Multidisciplinary approach for computing in Emilia Romagna (Italy)

Daniele Cesini (INFN-CNAF)
On behalf of the DUCK collaboration

EGI UF 2011

Overview

- DUCK collaboration
- Comput-ER infrastructure and the Italian Grid Infrastructure
- Ported Applications
  - Bioinformatics
    - protein annotation
    - genome sequencing
  - Geophysics
    - Ocean Ridge simulation
    - Ocean forecast simulation
  - Chemistry
    - Molecular Dynamics
- Conclusion

DUCK

(Distributed Unified Computing for Knowledge)

- Collaboration between *multidisciplinary* Academic and Research Institutions located in the Emilia Romagna region aiming to:
  - exchange knowhow to solve scientific computing problems
  - contribute to the development of software services and of distributed computing/data infrastructure in collaboration with the Italian Grid Infrastructure (IGI)
  - extend the services and the infrastructure to other communities in the region like medicine, public administrations and private organizations.

DUCK Partners

1. National Institute of Nuclear Physics, units INFN-BO and INFN-CNAF
2. National Institute of Geophysics and Volcanology (INGV), unit in Bologna
3. The Euro-Med Center for Climate Change (CMCC), unit in Bologna
4. Interdepartmental Center for Research on Cancer "GIORGIO PRODI" – University of Bologna
5. Bologna Computational Biology Network / Department of Evolutionary Experimental Biology, University of Bologna
6. CeIRG, Interdepartmental Center for Genomics Research, University of Modena and Reggio Emilia,
7. National Research Council,CNR, and National Institute of Astrophysics (INAF) in Bologna Area
8. Physics Department, University of Ferrara
9. Physics Department of Alma Mater Studiorum-University of Bologna

Comput-ER Sites and Resources
(Computing in Emilia Romagna)

Total Available Resources:

~1300 cores

~150 TB of storage space

7 SITES provided by DUCK partners

- **INFN-CNAF**
- **INFN-BOLOGNA**
  - INFN
  - Dip. Fisica
- **INFN-FERRARA**
  - INFN
  - Dip. Fisica
- **AREA-BO**
  - *IRA/INAF-BO*
  - *ISMAR/CNR-BO*
  - *ISOF/CNR-BO*
  - *OATS/INAF-BO*
- **INFN-PARMA**
- **INGV**
- **CMCC**

Comput-ER infrastructure

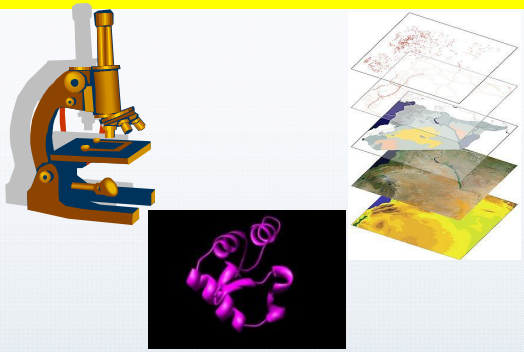
- Comput-ER is based mainly on commodity farms but new GPU-based servers for parallel applications are under testing.
- Comput-ER uses the gLite middleware.
- These resources are either based on real hardware systems, or are dynamically provisioned via virtual machines using the Worker Nodes on Demand Service (WNoDeS) grid/cloud virtualization system.
- A **Virtual Organizations** (comput-er.it) has been setup and deployed on the resource centers provided by the DUCK partners
- Comput-ER is foreseen to be used also by private organizations

Comput-ER and IGI

- **IGI** is the **Italian Grid infrastructure,** run for the benefit of the research and education communities in Italy and worldwide.
- IGI is the Italian NGI in EGI
- IGI offers more than 22000 computing cores and 30 PB of storage resources distributed across about 60 computing centres
- Comput-ER is embedded in IGI
  - sites are in the production BDII
  - central services are provided by IGI
  - IGI provided support and knowhow in application porting

A Consolidated Work-plan for application porting

Seek, identify, and closely support new communities to deploy their applications and create the computing model



Set-up and manage a Production Grid Infrastructure

Disseminate the work done to find new applications for long term sustainability

Train users and site administrators to use and operate the Infrastructure based on the EGI/EMI middleware

