

# Software Stack on FERMI

[e.rossi@cineca.it](mailto:e.rossi@cineca.it)



[www.cineca.it](http://www.cineca.it)



# Agenda

- Mathematical Libraries
  - Sequential
  - Parallel
  - ESSL, Lapack, ScaLapack
- Scientific Applications
  - NAMD
  - LAMMPS
  - QuantumEspresso
  - VASP
  - PLUTO

- For programmers
  - Mathematical libraries
  - I/O libraries
  - Tools for profiling and debugging
- For production
  - Scientific Applications



- **Mathematical libraries** • Scientific Applications
  - `essl/5.1`
  - `fftw/2.1.5, 3.3.2`
  - `gsl/1.15`
  - `lapack/3.4.1`
  - `petsc/3.3-p2`
  - `scalapack/2.0.2`
  - `blas/2007`
  - `mass/7.3`
  - `papi/4.4.0`
- I/O libraries
  - `netcdf/4.1.3`
  - `hdf5/1.8.9`
    - ✓ `gzip/2.1`
    - ✓ `zlib/1.2.7`
- Scientific Applications
  - `abinit/6.12.3`
  - `cpmd/v3.15.3`
  - `namd/2.9`
  - `amber/12`
  - `crystal09/1.01`
  - `QuantumEspresso/5.0bgq`
  - `bigdft/1.6.0`
  - `dl_poly/4.03`
  - `siesta/3.1`
  - `cp2k/2.3`
  - `gromacs/4.5.5`
  - `vasp/5.2.12, 5.3.2`
  - `cpmd/3.15.3_hfx`
  - `lammps/20120816`
  - `OpenFoam` (to be done)

- All libraries as available as module  
module avail
- Help on how to use library in the help of the module  
module help <name>
- Loading the module sets environment variables you have to use in the linking procedure  
-L\$(\*\_LIB) and -I\$(\*\_INCLUDE)
- The show command lists all defined variable  
module show <name>
- Link sequence important, <file>.o always before the libraries, sometimes double linking necessary



# General Info

```
[erossi00@fen07 ~]$ module avail
----- /cineca/prod/modulefiles/profiles -----
profile/advanced      profile/base(default) profile/front-end
----- /cineca/prod/modulefiles/base/environment -----
autoload/0.1
----- /cineca/prod/modulefiles/base/libraries -----
blas/2007--bgq-xl--1.0(default)      mass/7.3--bgq-xl--1.0
essl/5.1                             mpi4py/1.3--bgq-gnu--4.4.6
fftw/2.1.5--bgq-xl--1.0              netcdf/4.1.3--bgq-xl--1.0
fftw/3.3.2--bgq-xl--1.0              numpy/1.6.2--bgq-gnu--4.4.6
gs1/1.15--bgq-xl--1.0                papi/4.4.0--bgq-gnu--4.4.6
hdf5/1.8.9_par--bgq-xl--1.0          petsc/3.3-p2--bgq-xl--1.0
hdf5/1.8.9_ser--bgq-xl--1.0          scalapack/2.0.2--bgq-xl--1.0(default)
lapack/3.4.1--bgq-xl--1.0(default)    szip/2.1--bgq-xl--1.0
libint/2.0--bgq-xl--1.0(default)      zlib/1.2.7--bgq-gnu--4.4.6
libjpeg/8d--bgq-gnu--4.4.6
----- /cineca/prod/modulefiles/base/compilers -----
bgq-gnu/4.4.6      bgq-xl/1.0      front-end-gnu/4.6.3 front-end-xl/1.0
----- /cineca/prod/modulefiles/base/tools -----
cmake/2.8.8        irods/3.2      tau/2.21.4
globus/5.0.5       scalasca/1.4.2  tgftp/0.5.0
gtransfer/0.0.10a superc/1.0      uberftp/2.6
----- /cineca/prod/modulefiles/base/applications -----
abinit/6.12.3      cpmd/v3.15.3      namd/2.9
amber/12(default)  crystal09/1.01    qe/5.0bgq
bigdft/1.6.0       dl_poly/4.03(default) siesta/3.1
cp2k/2.3(default)  gromacs/4.5.5(default) vasp/5.2.12
cpmd/3.15.3_hfx(default) lammps/20120816    vasp/5.3.2
```



# General Info

```
[erossi00@fen07 ~]$ module help essl
```

```
-----  
Module Specific Help for /cineca/prod/modulefiles/base/libraries/essl/5.1:  
modulefile "essl/5.1"
```

```
essl-5.1  
Collections of state-of-the-art mathematical subroutines
```

```
-----  
License type: commercial  
Web site:    www.ibm.com  
Download url: www.ibm.com  
-----
```

license

```
-----  
ESSL and Parallel ESSL are collections of state-of-the-art mathematical subroutines  
specifically designed to improve the performance of engineering and scientific  
applications on the IBM POWER\u0022 processor-based servers and blades.  
-----
```

descr

```
*****  
PLEASE NOTE  
- This library is meant to be linked to a program that will be ran on the compute nodes.  
*****
```

```
libesslbg.a (MPI)  
libesslsmpbg.a (MPI & openmp)
```

example

```
Example of usage:
```

```
$ module load essl  
$ module load bgq-xl  
$ mpicc -o a.out foo.c -L$ESSL_LIB -lesslbg -lesslsmpbg  
$ mpiexec -o a.out foo.f90 -L$ESSL_LIB -lesslbg -lesslsmpbg  
-----
```

```
[erossi00@fen07 ~]$ module show essl
```

```
-----  
/cineca/prod/modulefiles/base/libraries/essl/5.1:
```

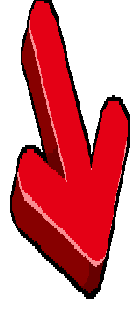
```
module-whatis    Collections of state-of-the-art mathematical subroutines  
conflict         essl  
setenv  ESSL_HOME      /opt/ibmmath/essl/5.1/  
setenv  ESSL_LIB        /opt/ibmmath/essl/5.1/lib64  
setenv  ESSL_INC        /opt/ibmmath/essl/5.1/include  
setenv  ESSL_INCLUDE    /opt/ibmmath/essl/5.1/include  
prepend-path    LIBPATH /opt/ibmmath/essl/5.1/lib64  
prepend-path    LD_LIBRARY_PATH /opt/ibmmath/essl/5.1/lib64 :
```

```
-----
```



- All libraries was compiled with XL compiler for the compute nodes  
`-O3 -g -qsimd=auto`
- Additional version compiled with the gnu compilers is sometimes available in the “advanced” profile
- See “module avail” for available versions
- Usually, only the most recent versions are available in the “base” profile

- **Vendor specific libraries**



- ESSL  
(Engineering and Scientific Subroutine Library), version 5.1
- MASS (IBM Mathematical Acceleration Subsystem)

- **Public domain Software**



- LAPACK (Linear Algebra PACKage)
- GSL (Gnu Scientific Library)
- BLAS (Basic Linear Algebra Subprograms)
- FFTW (Fast Fourier Transform)

- **Threaded Parallelism**



- ESSLsmp 5.1

- **MPI Parallelism**



- ScaLAPACK (Scalable Linear Algebra PACKage)
- FFTW (Fastest Fourier Transform of the West)
- PETSc (scalable solution by partial differential equations)



# Math lib: ESSL

**ESSL** - Engineering and Scientific Subroutine Library (IBM)  
(collections of state-of-the-art mathematical subroutines specifically designed to improve the performance of engineering and scientific applications on IBM platforms)

- High-performance math libraries
- Tuned for BlueGene/Q
- Can be called from C, C++ and Fortran
- Multi-threaded version (SMP) available
- Can be used together with BLAS or LAPACK libraries

**libesslbg.a** (serial)  
**libesslsmpbg.a** (OpenMP)

```
$ module load bgq-xl  
$ module load essl  
$ mpicc_r -o a.out foo.c -L$ESSL_LIB -lesslsmpbg  
$ mpixlf90_r -o a.out foo.f90 -L$ESSL_LIB -lesslbg
```

# Contents of ESSL Version 5.1

- **BLAS** level 1-3 and additional vector, matrix-vector, and matrix-matrix operations
- Sparse vector and matrix operations
- **LAPACK** computational routines for linear equation systems and eigensystems
- Banded linear system solvers, Linear Least Squares, Fast Fourier Transforms, Numerical Quadrature, Random Number Generation, Interpolation
- All routines are thread-safe, i.e. can be used within OpenMP threads

For further information see:

*IBM Engineering and Scientific Subroutine Library for Linux on POWER V5.1: **Guide and Reference***



# LAPACK

- **Solving dense linear algebra problems efficiently on HPC computers.**
- Developed by Argonne National Laboratory, supported by the National Science Foundation (NSF) and the United States Department of Energy (DOE)
- Performance issues are addressed by incorporating recent algorithmic improvements for linear algebra computation. Algorithms give nearly optimal performance.
- Public domain version 3.4.1
- Must be used together with ESSL (or ESSLsmp)
- Some routines already in ESSL
- Attention, some calling sequences are different!

```
module load essl
module load lapack
mpixlf_r name.f
-L$ESSL_LIB -lesslbfg
-L$LAPACK_LIB -llapack
-lesslbfg
```

