ENCCS-MaX School 2024

Overview of QUANTUM ESPRESSO suite of codes and main features

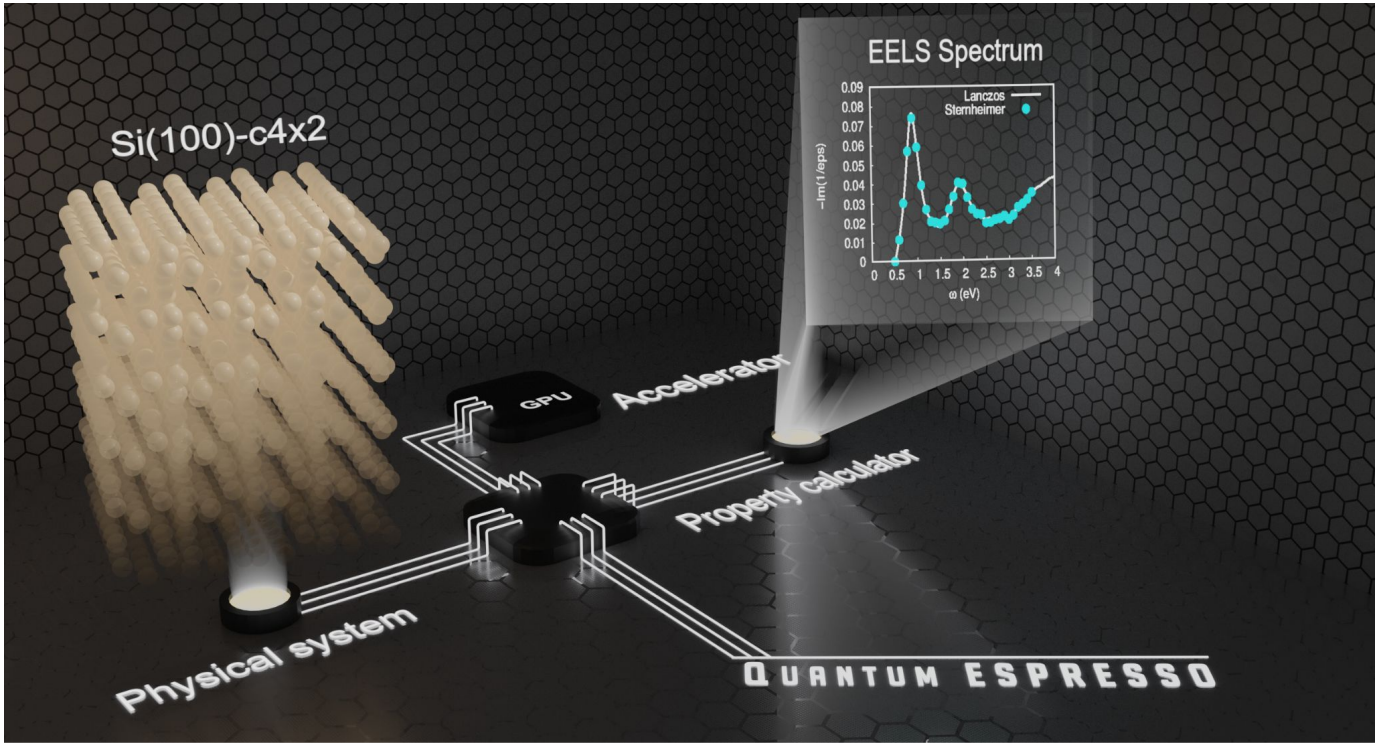
Pietro Delugas
Ivan Carnimeo

Scuola Internazionale Superiore di Studi Avanzati (SISSA)

Outline

1. QUANTUM ESPRESSO
   a. The project
   b. The Suite
   c. What QE do
2. Porting to heterogeneous architecture
3. Overall performance of the subprograms

The QUANTUM ESPRESSO project



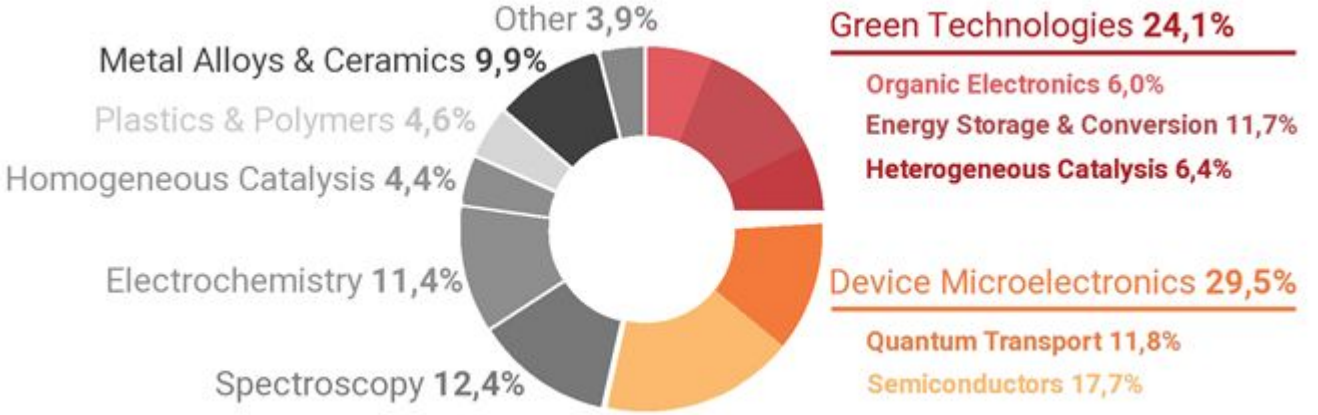
QUANTUM ESPRESSO is an integrated suite of **Open-Source** computer codes for electronic-structure calculations and materials modeling at the nanoscale. It is based on density-functional theory, plane waves, and pseudopotentials.

