High Performance Computing and Visualization at the School of Health Information Sciences

Stefan Birmanns, Ph.D.
Postdoctoral Associate
Laboratory for Structural Bioinformatics

Outline

- High Performance Computing
  - Supercomputer Architectures
  - SHIS Cluster Computer System
  - Applications
- Virtual Reality
  - Virtual Reality / Haptic Rendering
  - SHIS VR system
  - SenSitus

High Performance Computing

- Definition of "Supercomputer", HPC?
  - *Computational facilities substantially more powerful than current desktop computers*

- *Performance*
  - *Flop is a floating point operation per second*
  - *Clock speed*
  - *Peak performance = Maximal calculation speed of CPU*
  - *Actual performance depends on application, memory bandwidth, interconnection network, etc.*

- *Parallelism*
  - *Multiple calculation units within a CPU, multiple CPUs, etc.*
  - *Locality of problem*

High Performance Computing

Cray 1 (1976)

Cray T3E (1995)

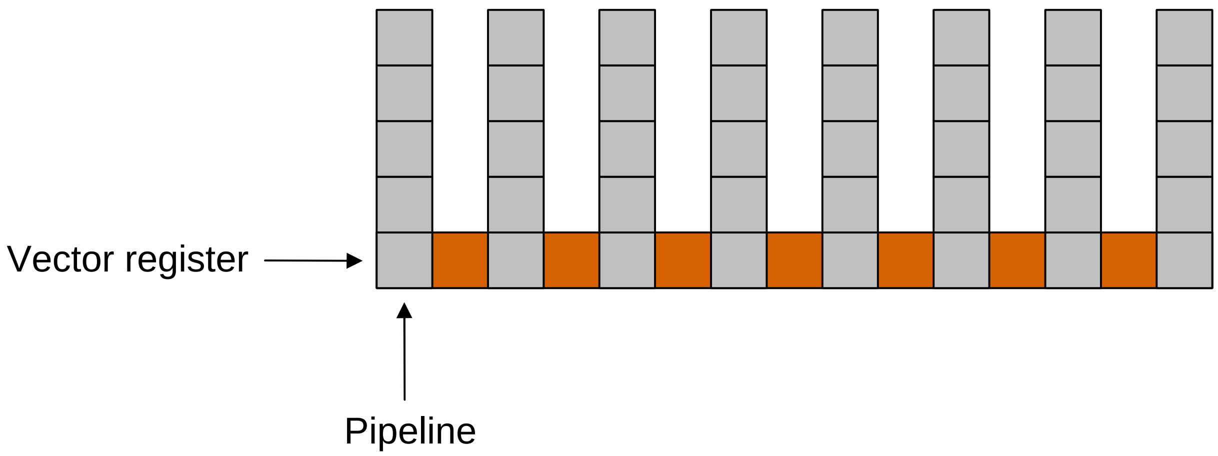
System X (2001)

Earth Simulator (2002)

Blue Gene (2004)

Vectorcomputer

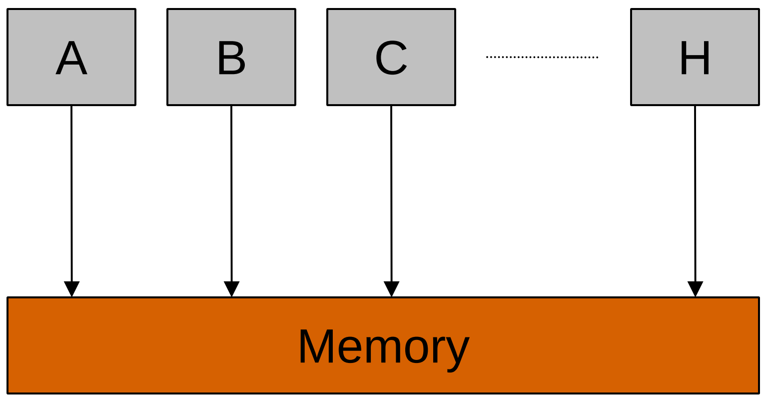
- SIMD parallelism
  - Bandwidth between memory and CPU dramatically increased



- Increases CPU size, only few special vector commands possible
- Easy to adapt existing code, but not all problems benefit of SIMD parallelism
- Expensive

