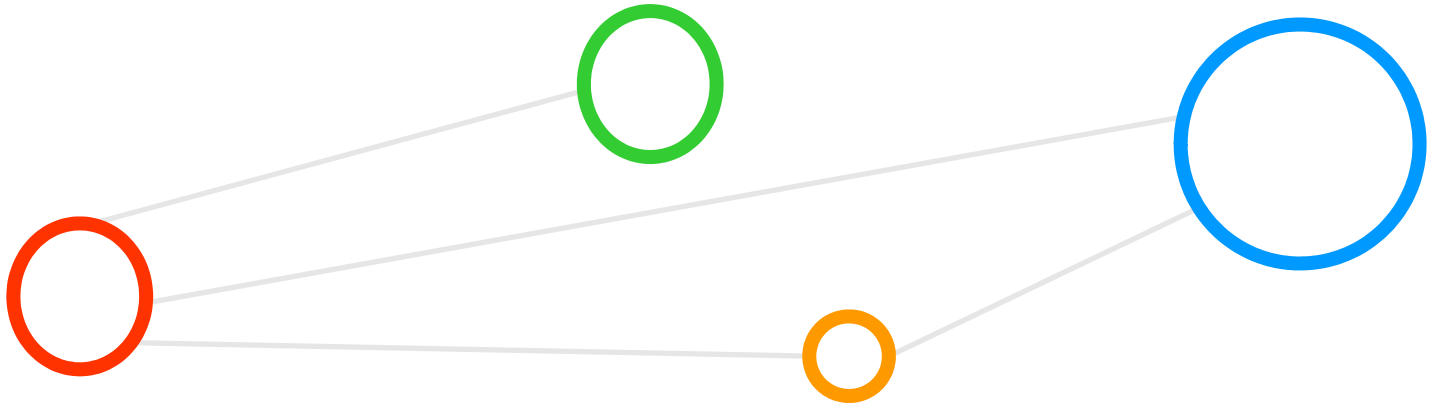
Outline

- Molecular Systems
  - Terminology & Motivation for Simulating Molecular Systems
  - Molecular Mechanics vs. Quantum Chemistry / Mechanics
  - Ab Initio Calculations & Examples in Application Fields
  - Molecular Docking to Determine Binding of Molecules
  - Molecular Dynamics to Simulate Physical Atom Movements
- Selected Methods & Libraries
  - Nanoscale Molecular Dynamics (NAMD) Simulations
  - Car Parinello Molecular Dynamics (CPMD) Simulations
  - Massively Parallel Multi-Particle Collision Dynamics (MP2C)
  - Assisted Model Building with Energy Refinement (AMBER)
  - Human Brain & Drug Development Application Example

- **Promises from previous lecture(s):**
- ***Practical Lecture 5:*** **Lecture 14 will give in-depth details on parallel and scalable molecular systems algorithms, tools, methods, and the use of libraries**
- ***Lecture 2:*** **Lecture 12 – 15 will offer more insights into a wide variety of physics & engineering applications that take advantage of HPC with MPI**
- ***Lecture 3:*** **Lecture 12 – 15 will provide details on applied parallelization methods within parallel applications & domain/functional decomposition**
- ***Short Lecture 13:*** **Lecture 14 will provide more pieces of information about molecular systems & libraries as well as selected parallel application methods**

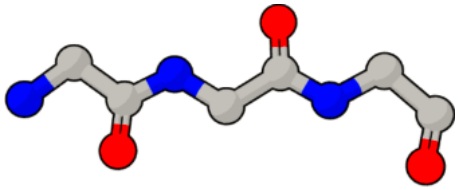
- **Note that this lecture is only a short lecture that usually needs a full course**
- **The goal is to understand selected HPC application fields & provide a few pointers to other advanced related university courses/topics/tutorials**

Molecular Systems

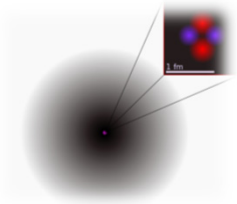


Molecular Systems – Terminology & Motivation

- Domain-specific usage
  - E.g. used to understand small chemical systems or large biological molecules (e.g. proteins, cf. Short Lecture 13)
  - E.g. used to study large material assemblies (with many thousands to millions of atoms)
- Different Approaches
  - Common is to enable atomistic level description of the molecular system
  - A wide variety of methods and libraries evolved over time
  - Parallelization techniques are used to make studying molecular systems possible



(molecular model of a protein)



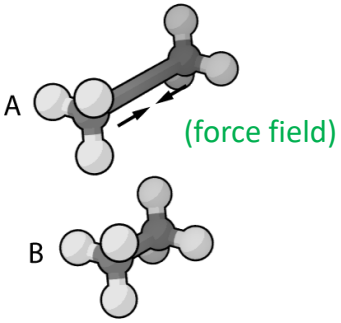
(atoms consists of 'electron cloud' & nucleus of protons and neutrons)

- **Molecular systems refers to techniques to model the behaviour (and dynamics) of molecules**
- **Molecules are a group of two or more atoms held together by chemical bonds**
- **Atoms are the smallest unit that defines the chemical elements: chemical substances consisting of a single type of atom**
- **Molecular Systems are used in many domains (e.g. computational chemistry, biology, materials science, drug design, etc.)**

[7] Wikipedia on 'Atoms' [6] Wikipedia on 'Molecule' modified from [8] Wikipedia on 'Molecular Modelling'

Molecular Systems – Terminology – Different Approaches

- Molecular Mechanics (aka ‘coarse granular‘)
  - Systems treat atoms as the smallest individual unit
  - Potential energy of systems is calculated using ‘force fields’
- Quantum Chemistry / Mechanics (aka ‘more fine granular‘)
  - Experiments of chemical systems with focus of quantum mechanics in physical models
  - Quantum mechanics provides a mathematical description (i.e. equations) of much of the behavior and interactions of energy and matter
  - Explicitly models electrons of each atom



(force field)