ESSL must be linked also after LAPACK to resolve references



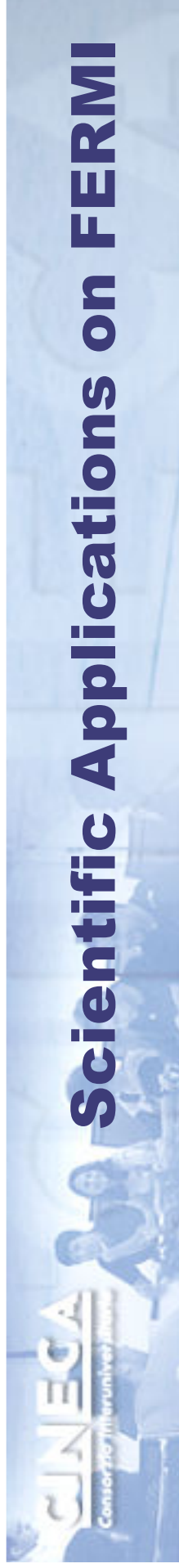
# ScaLAPACK

- Scalable Linear Algebra PACKage, or Scalable LAPACK
- Includes a subset of LAPACK routines redesigned for distributed memory MIMD parallel computers.
- Matrices are laid out in a two-dimensional block cyclic decomposition.
- Univ. of Knoxville, Oak Ridge National Lab, Univ. of Berkeley
- The BLACS is now part of ScaLAPACK, and is compiled into the ScaLAPACK library. It is no longer necessary to link against BLACS libraries.
- The current release 2.0.2 does no longer include sequential LAPACK routines needed, thus an LAPACK library has to be linked together with ScaLAPACK

```
$ module load scalapack
$ module load essl
$ module load lapack
$ module load bgq-xl

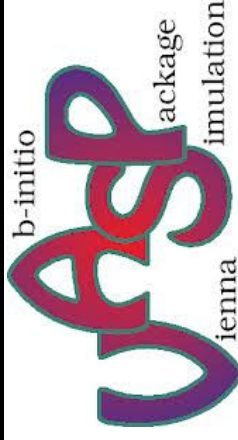
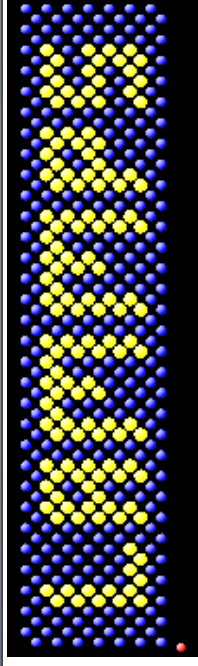
$ mpixlf90_r -o a.out foo.f90
-L$SCALAPACK_LIB -lscalapack -L$ESSL_LIB -lesslbg
-L$LAPACK_LIB -llapack -lesslbg
```





# Scientific Applications on FERMI

**NAMD**  
Scalable Molecular Dynamics



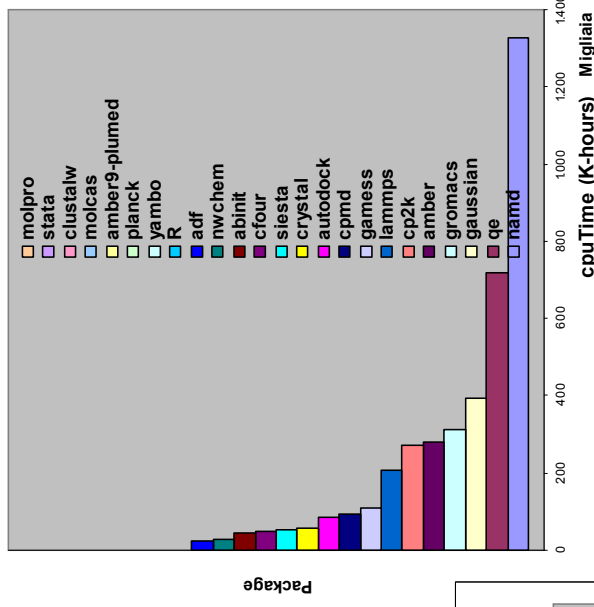
- How we decided about the applications to be install on FERMI (in the FERMI installation plan):
  - Select applications most used on the old system (SP6)
  - Verify if suitable for BGQ architecture
  - Add novel applications if known their suitability for BGQ
- Applications considered of general interest and well suited for FERMI are installed by CINECA through the “module” environment
- Both licensed and open-source applications. Even if licensed, normally users are allowed to use them freely (site licenses)
- We normally maintain the applications installing new releases as available and stay in touch with providers



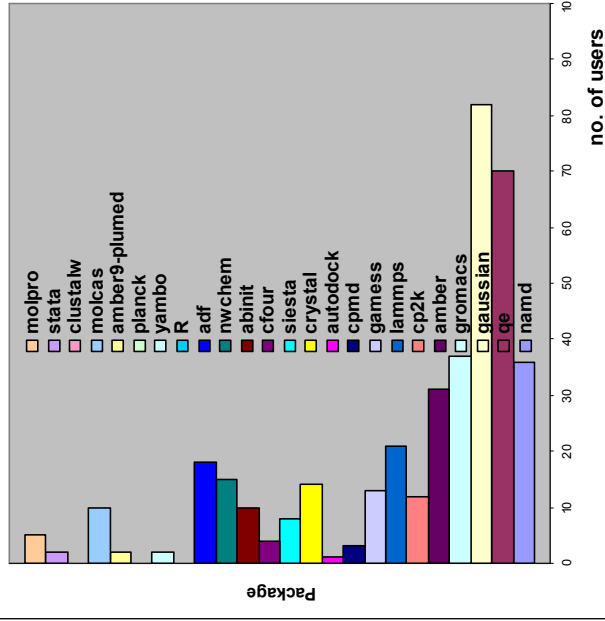
# Packages most used in 2011 (on SP6)

Among the most used packages, several are of interest for Chemistry, Material Science, ...

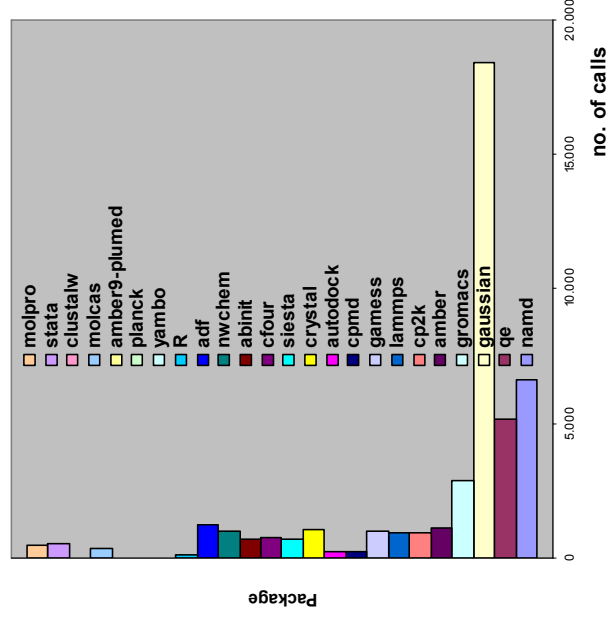
cpuTime per Package



distinct users (index) per Package



no. of calls per Package



- Quantum Chemistry
  - Gaussian
  - Gamess
  - Crystal
  - CFour
  - ...
- Molecular Dynamics
  - NAMD
  - Amber
  - Gromacs
  - LAMMPS
  - Ab-initio Dynamics
    - QuantumEspresso
    - CP2K
    - CPMD



# The Migration plan

Package	%core/hours	Num.calls%	Num. users	Max_PE	Science domain	where
NAMD	10.1	19.3	<b>M</b>	1156	CompBio	BGQ
QE	4,1	11.1	<b>L</b>	1024	CondMatter	BGQ
CP2K	3.7	1.2	<b>S</b>	256	CondMatter	?
Gaussian	2.5	25.8	<b>L</b>	64	CompChem	PLX (openMP)
Gromacs	2.5	10.4	<b>M</b>	512	CompBio	?
Amber	0.9	2.0	<b>M</b>	256	CompBio	PLX (gpu)
CPMD	0.9	1.0	<b>S</b>	256	CondMatter	?
LAMMPS	0.8	2.1	<b>M</b>	128	CondMatter/ CompBio	BGQ
Gadget	-	-			astro	?
Pluto/ Enzo	-	-	<b>M</b>		astro	?
OpenFoam	-	-	<b>S</b>		CompFluidDyn	PLX/BGQ?