SMP

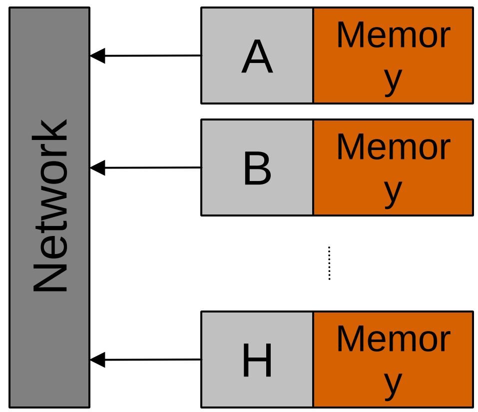
- Multiprocessor Supercomputers
- Shared memory
  - Multiple CPUs have access to global memory



  - Problem: Conflicts when accessing the same memory location
  - Fast communication, fast memory access, easy to program
  - Complex system architecture, limits the number of CPUs

Distributed Memory HPC

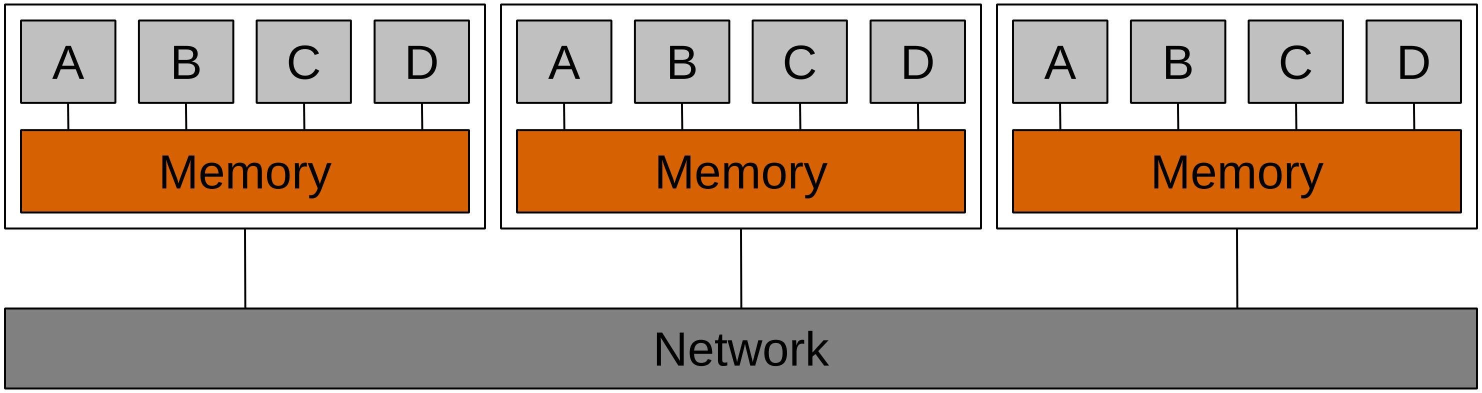
- Distributed memory
  - CPUs have local memory



  - Parallelization complicated
  - Communication bottleneck
  - Advantage:
    - Simple system design,
    - Hardware scales very well in respect of number of CPUs (“massive parallel” systems)

High Performance Computing

- Hybrid designs
  - E.g. cluster systems with SMP nodes



  - High CPU count and faster communication – but optimization difficult
  - Earth Simulator (Japan): 640 nodes
  - Nodes: 8 vector processors with shared memory