[9] Wikipedia on ‘Molecular Mechanics’

- **Molecular mechanics refer to coarse granular simulations of molecular systems**
- **Quantum mechanics use the Schroedinger equation for a more fine granular simulation of molecular systems**
- **The Schroedinger equation is a linear partial differential equation that describes the state function of a quantum-mechanical system**
- **One Example of using the Schroedinger equation is to describe the behaviour of a particle in a field of force**

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r},t) = \hat{H}\Psi(\mathbf{r},t)$$

(Schroedinger Equation)

[23] Wikipedia on ‘Schroedinger Equation’ [10] Wikipedia on ‘Quantum Chemistry’ [11] Wikipedia on ‘Quantum Mechanics’

Schroedinger Equation

- Quantum Theory in Physics
  - Foundation of quantum mechanics
  - Their consequences is the well-known Schroedinger equation for stationary states of the molecule
  - Stationary and evolution of non-stationary states depend on the energy operator
  - Schroedinger equation also describes the evolution of a given wave function
  - The wave function is a central notion in quantum mechanics
- Two Types of the Equation
  - Time-independent Schroedinger equation: stationary states can be produced as solutions of the equation
  - Time-dependent Schroedinger equation: plays a pivotal role as the equation of motion
- Usage in simulations
  - To obtain a special solution to such equations, one has to insert particular boundary conditions to be fulfilled

$$\left[\frac{-\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right]\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r},t) = \left[\frac{-\hbar^2}{2m}\nabla^2 + V(\mathbf{r},t)\right]\Psi(\mathbf{r},t)$$

[23] Wikipedia on 'Schroedinger Equation'

Molecular Systems – Terminology – Ab Initio Calculations

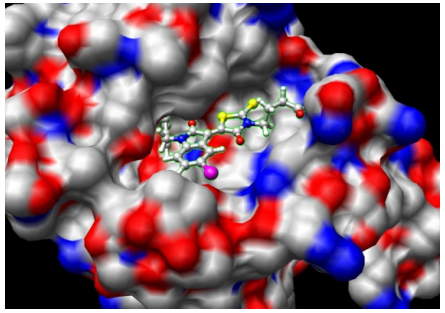
- Term often used in science & engineering
  - Also known as ‘from first principles calculations‘ or ‘from the beginning‘
- E.g. Terrestrial Systems
  - Terrestrial systems: ab initio calculation of liquid water includes the properties of the constituent hydrogen and oxygen atoms
  - Calculation use these properties with the established laws of electrostatics and quantum mechanics (e.g., 1 water molecule → groups of them → water)
- E.g. Bioinformatics / Biophysics
  - Bioinformatics: term used to define methods for making predictions about biological features using a computational model
  - Biophysics: term used to define methods for the prediction of protein structures in protein folding (cf. Short Lecture 13) as computational model

- **Ab initio calculations rely on basic & established laws of nature w/o additional assumptions**
- **Ab initio calculations in computational chemistry refer to methods based on quantum chemistry**

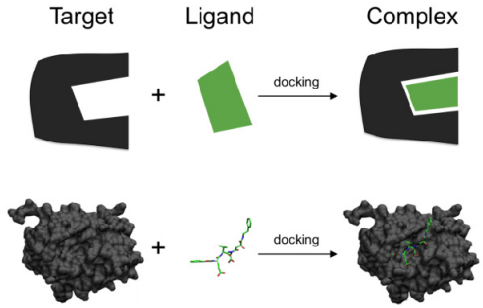
modified from [12] Wikipedia on ‘Ab initio’

Molecular Systems – Terminology – Molecular Docking

- Approach
  - E.g. Geometric matching shape complementarity methods describe the protein (cf. Short Lecture 13) and ligand as a set of features that make them dockable
- Computational complexity
  - Relatively small (simulation of docking becomes more larger)
  - E.g. nicely parallizable (i.e. HTC)
- Selected applications fields
  - Used to predict the binding orientation of small molecule drug candidates to their protein targets
  - The binding orientation in turn predict the affinity and activity of the small molecule



(docking of a small molecule ligand to a protein receptor to produce a complex)



(once docked a simulation of time is useful to apply)

- **Molecular docking is a method to determine whether one molecule can bind to another molecule**
- **Predicting strength of the association or binding affinity between two molecules is performed with a score function**

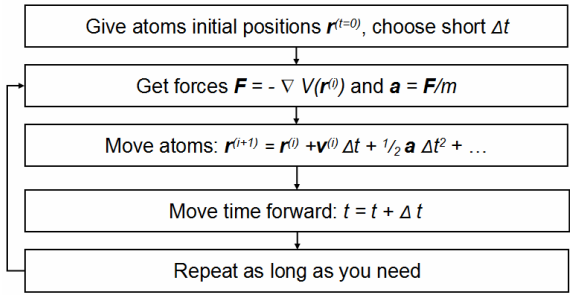
modified from [13] Wikipedia on 'Molecular Docking'

Molecular Systems – Terminology – Molecular Dynamics

- Approach (simplified)
  - Trajectories of atoms/molecules are determined by numerically solving Newton's equation of motion for a 'system of interacting particles'
  - Forces between particles and potential energy are defined by molecular mechanics force fields
- Computational complexity
  - Complex molecular systems consist of a vast number of particles
  - Impossible to find solutions analytically
  - Use numerical methods (cf. Short Lecture 12)
- Selected application fields
  - Chemical physics, materials science, biomolecule modelling, drug development, etc.