- Scientific Applications available on FERMI as Nov2012

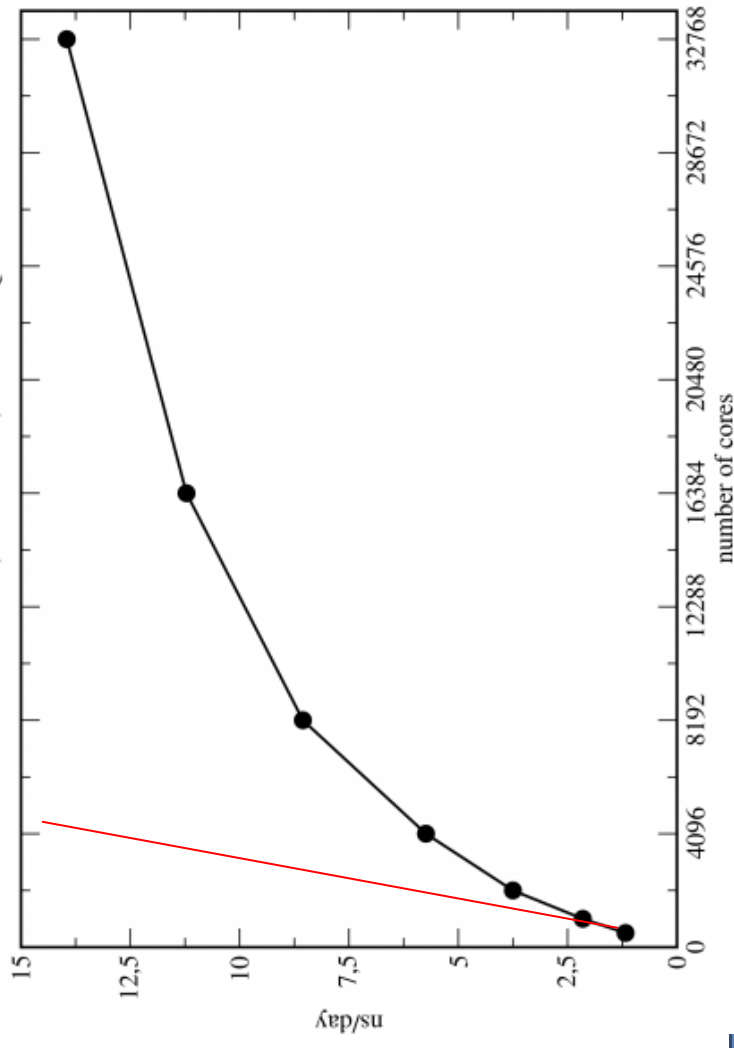
- abinit/6.12.3
- CPMD/3.15.3
- **NAMD/2.9**
- Amber/12
- Crystal09/1.01
- **QuantumEspresso/5.0bqq**
- bigDFT/1.6.0
- DL\_Poly/4.03
- Siesta/3.1
- CP2K/2.3
- Gromacs/4.5.5
- **Vasp/5.2.12, 5.3.2**
- **Lammps/20120816**
- **PLUTO 4.0**
- OpenFoam (to be done)

Statistics not yet available

## NAMD Scalable molecular dynamics

- parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems.
- scales to hundreds of processors on high-end parallel platforms
- Compiled with XL compilers, mixed parallelism MPI + multi-threads, uses FFTW2 (single-precision), BG/Q optimised by developers
- Benchmarks use STMV (virus) (1,066,628 atoms, periodic, PME), as defined on the NAMD web-site

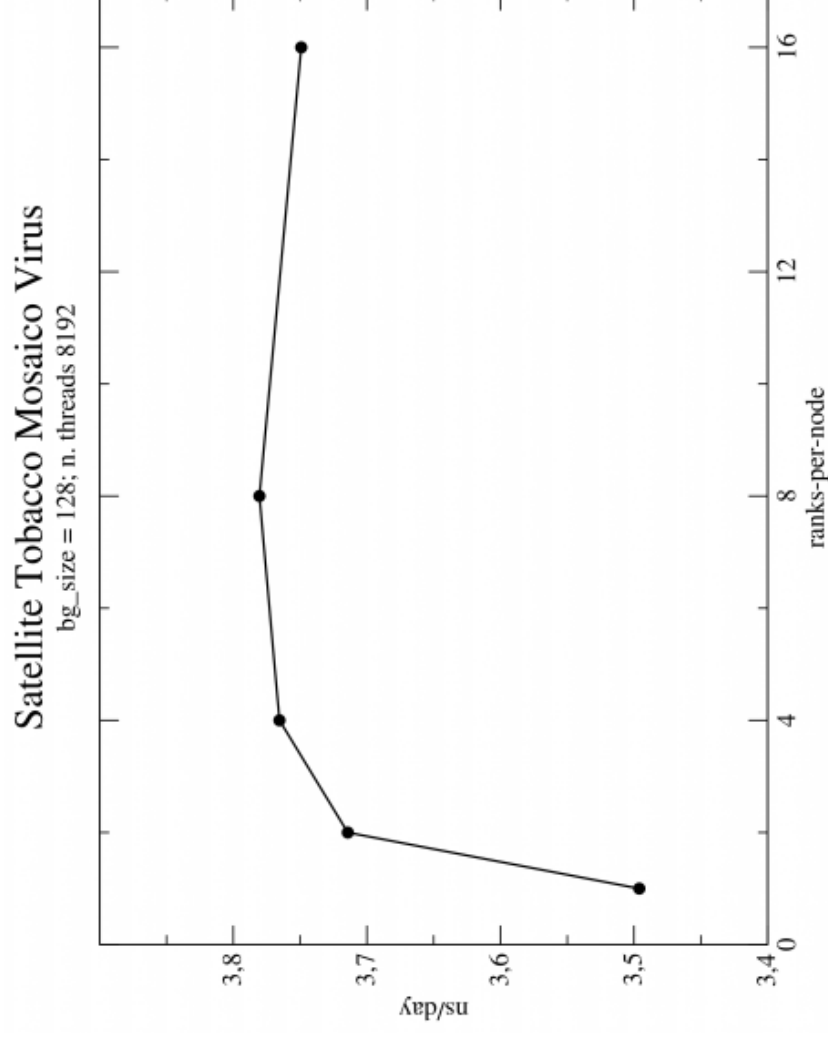
Satellite Tobacco Mosaic Virus  
NAMD 2.9, 1.2 MAtoms, BlueGene/Q



- **Graphic 1:** the NAMD performance (simulation time in ns/day) is reported vs. the increasing number of nodes (fixed ranks-per-node=4 and ppn=4)
- bg\_size = from 64 to 2048
- ranks-per-node = 4
- ppn =4
- num. of core =  
from  $64 \times 4 \times 4$  to  $2048 \times 4 \times 4$   
→ from 1.024 to 32.768

- **Graphic 2:** the NAMD performance (simulation time in ns/day) is reported vs. the increasing no. of ranks-per-node and decreasing ppn (fixed bg\_size=128 and 64 threads/node)

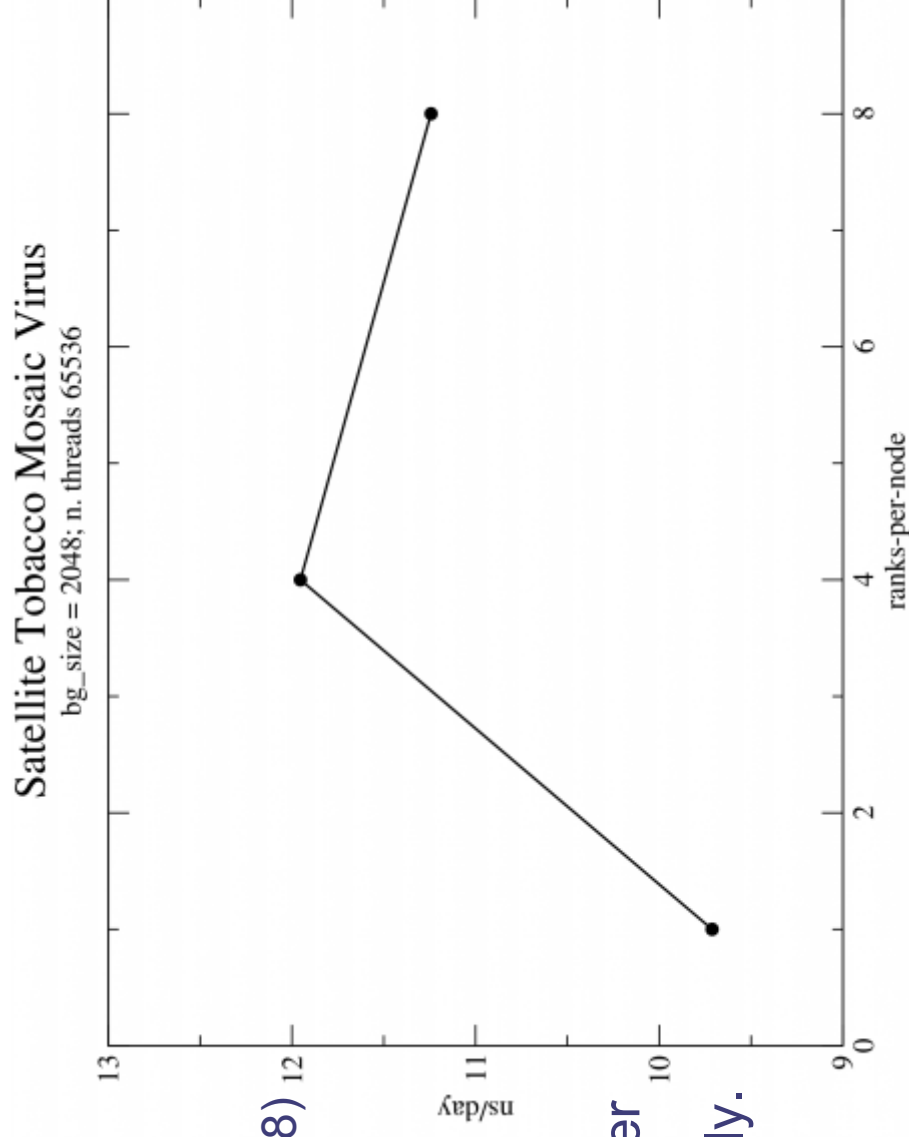
- bg\_size = 128
- ranks-per-node = from 1 to 16
- ppn = from 64 to 4
- Ranks-per-node \* ppn = 64
- Total MPI pes = 128\* (from 1 to 16)  
--> from 128 to 32
- Total threads = 8.192



- The performance reaches a maximum value for 4-8 ranks-per node.

- **Graphic 3:** the NAMD performance (simulation time in ns/day) is reported vs. the increasing no. of ranks-per-node and decreasing ppn (fixed bg\_size=2048 and 32 threads/node)

- bg\_size = 2048
- ranks-per-node = from 1 to 8
- ppn = from 32 to 4
- Ranks-per-node \* ppn = 32
- Total MPI pes = 2048\* (from 1 to 8)  
--> from 2048 to 16.384
- Total threads = 65.536



The performance reaches a maximum value for 4 ranks-per node; for higher values, the performance decreases rapidly.