High Performance Computing

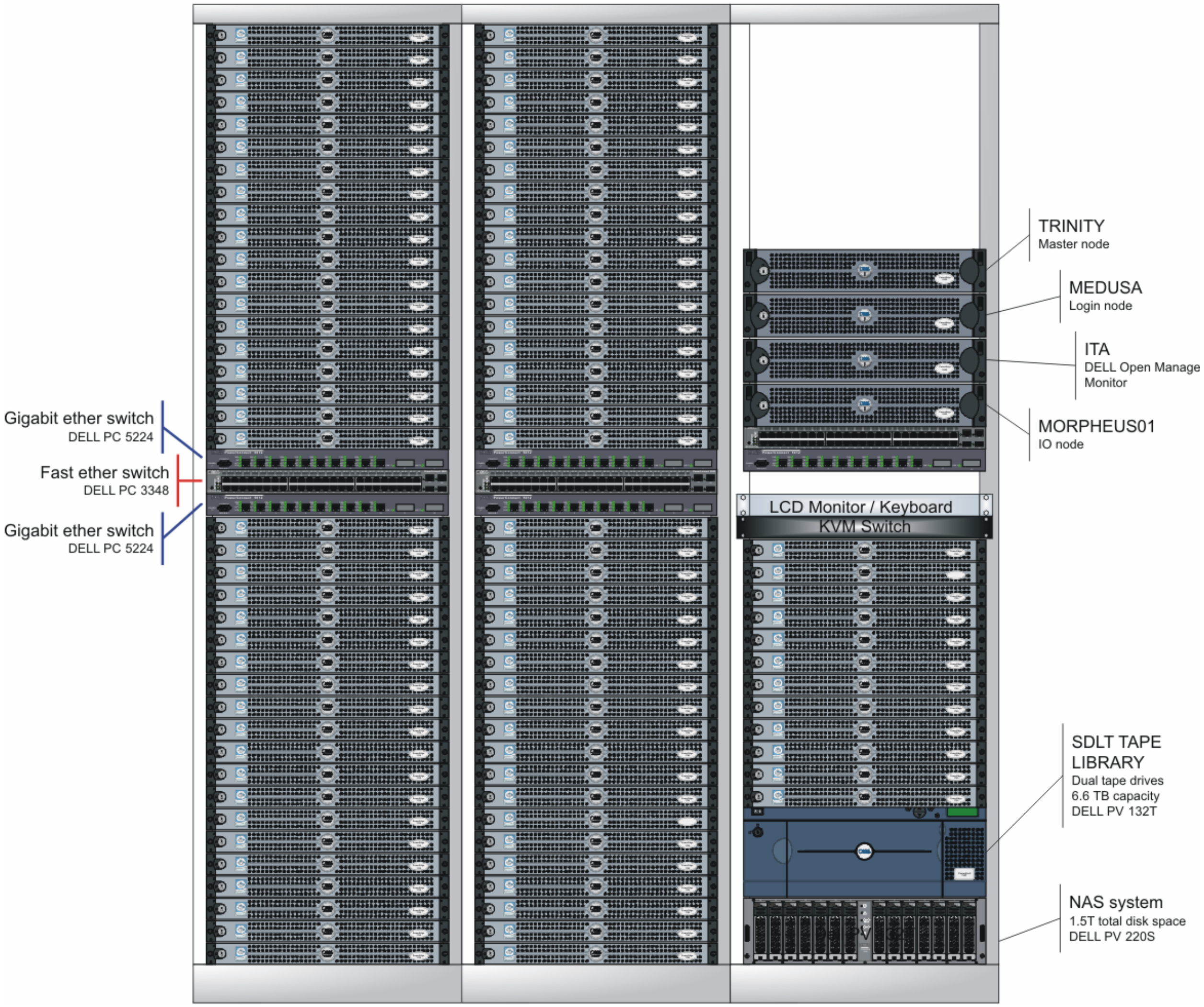
- Research areas / trends in HPC:
  - Unlike earlier a lot of HPCs are built using of-the-shelf-hardware
    - Viginia Tech “Big Mac” (1100 Apple Dual G5)
    - PC cluster systems
    - Problems:
      - Space and power consumption, heat
      - Interconnection networks (bandwidth, latency, cpu overhead)
      - Reliability
    - This made HPC affordable for smaller institutions!
  - Performance analysis
    - Why does a program not scale well?
    - What is the “speed” of a supercomputer?
  - Strategies for problem decomposition
    - Make a parallel code scale better
  - GRID
    - Provide transparent access to supercomputer resources

SHIS HPCC

- SHIS Cluster System
  - PC cluster computer with SMP nodes:
    - 90 nodes with 2 Xeon CPUs each (2.8 GHz)
    - CPUs support Hyperthreading
    - 2 GB RAM each node
    - 80 GB HD each node
  - Gigabit ethernet interconnection network
  - 1.5 TB global harddrive space
  - 6.8 TB SDLT tape library

- Performance:
  - 0.59 TFlops
  - 1 TFlops (Peak)

SHIS HPCC



Gigabit ether switch
DELL PC 5224
Fast ether switch
DELL PC 3348
Gigabit ether switch
DELL PC 5224
TRINITY
Master node
MEDUSA
Login node
ITA
DELL Open Manage
Monitor
MORPHEUS01
IO node
LCD Monitor / Keyboard
KVM Switch
SDLT TAPE
LIBRARY
Dual tape drives
6.6 TB capacity
DELL PV 132T
NAS system
1.5T total disk space
DELL PV 220S

SHIS HPCC

- Software
  - Linux RedHat 9 Operating System
  - Message Passing Interface - MPI (MPICH[1] and MPIpro[2])
  - Portable Batch System – PBSPro[3]
  - Ganglia Performance Monitor[4]

- SHIS Team
  - HPCC Committee
    - Dr. Jiajie Zhang
  - Support:
    - David Ha