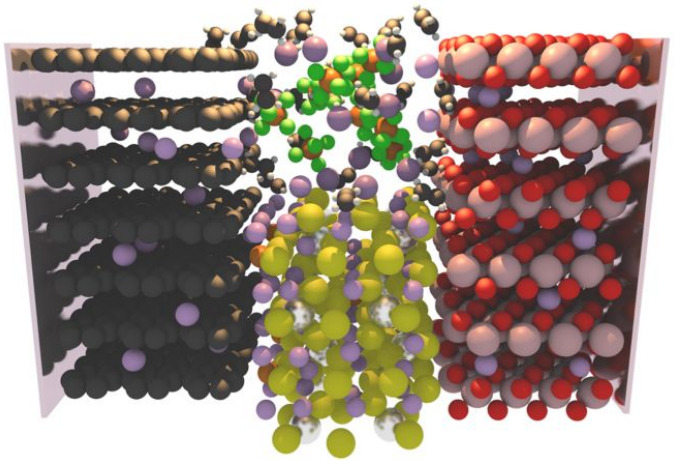
The QUANTUM ESPRESSO project

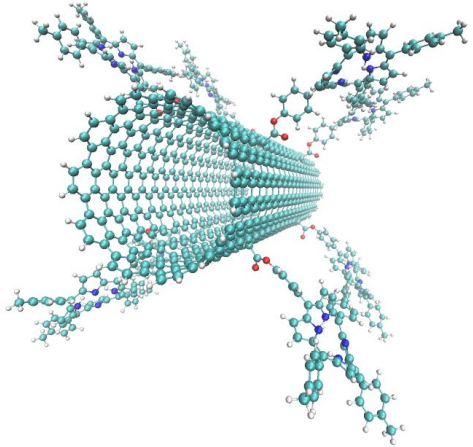
- QUANTUM ESPRESSO is Quantum op**E**n-**S**ource **P**ackage for **R**esearch in **E**lectronic **S**tructure **S**imulation and **O**ptimization
- exists since 2002, resulting from the merge of pre-existing packages; some core components have been under development for ~ 30 years
  - **PWscf** and **PHonon** (Baroni, De Gironcoli, Dal Corso, Giannozzi and others ...)
  - **CP/FPMD** (Pasquarello, Laasonen, Trave, Car, Marzari, Cavazzoni, Scandolo and others ...)
- Main goals of QUANTUM ESPRESSO are:
  - **innovation** in theoretical methods and numerical algorithms
  - **efficiency** on modern computer architectures