(electron – sub-atomic particle)

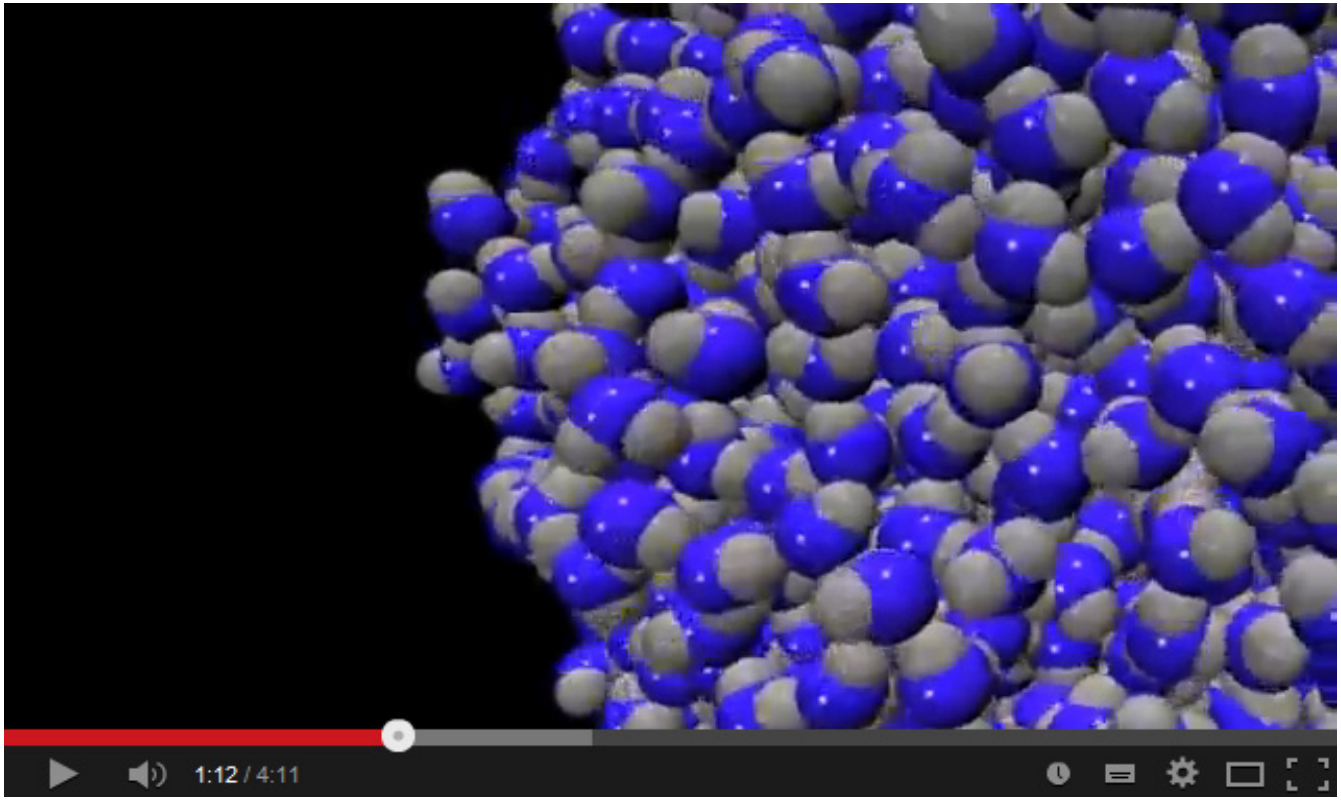
(atoms/molecules – microscopic particle)



- **Molecular dynamics (MD) refers to the simulation of physical movement of atoms and molecules**
- **Atoms (and whole molecules) interact for a period of time to simulate the motion of the atoms**
- **MD is an iterative HPC simulation method calculating & solving the equations of motion based on the acceleration obtained from the new forces**

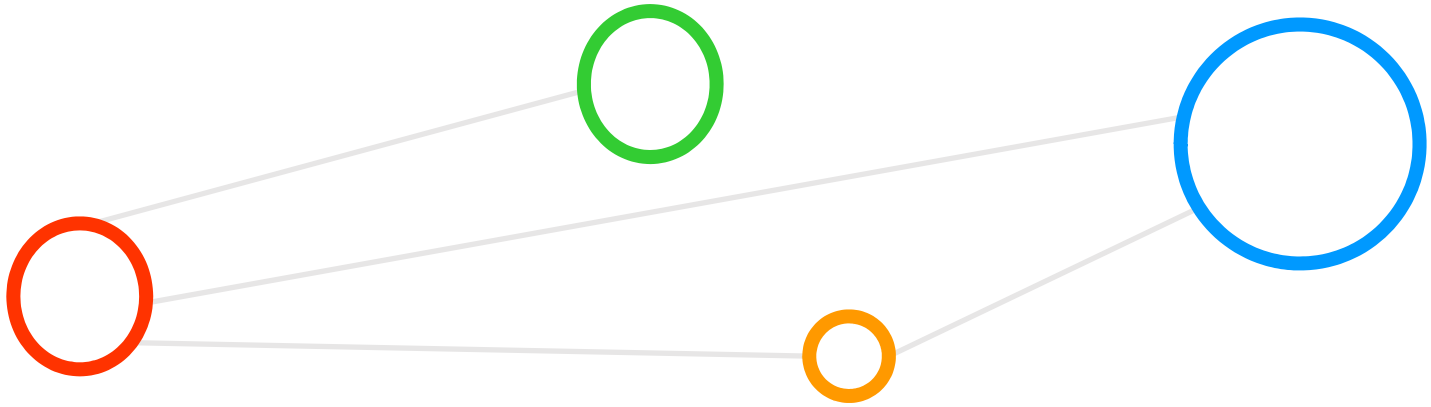
modified from [14] Wikipedia on 'Molecular Dynamics'

