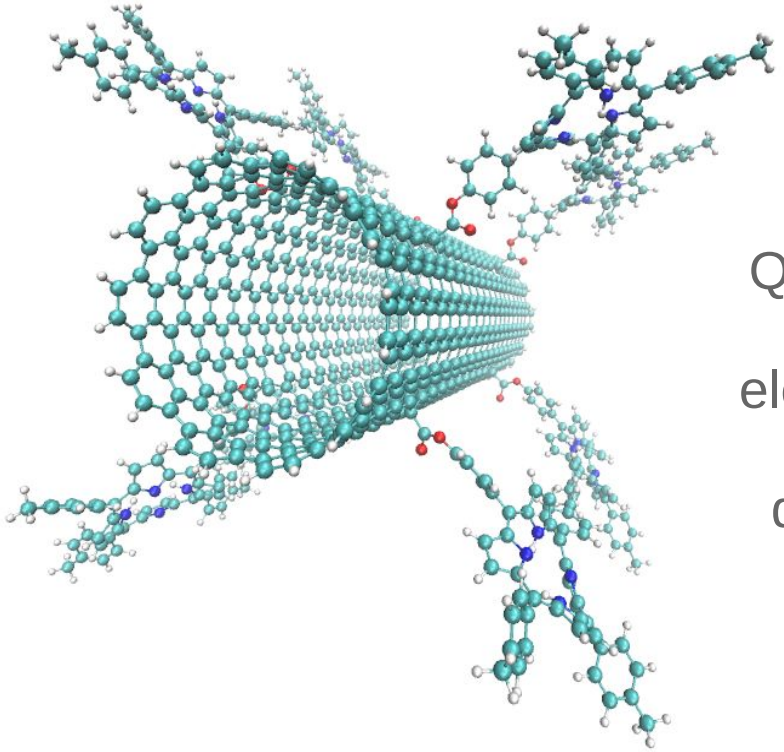
QUANTUM ESPRESSO on GPU: from energies to molecular properties

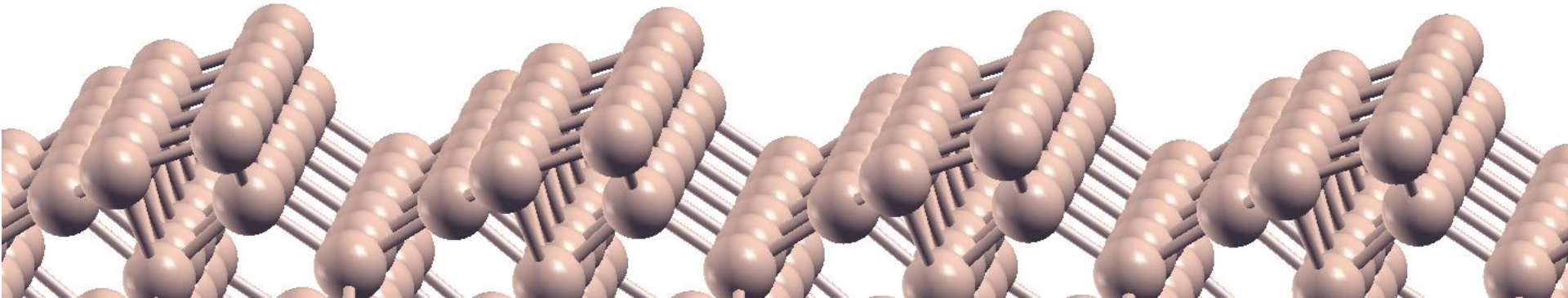
Ivan Carnimeo

Scuola Internazionale Superiore di Studi Avanzati (SISSA) – Trieste, Italy

QUANTUM ESPRESSO project

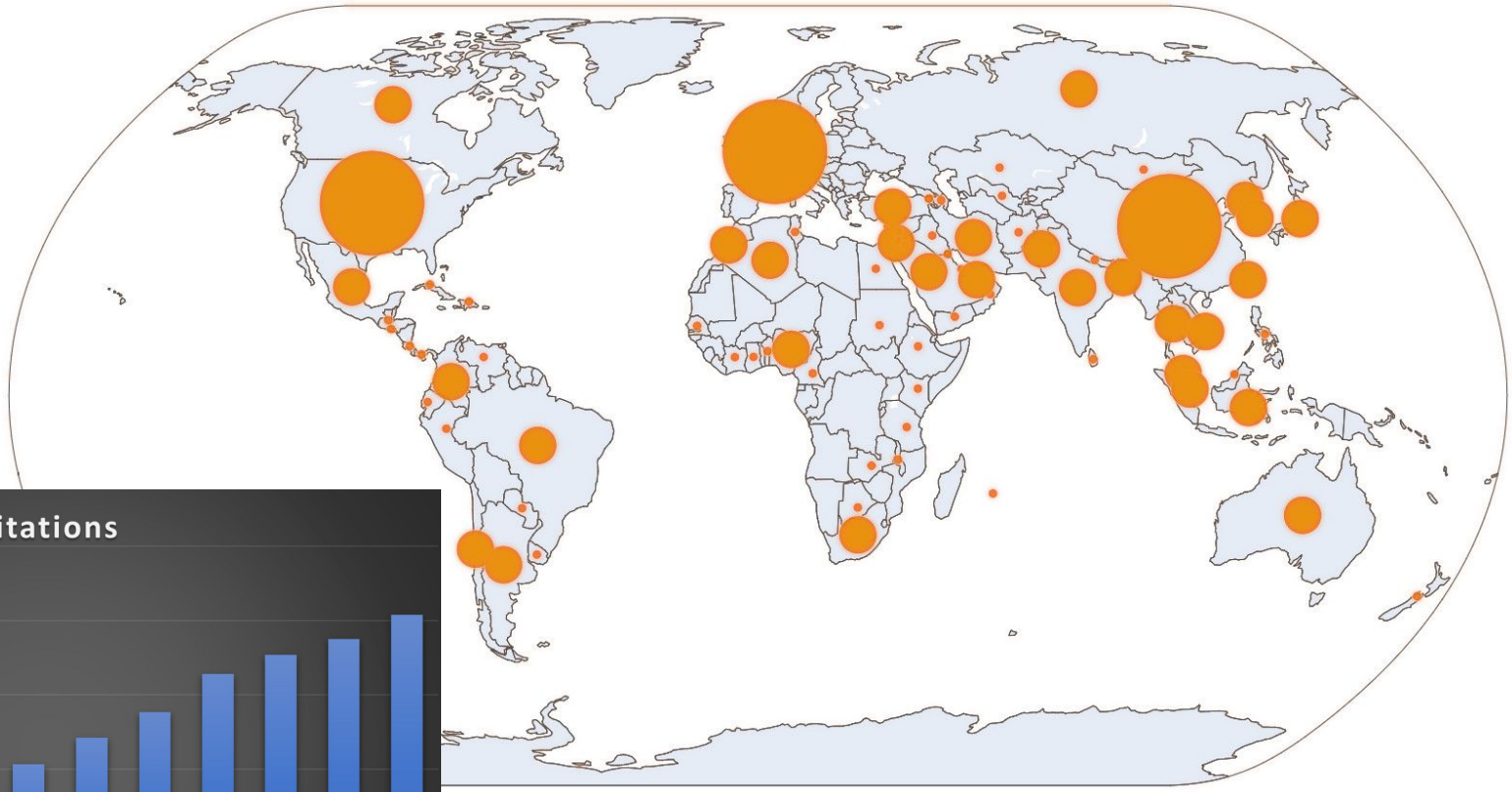


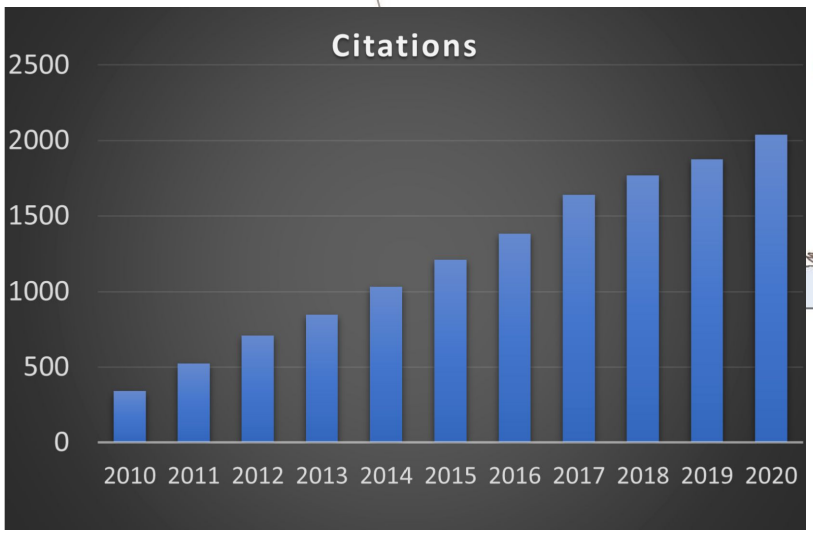
QUANTUM ESPRESSO™ is an integrated **suite of Open-Source computer codes** for electronic-structure calculations and materials modeling at the nanoscale. It is based on density-functional theory, plane waves, and pseudopotentials.