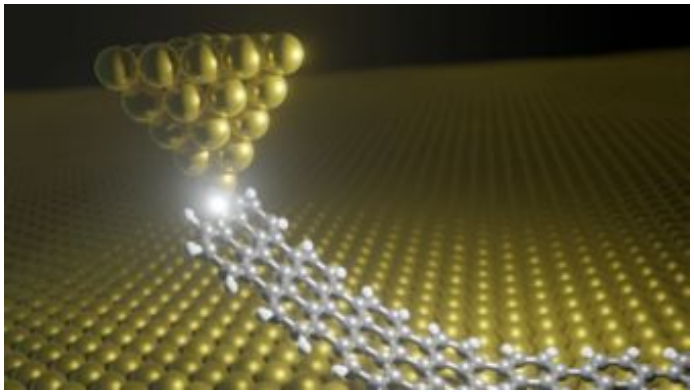
The QUANTUM ESPRESSO project

QUANTUI **complex**
atomisti cial role to
enhance





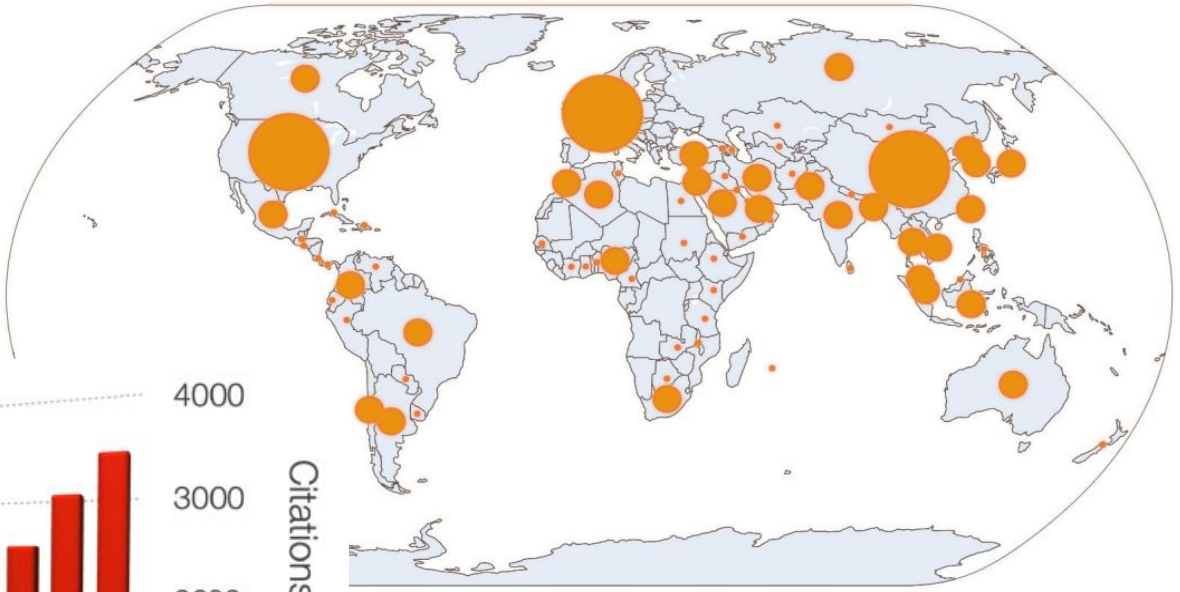




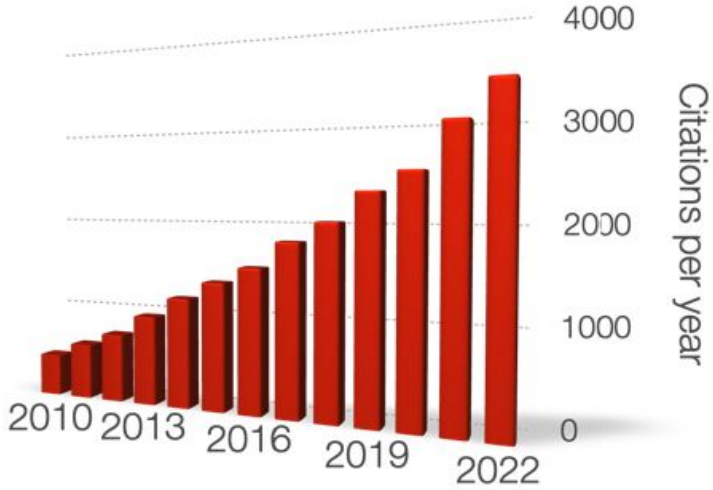
MAX DRIVING THE EXASCALE TRANSITION

The QUANTUM ESPRESSO project

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



Geographic distribution of the authors of articles citing the main reference articles as QUANTUM ESPRESSO