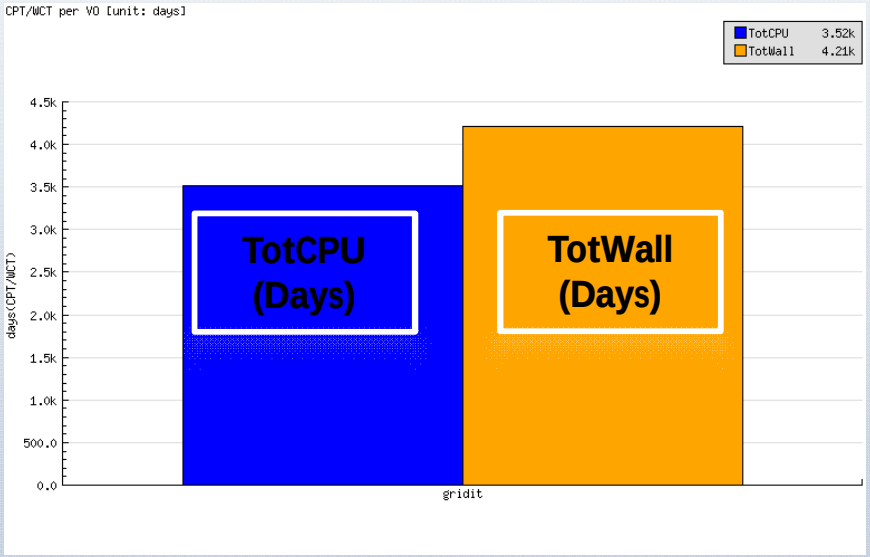
Applications on Comput-ER

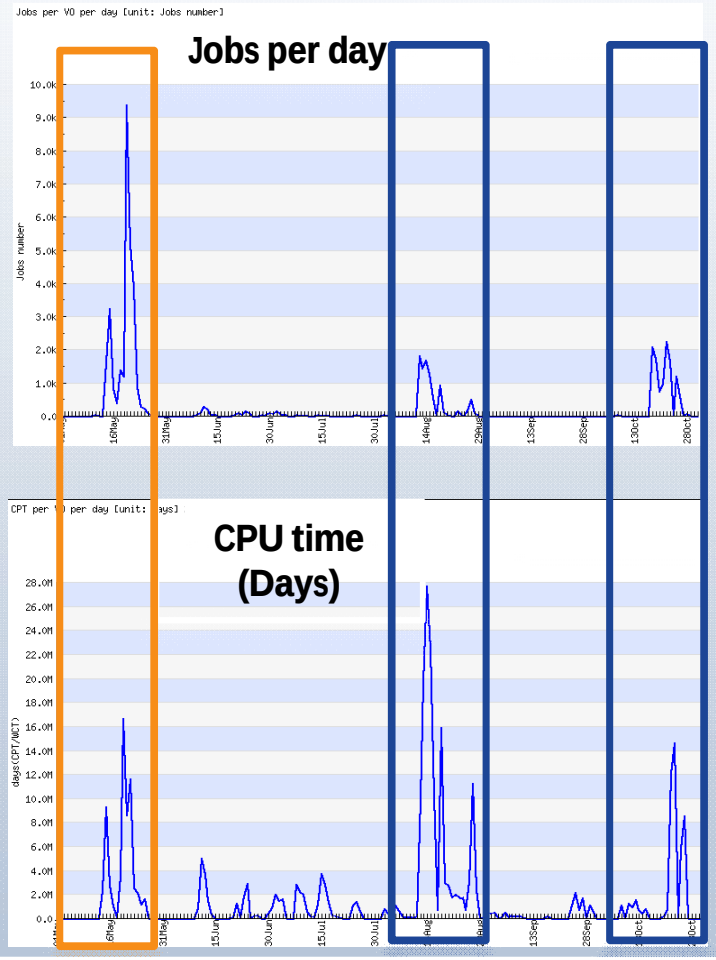
- Common characteristics :
  1) Small communities (even if belonging to national institution) owing limited hw resources compared to their computational needs
  2) Localized in Emilia Romagna
  3) No or very limited previous Grid experience
- Research fields:
  - Biochemistry: protein annotation, genome sequencing
  - Geophysics: mid ocean ridge processes, ensemble methods for ocean forecasting
  - Computational Chemistry

Accounting data

Accounting data about test and application porting (high job number rate, low cpu time consumed) and first productions (higher cpu utilization)