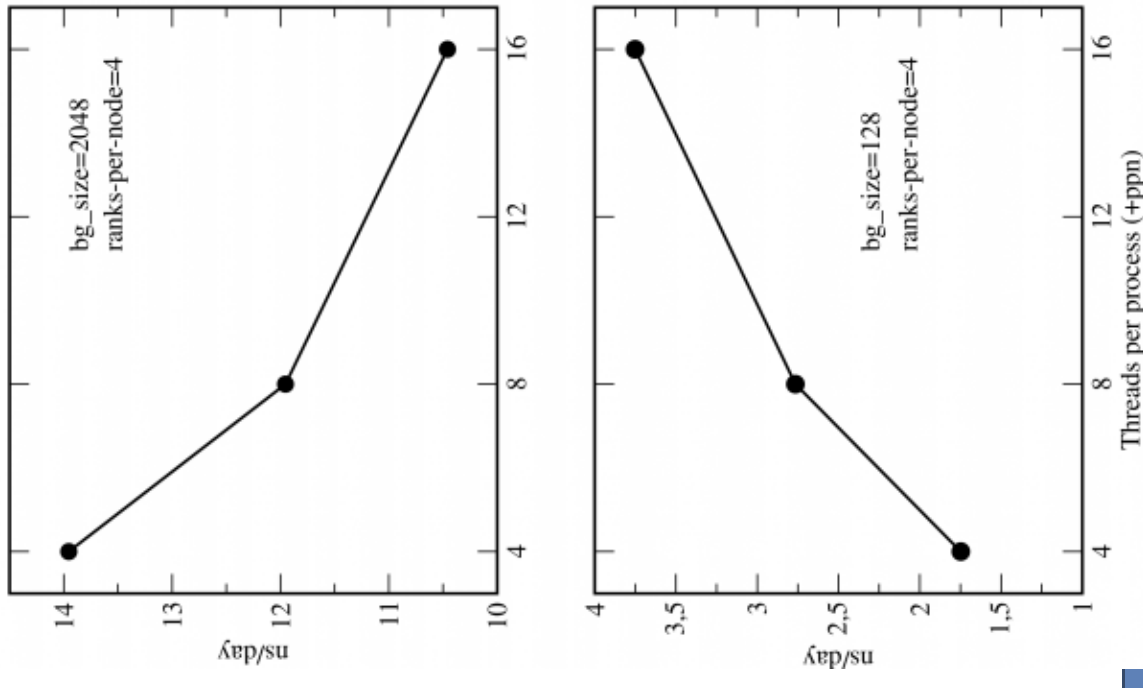
# NAMD 2.9 -4

- **Graphic 4:** the NAMD performance (simulation time in ns/day) is reported vs. the increasing number of threads per MPI process (ppn) (fixed bg\_size=128 and 2048, fixed ranks-per-nodes=4)

- bg\_size = 128 (2048)
- ranks-per-node = 4
- ppn = from 4 to 16
- Total MPI processes = 512 (8.192)
- Total threads = from 2048 to 8192 (32K to 128K)

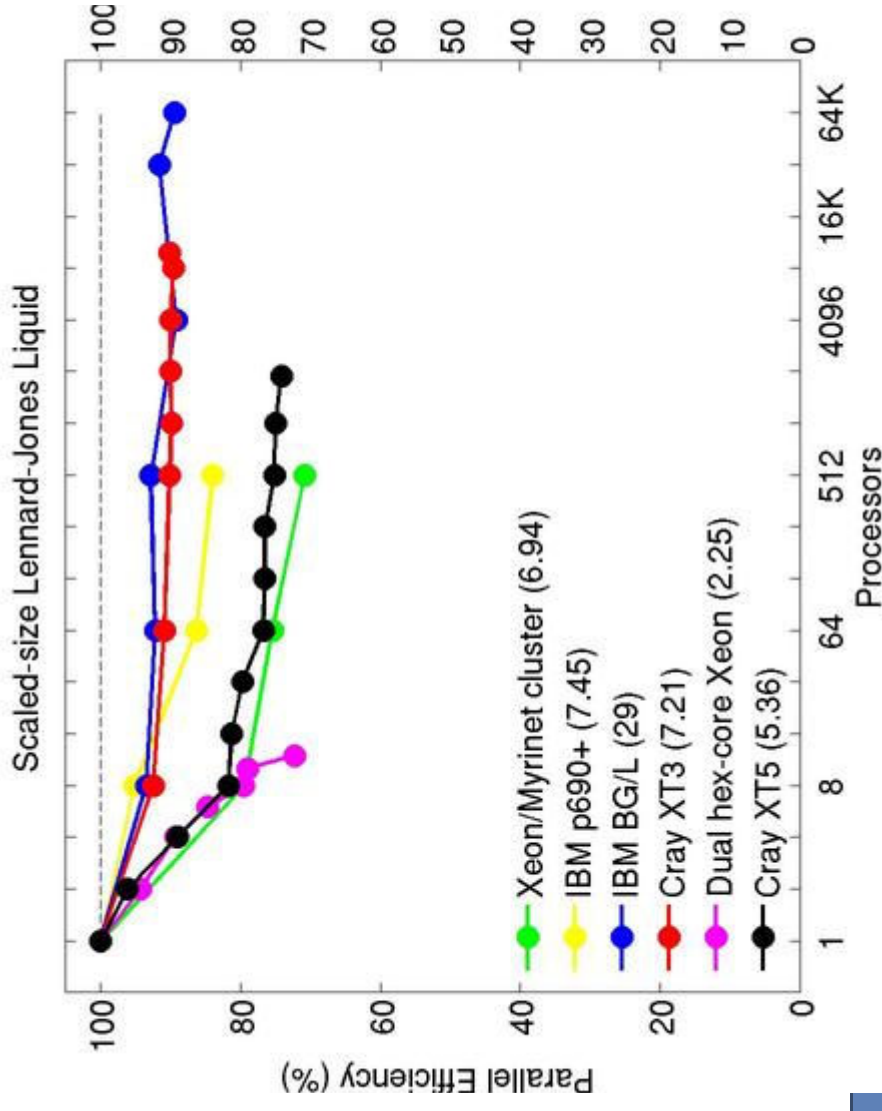
- The first graph is due to scalability reaching its limit for the given input

Satellite Tobacco Mosaic Virus  
NAMD 2.9, BlueGene/Q, ~1.2 MAtoms



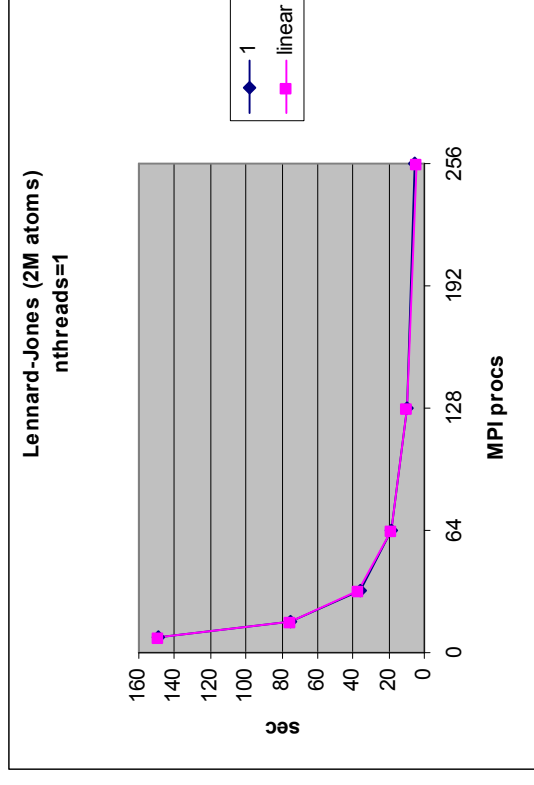
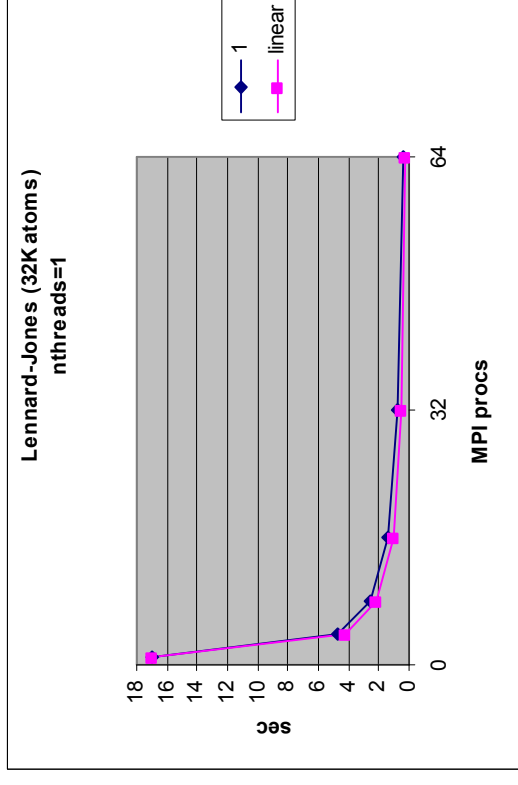


- classical molecular dynamics code that models an ensemble of particles in a liquid, solid, or gaseous state. It can model atomic, polymeric, biological, metallic, granular, and coarse-grained systems using a variety of force fields and boundary conditions.
- LAMMPS runs in parallel using MPI and a spatial-decomposition of the simulation domain. Each MPI task can distribute work over several OpenMP threads.
- Benchmarks use two different kinds of input
  - **Protein:** Rhodopsin protein in solvated lipid bilayer
  - **Lennard Jones:** atomic fluid with Lennard-Jones potential
- For both types, input dimension range from 32K to 2M atoms