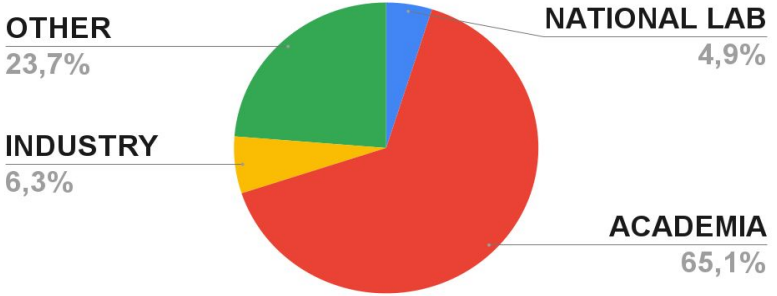


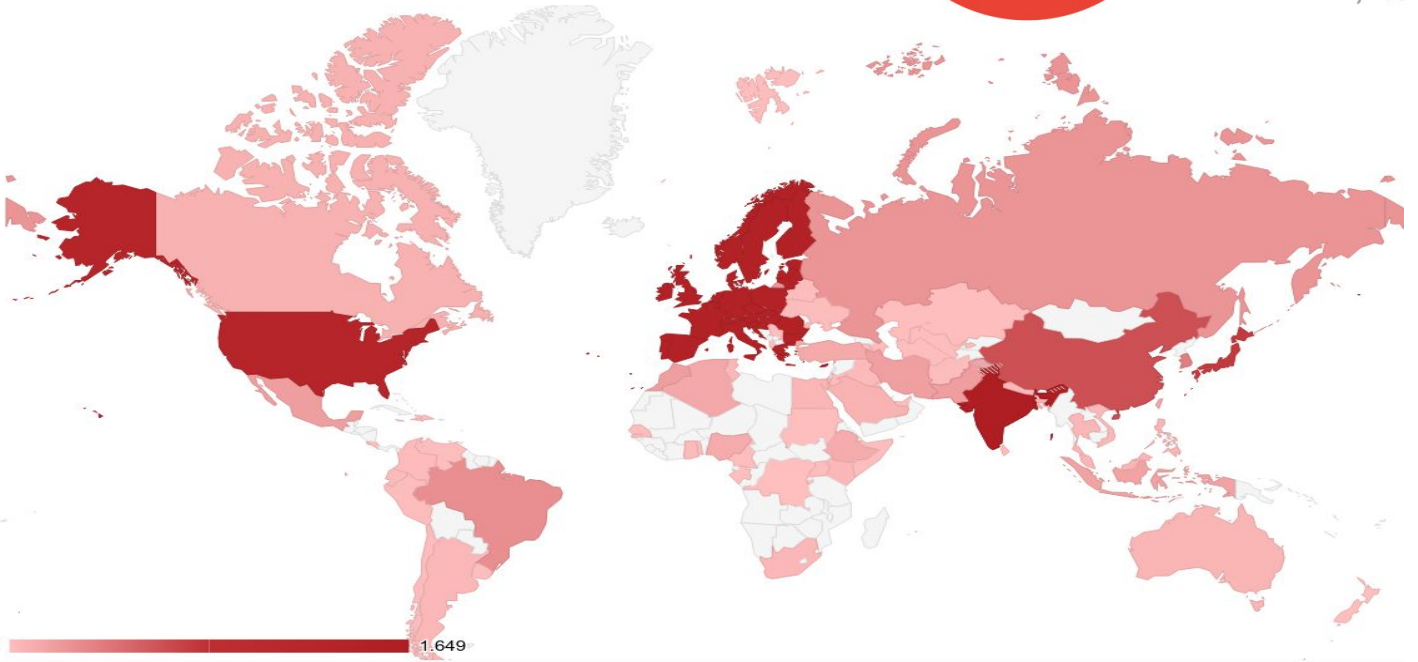
Data provided by courtesy of the QUANTUM ESPRESSO Foundation

The QUANTUM ESPRESSO project

Geographic distribution of downloads from the QE website since the beginning of 2022

35000+ downloads
20000+ citations
50000+ authors





Data provided by courtesy of the QUANTUM ESPRESSO Foundation

The QUANTUM ESPRESSO project

Several software houses and service providers include QUANTUM ESPRESSO in their proprietary packages or in the services they provide

MATERYS
SCIENOMICS delivering science
materials square
SCM
MATERA
Schrödinger
rescale

QUANTUM ESPRESSO is being used for validation and co-design by major hardware vendors

intel
NVIDIA
AMD
arm

Leading companies world-wide use QUANTUM ESPRESSO as a research tool

BOSCH LOCKHEED MARTIN TOYOTA DELL hp Eni HUAWEI TotalEnergies mazda lenovo KOBELCO CLARIANT BRITISHVOLT kao Vulcaflex
HITACHI DAIKIN Nikon Panasonic SAMSUNG SONY TOSHIBA NISSAN FUJIFILM JM Nippon Electric Glass JEOL idener NASA
FUJITSU TDK Western Digital RICOH DANIELI NEC KSOE xerox Terramera Western Digital Thermo-Calc Software and many others …