First 6 months of DUCK
May-October 2010





~47000 jobs
~3500 days of CPU time

Application Ported on Comput-ER

Protein Annotation

(Biocomputing Group, University of Bologna)

Protein annotation is essential for understanding diseases at the molecular level

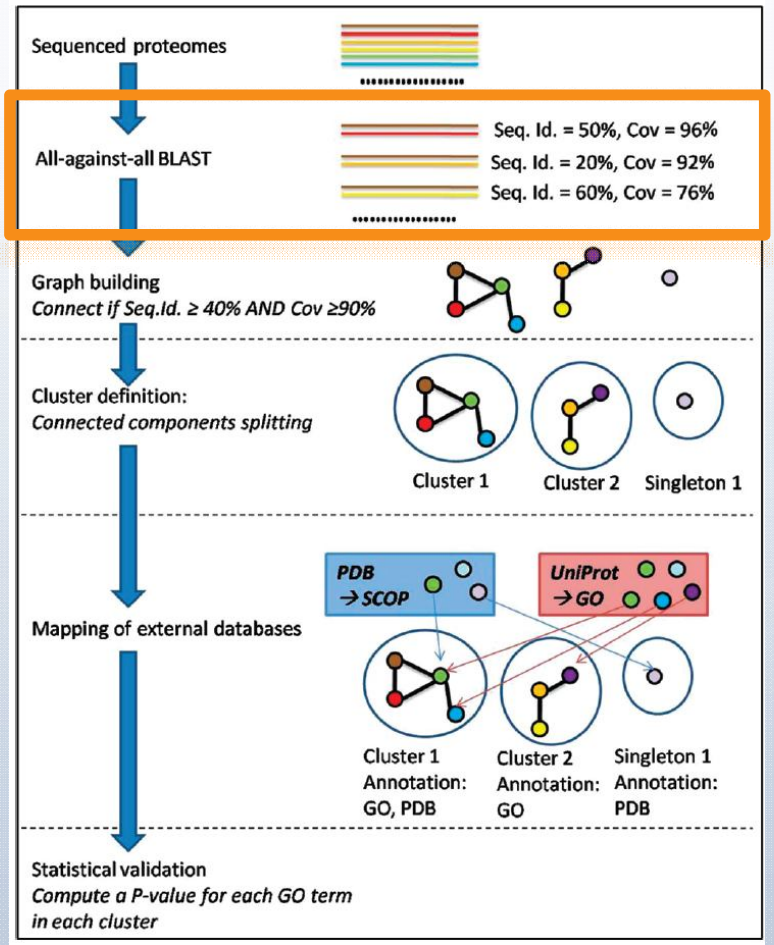
Bologna Annotation Resource (BAR) method was used

BAR uses the BLAST software for aligning the sequence: the number of computations is the cartesian product of the number of the input sequences.

~ 17 X 10^{13}

Problem: the amount of time to do these alignments is estimated in seven years on the local resources of the group