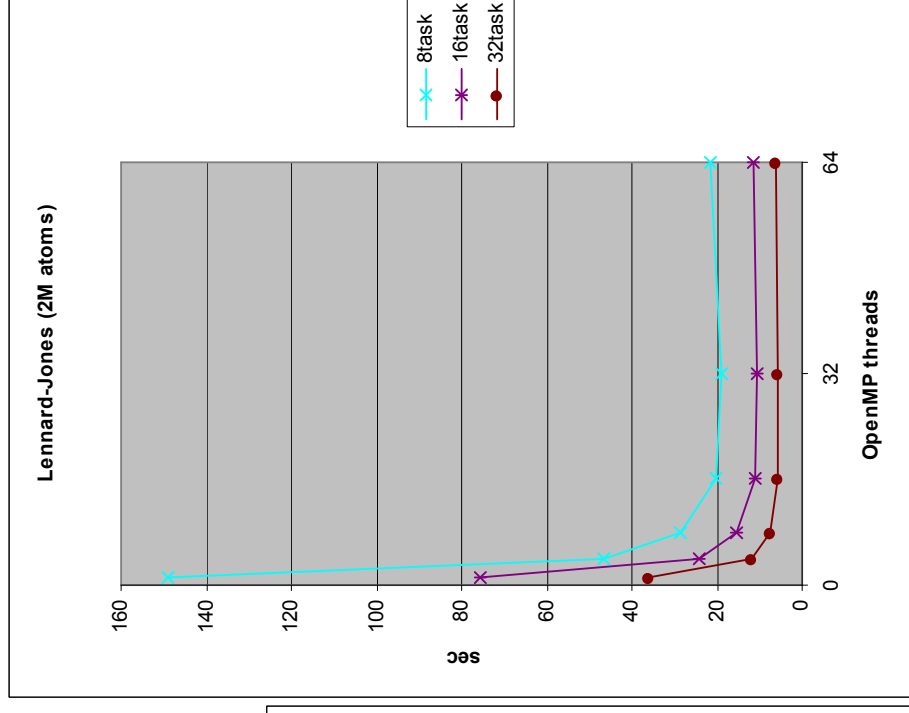
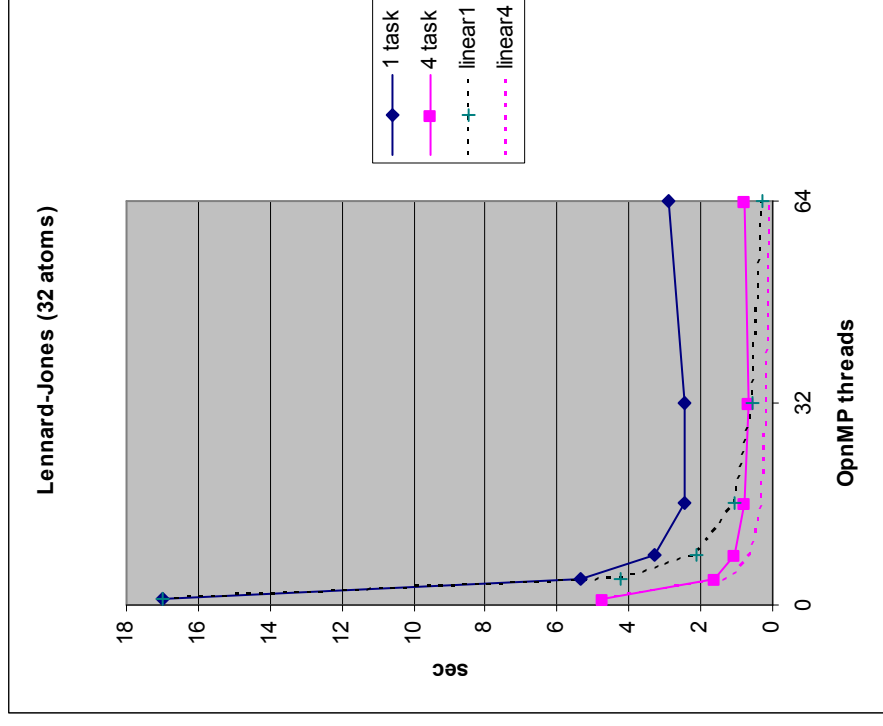
- **LAMMPS** execution time (sec) is reported vs. the increasing number MPI tasks (no OpenMP).

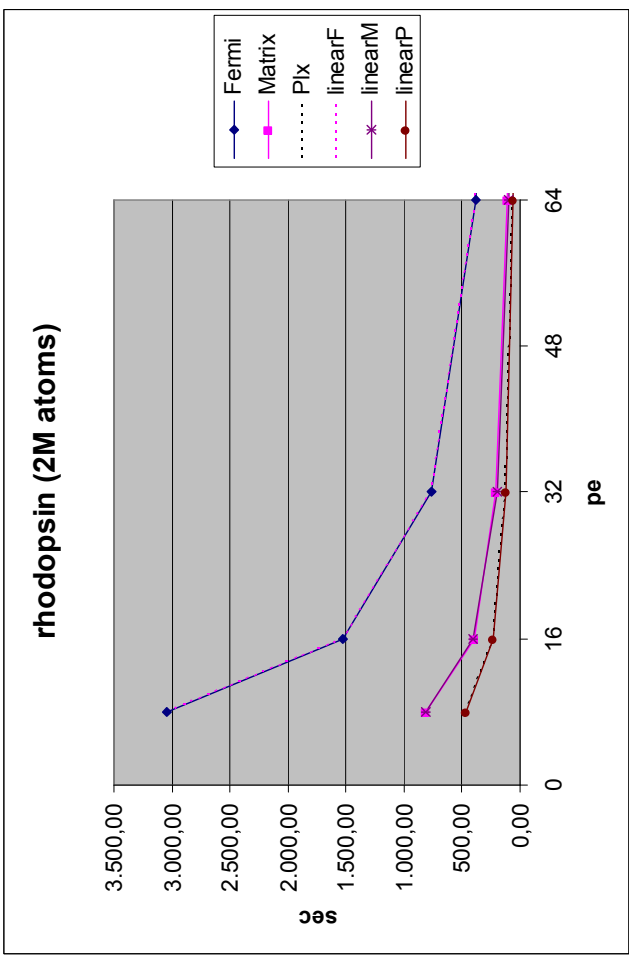
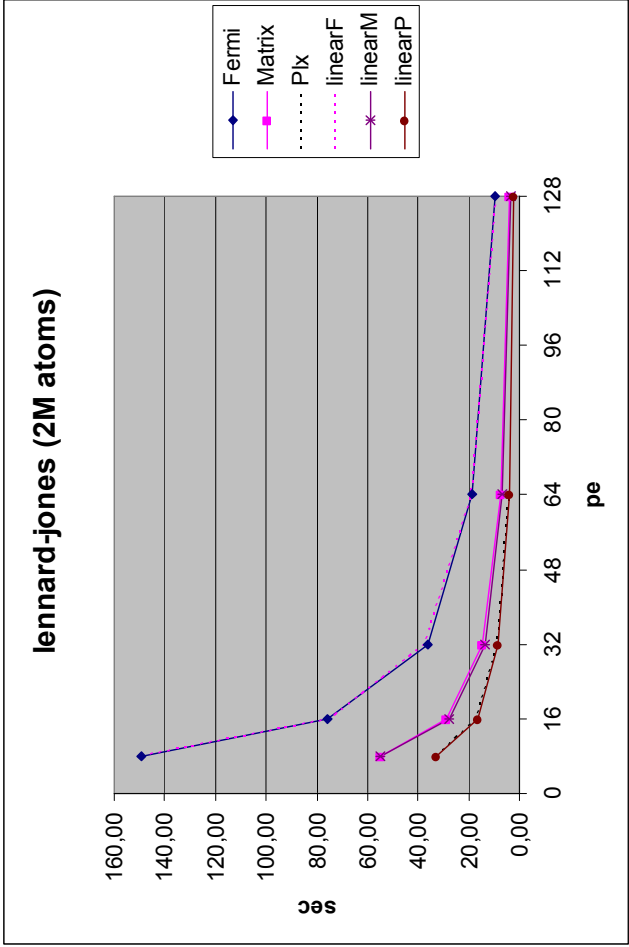


- Scalability is nearly perfect for all the tested cases (small and large, LennardJones and Rhodopsin)

- Study of the scalability in terms of OpenMP threads
- The number of MPI tasks are fixed (1 and 4, on different nodes) ranks-per-node=1, OpenMP threads range from 1 to 64

Scalability is rapidly decreasing.  
Seems to be no gain from SMT.





