A toolkit for computational (atomic scale) biophysics

G. La Penna - CNR-ICCOM, Firenze
V. Minicozzi, S. Morante, G.C. Rossi - UNI-TOV, Roma
P. Giannozzi - UNI-UD, Udine

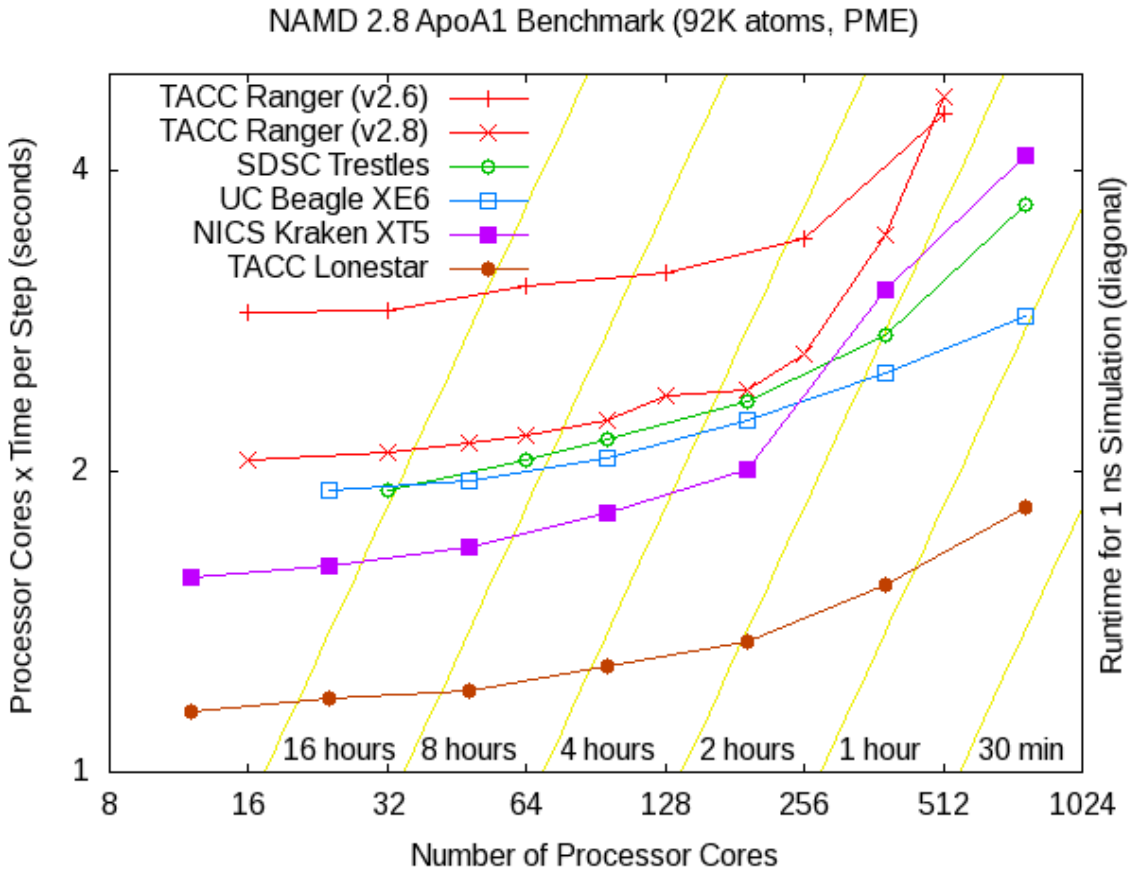
SUMA kick-off meeting, Bologna, 11 december 2012

$\sim$100,000 atoms

Interatomic forces at empirical level (classical molecular dynamics)

NAMD Not Another Molecular Dynamics - University of Illinois at Urbana-Champaign (USA)