BAR Method

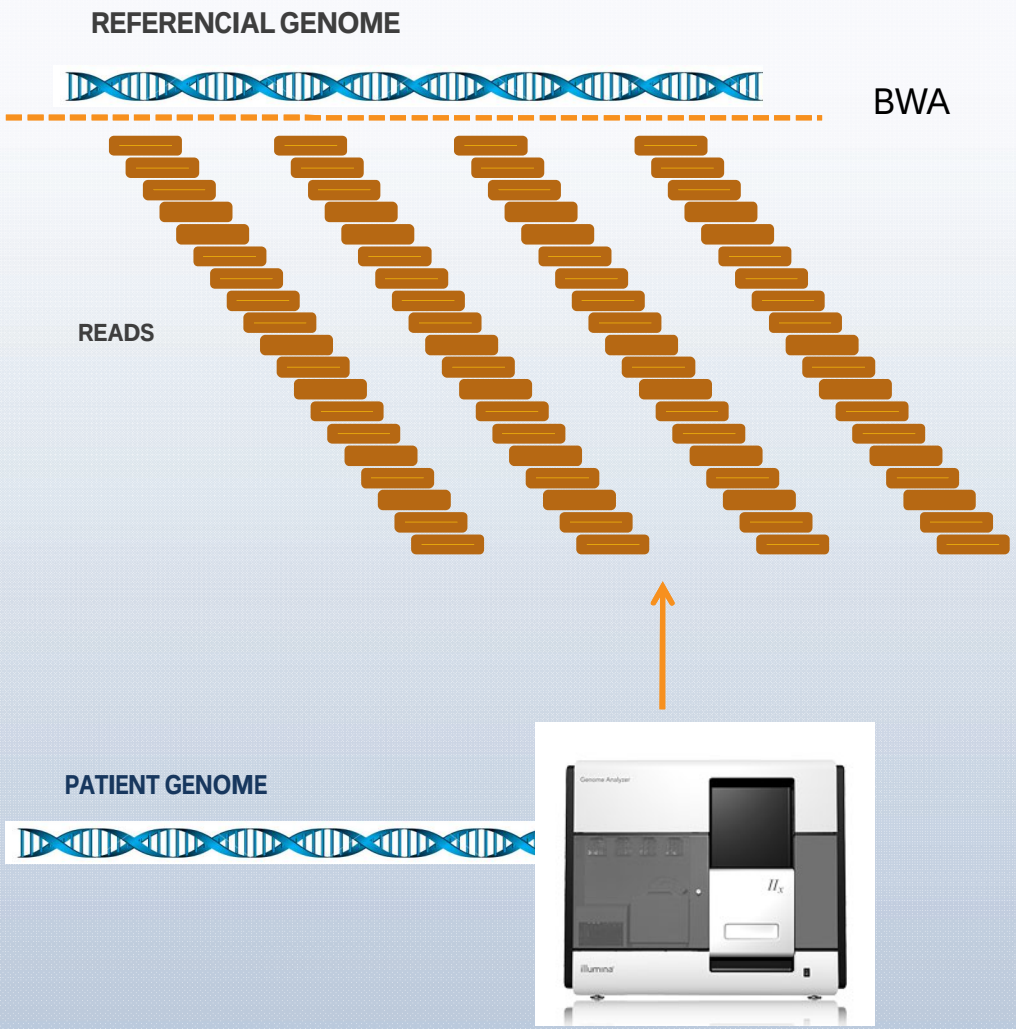


The alignment step was the griddified one

Massively Parallel Genome Sequencing

(Biocomputing Group, University of Bologna)

- Used in the study of cancer diseases
- Next generation sequencing technology allows a massive amount of DNA or RNA fragments to be sequenced in a single experiment.
- For the *massively parallel sequencing it is used* **BWA** tool (Burrows-Wheeler Aligner) for indexing and alignment
  - Memory request~ 3,5 GB
  - Total time ~ 50h using the group local resources
  - Problem: it is more the time to elaborate the data than the time to produce them. With the consequence to accomulate data in storage element.



Results for BLAST and BWA

Application	Local	DUCK Grid	GPU preliminary
BLAST	**7 yrs**	**1/120 (3 weeks)**	½ CPU result (1 cpu core vs 1cpu core + 1 TESLA GPU) (high memory demanding)
BWA	**50 h**	**1/4**	

Same solution parametric jobs were used to parallelize the computation. In this way is the WMS to do the most of the work.

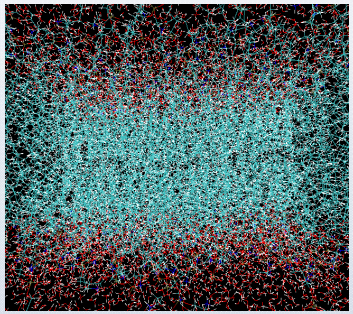
Authors: Biocomputing Group, University of Bologna

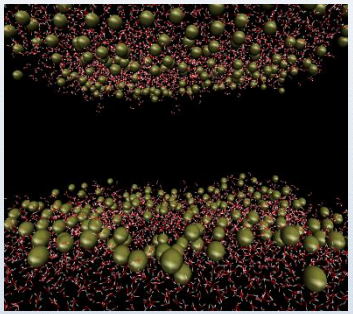
CASADIO Rita , INDIO Valentina, FARISELLI Piero, GUERZONI Daniele, MARTELLI Pier Luigi, PIOVESAN Damiano, ROSSI Ivan