QUANTUM ESPRESSO project

QUANTUM ESPRESSO™ is an open initiative involving a **large community** of developers and contributors from different regions of the world



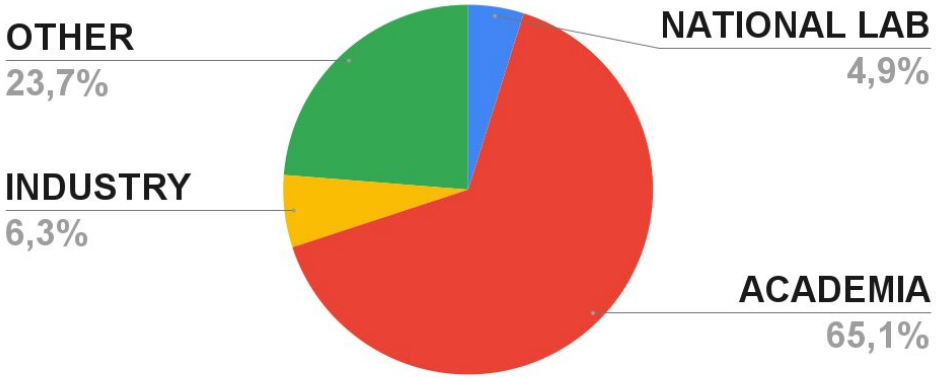


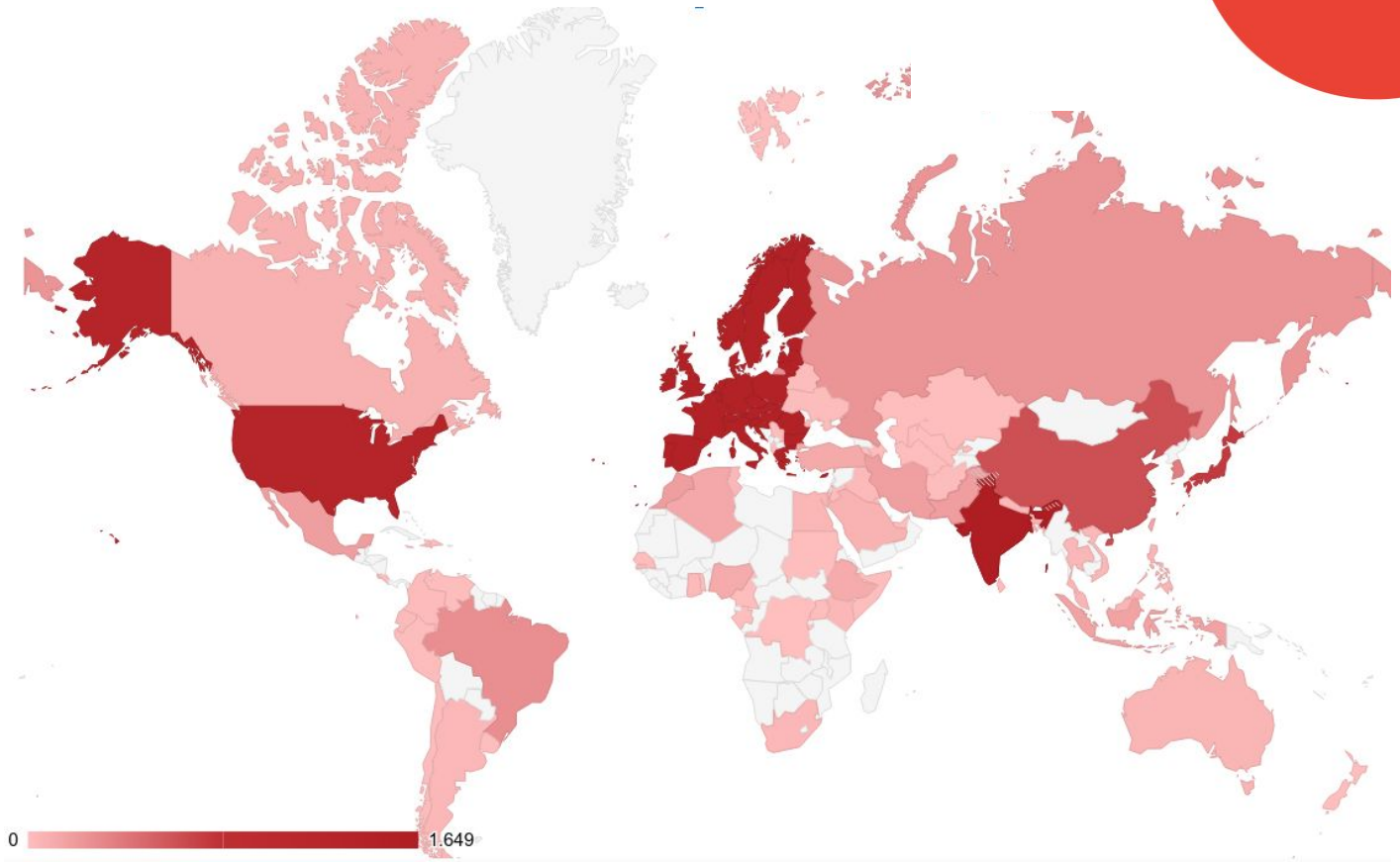
Geographic distribution of the authors of the articles citing the main reference articles of QE

Data provided by courtesy of the QUANTUM ESPRESSO Foundation

QUANTUM ESPRESSO project

Nearly **10000 download** of the code from the website since the beginning of 2022, mostly from Europe, USA, India and China