[Video] Molecular Dynamics Summary



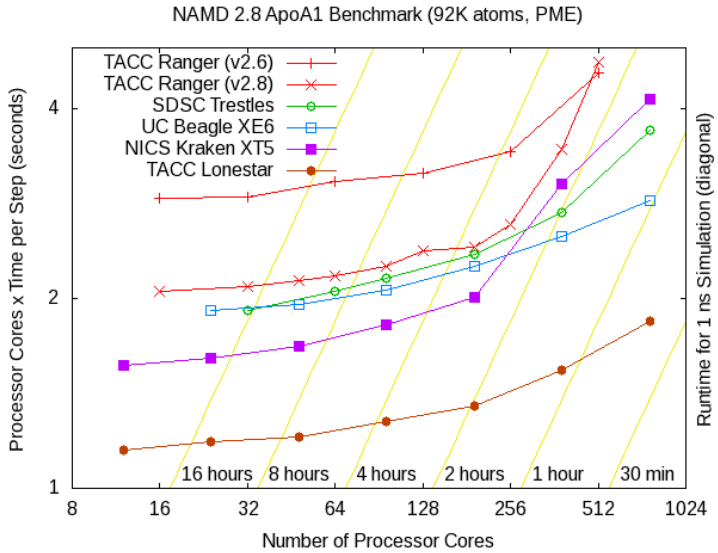
[15] YouTube Video, Introduction to MD

Selected Methods & Libraries



Selected Methods & Libraries – NAMD

- Selected Facts
  - Uses the molecular graphics program VMD for simulation setup
  - Free software package and open source
- Parallelization
  - Based on Charm++
    (abstraction of MPI, cf. Lecture 2)
  - Scalable to hundreds of cores
    (beyond 200,000)
- Example
  - 92224 atoms simulation
    on a 1ns simulation timescale

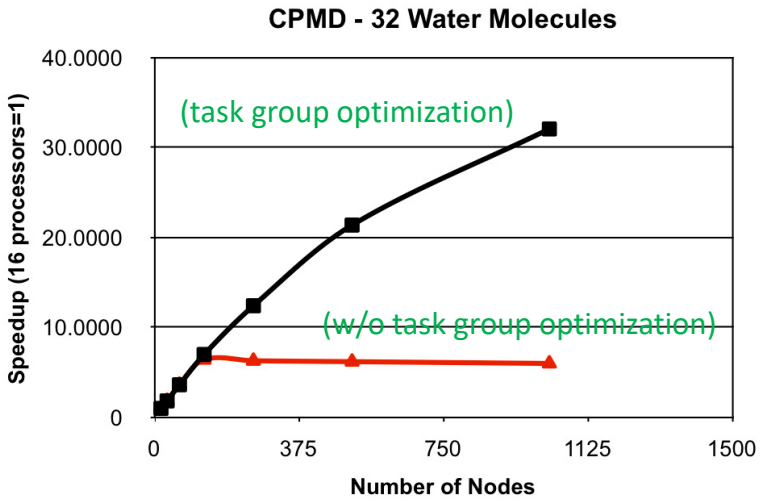


- **Nanoscale Molecular Dynamics (NAMD) is a parallel code designed for HPC simulations of large biomolecular systems to analyse trajectories and study molecular systems**

modified from [17] NAMD Webpage

Selected Methods & Libraries – CPMD

- Selected Facts
  - Plane wave / pseudopotential implementation of density functional theory (DFT) designed for ‘ab initio molecular dynamics‘
  - Free software package (for non-profit organization)
- Parallelization
  - Shared memory with OpenMP (cf. Lecture 6)
  - Distributed memory with MPI (cf. Lecture 2)
- Example
  - Detailed parallel simulation of 32 water molecules

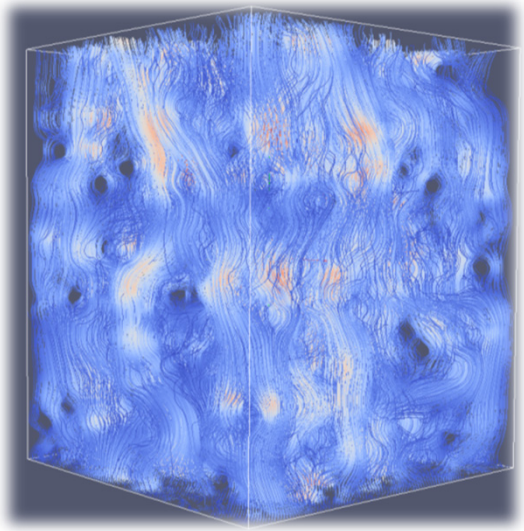


- **Car Parinello Molecular Dynamics (CPMD) is a parallel program used for ab initio electronic structure & molecular dynamics simulations**

modified from [18] CPMD Webpage

Molecular Systems Parallel Algorithms – Revisited (cf. Lecture 5)

- Scientific case:
  - Understanding physical movements of atoms and molecules in the context of n-body simulations
- Molecular dynamics algorithms for interacting ‘particles‘
  - Determine trajectories of atoms and molecules
  - Numerically solving the Newton‘s equations of motion
  - Forces between particles and potential energy is parallel computed according to molecular mechanics force field methods
- Using a library
  - E.g. MP2C code: particle-based hydrodynamics (fluid simulations)

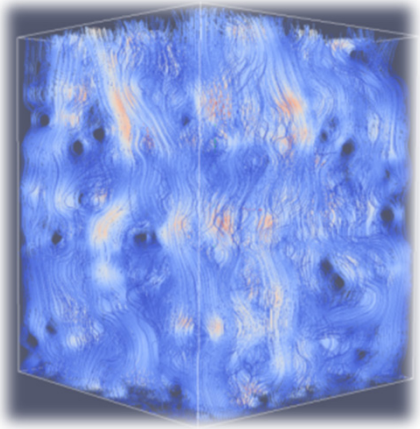


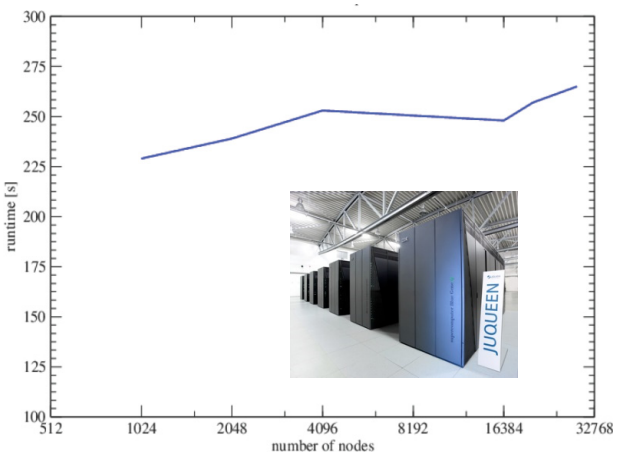
Flow field in a gas diffusion membrane

Selected Methods & Libraries – MP2C

- Selected Facts
  - Implements a hybrid representation of solvated particles in a fluid
- Parallelization
  - Based on MPI (cf. Lecture 3)
  - Scalable parallel/IO (via SionLib library)
- Example
  - Flow field in gas diffusion membrane
  - Particle-based hydrodynamics application (highly scalable fluid simulation)