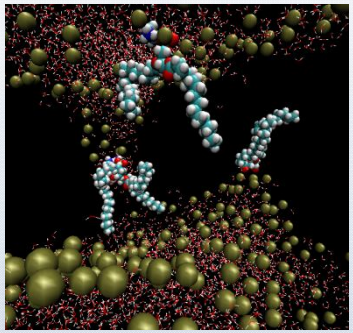
Molecular Dynamics

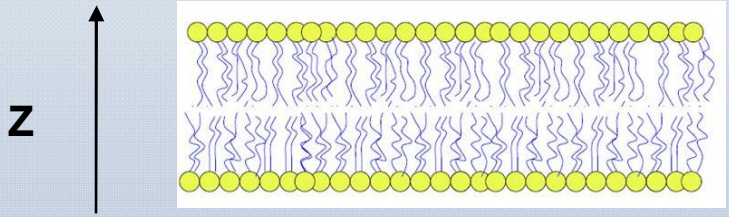
(ISOF-CNR - National Research Council-Institute for the OrganicSynthesis and Photoreactivity, Bologna Unit)

Dynamics of a lipid bilayer induced by electric fields: These processes are the basis of important properties and functions of cell membranes





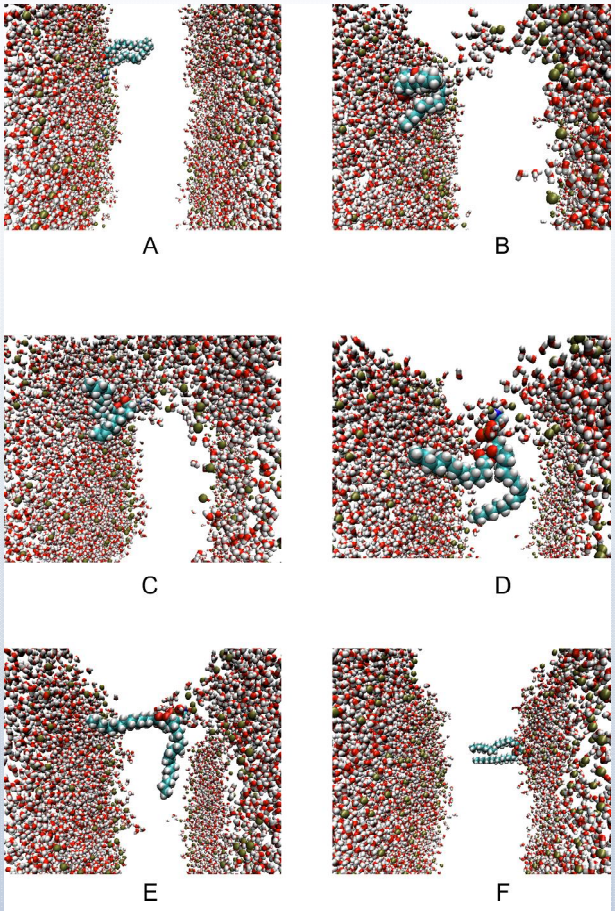




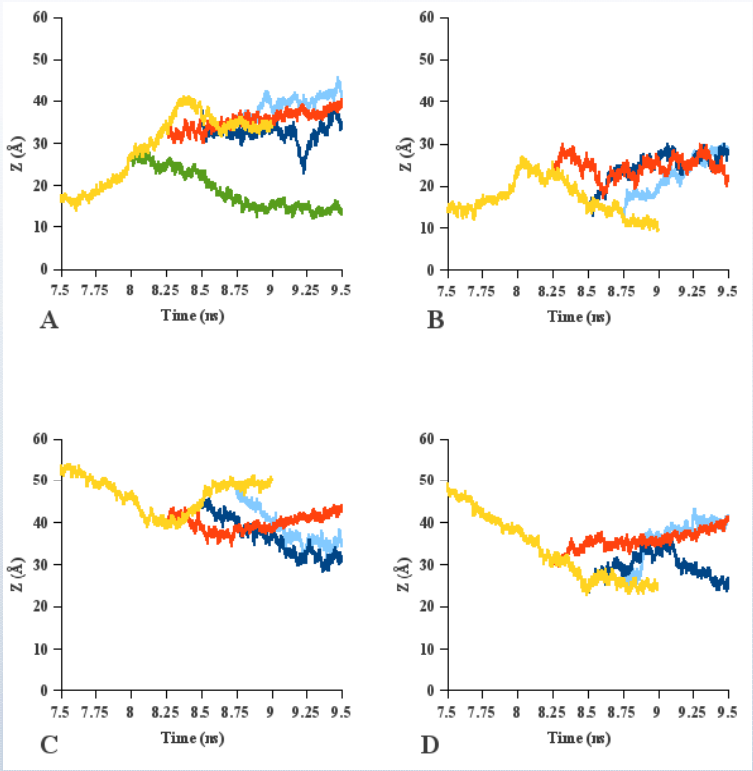
- External field of **0.35, 0.5 and 1.0 V/nm** was applied parallel to the membrane normal z