1) http://www-unix.mcs.anl.gov/mpi/mpich

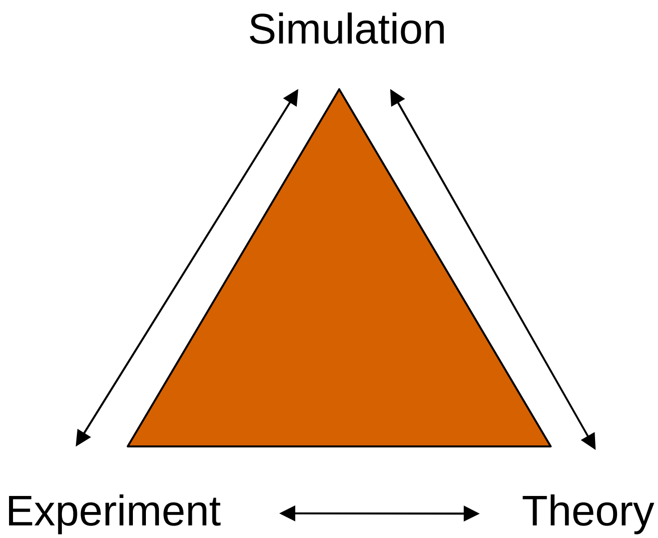
2) http://www.mpi-softtech.com

3) http://www.pbspro.com

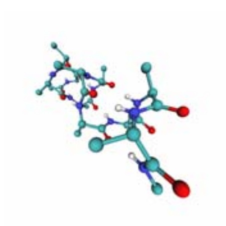
4) http://ganglia.sourceforge.net

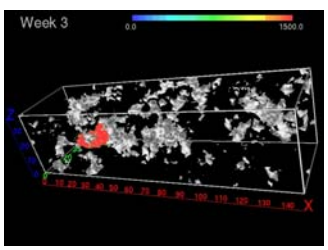
SHIS-HPCC

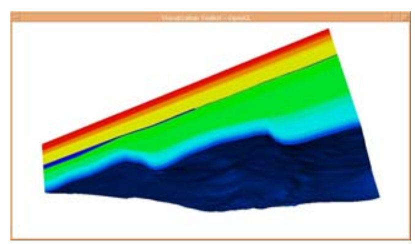
Applications

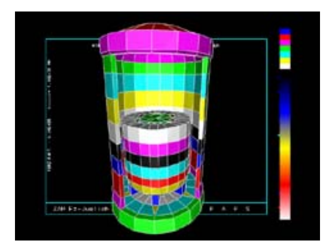


- Simulation
  - Biophysics
  - Meterology
  - Fluid dynamics
  - Finite element calculations
  - Traffic simulations
  - Artificial life









Applications

- How to develop applications for a supercomputer?
- Automatic parallelization
- Programming languages
  - Fortran, C, C++
  - Interpreted languages problematic (Java, Python, Perl, …)
  - Optimization
- Programming model depends on architecture
  - MPI (distributed memory)
  - OpenMP (shared memory)