$\sim$ 92,000 atoms



GROMACS GRoningen MAchine for Chemical Simulations -

F. Affinito et al., PRACE report 25/06/2012, Test on 1,800,000 atoms

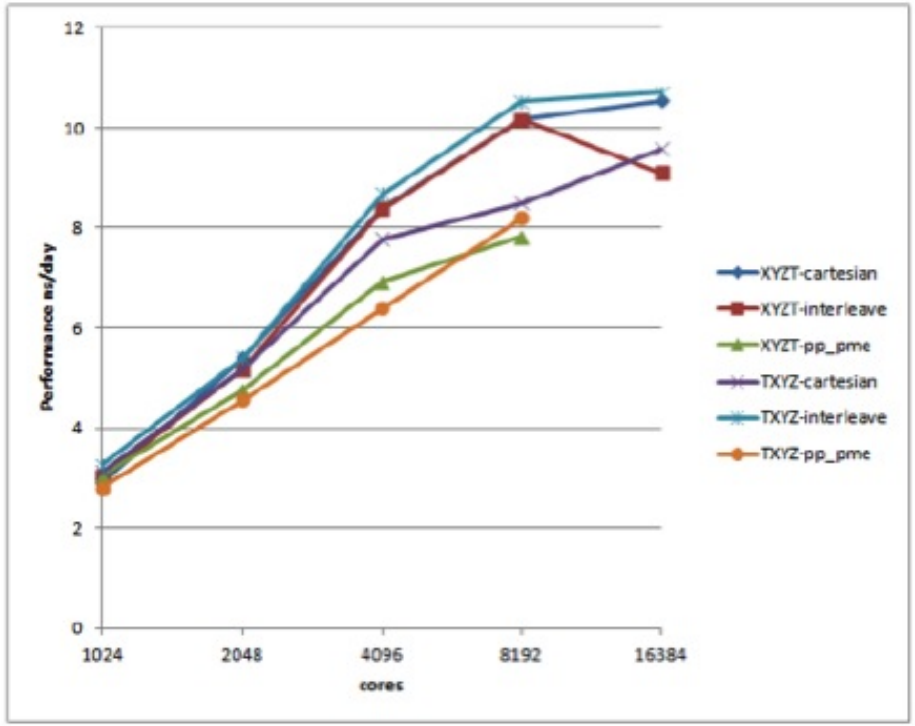


Figure 1: Performance as a function of number of cores for various GROMACS pp-pme partition schemes and Blue Gene/P MPI rank mappings. See text for an explanation of the legend.

$\sim$10000 atoms

DFTB *Density Functional Tight-Binding (DFTB+)*

Interatomic forces with explicit electrons:
semi-empirical approximation of $\mathcal{H}$

99% time in (sparse) matrices diagonalization

$\rightarrow$ Test of diagonalization routines

∼1000 atoms

Interatomic forces with explicit electrons: density-functional theory approximation

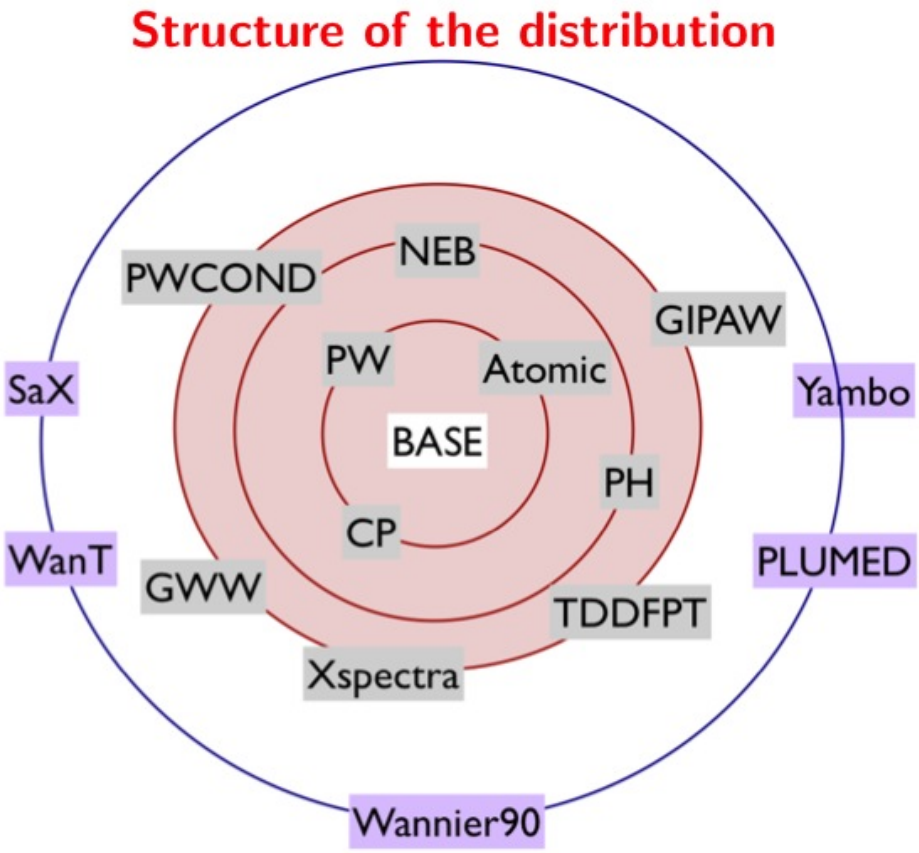
QUANTUM ESPRESSO *Quantum opEn-Source Package for Research in Electronic Structure, Simulation, and Optimization*

- `PWscf`: self-consistent electronic structure, structural optimization, molecular dynamics on the electronic ground state
- `CP`: variable-cell Car-Parrinello molecular dynamics

They share a common installation method, input format, PP format, data output format, large parts of the basic code. Also included:

- `PostProc`: graphical and postprocessing utilities (band structure, density of states, STM maps, ...)