Molecular Dynamics Results

Lipid translocation in a bilayer can be driven by switching on and off an electric field



A lipid translocation by switching on and than off the electric field.



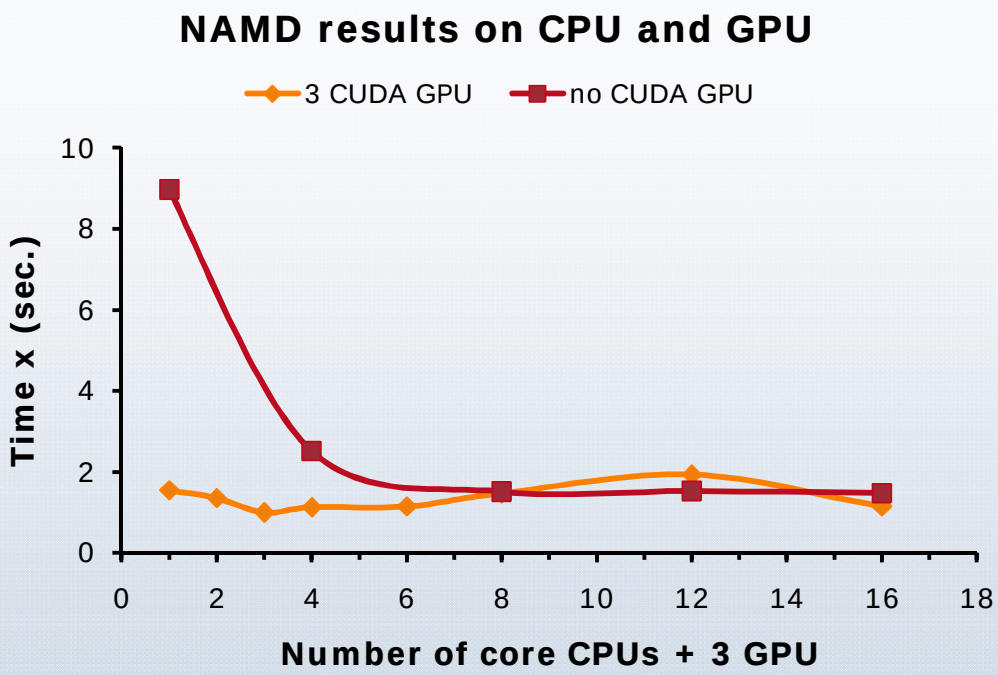
Time evolution of the position of the centre of mass of four lipids.

Authors and ref.: A. Venturini and F. Zerbetto, Phys. Chem. Chem. Phys. 2011, DOI: 10.1039/c1cp20322c

Molecular Dynamics Computational Problem

- Huge molecular systems to be simulated: from 35 to 65 katoms
- NAMD package used (http://www.ks.uiuc.edu/Research/namd/)
- OPENMPI long parallel runs needed
  - Single run too long for a Grid job
    - 41 days (35katoms – 25 ns sym_time) on 8 cores
    - 18 days (35katoms – 25 ns sym_time) on 32cores
  - Computation split into N sequential MPI jobs
  - Two parameters influence the computational time:
    - N: number of jobs – set to 50
    - CPU_NUM: number of required cores – set to 32
  - 1 day lost in Grid with respect to local run
    - Depends on the Grid load
  - Need bigger MPI enabled site