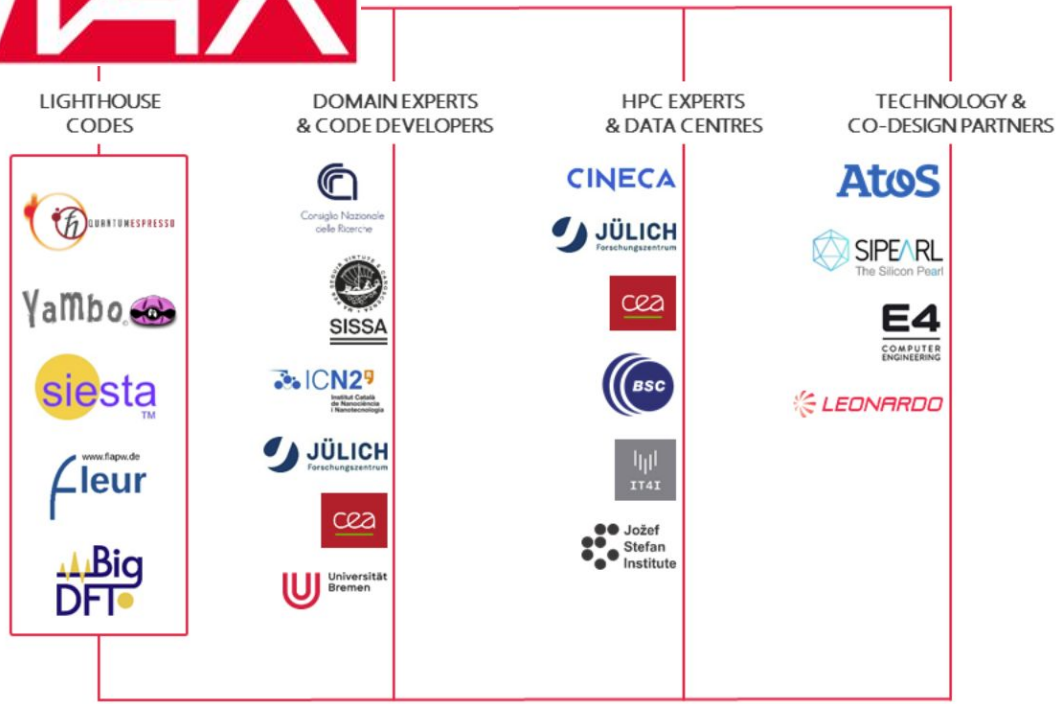
SISSA

SISSA is probably the major funder of the QUANTUM ESPRESSO project, thanks to the contributions of its research staff

Materials design at the Exascale

MAX

CoE for HPC applications in materials science



exploit **frontier HPC** for material science research in strong link with **scientific communities**