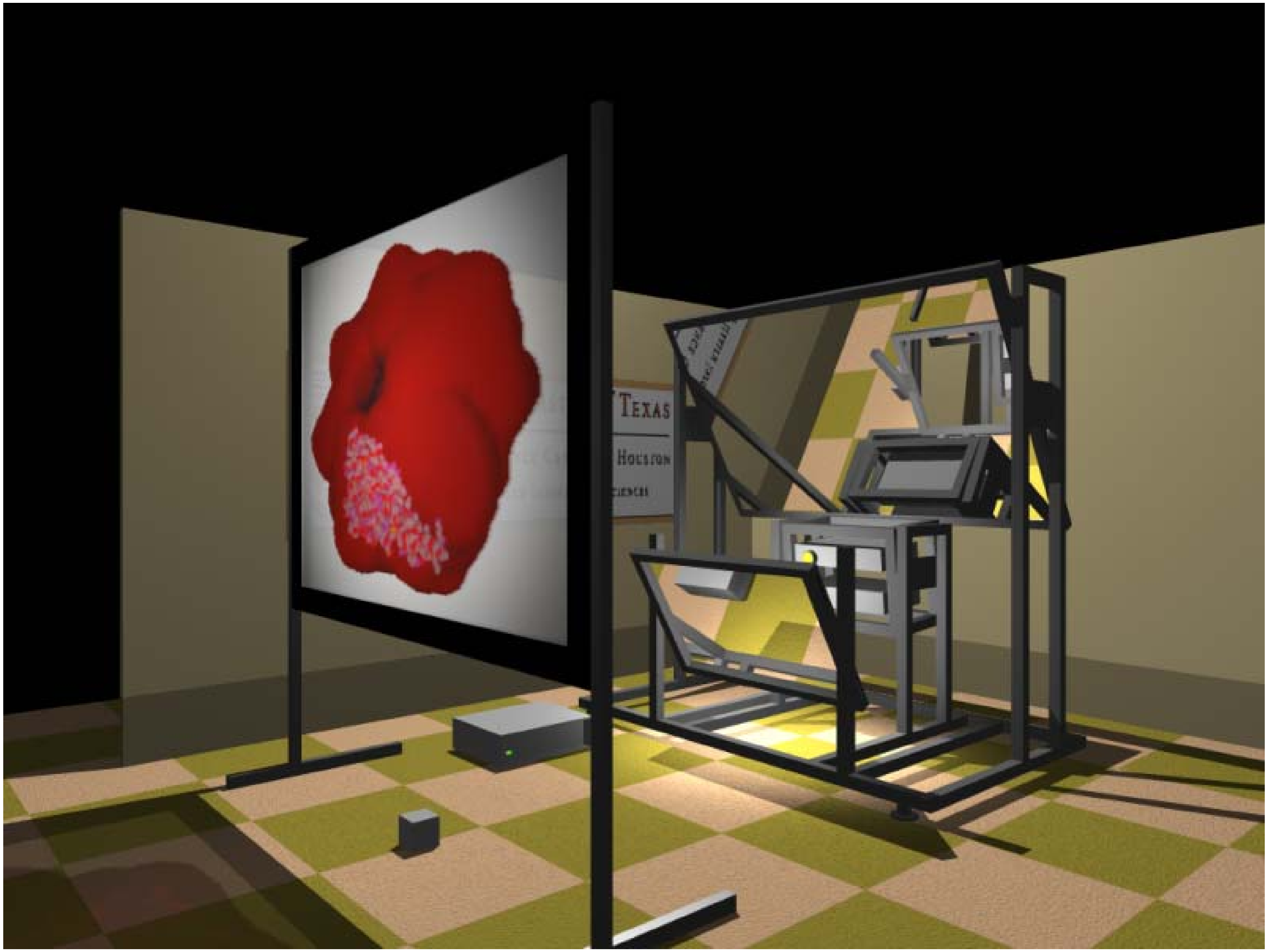
Visualization

- Visualization essential to analyze data sets from HPC simulations
  - Exploration of datasets
    - Discover
    - Decide
    - Explain
- Online supervision and / or steering of simulations

- Challenges
  - Interactive framerates ~30 FPS
  - Size of datasets is increasing dramatically every year
  - Development of special rendering techniques necessary
  - Network bandwidth limited or latency problems
  - Find useful representations for multidimensional datasets

Virtual Reality

- Goal: Interact with virtual objects like with real objects
- VR systems:


Texas
Houston

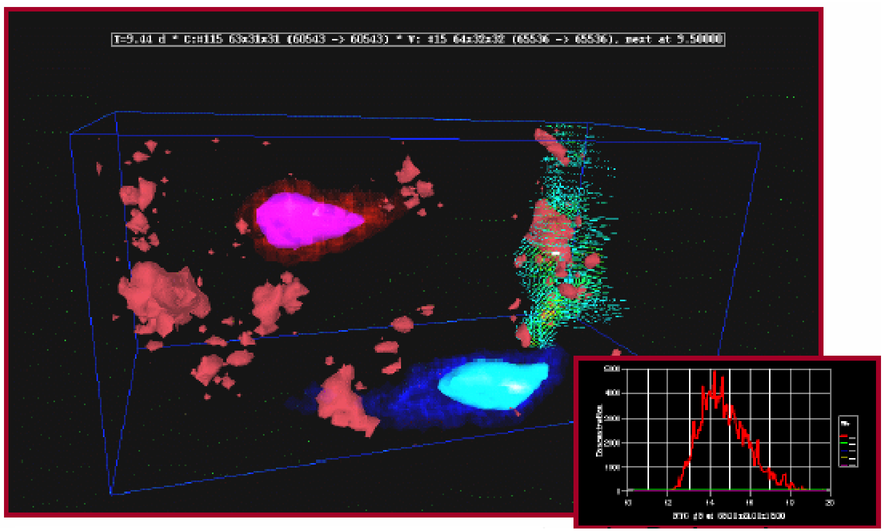
SHIS VR System

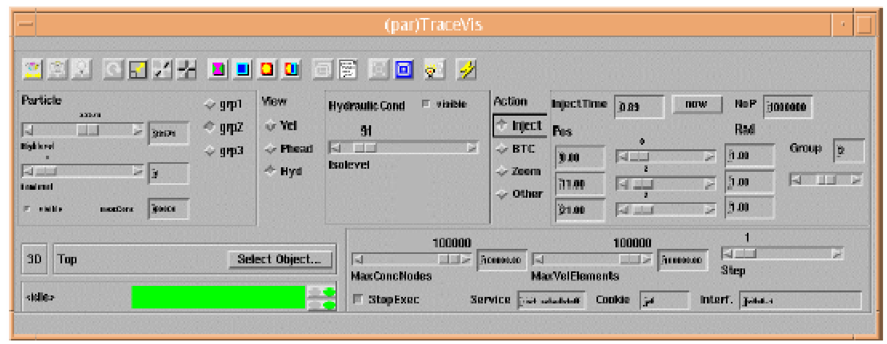
- Components
  - 2 DLP InFocus LP530 projectors
  - Polarization filters for passive stereo
  - Steward (polarization preserving) screen and mirror system
  - Polhemus Fastrak electro-magnetic position tracker
    - 3 standard sensors and 1 stylus-like sensor
  - Computer system
    - Dual Xeon 3 GHz
    - nVidia Quadro FX 2000
    - 2 GB Ram
  - RedHat / Fedora Core Linux
    - Supports OpenGL and passive stereo

Polhemus Fastrak

Virtual Reality

- Steering of supercomputer applications
  - Solute transport in variable saturated porous media
  - Simulation "TRACE"[1] (environmental research)
  - Online supervision of simulation running on massive parallel HPC with TraceVis[2]
  - Steering: Injection of solute into simulated area during simulation run





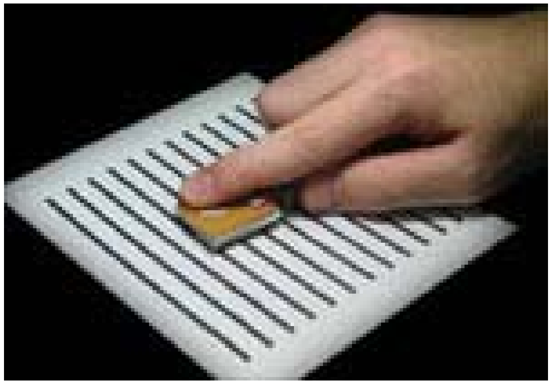
1) http://www.fz-juelich.de/icg/icg4

2) http://www.fz-juelich.de/zam

Haptic Rendering

- Haptic Rendering
  - *Haptesthai* – greek “to touch”
  - Create an artificial tactile sensation
- Applications:
  - Experience surface / mass of virtual objects
  - Teleoperation / telerobotics
  - Exploration of multidimensional datasets
- Challenges:
  - Design of haptic devices
  - High temporal bandwidth:
    ~1000 force updates per second

Haptic Devices



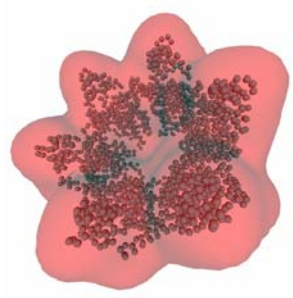
- Sensable Phantom 6DOF
  - Original device developed at the MIT
  - Indirect haptic device
  - Translational forces and torques
  - 6D position / orientation sensors

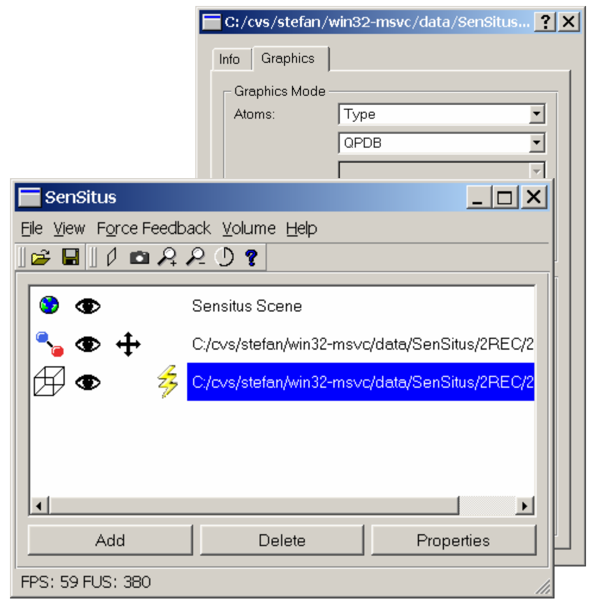
SensAble technologies

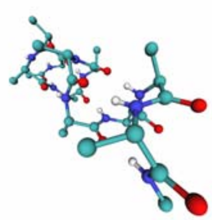
Sculptor

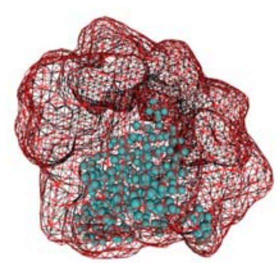
- Sculptor
  - Interactive multiresolution fitting by haptic rendering
  - Visualization of biophysical datasets
  - Support of VR systems