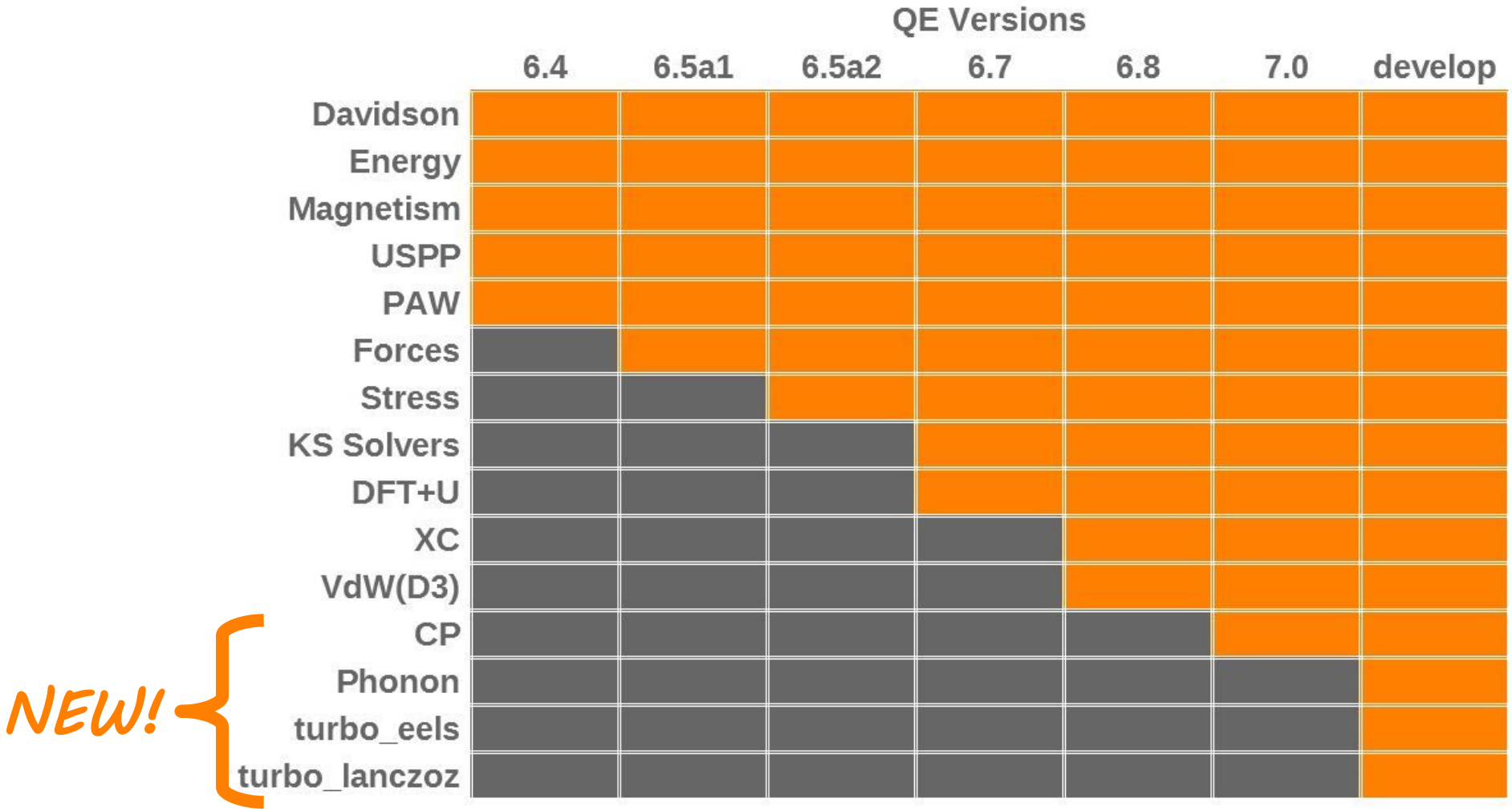


Geographic distribution and main professional fields of people who have downloaded QE from the website since the beginning of 2022

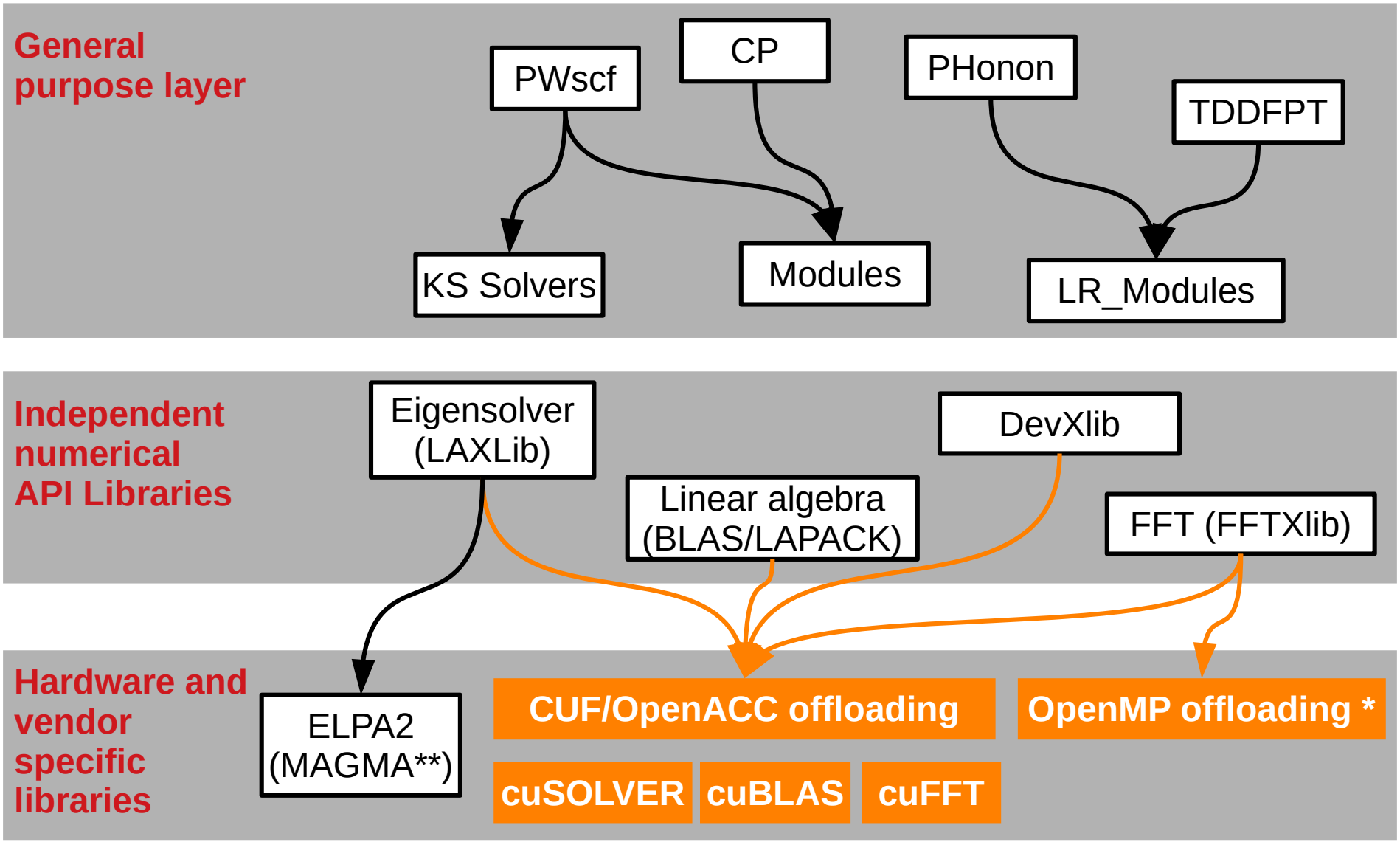
Data provided by courtesy of the QUANTUM ESPRESSO Foundation

QE on GPU (portability)

The QUANTUM ESPRESSO™ codes have been accelerated using a mixed **CUDA Fortran/OpenACC** scheme. Recently also the **OpenMP** offloading begins to be used in some modules (FFTXlib), to enhance portability to hardware from different vendors



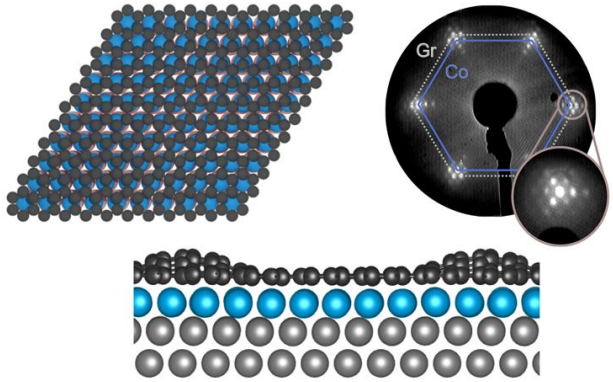
QE on GPU (portability)



* OpenMP offloading in FFTXlib under development by Giacomo Rossi, Intel
** MAGMA and AMD technologies have been tested during the MaX Hackathon, January 2022

PWscf on GPU

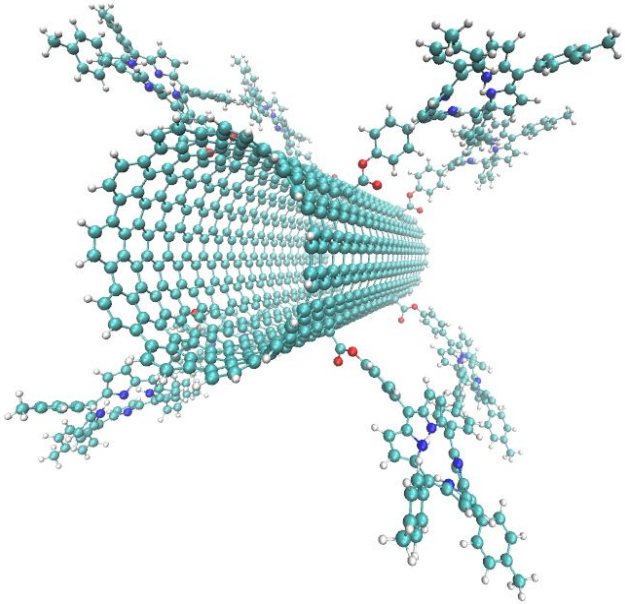
PWscf, the core of QE, computes the electronic structure of molecules and materials by solving the Kohn-Sham (KS) equations, within the **Density Functional Theory** framework:



$$\hat{H}^{KS}\psi_{nk}(\mathbf{r}) = \varepsilon_{nk}\psi_{nk}(\mathbf{r})$$

using Plane-Wave and Pseudopotential methods:

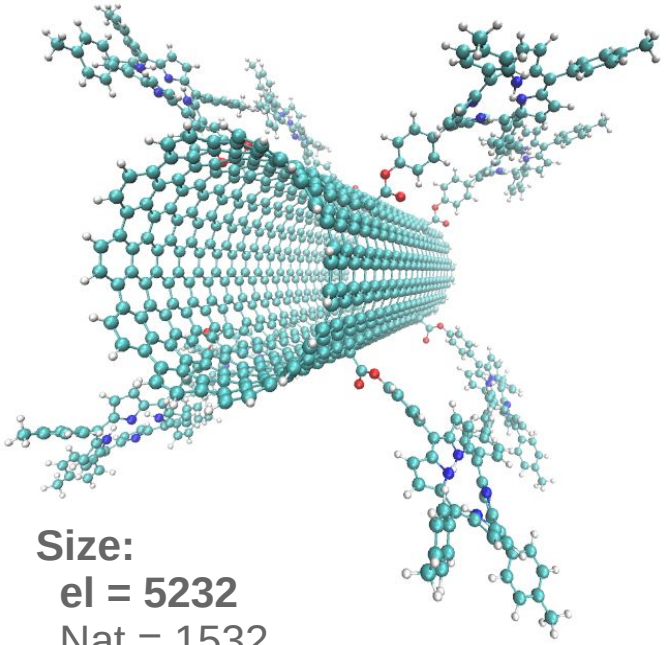
$$\psi_{nk}(\mathbf{r}) = \sum_{\mathbf{G}}^{N_{PW}} C_{\mathbf{G},nk} \frac{e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}}{\sqrt{\Omega}}$$



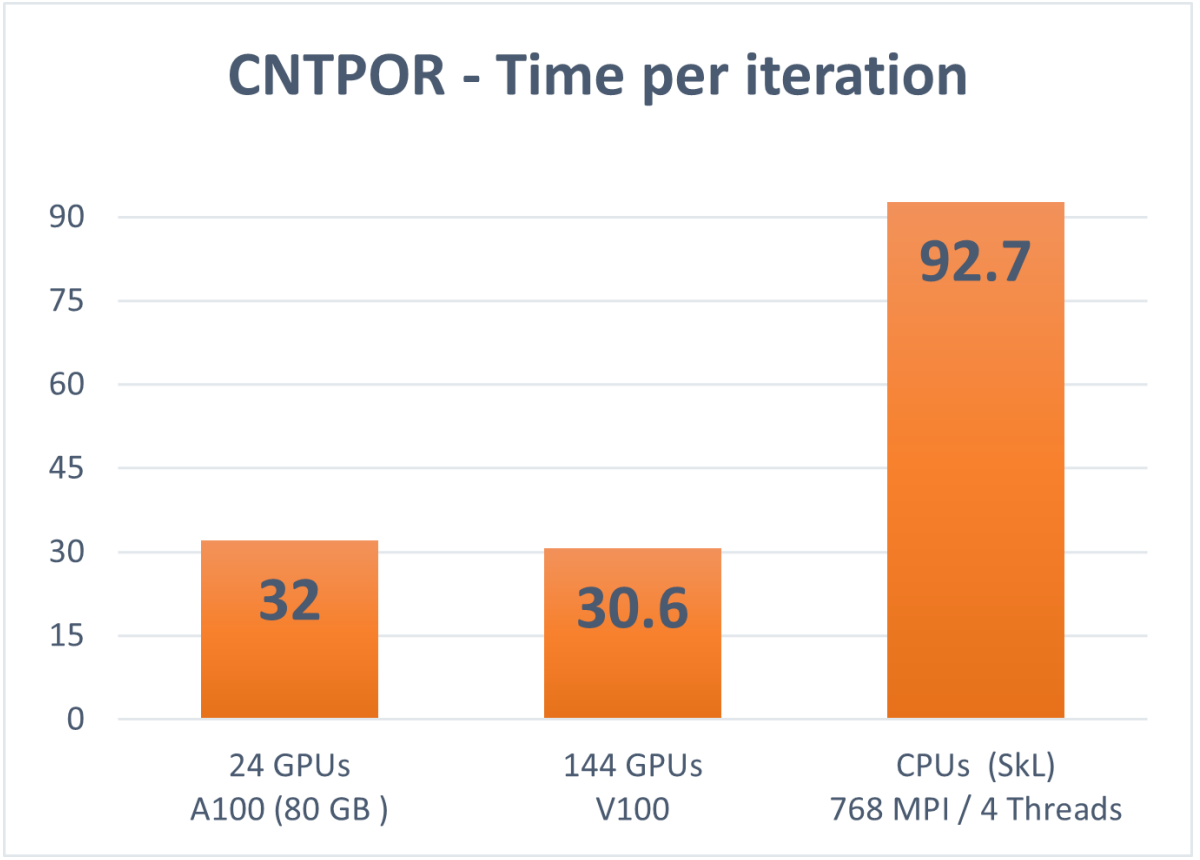
P. Giannozzi, J. Chem. Phys. 152, 154105 (2020)
P. Giannozzi et al 2017 J. Phys.: Condens. Matter 29 465901
P. Giannozzi et al 2009 J. Phys.: Condens. Matter 21 395502
S. Scandolo et al. Zeitschrift fur Kristallographie, 220, 574-579 (2005)

PWscf on GPU

Given the very good performance of core part of QE, **PWscf**, on GPU-accelerated hardware, in the last months we accelerated other relevant codes of the QE suite

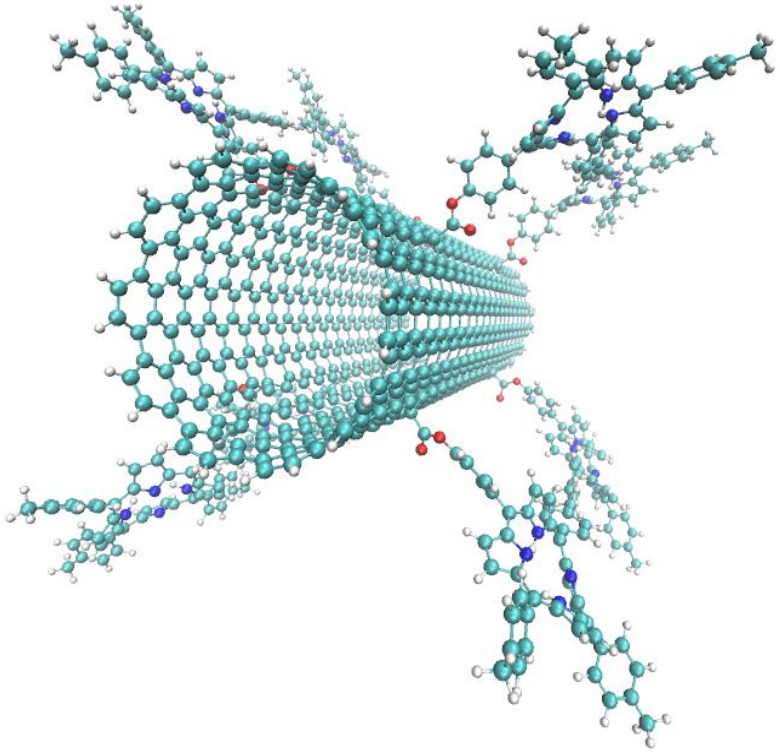


Size:
el = 5232
Nat = 1532
Ecut = US/25/200 Ry
Nks = Gamma only



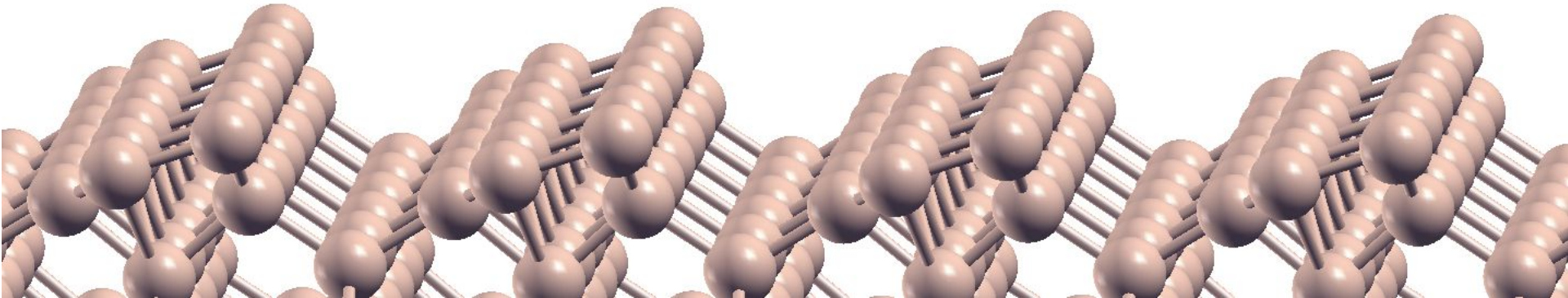
Calculations on A100 provided by courtesy of NVIDIA