CODE PORTING

↓

CO-DESIGN

↓

HTC ECOSYSTEM

http://www.max-centre.eu/

QUANTUMESPRESSO

MAX DRIVING THE EXASCALE TRANSITION

Spoke 7 – Flagship codes

QUANTUMESPRESSO

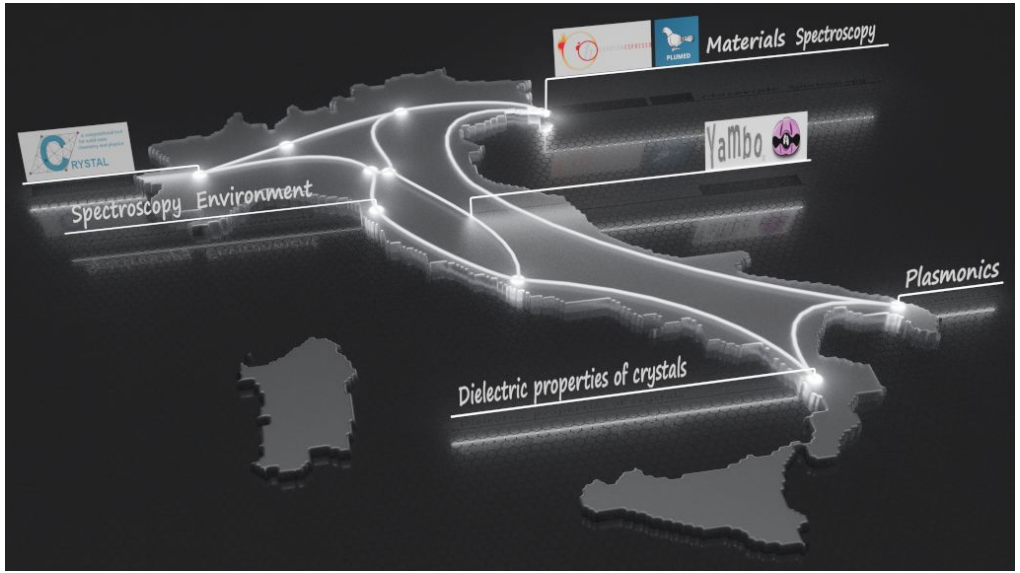
YaMBo

a computational tool for solid state chemistry and physics
CRYSTAL

PLUMED

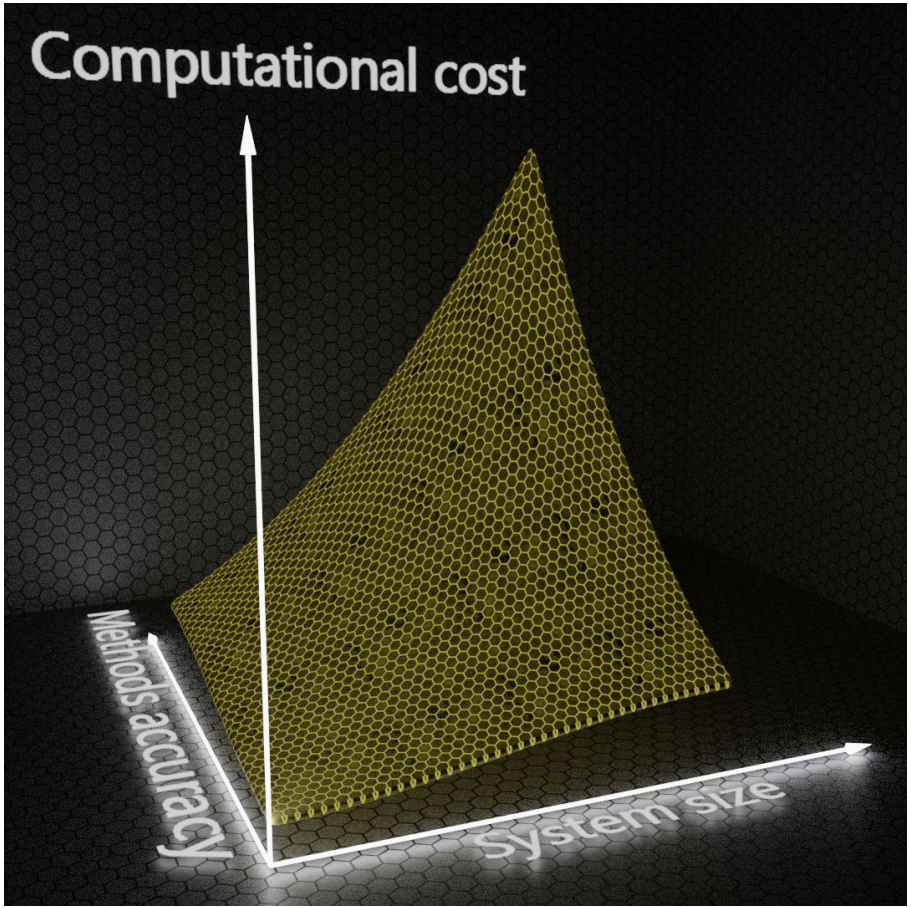
OpenMMPol

LibOpt



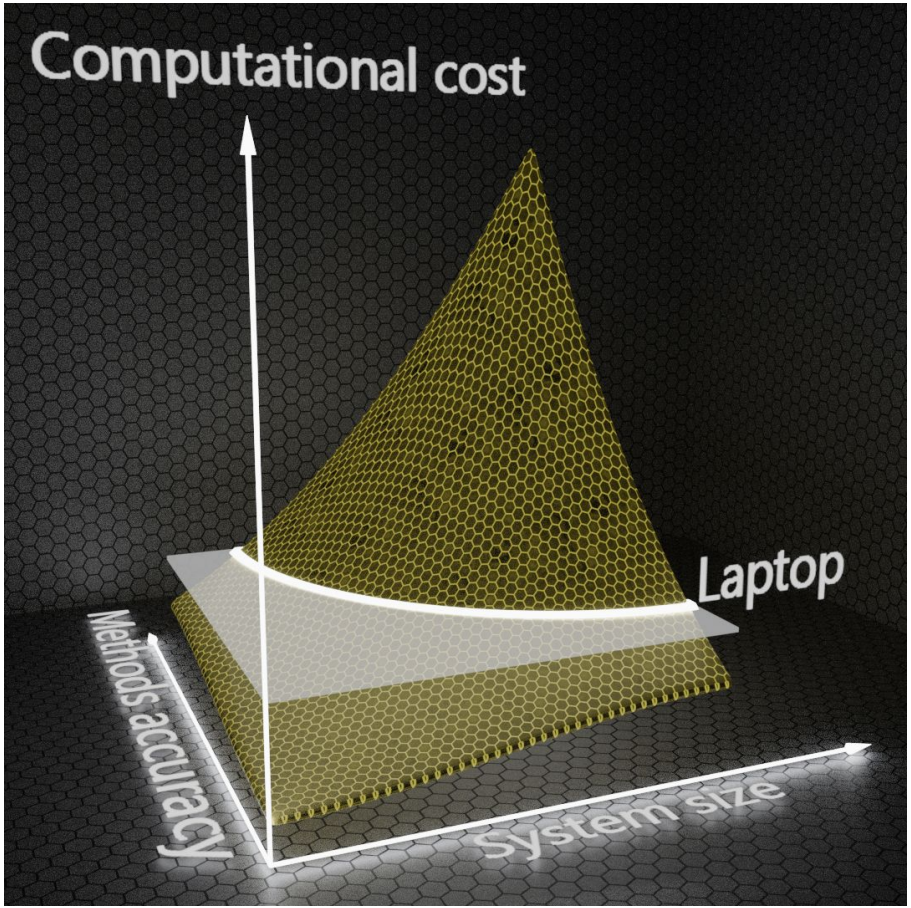
ICSC
Centro Nazionale HPC,
Big Data e Quantum Computing