Molecular Dynamics GPU tests



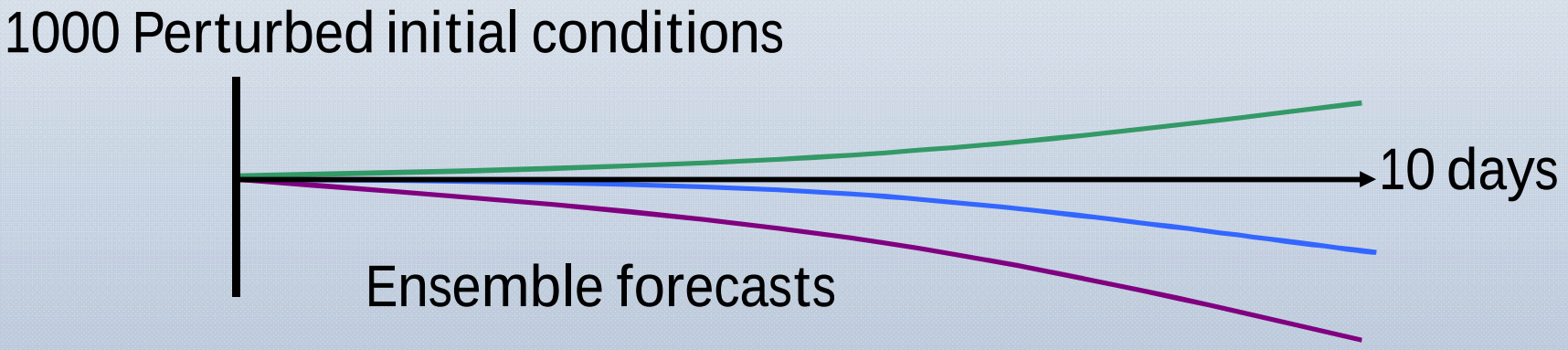
Preliminary results using a 16 cores + 3 TESLA GPU machine

Need further investigatigation

Ocean Ensemble Forecasting

(INGV-National Institute of Geophysics and Volcanology – Bologna Unit)

- Ensemble method techniques has been used to quantify the forecast uncertainty in short-term ocean forecasting systems
- Grid allowed us to perform several ensemble forecast experiments with 1,000 members: they are completed within 5 hours of wallclock time after their submission



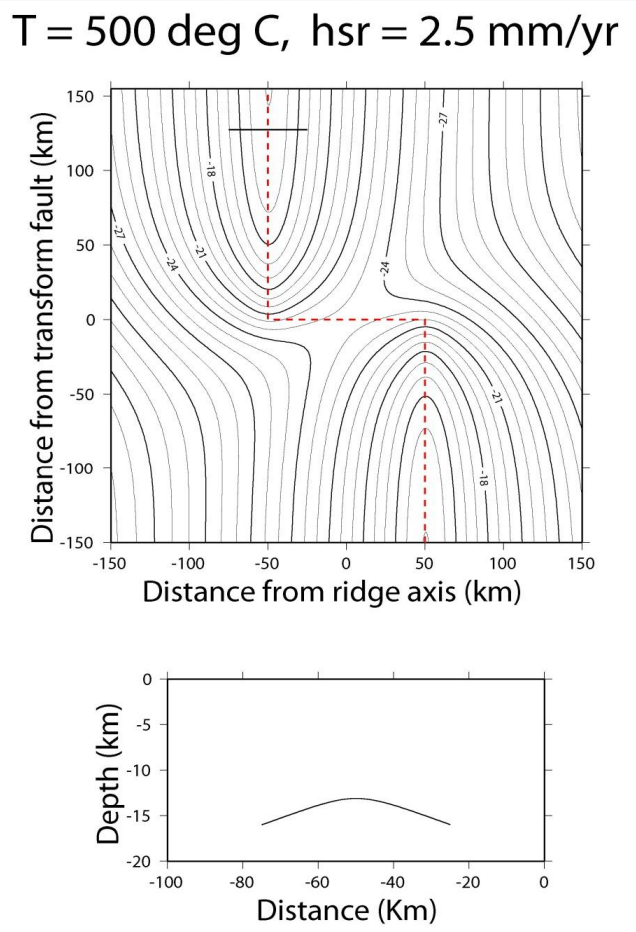
AUTHORS: INGV-Bologna Unit: PINARDI Nadia, SCOCIMARRO Enrico

