Computational Sciences at Uni Computing

Visual Computing Forum
2012-04-13

Klaus Johannsen, Uni Computing, Director
Overview

- Introduction
- Activities
- Visualization
- Todo
Introduction

Uni Computing is a department of Uni Research AS, the research company associated with the University of Bergen.

Uni Computing is organized in five groups with more than 60 staff.

Our Vision

Uni Computing carries out research and development in basic and applied areas with a focus on computational techniques.

We seek the comprehensive uptake of our results, methods, services and competences within science, industry and wider society.
Introduction (cont'd)

- Uni Computing is organized in five groups:
  - CBU: *Bio-informatics, research in molecular biology and genomics*
  - CEU: *Computational ecology, individual based population dynamics*
  - CLU: *Language technology, computational linguistics, Lexicography, computational media analysis*
  - EFG: *Oceanography, meso- and micro-climatology, CFD, wave-modeling, artificial intelligence*
  - Parallab: *High Performance Computing, e-Infrastructures, Programming*
Introduction (cont'd)

- Uni Computing research activities are pretty heterogeneous

Proteinase 3: Key amino acids for ligand recognition and membrane binding (Reuter et.al., CBU)

Cod: Spawning distribution along the Norwegian coast, 1910 and 1948 (Jørgensen et.al., CEU, 2008)

A graph displaying blog posts collected around the topic of climate change (Salway et. al., CLU, 2012)
Introduction (cont'd)

- Uni Computing research ...

Simulation of oil spill, Rocknes accident at Vatlestraumen (Torsvik et.al., EFG, 2009)

Ormen Lange process studies (Avlesen et.al., EFG, 2005)

Hexagon: Supercomputer operated in collaboration with II/UiB (Parallab, 2012)

HPC-Europa2: Usage pattern of a postprocessing e-infrastructure (Anderlik et.al., Parallab, 2011)
Activities: The WWW-Column (EFG)

- Focusing on what makes sense (in Norway). Integrating
  - Wind, waves & currents in fjords
  - Atmospheric & marine dispersion
  - Marine physics & ecology

- With application to
  - Offshore wind: resource assessment & forecasting
  - Env. management of aquaculture
  - Marine oil spills & releases
  - Subsea CO₂ storage: hazards & impact assessment
Activities: NORCOWE (EFG)

- WP1 - Wind and ocean conditions (Uni lead)
  - Climatology of met / ocean conditions
  - Modelling of the atmospheric boundary layer over sea
- WP2 - Offshore wind technology & innovative concepts
- WP3 - Offshore deployment & operation
- WP4 - Wind farm optimisation
- WP5 - Common themes
  - Education
  - Impact assessment
  - Infrastructure
  - Data storage & management
Activities: AI & Floating Wind Farms (EFG)

- A network of *Artificial Neural Networks* to describe floating wind farms using
  - Simplified wind turbine models
  - Empirical models for turbine-turbine/wake-wake and other complex interactions

- With applications to
  - Short term power forecast
  - Optimal operational strategies

- Using/enhancing our knowledge about text-processing to form semantic units (e.g. key-statements)
- Relate units to form a knowledge-base to be analysed by
  - Social science researchers
  - Media monitoring companies
- Interactive visualization to understand the dynamics of human society
Activities: INESS (CLU)

- Norwegian Infrastructure to Explore of Syntax and Semantics
- Interactive, language independent system for hosting, building, accessing and exploiting treebanks
- Build a 50 million sentence treebank of Norwegian
- Step towards developing the next generation of language technology applications
Activities: New Services (Parallab)

- Anticipating the future to be ready when people (scientists and others?) need us
  - Intelligent vertical e-infrastructures (interactive end-points)
  - GPU-programming support
  - Cloud services
Activities: Cloud Computing (Parallab)

- Use cloud computing/virtual machines for scaling up scientific applications
- Access to on-demand computing ready to use. Highly configurable with respect to operation system, memory and processor(s)
- Pilot project: use Amazon Cloud to run R(r-project.org) based scripts for statistical analysis. Used e.g. in eSysBio
Activities: StoreBioinfo (CBU)

- StoreBioinfo has two aims
  - Together with NorStore develop data storage policies and govern a large block allocation of storage dedicated to Life Sciences
  - Establish e-services providing Life Science users integrated access to storage and computational resources from NorStore/Notur

- E-services used to integrate Genomic HyperBrowser for analysis on data stored in StoreBioinfo

- Portal in production (storebioinfo.norstore.no)
Activities: HPC-Europa2 (Parallab)

- Common requirements for many scientific applications: large datasets, many files, need for metadata, and post-processing

- Established distributed storage infrastructure (iRODS)

- Developed advanced clients featuring: data and metadata management, search, filtering, post-processing and visualization of the data
Activities: Protein Dynamics (CBU)

- **Aim:** Drug discovery
- **Simulations are based Newton's mechanics:** Atoms in force-fields, formally ODEs.
- **Simulations calculate the trajectories of all atoms. Typically**
  - 100k atoms
  - T=200ns (10 days · 500 cores)
  - Output: 5GB
- **From output calculate molecular properties**
Activities: Fruitful Collaborations (CBU)

- We highly appreciate fruitful collaborations like the one Július Parulek – Natalie Reuter:

  *Implicit Representation of Molecular Surfaces, Július Parulek et.al.*

- The tool enables much faster scientific progress on CBU's side

(ː BTW: Natalie says thanks ː-)
Activities: Time Scales (Sci Comp)

- Scientific Computing approach to quantify uncertainties in unstable dynamics
- Statistics over large ensembles of parallel computations
- Post-processing using a vertical-e-infrastructure
- The basis for homogenization?