Why exascale



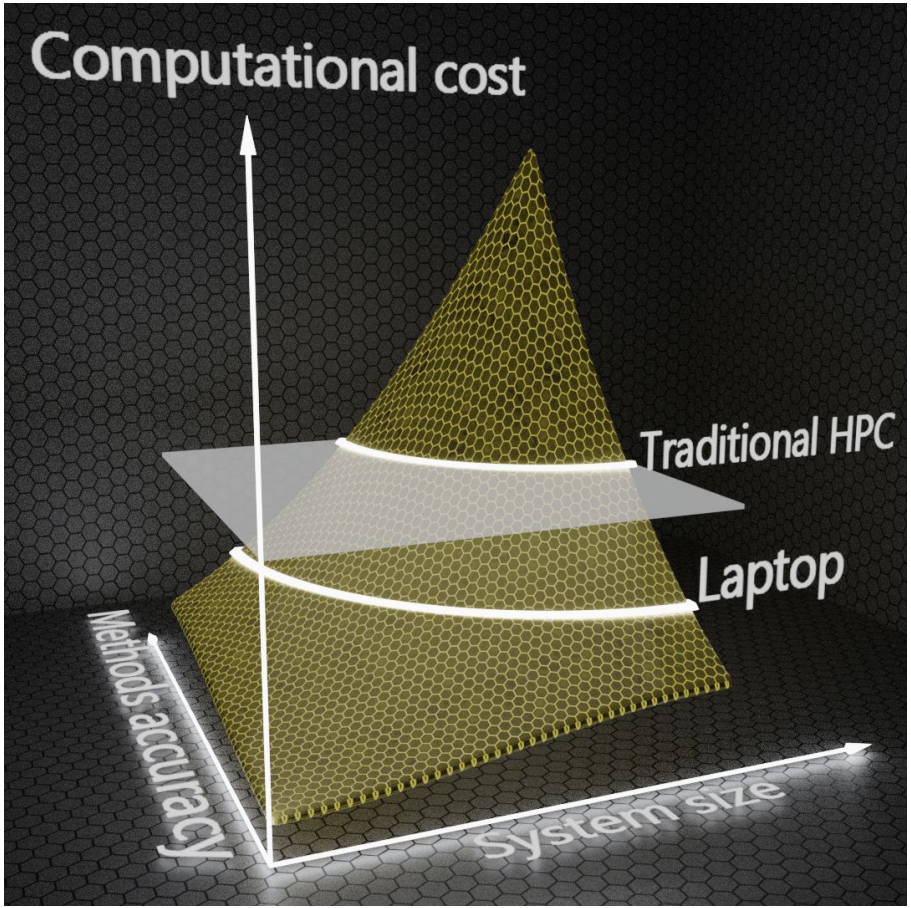
HPC resources play an important role in order to give code users and developers the freedom to choose proper levels of theory for simulations of molecules and materials

Why exascale



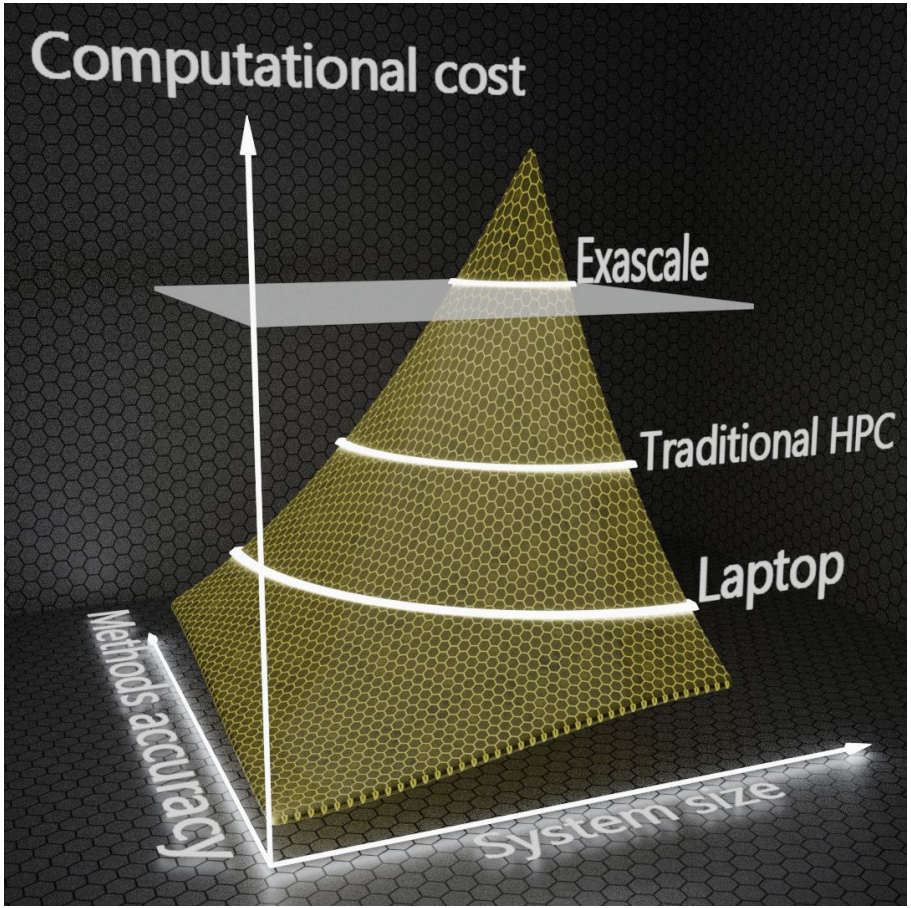
HPC resources play an important role in order to give code users and developers the freedom to choose proper levels of theory for simulations of molecules and materials

Why exascale



HPC resources play an important role in order to give code users and developers the freedom to choose proper levels of theory for simulations of molecules and materials

Why exascale



HPC resources play an important role in order to give code users and developers the freedom to choose proper levels of theory for simulations of molecules and materials