QUANTUM ESPRESSO project



Some important developments done in the last year on the code include:

- Efficient EXX code for band structures with hybrid functionals
- CP porting to GPU
- Phonon porting to GPU
- TDDFPT porting to GPU



PWscf on GPU (EXX)

Band structures with hybrid functionals

relax/
vc-relax

↓

scf

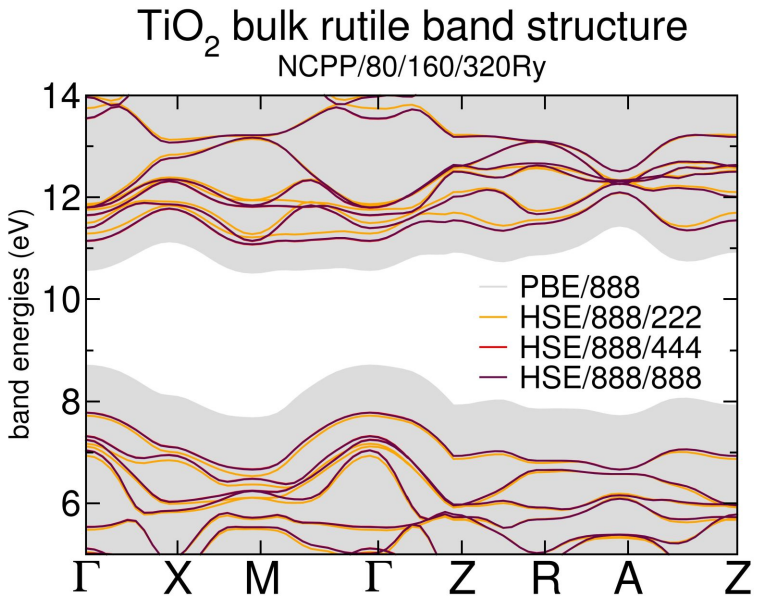
↓

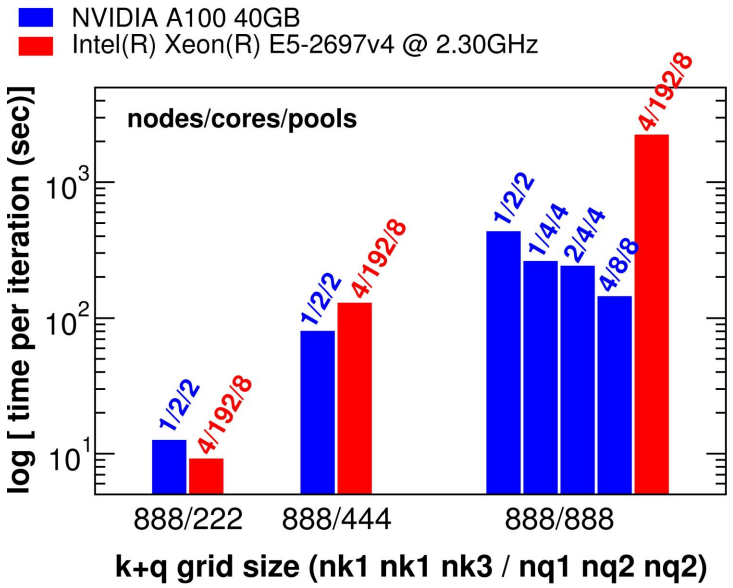
bands

Hybrid functionals allow to correctly reproduce the band structure of many molecular systems, e.g. oxides, but they are often prohibitive for periodic systems with PW, due to **k k'** mixing

Last developments include:

- **NSCF + EXX**, to separate occ and virt orbitals calculations;
- **interpolation scheme**, to get bands from uniform grids;
- **GPU acceleration**





Calculations on A100 performed during the NVIDIA UK Hackathon, Feb-Mar 2022

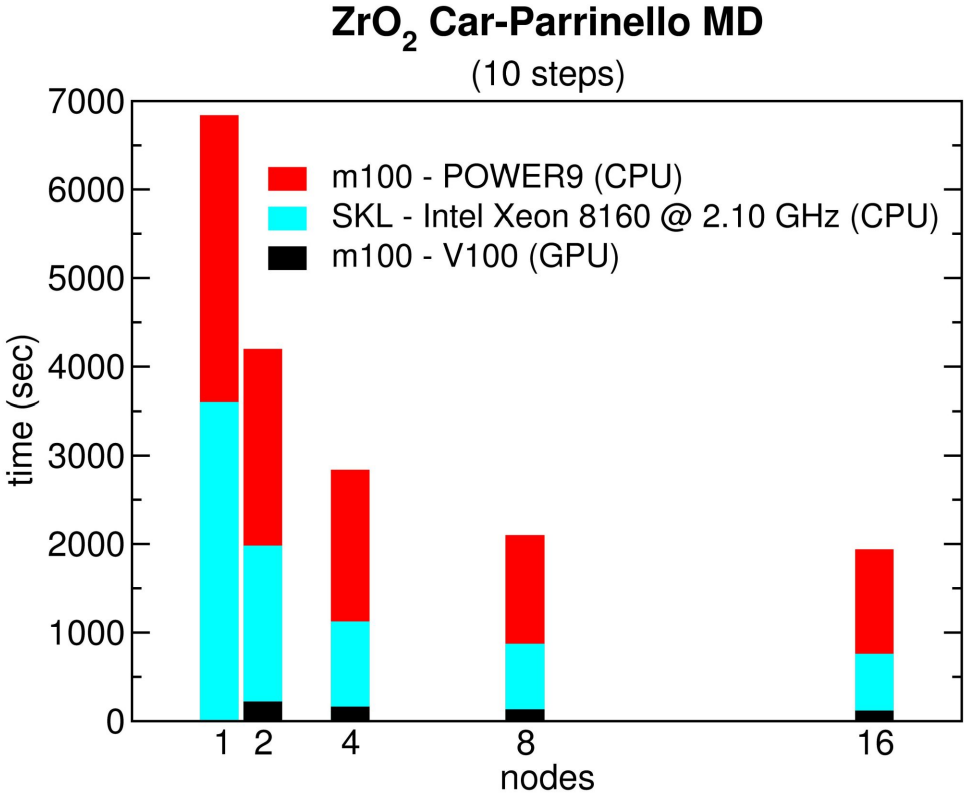
CP on GPU (CPMD)

Molecular dynamics of large systems

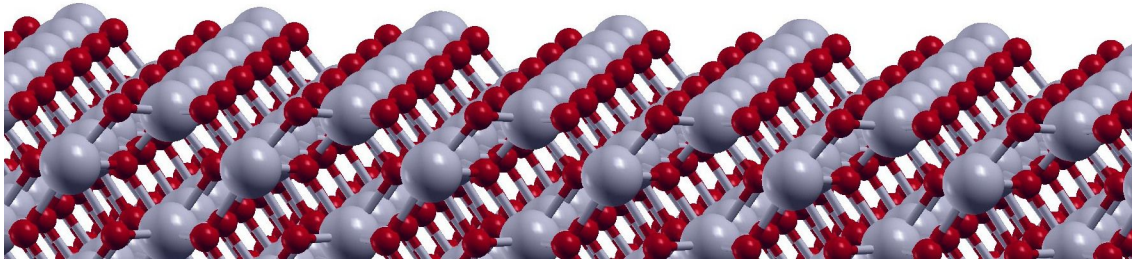
cp

analysis

Each step of the accelerated **Car-Parrinello** MD takes a very small fraction of the equivalent calculation run on CPU, allowing significantly longer dynamics