![Fig. 4 Concentration of solute in the domain at three times. The first visible signs of instability appear around $t = 4.1 \cdot 10^3$. At the nonlinear onset time (center figure), 6 fingers are clearly visible. The wavelengths are therefore approximately $1000/6 = 170$.](image-url)
Visualization

What does mean visualization to us?

For us, non visualization-people, visualization

a) Is interesting science (which we don't do)
b) Provides SW that enables new developments in our field
c) Provides SW that helps us with presentations, easing our life

So essentially, when we use it, it is a SW

• Vertical, as it is specific
• Similar to MW and e-IS

Hence, when the SW becomes mature it is difficult to finance
Visualization (cont'd)

- To complete the picture, let's plot science vs. services (for Uni Computing)

<table>
<thead>
<tr>
<th></th>
<th>Bioinformatics</th>
<th>Env. Flow</th>
<th>Ecology</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT (hor)</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>SW (ver)</td>
<td>D,S</td>
<td>D,S</td>
<td>D,S</td>
<td>D,S</td>
</tr>
<tr>
<td>MW/e-IS (ver)</td>
<td>D,S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Viz (ver)</td>
<td>S</td>
<td>T</td>
<td>S</td>
<td>S</td>
</tr>
</tbody>
</table>

T – Tool
D – Development
S – Science-enabling

Green – ongoing
Red – potentially
Visualization (cont'd)

- Uni Computing has a big need for visualization
  - Bioinformatics: *Has shown already in (at least) two collaborations with Viz/II its potential*
  - Env. Flow: *Has a need for visualization tools (maybe not research for Viz/II here?). More the standard (CFD, …)*
  - Ecology: *Needs visualization, which likely will enable new ecology-type research*
  - Language: *Needs visualization, which likely will enable new research*

- Let's have a look to some details
Visualization: Bioinformatics

- Július Parulek/Viz has developed an *Implicit Representation of Molecular Surfaces* to visualize the surface of proteins.

- The method allows for fast rendering of surfaces and allows the researcher to identify important geometric properties.

- A production-type software-tool with these features would be of very high value to Bioinformatics/II and to the community.

- A linux port would be desirable.
Visualization: Bioinformatics II

- To help data interpretation

Heatmap – gene expression profiles (genes – rows; samples columns – clustered two-ways; red: over-expression; green: under-expression) 
Eisen et al 1998

Promoter-enhancer interactions – plot showing expression change (vertical) and binding of regulatory proteins (x-axis) – size of circle – amount of binding
Lenhard.
Visualization: Environmental Flow

- The Env Flow Group is working towards an integrated Wind-Water-Wave model to simulate dispersion, marine physics and ecology.
- Integrated visualization of wind, currents and waves together with the distribution of e.g. pollutants would be highly desirable.
- Desirable features:
  - Automatic image- and movie generation
  - Client-server implementation
  - Possible integration/visualization of external data
Visualization: Evolutionary Ecology

- **Individual-based evolutionary ecology** derives the dynamics of a population from the individual level.

- The population of e.g. fish is described by a dynamical system of 50-100k individuals described by a set of continuous (age, weight, location, ...) and discrete (gender, ...) parameters subject to a set of environmental conditions.

- To gain further insight into the dynamics, we need a data analysis tool able to visualize:
  - Env conditions (temp, salinity, flow, nutrition, ...)
  - Swarm properties (age, weight, ...)

simultaneously in space and time.

- To analyze efficiently an ensemble populations.
Visualization: Analyzing the Blogosphere

• Visualize the distribution, flow and development of knowledge and opinions across online social networks

• The example on the right shows the mockup of a SW displaying 56 blogs about Climate Change and their inter-relation (over time)

• A realistic corpus of blogs will have 1-10M blogs, to be analyzed

• We need a visualization able to handle and analyze large graphs interactively
Visualization: ANN Learning Process

- We need to understand Artificial Neural Network learning processes
  - Understand the dynamics of the learning process
  - Show under- and overfitting effects
  - Compare quality of different networks' architectures
  - Identify input space regions with potential problems
  - Visually compare various optimization procedures
  - Investigate stability of network classification
  - Estimate confidence in classification
Todo

- There is something to do
  
  *For the scientists on both sides: Identify areas of collaboration.*
  *Then to write research applications … That's the easy part ;-)*

But there are some problems

- Uni C needs probably research on the Viz-side of things. But sure, we need mature SW as well. May or may not be the interest of Viz.
- Uni C would probably be interested to develop SW … with some help.
  
  *We shall both think if we have common ground. And then, in case, think how to realize that.*
Thank you
for your attention