- **Comparison with other platforms:**
  - Matrix (in Rome) → 1: 2-4
  - PLX (in Bologna) → 1: 4-8
- Only pure MPI
- Perfect scalability on all platforms

- Different types of run on FERMI (3 nodes= 48 cores)
  - Only MPI, no SMT: 3\*16 task MPI
  - Only MPI, 2-way SMT: 3\*16\*2 MPI tasks
  - Only MPI, 4-way SMT: 3\*16\*4 MPI tasks
  - Max OpenMP, no SMT: 3 MPI tasks, 16 threads/each
  - Max OpenMP, 2-way SMT: 3 MPI tasks, 16\*2 threads/each
  - Max OpenMP, 4-way SMT: 3 MPI tasks, 16\*4 threads/each

#	2M l.j
28.4018	192 # (48 MPI x 1 OpenMP)
17.6195	96 # (48 MPI x 2 OpenMP)
13.0766	192 # (48 MPI x 4 OpenMP) <---
50.7467	48 # (3 MPI x 16 OpenMP)
48.016	96 # (3 MPI x 32 OpenMP)
53.1676	192 # (3 MPI x 64 OpenMP)
#	2M rhodo
578.888	192 # (48 MPI x 1 OpenMP)
297.991	96 # (48 MPI x 2 OpenMP)
189.945	192 # (48 MPI x 4 OpenMP) <---
710.398	48 # (3 MPI x 16 OpenMP)
513.015	96 # (3 MPI x 32 OpenMP)
426.492	192 # (3 MPI x 64 OpenMP)



# LAMMPS

To understand the “real” ratio, we run LAMMPS on the same resource amount on PLX

- Different types of run on PLX (4 nodes=48 cores)
  - Only MPI: 4\*12 task MPI
  - Max OpenMP: 4 MPI tasks, 12 threads/each
  - Mixed OpenMP:8 MPI tasks, 6 threads/each

# 1M lj	#	(48 MPI x 1 OpenMP)	<---
2.65876	#	(4 MPI x 12 OpenMP)	
5.01952	#	(8 MPI x 6 OpenMP)	
3.37871	#		
# 1M rhodo	#	(48 MPI x 1 OpenMP)	<---
42.9256	#	(4 MPI x 12 OpenMP)	
56.748	#	(8 MPI x 6 OpenMP)	
39.1213	#		

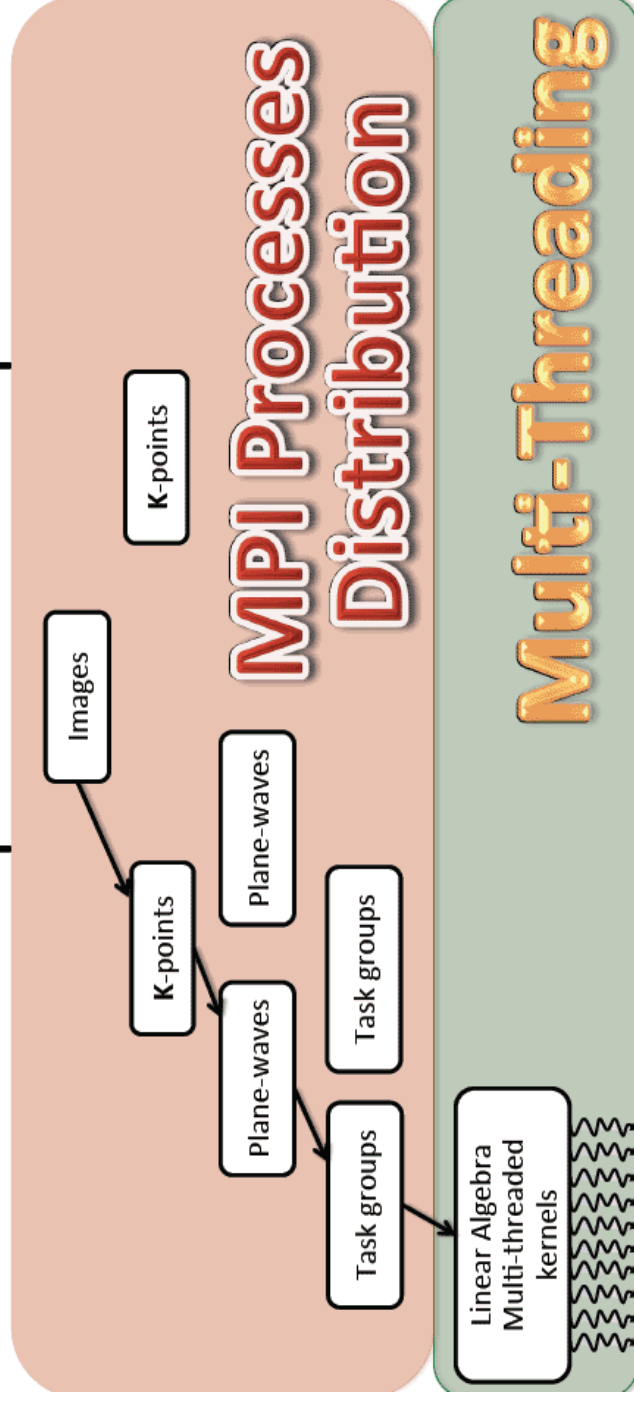
# 2M lj	#	(48 MPI x 1 OpenMP)	<---
5.49188	#	(4 MPI x 12 OpenMP)	
12.2105	#	(8 MPI x 6 OpenMP)	
6.78751	#		
# 2M rhodo	#	(48 MPI x 1 OpenMP)	<---
75.5071	#	(4 MPI x 12 OpenMP)	
109.026	#	(8 MPI x 6 OpenMP)	
80.943	#		

The “real” ratio, is about **2.5** (using the same resources, FERMI is about 2.5 times slower than PLX)

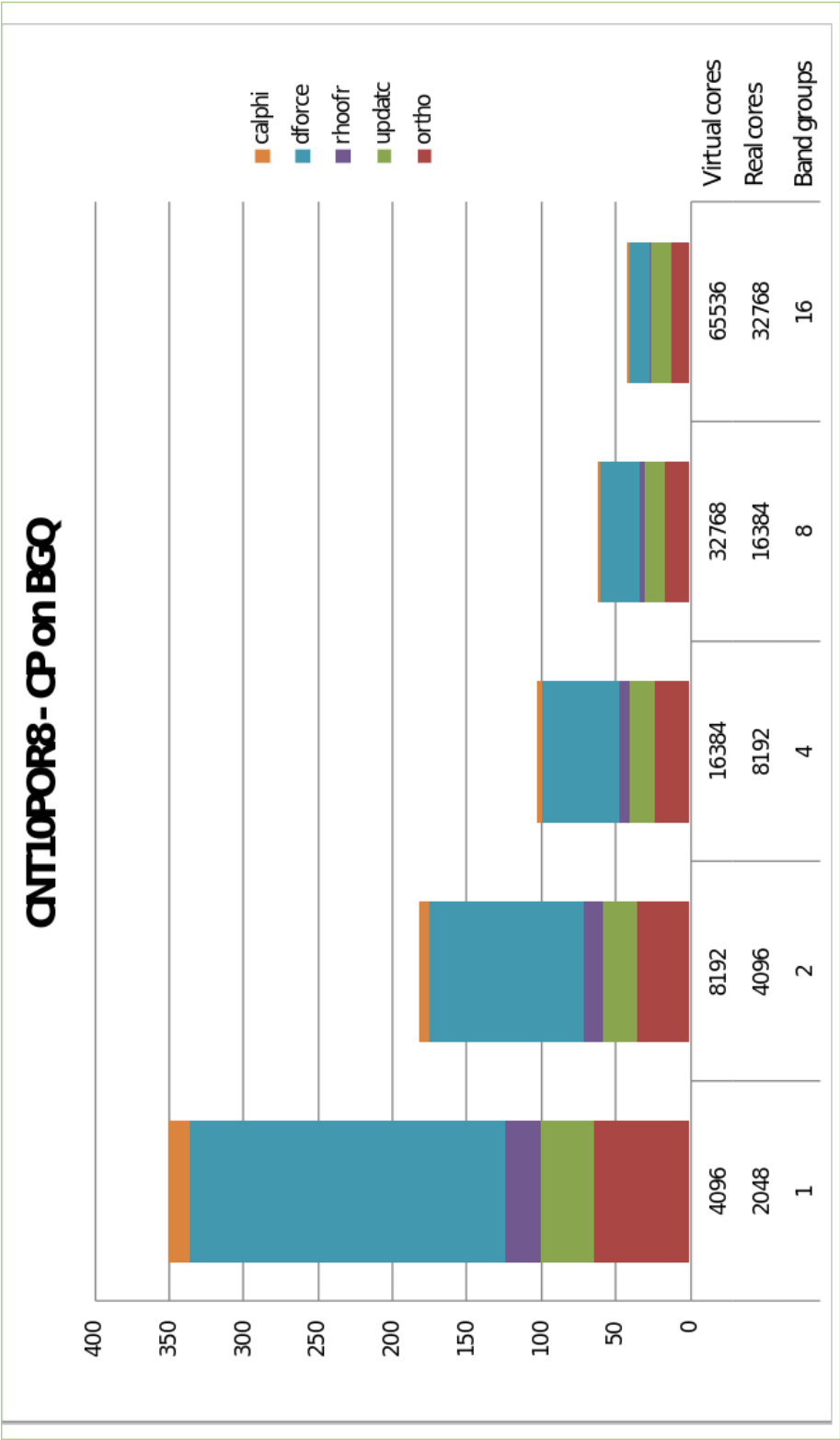
- It was 4-8
- Official ratio is 5
- FERMI can scale to much higher number of processors

Quantum ESPRESSO is an integrated suite of computer codes for electronic-structure calculations and materials modeling at the nanoscale. It is based on density-functional theory, plane waves, and pseudopotentials (both norm-conserving and ultrasoft). It is written in Fortran/C and is parallelised at different levels (mixed parallelism MPI + OpenMP)