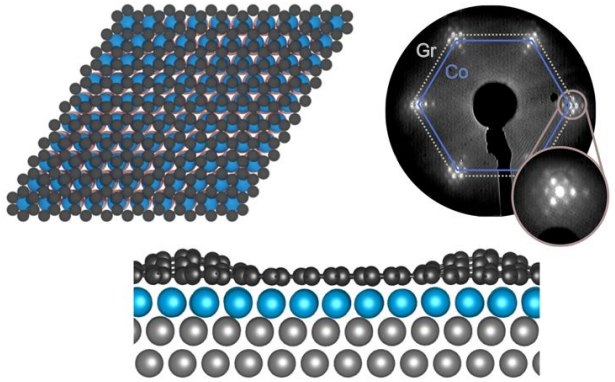
Size:
el = 6336
Nat = 792
Ecut = US/30/200 Ry
Nks = Gamma only



Porting to GPU carried out jointly with CINECA HPC Staff

PHonon on GPU

PHonon code computes the response of the system to static perturbations, such as a static electric field and nuclear displacements, giving access to vibrational properties, such as IR and Raman spectra and phonon dispersion curves



$$(H_{\mathrm{SCF}} + \alpha P_v - \varepsilon_n)\,|\Delta\psi_n\rangle = -P_c \Delta V_{SCF} |\psi_n\rangle$$

The core part of the code is the solution of one set of self consistent **Sternheimer equations** for each perturbation, using iterative (Conjugate Gradient) algoritms and linear algebra routines for solution of linear systems

S. Baroni, S. de Gironcoli, A. Dal Corso, P. Giannozzi, Rev. Mod. Phys. 73, 515, 2001
Porting to GPU carried out jointly with CINECA HPC Staff

PHonon on GPU

Vibrational and optical properties with DFPT

relax/
vc-relax

↓

scf

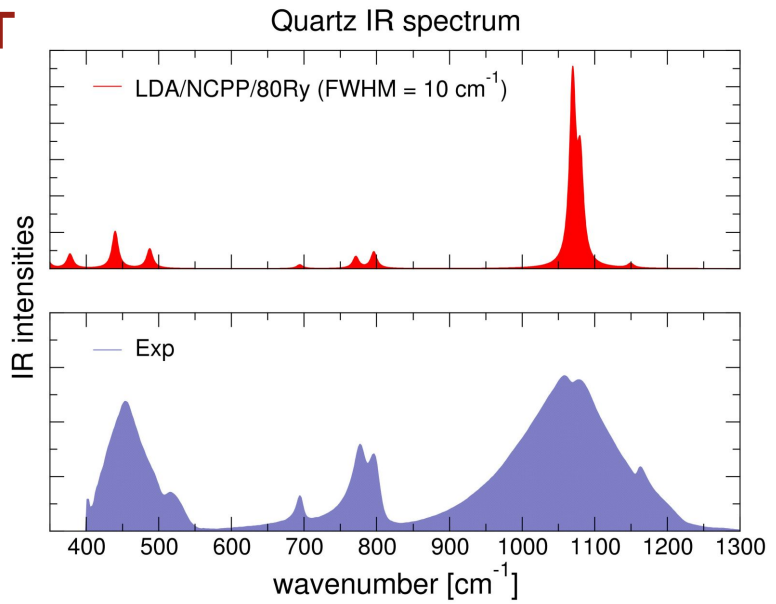
↓

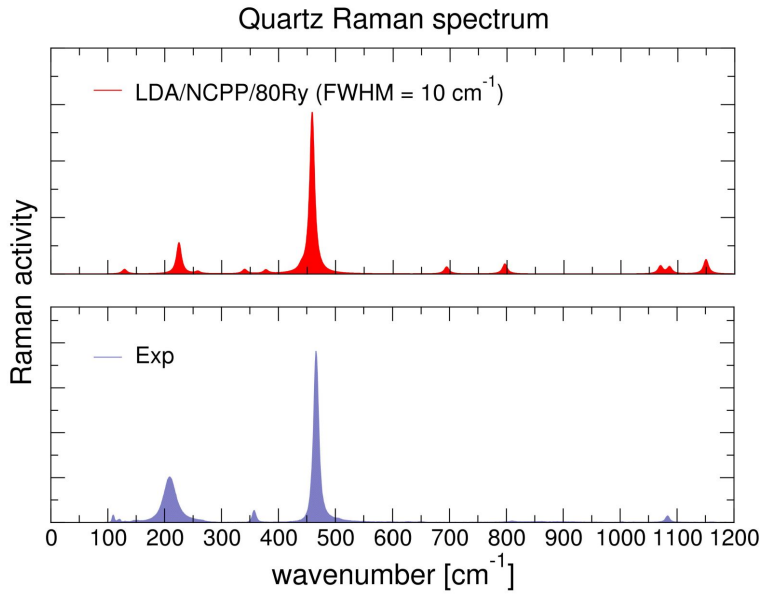
Phonon
(+recover)

↓

dynmat

Phonon code is based on the solution of Sternheimer (CPHF) equations, and provides vibrational properties, IR and Raman spectra, phonon dispersion





PHonon on GPU

Vibrational and optical properties with DFPT

relax/
vc-relax

↓

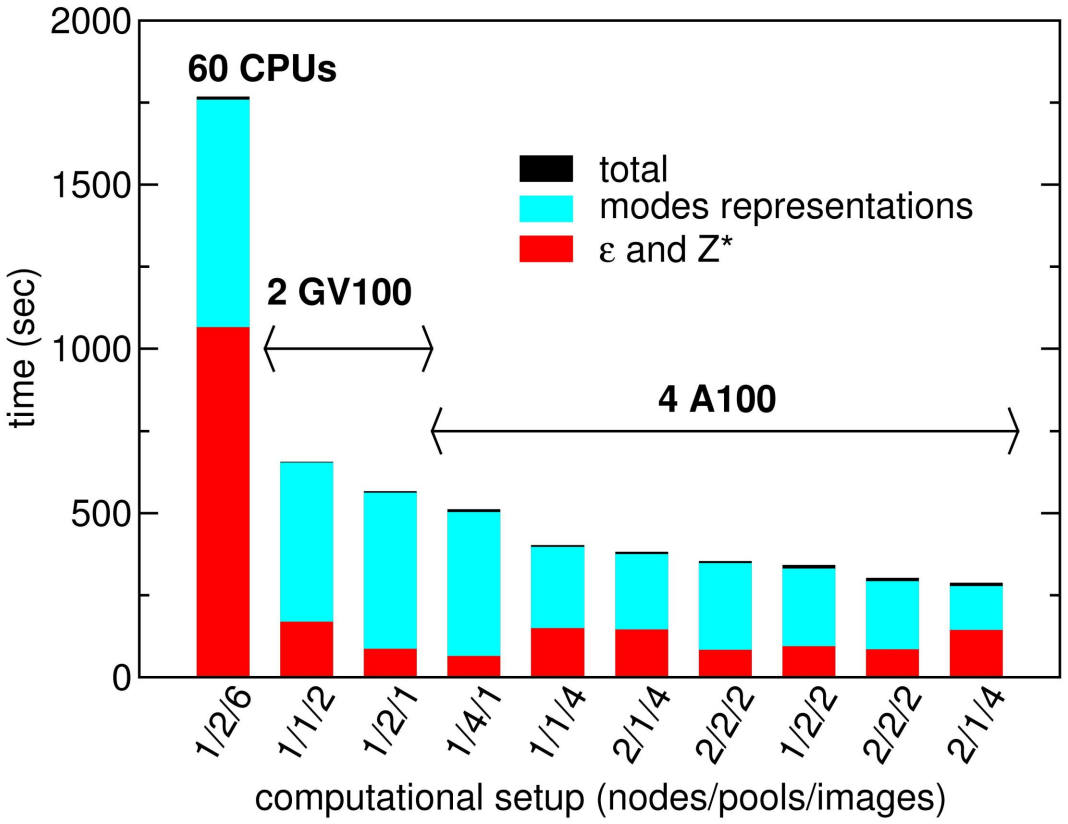
scf

↓

Phonon
(+recover)

↓

dynmat



Calculations on A100 performed during the NVIDIA UK Hackathon, Feb-Mar 2022

The **total computational time** for Phonon calculations is almost entirely due to the solution of one set of Sternheimer (CPHF) equations for each perturbation, i.e. **electric field components** and **nuclei displacements**