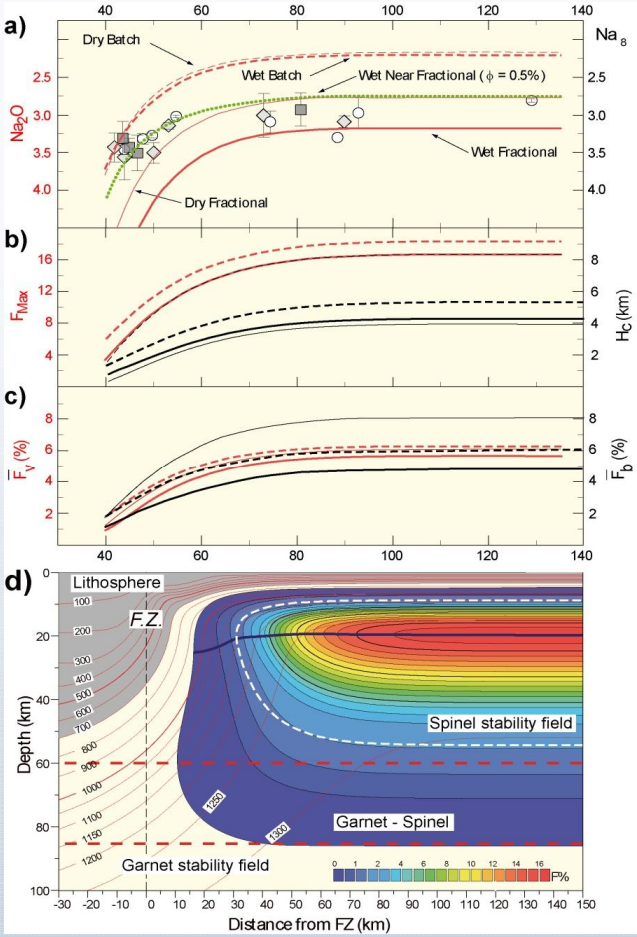
Mantle rock serpentinization

(ISMAR- Italian Institute of Marine Science)

- Simulation of mantle flow dynamics to study mantle rock serpentinization, a process central to a variety of chemical exchanges between solid earth, hydrosphere and biosphere.
- Hydration/ serpentinization of mantle rocks influences global cycles of H_2O, Carbon and Mg, affecting the composition of sea water and Earth's climate
- Sub-lithospheric mantle flow and mantle thermal structure are obtained from the solution of Stokes and heat equations by semi-analytical-pseudo-spectral and finite difference techniques, respectively.

Serpentinization Model Results





Model results: Depth of 500 °C isotherm and along axis crustal production

Parametric jobs exploited

Authors: CNR-ISMAR - Bologna Unit: BORTOLUZZI Giovanni, LIGI Marco

Conclusion and future works

- DUCK and comput-ER allowed to share computing resources between small, local communities
  - without a regional project probably would not have joined the national Grid
- The comput-ER infrastructure together with the support of the NGI staff demonstrated to be a powerful and flexible tool for their research
  - Applications got benefits not only in sharing resources but also in knowledge-sharing from a multidisciplinary scientific computing environment
- For most of the applications the key solution is based on their execution as loose parallel applications, i.e. parallel and DAG jobs through the gLite WMS
- For some of them, MPI and GPU has been investigated, good results have been obtained but
  - more testing is required with the GPU
  - bigger MPI enabled sites are needed
- DUCK will try to open its resources to research groups from the industry
- DUCK gives to the possibility to ask for regional public funding

Authors

- **Biocomputing Group, University of Bologna**
  - CASADIO Rita , INDIO Valentina, FARISELLI Piero, GUERZONI Daniele, MARTELLI Pier Luigi, PIOVESAN Damiano, ROSSI Ivan
- **CMCC(Euro-Mediterranean Centre for Climate Change), Bologna Italy**
  - DOBRICIC Srdjan, NAVARRA Antonio
- **CNR/ISMAR(National Research Council-Institute of Marine Sciences), Bologna Italy**
  - BORTOLUZZI Giovanni, LIGI Marco
- **CNR/ISOF (National Research Council-Institute for the OrganicSynthesis and Photoreactivity), Bologna Italy**
  - OTTANI Stefano, VENTURINI Alessandro
- **INFN-CNAF (National Institute of Nuclear Physics – National Computer Science Center), Bologna Italy**
  - BENCIVENNI Marco, CESINI Daniele, GHISELLI Antonia, SALOMONI Davide, VERONESI Paolo
- **INGV(National Institute Geophysics and Volcanology), Bologna Italy**
  - PINARDI Nadia, SCOCIMARRO Enrico