(Flow field in a gas diffusion membrane)





- **MP2C stands for Massively Parallel Multi-Particle Collision Dynamics**
- **MP2C is a molecular dynamics code that focuses on mesoscopic particles**

[19] MP2C Web page

➢ **Lecture 15 will give further details on computational fluid dynamics (CFD) techniques & codes including Finite Elements Method (FEM)**

Computational Fluid Dynamics – Methods Overview

Real Fluid
Free molecules in continous space

Continuum Kinetic Theory
Microscopic particles (Boltzmann Equation)

Conventional CFD Methods

Construction of fluid equations
Navier-Stokes equations (PDE)

Discrete approximation of PDE
Finite difference, finite element, etc

Numerical integration
Solve the equations on a given mesh and apply PDE boundary conditions

Lattice Based Method

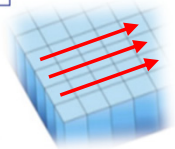
Discrete formulation of kinetic theory
Lattice Boltzmann equations

No further approximation
The equations are already in discrete form

Numerical integration
Solve on lattices and apply kinetic based BC

Simple conversion to fluid variables
These are theoretically shown to obey the required fluid equations

Results
Fluid dynamic quantities at discrete points in space and time

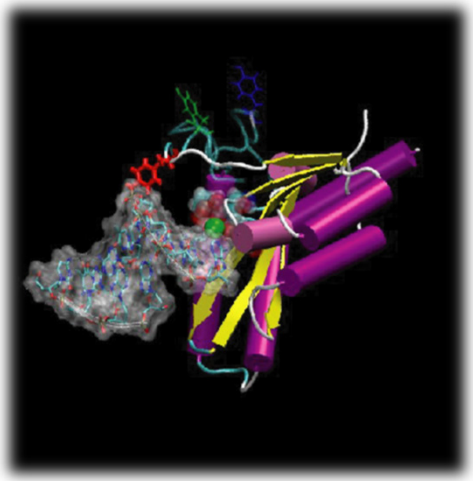


[22] S. Orszag et al.

➢ **Lecture 15 will give further details on computational fluid dynamics (CFD) techniques & codes including Finite Elements Method (FEM)**

Selected Methods & Libraries – AMBER

- Selected Facts
  - AMBER is a MD suite of several programs
- Parallelization
  - Full 3D domain decomposition (cf. Lecture 3)
  - Parallel NetCDF (cf. Lecture 4)
  - Offers CUDA GPU use (cf. Lecture 7)
- Example
  - Simulation of HIV-1 integrase enzyime
  - Understanding challenging diseases like retroviruses (e.g. HIV)



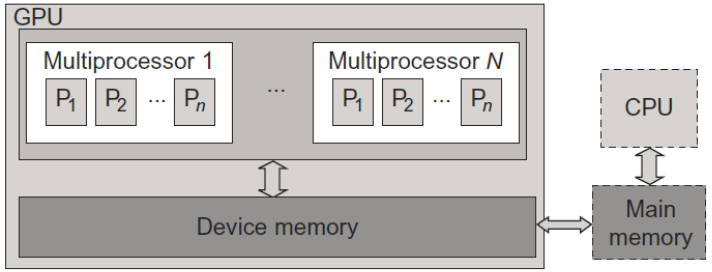
(Integrase is one of the three essential enzymes required for replication of the HIC-1 virus)

- **AMBER stands for Assisted Model Building with Energy Refinement**
- **AMBER is a set of molecular mechanical force fields for simulating biomolecules & MD software**

[20] AMBER

Many-core GPGPUs – Revisited (cf. Lecture 1)

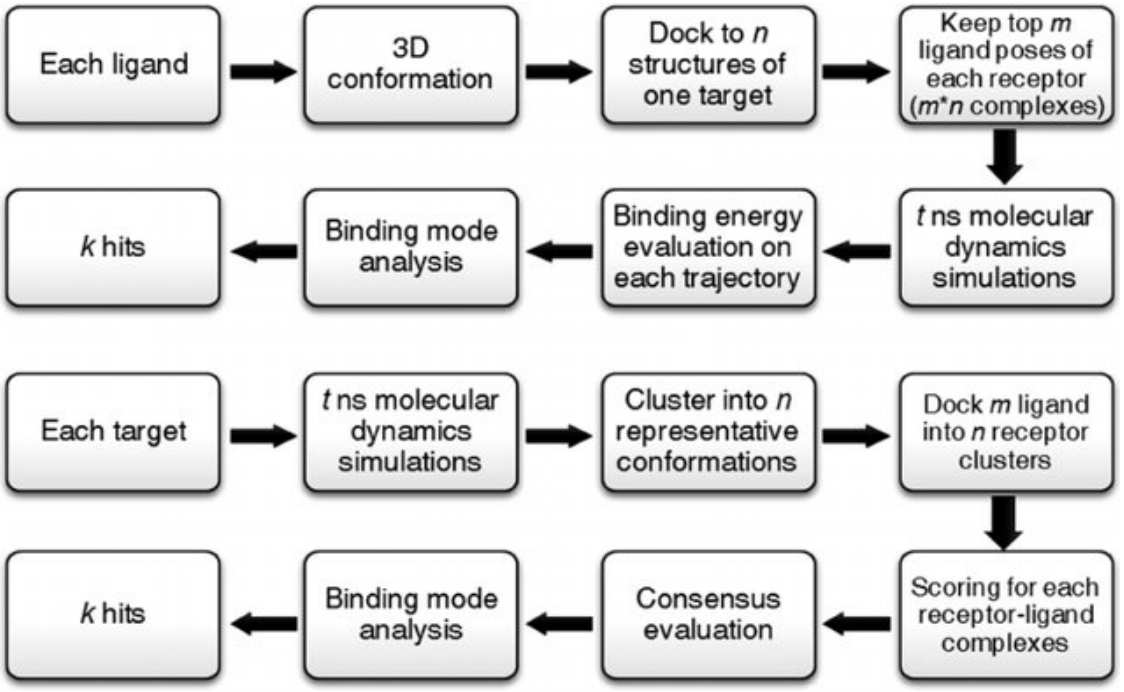
- Use of very many simple cores
  - High throughput computing-oriented architecture
  - Use massive parallelism by executing a lot of concurrent threads slowly
  - Handle an ever increasing amount of multiple instruction threads
  - CPUs instead typically execute a single long thread as fast as possible