## Levels of parallelism in pw.x

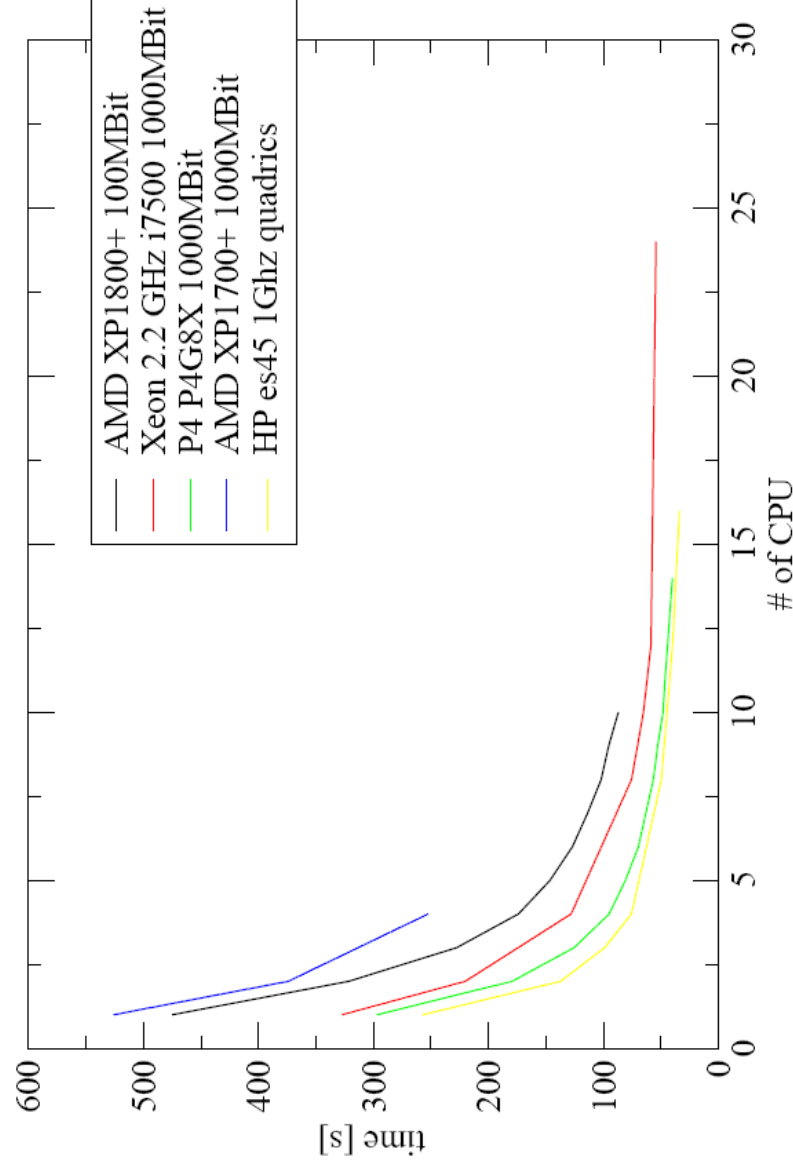


- This benchmark refers to a CP run on a sysytem of 1532 atoms.
- The input file can be downloaded from
- <http://qe-forge.org/gf/download/frsrelease/49/63/CNT10POR8.tgz>



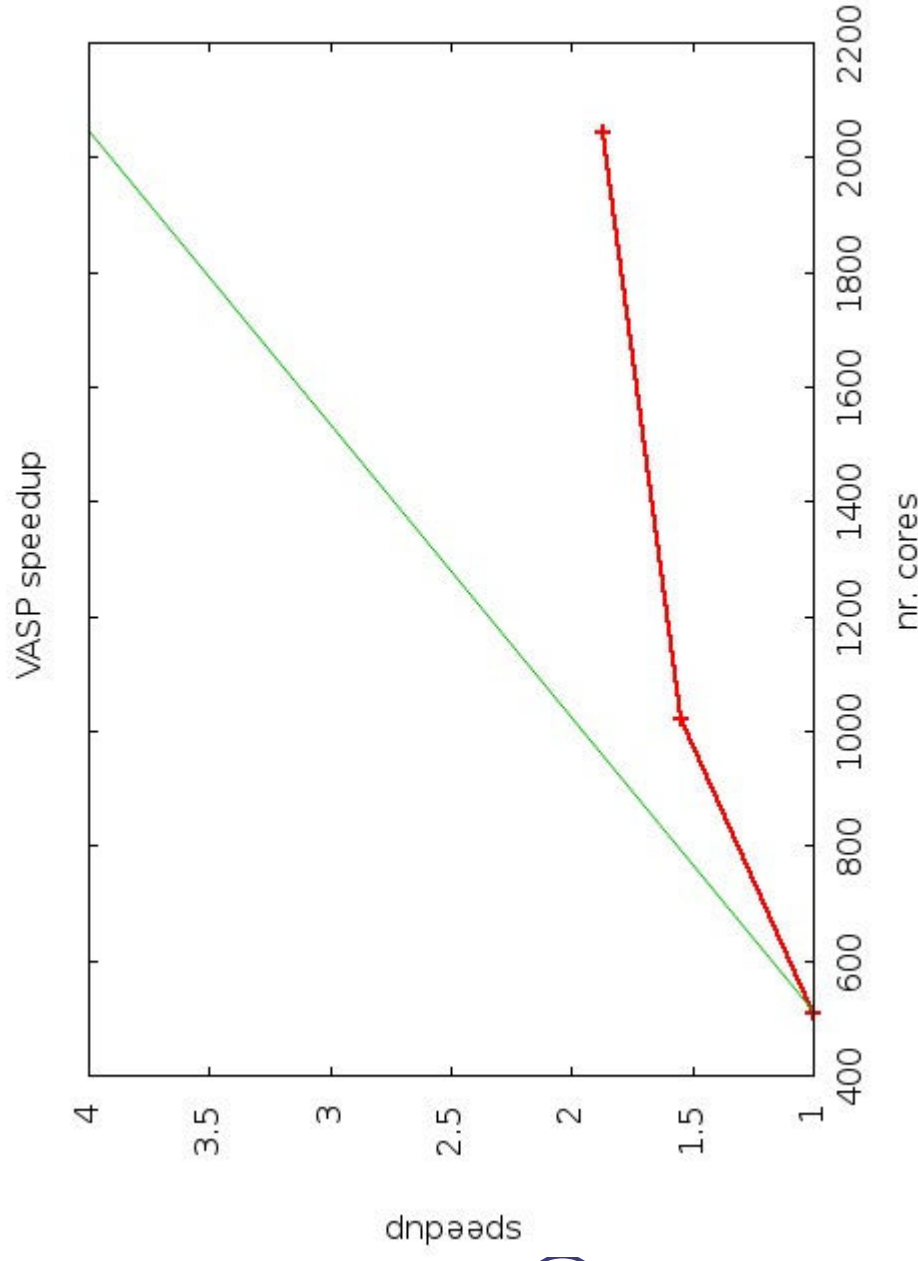


- Vienna Ab initio Simulation Package: a computer program for atomic scale materials modelling (electronic structure calculations and quantum-mechanical molecular dynamics, from first principles)
- Based on DFT or HF and hybrid functionals, Green's functions methods and many-body perturbation theory
- Uses plane wave basis sets, norm-conserving or ultrasoft pseudopotentials (projector-augmented-wave method)
- Fortran code, parallelisation via MPI library



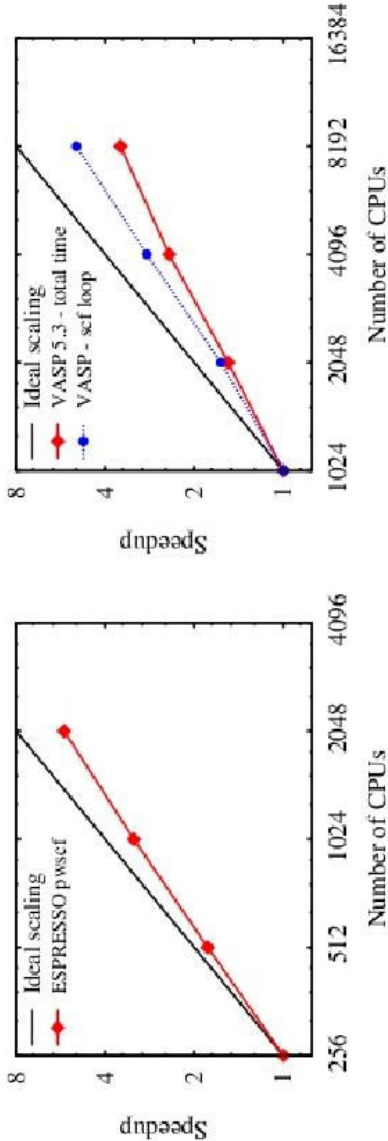


- Test performed within a Prace project. It shows a quite good speedup up to 1 thousand processors.
- Scalability depends on the method and input data.
- Memory is critical: for large cases it is better to use ranks-per-nodes= 8 or lower
- Modifying input parameters allow for important gains in performance (50-80%)



- Comparison with QuantumEspresso shows slightly lower scalability

Data in the table refers to VASP run



# cores	absolute timing (s)	speedup
1024	7112	1
2048	4633.5	1.5
4096	2930.1	2.4
8192	2003.7	3.5



## PLUTO 4.0

- Numerical solution of mixed hyperbolic/parabolic systems of partial differential equations (conservation laws) targeting high Mach number flows in astrophysical fluid dynamics.
- Equations are discretized and solved on a structured mesh that can be either static or adaptive. The static grid version of PLUTO is entirely written in C and is based on MPI to achieve highly scalable parallel performance.
- PLUTO is freely-distributed and is developed at the Dipartimento di Fisica, Torino University in a joint collaboration with INAF, Osservatorio Astronomico di Torino and CINECA.
- <http://plutocode.ph.unito.it/>