PHonon on GPU

Vibrational and optical properties with DFPT

relax/
vc-relax

↓

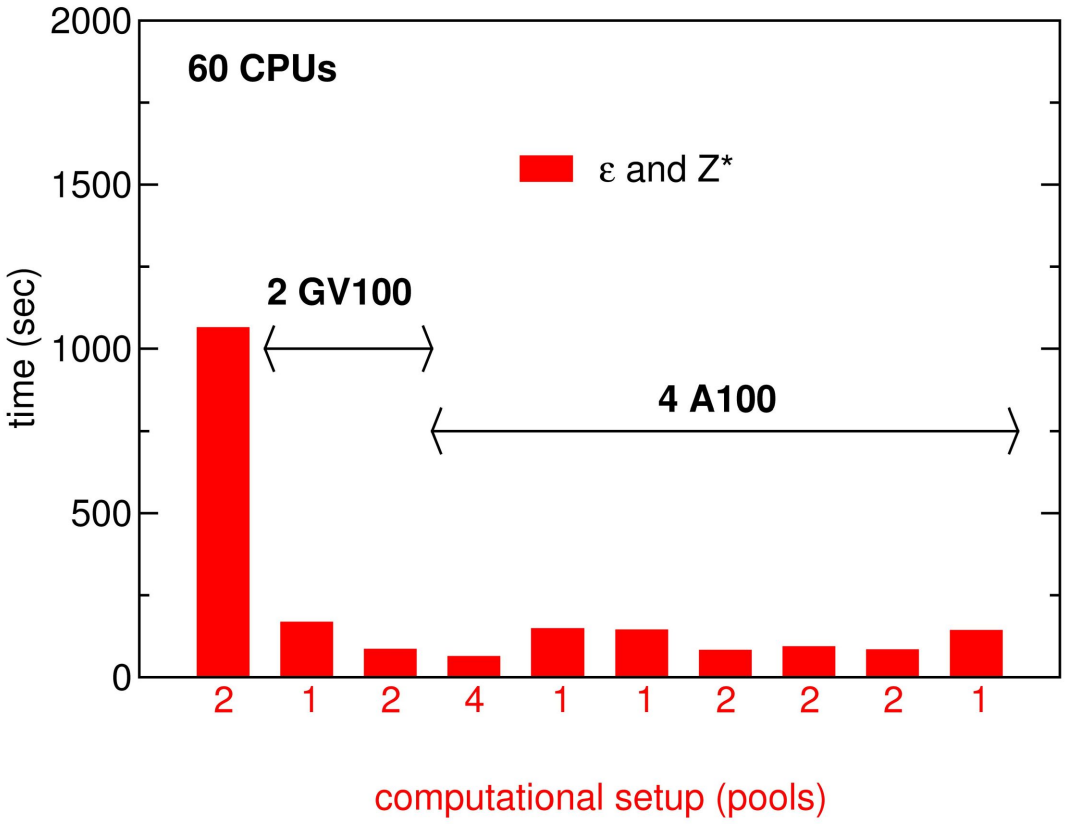
scf

↓

Phonon
(+recover)

↓

dynmat



Pools (k-points) parallelization is very effective for the **electric field components**

PHonon on GPU

Vibrational and optical properties with DFPT

relax/
vc-relax

↓

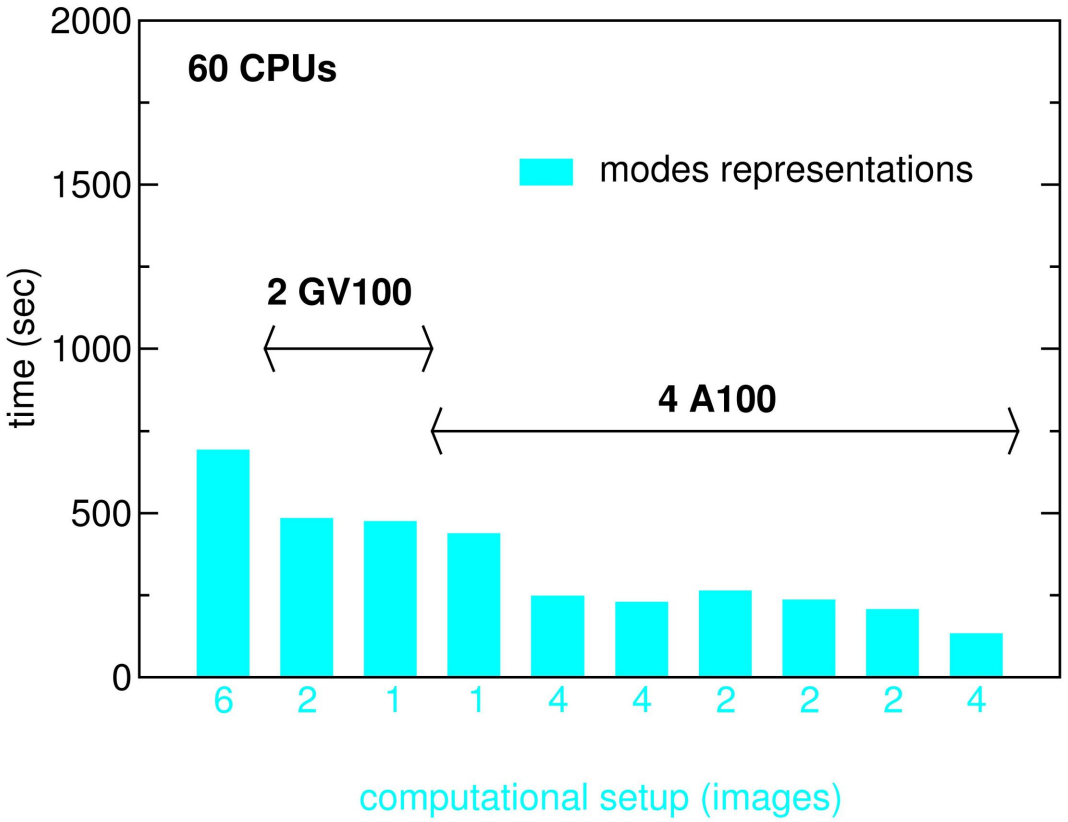
scf

↓

Phonon
(+recover)

↓

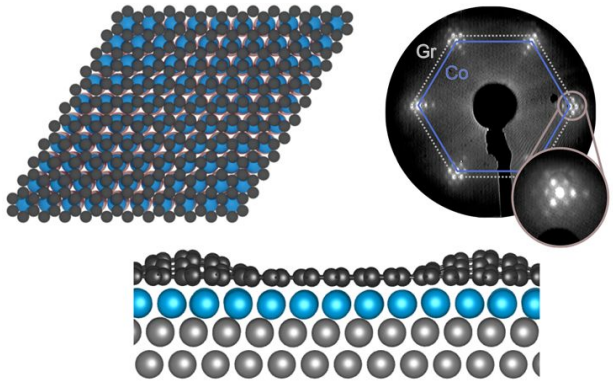
dynmat



The normal modes can be also distributed among tasks as images, so that each image of processors solves the Sternheimer equations for a subset of irreducible representations of **nuclei displacements**

TDDFPT on GPU

TDDFPT codes (turbo_lanczos, turbo_eels) compute the response of the system to time dependent perturbations, giving access to electronic optical properties, electronic scattering properties, useful for various types of spectroscopies



$$\left[\hat{H}^{\circ}_{\mathbf{k}+\mathbf{q}} - (\varepsilon_{\mathbf{k},i} - \omega)\hat{S}_{\mathbf{k}+\mathbf{q}}\right] |\hat{P}_c^{\mathbf{k}+\mathbf{q}} u'^{*}_{-\mathbf{k}-\mathbf{q}}(-\omega)\rangle = -\hat{P}_c^{\dagger\,\mathbf{k}+\mathbf{q}}\,\hat{v}'_{\mathbf{q}}(\omega)|u^{\circ}_{\mathbf{k},i}\rangle$$

These codes can be run using either **Sternheimer** or **Lanczos based algotihms**

O. Motornyi, N. Vast, I. Timrov, O. Baseggio, S. Baroni, A. Dal Corso, Phys. Rev. B 102, 035156 (2020)
O. Motornyi et al 2018 J. Phys.: Conf. Ser. 1136 012008
Porting to GPU carried out jointly with CINECA HPC Staff