- Many-core GPUs are used in large clusters and within massively parallel supercomputers today
  - Named General-Purpose Computing on GPUs (GPGPU)
  - Different programming models emerge

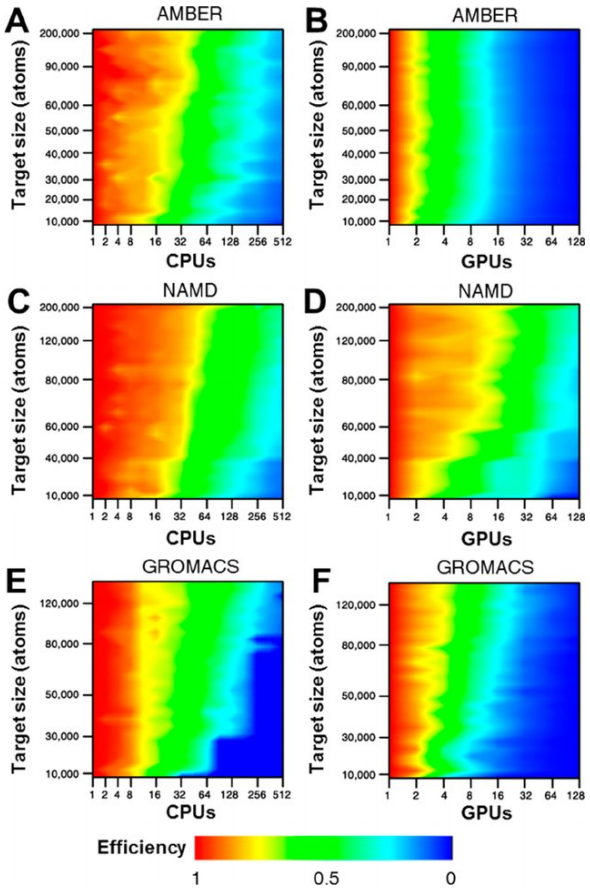
- **Graphics Processing Unit (GPU) is great for data parallelism and task parallelism**
- **Compared to multi-core CPUs, GPUs consist of a many-core architecture with hundreds to even thousands of very simple cores executing threads rather slowly**

Selected Comparisons – Application



[24] Hu Ge et al.

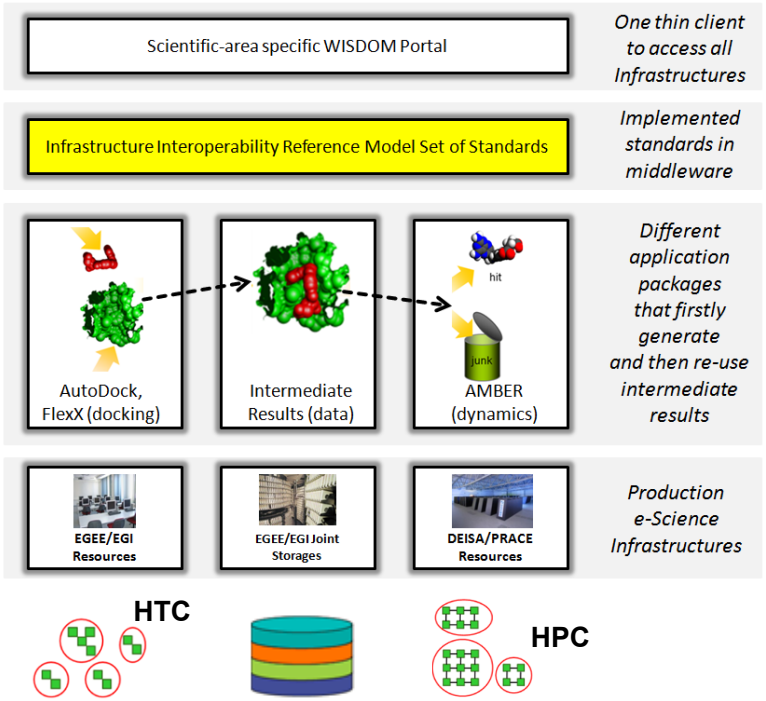
Selected Comparisons – Packages using CPUs & GPUs

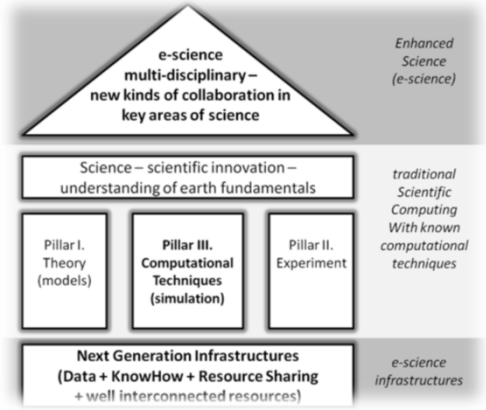


[24] Hu Ge et al.