- Research funded by: Human Frontier Science Program
  - UTH/SHIS Laboratories for Biocomputing and Imaging
    - Willy Wriggers (USA)
  - School of Science and Engineering
    - Takeyuki Wakabayashi (Japan)
  - CRNS, Laboratoire de Genetique des Virus
    - Jorge Navaza (France)
  - RCJ, John von Neumann Institute for Computing
    - Herwig Zilken (Germany)



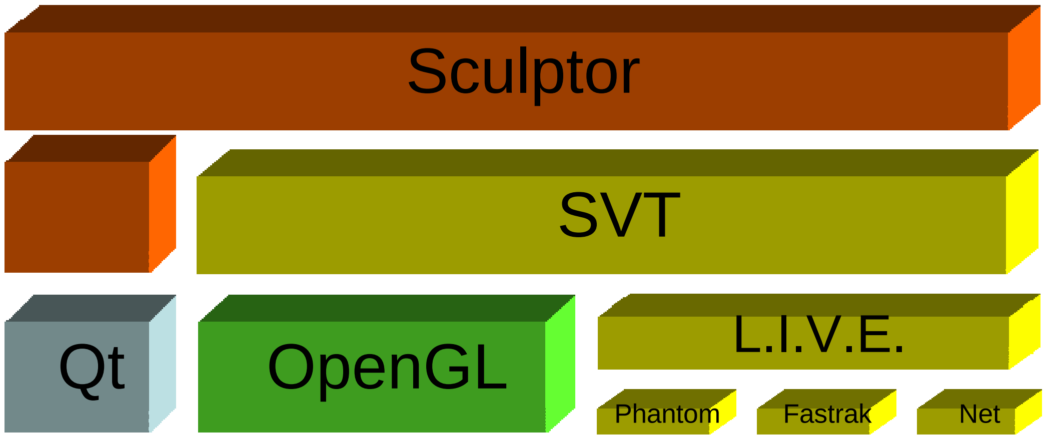






sculptor.biomachina.org

Sculptor

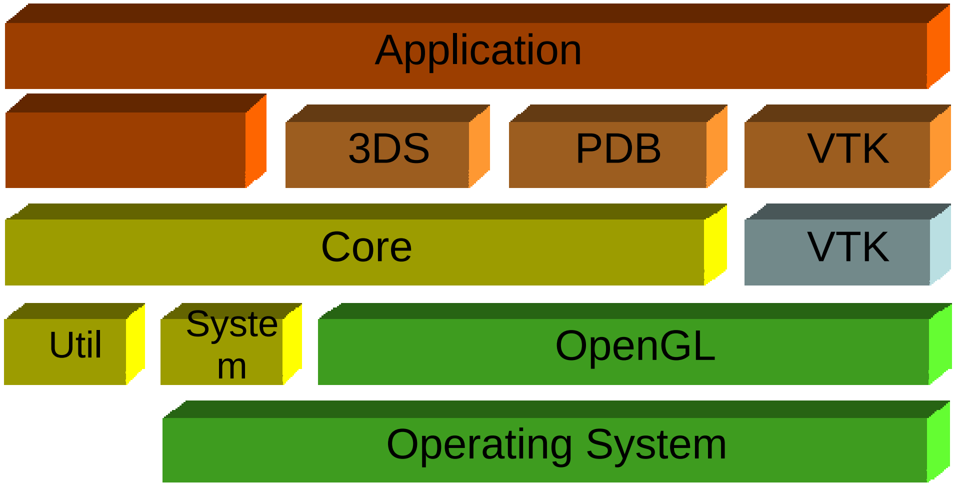


- Sculptor:
  - Qt[1] GUI library
  - OpenGL[2] 3D graphics library
  - SVT - Underlying VR and visualization toolkit
  - Multiplatform (Linux, SGI, Windows)

1) http://www.opengl.org

2) http://www.trolltech.com

Sculptor



- SVT:
  - Multi-Display VR environments
  - No dependencies to other libraries
  - Encapsulation of all system dependent calls