Packages



http://www.quantum-espresso.org

Packages

A set of packages, using routines from the core distribution, can be downloaded and installed on demand:

- `PWGui`: a Graphical User Interface for production of input files
- `PHonon`: linear-response calculations (phonons, dielectric properties)
- `NEB`: reaction pathways and barriers using Nudged Elastic Band
- `atomic`: atomic calculations, pseudopotential generation
- `PWcond`: ballistic conductance
- `XSpectra`: X-ray near-edge adsorption spectra (XANES) shifts
- `TDDFPT`: Time-Dependent Density-Functional Perturbation Theory

Plugins

More advanced features, interoperable "external" packages:

- `GIPAW`: Gauge-Independent PAW for EPR and NMR chemical shift
- `GW`: GW band structure with ultralocalized Wannier functions
- `PLUMED`: plugin for metadynamics calculations
- `WanT`: Transport using maximally Localised Wannier functions
- `W90`: package for Maximally Localised Wannier functions
- `SaX`: electronic excitations from GW band structure
- `Yambo`: GW band structure, Bethe-Salpeter equations

What can QUANTUM ESPRESSO do?

- Structural modeling (equilibrium structures of molecules, crystals, surfaces)
- Dynamical modeling (first-principles molecular dynamics) either on the electronic ground state (Born-Oppenheimer) or with fictitious electronic kinetic energy (Car-Parrinello)
- Chemical reactivity and transition-path sampling, using Nudged Elastic Band (NEB) method
- Linear response functions (vibrational and dielectric properties), plus some non-linear ones (third-order force constants and dielectric response, non-resonant Raman)
- Computational microscopy (STM)

Advanced QUANTUM ESPRESSO **capabilities**

- Beyond simple DFT: DFT+U, nonlocal (vdW-DF), hybrid functionals (PBE0, B3LYP), meta-GGA
- Ballistic transport, transport with Wannier functions
- Free-energy sampling (metadynamics, with PLUMED plugin)
- Computational spectroscopy:
  - lattice and molecular vibrations: Raman, Infrared, Neutrons
  - magnons and spin excitations
  - photoemission (MBPT)
  - optical/UV absorption (TDDFT, MBPT soon to come)
  - NMR chemical shifts
  - X-ray spectra, core level shifts

Technical features (algorithms)

- use of iterative techniques: the Hamiltonian is stored as operator, not as matrix. All standard PW technicalities: FFT, dual-space, etc., are used. Iterative diagonalization used whenever it is useful.

- fast "double-grid" implementation for ultrasoft PPs: the cutoff for the augmentation part can be larger (the corresponding FFT grid denser in real space) than the cutoff for the smooth part of the charge density.

- Many parallelization levels: on images, on **k**-points, on PW's and FFT grids, on Kohn-Sham states in FFTs, on orthonormalization or subspace diagonalization, for extended scalability up to O(1000) processors.

Technical features (coding)

- written mostly in Fortran-90, with various degrees of sophistication (i.e. use of advanced f90 features) – no dirty tricks, "dusty decks"
- use of standard library routines (lapack, blas, fftw) to achieve portability – Machine-optimized libraries can be used if available
- C-style preprocessing options for machine dependencies (e.g. to select machine-optimized libraries) allow to keep a single source tree
- parallelization via MPI calls, hidden into calls to very few routines – (almost) unified serial and parallel versions. Unless something special is desired, there is no need to know the internals of parallelization in order to write parallel code. OpenMP parallelization is also present.

Easy (or not-so-difficult) installation via the GNU utility `configure`

Computer requirements

Quantum simulations are both CPU *and* RAM-intensive.
Actual CPU time and RAM requirements depend upon:

- *size of the system under examination:* As a rule of thumb, CPU $\propto N^{2\div3}$, RAM $\propto N^2$, where $N =$ number of atoms

- *kind of system:* type and arrangement of atoms, influencing the number of plane waves, of electronic states, of **k**-points needed...

- *desired results:* computational effort increases from simple self-consistent (single-point) calculation to structural optimization to reaction pathways, molecular-dynamics simulations, ...

CPU time mostly spent in FFT and linear algebra.
RAM mostly needed to store Kohn-Sham orbitals.

Parallelization levels for PW/CP

grouping	distributed quantities	communications	performances
plane-wave	PW, **G**-vector coefficients, **R**-space FFT arrays	high	good CPU scaling, good load balancing, distributes most RAM
Electron WF	Kohn-Sham states	high	distributes RAM & computation
task	FFT on electron states	high	improves load balancing
linear-algebra	subspace hamiltonians and constraints matrices	very high	improves scaling, distributes more RAM
OpenMP	FFT, libraries	intra-node	extends scaling on multicore machines
Configurations	Sampling	very low	distributes trajectories

Parallelization levels for CP

MPI+OpenMP parallelization *crucial* for good BG/Q performances

PRACE project (4th-call):
16 replicas, $\sim$1000 atoms, L $\sim$2 nm, $\sim$3000 electrons

set-up for 16384 cores:

grouping	distribution
PW - electron WF	1 KS-states / 2 tasks MPI
task	2 FFT-slices / task
linear-algebra	1 matrix sub-block / 128 tasks
OpenMP	16 OpenMP threads / task
Configurations	1 replica / 1024 tasks

1 fs = 275 s $\rightarrow$ 1 ps in 3 days