Bioinformatics – Drug Development using Scalable Computing Infrastructures

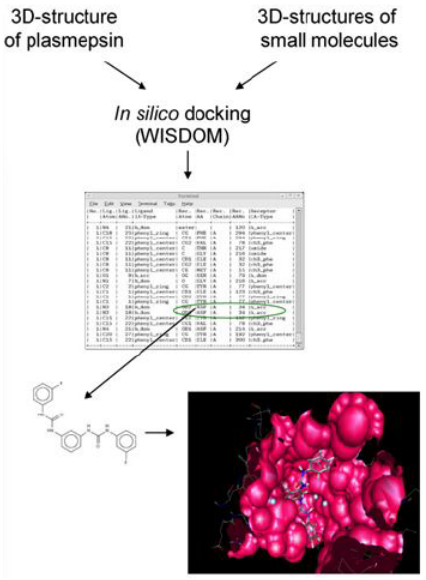
- Use of Interoperability of two different computing paradigms
  - E.g. use of HTC in EGI & HPC in PRACE infrastructures (cf. Short Lecture 11)



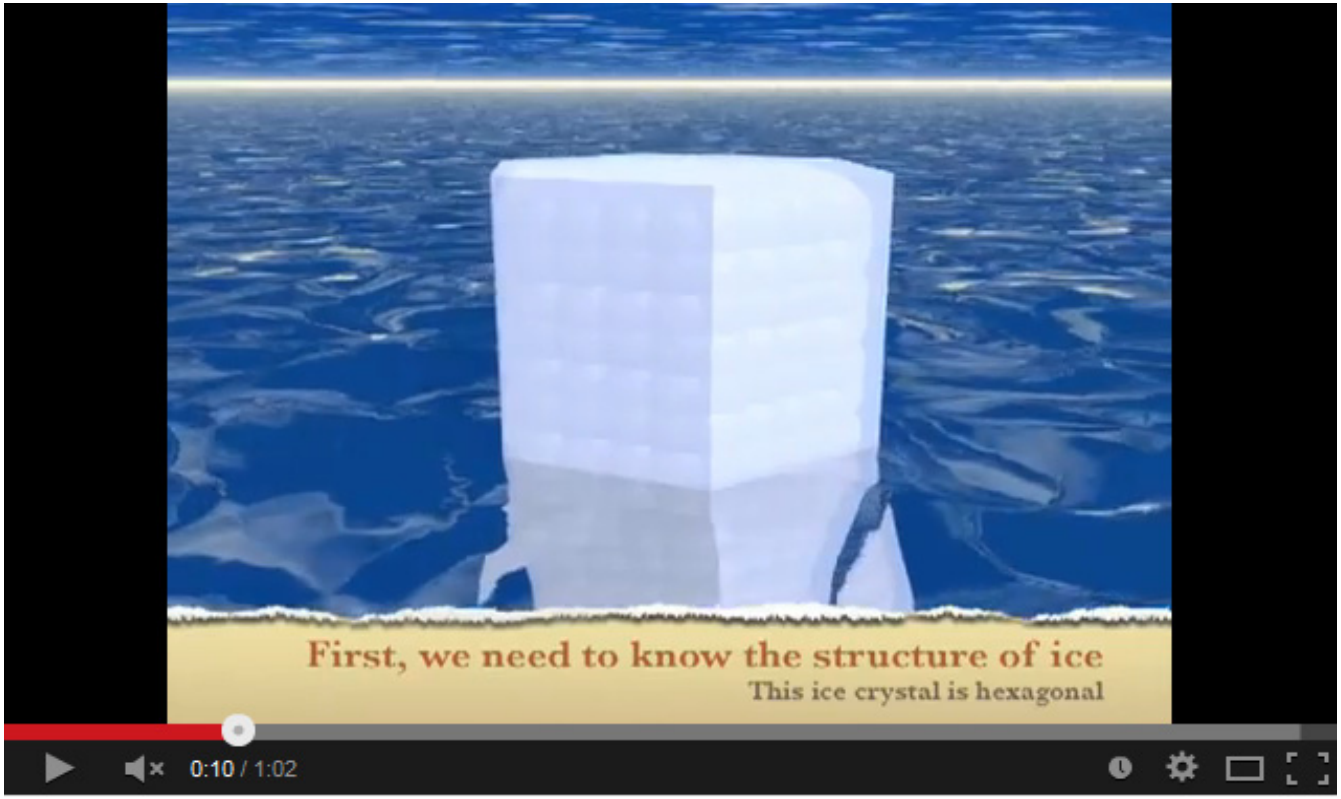


[21] M. Riedel et al., 'Research Advances by using interoperable e-Science Infrastructures', 2009

(Web-based portal make complex infrastructure access easy for biological domain scientists)