TDDFPT on GPU

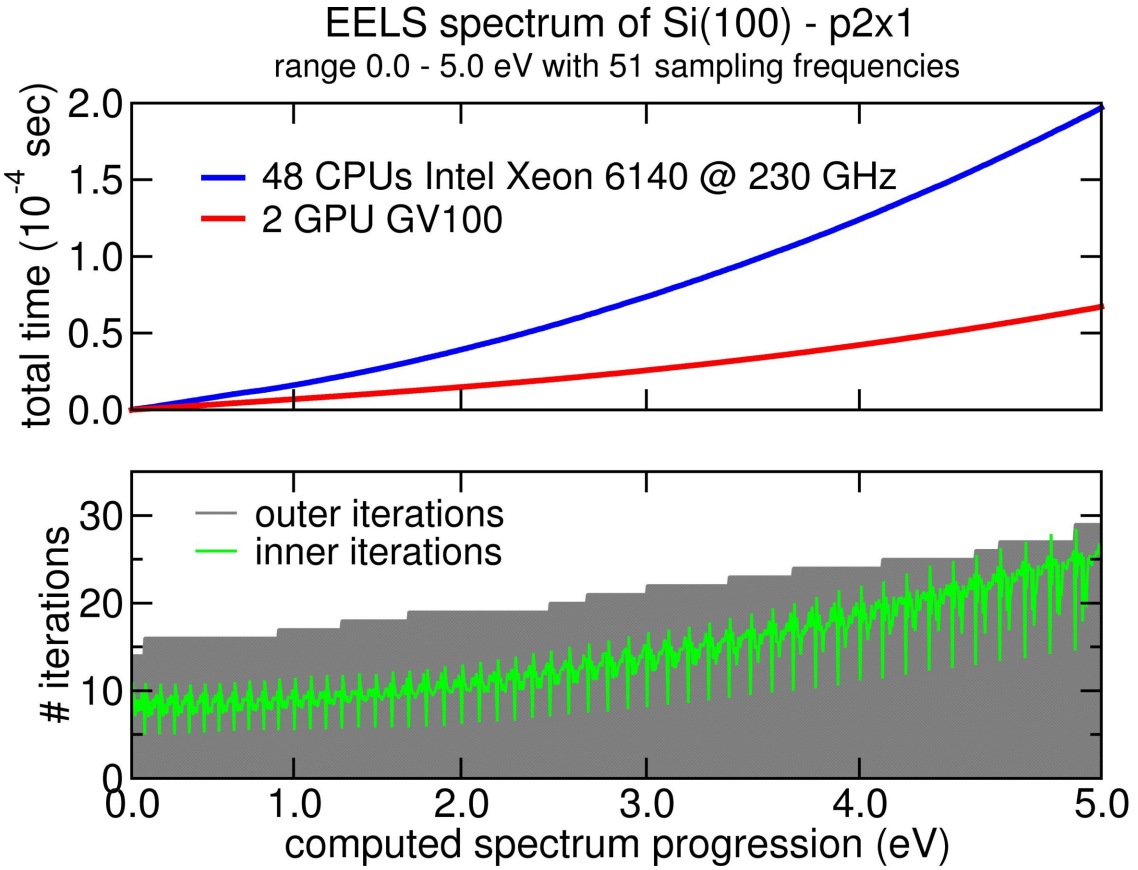
Electronic properties with TD-DFPT

relax/
vc-relax

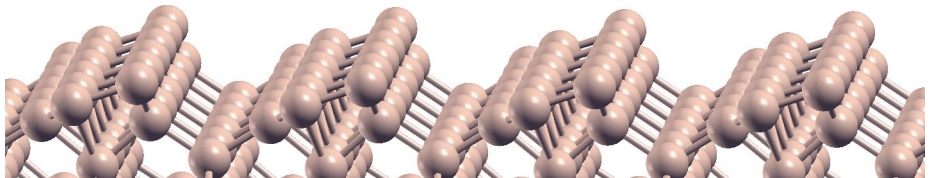
↓

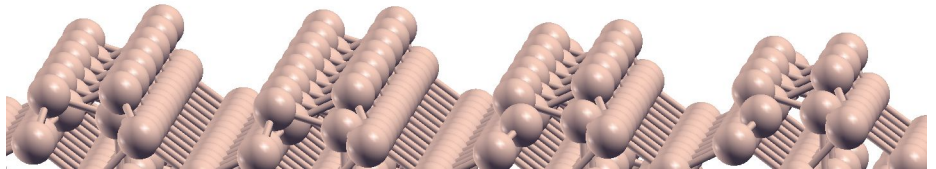
eels/
lanczos

The GPU acceleration allows to compute the EELS spectrum in a fraction of the time of the CPU code

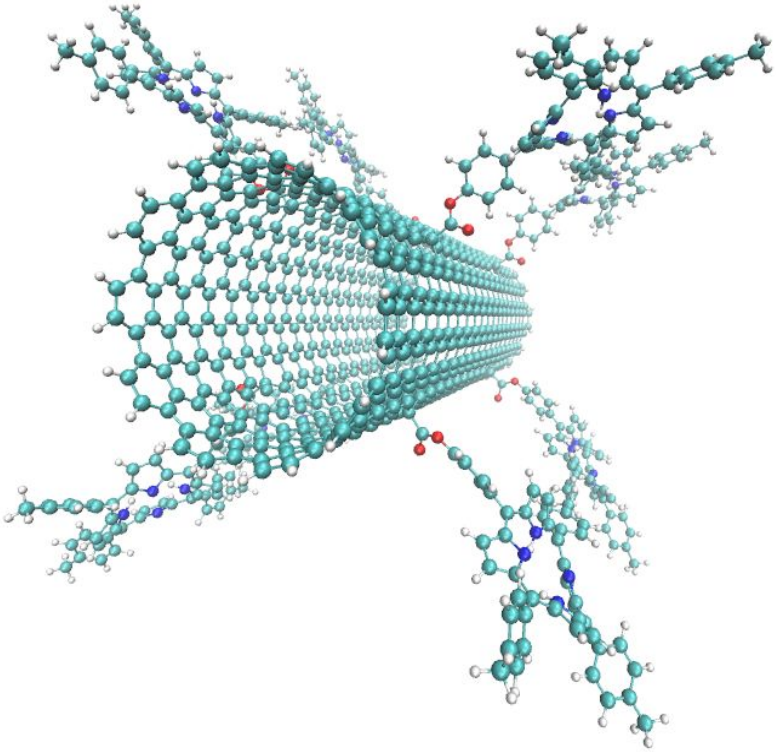


Si(100) – p2x1 and p2x2 reconstructions

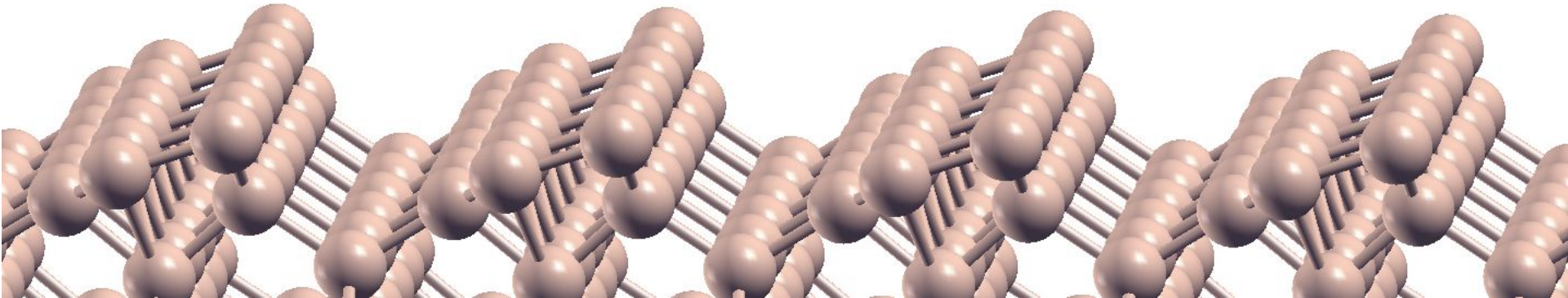




Conclusions and perspectives



- The current diverse panorama of GPU hardware poses a challenge for code portability
- We will continue to reduce the CUF directives in favor of OpenACC, also in view of an ever greater integration of OpenMP
- Better data management and code reorganization



Acknowledgements

QUANTUM ESPRESSO development group

- Pietro Delugas, SISSA
- Ivan Carnimeo, SISSA
- Fabrizio Ferrari Ruffino, CNR-IOM
- Oscar Baseggio, SISSA
- Elena De Paoli, CNR-IOM
- Stefano Baroni, SISSA, CNR-IOM
- Paolo Giannozzi, UniUD, CNR-IOM

CINECA

- Fabio Affinito, CINECA
- Sergio Orlandini, CINECA
- Laura Bellentani, CINECA

QUANTUM ESPRESSO Foundation

- Francesca Garofalo, QEF
- Enrico Siagri, QEF

Nvidia (technical support)

- Filippo Spiga, Nvidia
- Louis Stuber, Nvidia

Intel (technical support)

- Giacomo Rossi, Intel