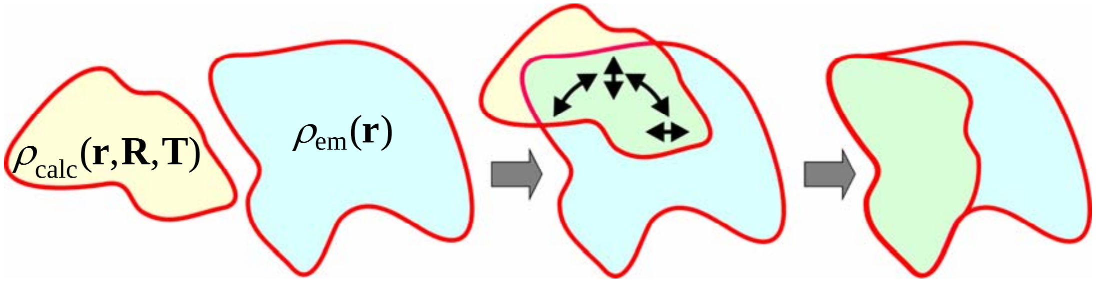
Interactive Multi-Resolution Fitting

- Fitting of high-resolution crystal structures into low resolution electron density maps
  - High-resolution molecular structures obtained by x-ray crystallography
  - Low-resolution electron microscopy volumetric maps
- New interactive fitting approach using haptic rendering
- Force calculation
  - Gradient of the cross correlation coefficient
  - Guide user to better fitting location

Interactive Multi-Resolution Fitting

- Cross correlation coefficient

$$C(R,T) \propto \int \rho_{calc}(\mathbf{r}, R, T) \cdot \rho_{em}(\mathbf{r})\, d^3\mathbf{r}$$

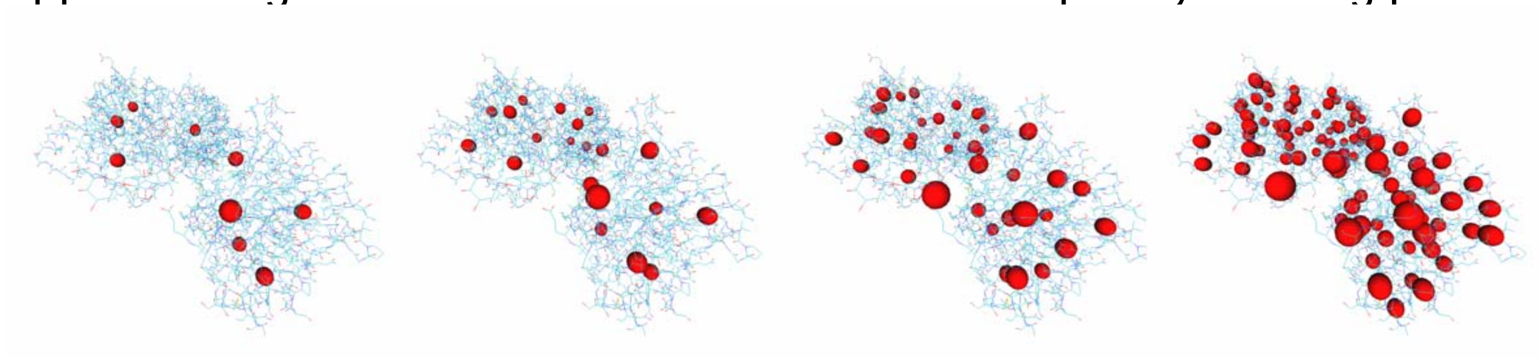


- Problem: Not time efficient enough

Reduced Fitting Criterion
Vector Quantization

- Vector Quantization
  - Popular method in signal processing
  - Replace complex function by small number of feature vectors
  - Topology Representing Networks (Martinez, Schulten)

- Applied to high-resolution structure to reduce complexity of fitting problem:

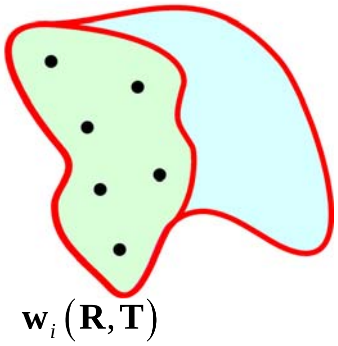


10CV 20CV 40CV 100CV

Reduced Fitting Criterion
Cross Correlation

- Replace complex crystal structure by feature vectors

$$\rho_{calc}(\mathbf{r}) \equiv \sum_{i=1}^{k} \delta(\mathbf{r} - \mathbf{w_i})$$



- Reduced cross correlation criterion:

$$C(R,T) \propto \sum_{i=1}^{k} \rho_{em}(\mathbf{w_i}(R,T))$$

- Reduced criterion is time efficient enough for haptic rendering

Reduced Fitting Criterion
Cross Correlation

- By using this reduced fitting criterion we were able to achieve update frequencies >1KHz