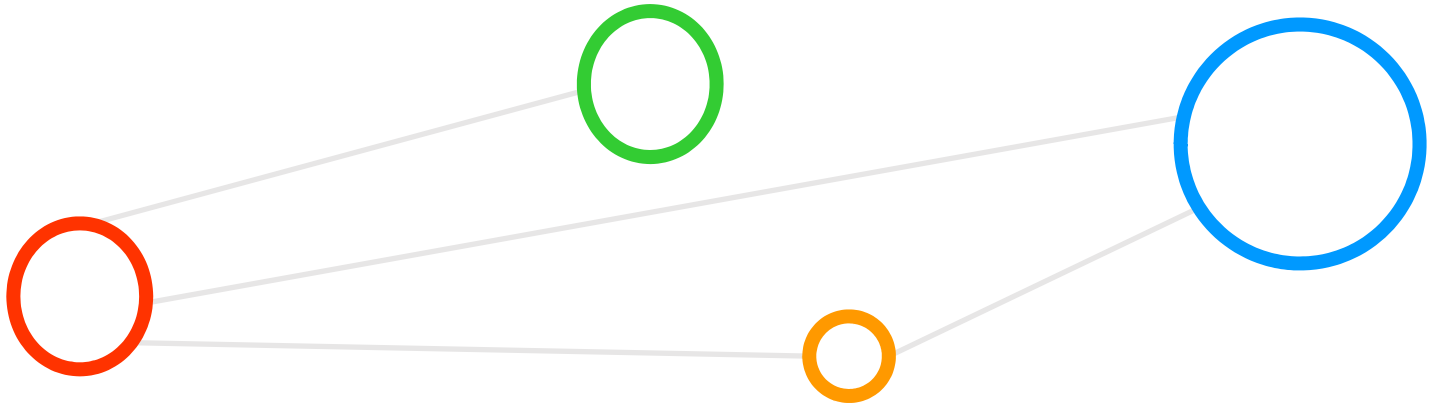


[Video] Towards Simulating Jökulsárlón



[16] YouTube Video, A molecule's eye view of ice melting

Lecture Bibliography



Lecture Bibliography (1)

- [1] Wikipedia on 'Systems Biology', Online: http://en.wikipedia.org/wiki/Systems_biology
- [2] Wikipedia on 'Protein', Online: http://en.wikipedia.org/wiki/Protein
- [3] JSC Simulation Lab Biology, Online: http://www.fz-juelich.de/ias/jsc/EN/AboutUs/Organisation/ComputationalScience/Simlabs/slbio/_node.html
- [4] Folding@Home, Online: http://folding.stanford.edu/
- [5] J. Archuleta et al., 'A Maintainable Software Architecture for Fast and Modular Bioinformatics Sequence Search, Sequence Searching, 2007
- [6] Wikipedia on 'Molecule', Online: http://en.wikipedia.org/wiki/Molecule
- [7] Wikipedia on 'Atom', Online: http://en.wikipedia.org/wiki/Atom
- [8] Wikipedia on 'Molecular Modelling', Online: http://en.wikipedia.org/wiki/Molecular_modelling
- [9] Wikipedia on 'Molecular Mechanics', Online: http://en.wikipedia.org/wiki/Molecular_mechanics
- [10] Wikipedia on 'Quantum Chemistry', Online: http://en.wikipedia.org/wiki/Quantum_chemistry
- [11] Wikipedia on 'Quantum Mechanics', Online: http://en.wikipedia.org/wiki/Quantum_mechanics
- [12] Wikipedia on 'Ab initio', Online: http://en.wikipedia.org/wiki/Ab_initio

Lecture Bibliography (2)

- [13] Wikipedia on ‘Molecular Docking‘, Online: http://en.wikipedia.org/wiki/Docking_(molecular)
- [14] Wikipedia on ‘Molecular Dynamics‘, Online: http://en.wikipedia.org/wiki/Molecular_dynamics
- [15] YouTube Video, ‘An Introduction to Molecular Dynamics‘, Online: http://www.youtube.com/watch?v=lLFEqKl3sm4
- [16] YouTube Video, ‘A molecule's eye view of ice melting‘, Online: http://www.youtube.com/watch?v=CDTZoFGmZoc
- [17] NAMD Web page, Online: http://www.ks.uiuc.edu/Research/namd/
- [18] CPMD Web page, Online: http://www.cpmd.org/
- [19] MP2C Webpage, Online: http://www.fz-juelich.de/ias/jsc/EN/Expertise/High-Q-Club/MP2C/_node.html
- [20] D.A. Case, et al., (2014), AMBER 14, University of California, San Francisco
- [21] M. Riedel et al., ‘Research Advances by using Interoperable e-Science Infrastructures’, Journal of Cluster Computing, 12(4):357–372, 2009, Online: https://www.researchgate.net/publication/220405901_Research_advances_by_using_interoperable_e-science_infrastructures
- [22] Steven Orszag et al., ‘Lattice Boltzmann Methods for Fluid Dynamics‘, Online: http://physics.wustl.edu/nd/event/qmcd09/Presentations/qmcd09Talks/orszag.pdf

Lecture Bibliography (3)

- [23] Wikipedia on ‘Schroedinger Equation‘, Online: https://en.wikipedia.org/wiki/Schr%C3%B6dinger_equation
- [24] Hu Ge et al., ‘Molecular Dynamics-Based Virtual Screening: Accelerating the Drug Discovery Process by High-Performance Computing‘ Journal of Chemical Information and Modeling 53(10), Online: https://www.researchgate.net/publication/256424059_Molecular_Dynamics-Based_Virtual_Screening_Accelerating_the_Drug_Discovery_Process_by_High-Performance_Computing


services
measurement
within
fusion
new
funding
Policy-based
concepts
device
Analysis
forms
cross-disciplinary
resources
Centre
climate
Computer
expertise
Approaches
construction
Fusion
year
like
Computational
Test
E.g
Computing
disciplines
results
Enable
different
Cross-Disciplinary
methods
High
big
nodes
PB
community
Environment
Modelling
EU
Science
centers
Key
scientific
important
files
model
international
SMAQ
engineering
References
etc
DKRZ
Use
storage
basis
Infrastructure
simulations
ENES
technologies
formats
brain
increasing
via
EUDAT
physical
IPCC
Big
Project
Data
processing
Common
TB
Methods
MapReduce
create
computing
Using
European
databases
GB
Supercomputing
Work
images
codes
analysis
Large
Summary
data-intensive
Service
Scientific
levels
often
DLCL
stored
Resources
analyze
external
use
Level
Research
Matters
Concrete
performance
JSC
computational
research
e.g
science
HPC
Juelich
system
datasets
acquisition
large-scale
compute
Landscape
management
System
Climate
modelling
AR5
Hardware
distributed
cores
access
hundreds
Nutshell
Services
DLCLs
Understanding
structures
Technologies
Earth
Structure
current
network
Germany
Simulation
directory
project
General
Provide
NASA
Energy
systems
day
Processing
Health