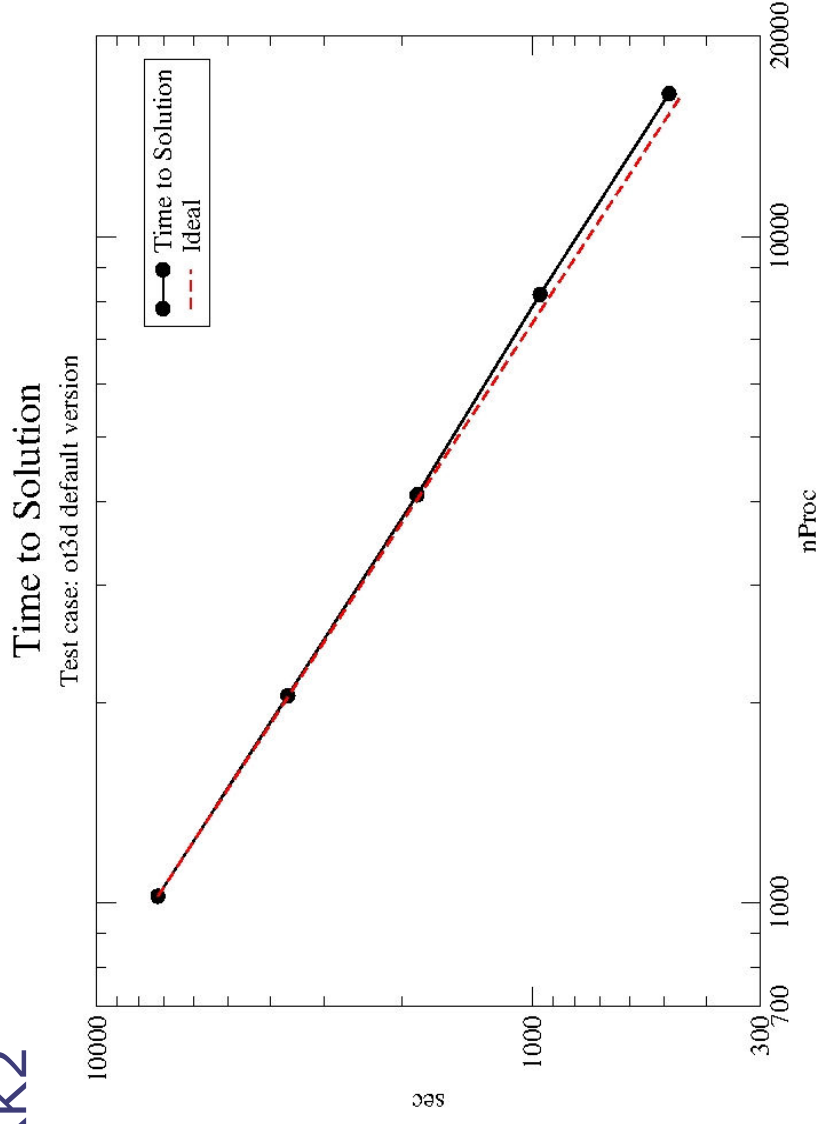
### First test case: Orszag-Tang vortex system

- The compressible Orszag–Tang vortex system describes a triply periodic fluid configuration undergoing supersonic MHD turbulence in three dimensions.
- The domain is the box  $[0, 1]^3$  with periodic boundary conditions imposed on all sides.

• **Default algorithm** configuration:

• **TIME\_STEPPING** = RK2

- Strong scaling up to 16K cores (bg\_size=1024)



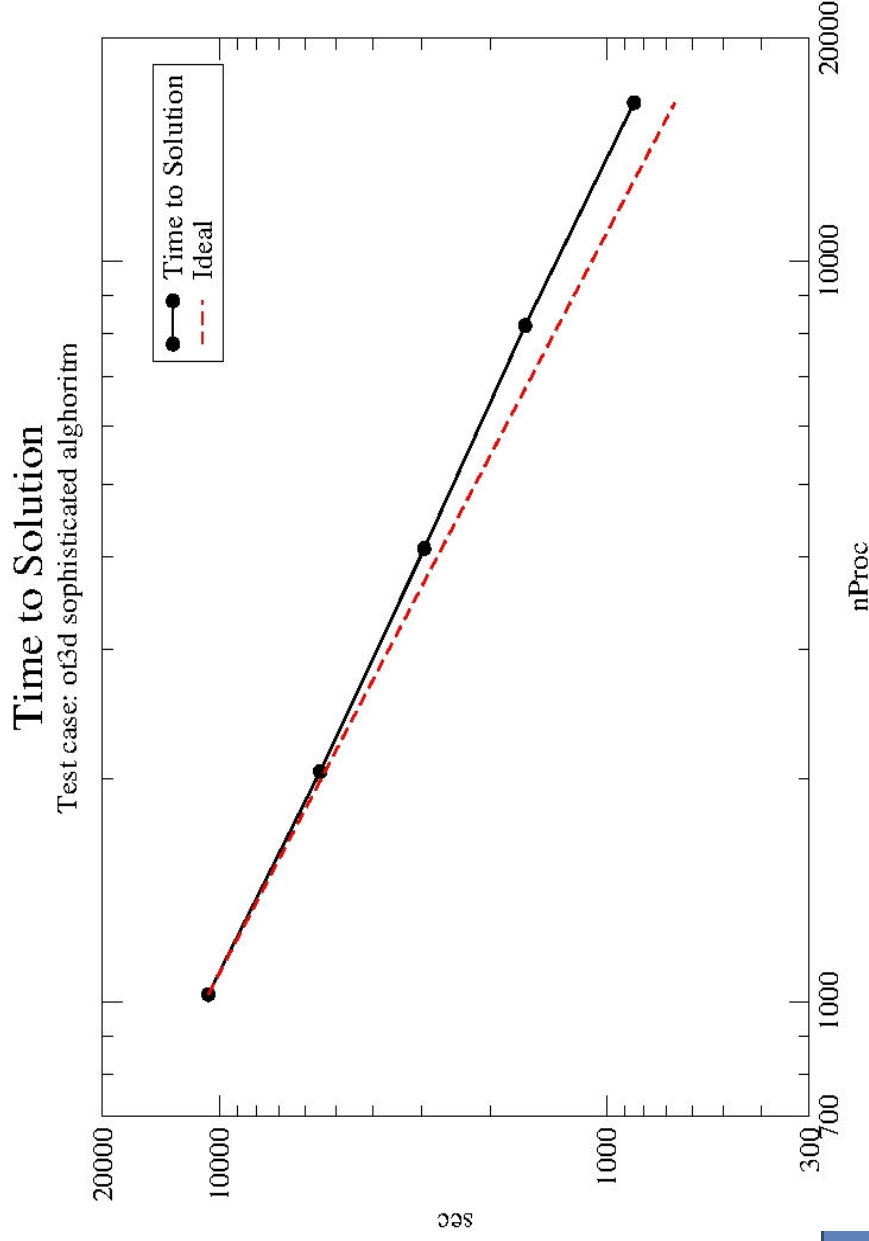
## Second test case: Orszag-Tang vortex system

- The compressible Orszag–Tang vortex system describes a triply periodic fluid configuration undergoing supersonic MHD turbulence in three dimensions.
- The domain is the box  $[0, 1]^3$  with periodic boundary conditions imposed on all sides.

- **Sophisticated algorithm** configuration:

• TIME STEPPING =  
HANKOK

- A more sophisticated algorithm generates unbalanced workload with a limited loss of performance
- strong scaling remains linear





# documentation

Home > Resources > Software

## Application Software for Science

### Content:

- Chemistry
- Physics
- Life Science
- Engineering
- Astronomy
- Visualisation
- Maths Libraries
- Data Libraries
- All Software

### Application Software for Science & Technology

Cineca offers a variety of third-party applications and community codes that are installed on its HPC systems. Most of the third-party software are installed using the software modules mechanism. The packages available and detailed descriptions of them can be viewed for discipline by selecting the menu on the left. If you want to see the full catalog, please get [all Software in alphabetical order](#).

The information in this list may not reflect all software products available at Cineca. If you do not see an application you are interested in, or if you have questions about software that is currently available, please contact the [Help Desk](#)

### Resources

- Hardware
- Software
  - Application Software
    - Chemistry
    - Physics
    - Life Science
    - Engineering
    - Astronomy
    - Visualisation
    - Maths Libraries
    - Data Libraries
    - All Software
  - Production Environment
    - Reports on systems usage



## For users

- My portal
- Getting started
- Get in touch**
- Help desk
- Documentation

## Help desk



Service Closed

## Center news

23/11/2012

Power outage

23/11/2012

Reboot of FERMI login node fen01

20/11/2012

FERMI shut down for management operation

14/11/2012

Call for proposals for DECI-10 call for Tier-1 resources

» more Center News

# Information and announcements

Home > For users

## Get in touch

### How to get Center's announcements (HPC-news)

We manage a mailing list (HPC-news) for posting announcements, scheduled downs, software updates, any problems and so on, about our HPC computing resources. It is advisable for HPC users to be included in that list!

You can subscribe (or unsubscribe) to HPC-news by sending an email from the address you want to subscribe. You can consult the archive browsing the archive web site.

#### To subscribe to HPC-news:

send a mail to [listsenv@list.cineca.it](mailto:listsenv@list.cineca.it)  
in the body --> "subscribe hpc-news"  
in the subject --> any string...

#### To unsubscribe to HPC-news:

send a mail to [listsenv@list.cineca.it](mailto:listsenv@list.cineca.it)  
in the body --> "unsubscribe hpc-news"  
in the subject --> any string...

#### To consult the archive:

for the recent archive go to → [Center News](#)  
for the full archive go to → <http://list.cineca.it/archives/hpc-news.html>

FERMI status



PLX status