The QUANTUM ESPRESSO Suite

Libraries

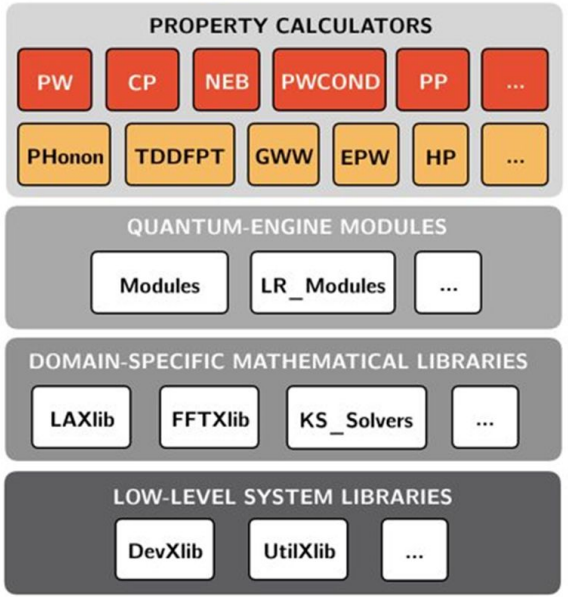
- Encapsulated and reusable
- Included in MaX Libraries

Modules

- Encapsulated and self-contained. Share part of the QE data structure

Applications

- Quantum Engines
- Property calculators



3 Releases

- qe-7.2 (July 2023)
- qe-7.3 (Dec 2023, D1.2)
- qe-7.3.1 (Feb 2024)

The QUANTUM ESPRESSO Applications

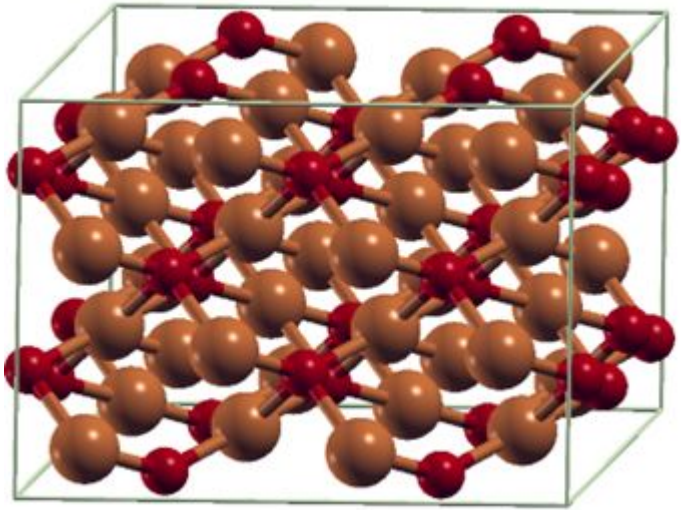
- **PWscf**:
  - Total energies, forces stresses using plane waves + pseudopotentials or PAW
  - L(S)DA , GGA , metaGGA and many other advanced functionals
  - collinear and noncollinear spin density
  - much more ...
- **CP** - Car-Parrinello molecular dynamics
- **NEB** – Nudged elastic band
- **PHonon**: vibrational frequencies, dielectric response, anharmonic terms and many more with linear response
- **HP** - Computes Hubbard parameters with linear response
- **TDDFPT**: Optical spectra and collective excitations.
- **QEHeat**: energy flux.
- **PP**: post-processing utilities
- **EPW**: electron phonon with Wannier Function
- **KCW**: Koopmans-Compliant band structure calculations with Wannier Functions
- Interoperability with many other external packages ...

PW:

- SCF calculations:
  - spin-degenerate, spin-polarized, noncollinear (+SOC)
  - Standard DFT functionals
  - metaGGA
  - Non-local functionals
  - Hubbard (onsite and intersite) corrections
  - Hybrid Functionals
- Forces and Stress
  - Relaxations
  - NEB
  - Molecular dynamics
- Isolated systems
  - Clusters and Surfaces
  - Dipole corrections for 2D system
- Electric Fields
- …

Input of PW

- The input is organized in a group of Fortran namelists + some formatted cards



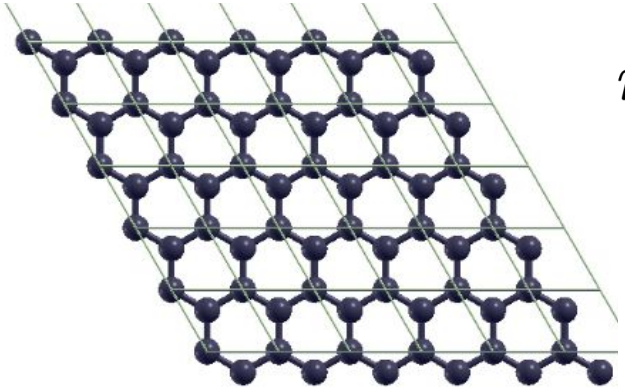
```
&control
    prefix='pwscf'
    calculation='scf',
    restart_mode = 'from_scratch'
    outdir='./out'
    pseudo-dir='pseudo'
    outdir='./'
 /
 &system
    ibrav = 0, nat=64, ntyp=2,
    ecutwfc = 35,
    ecutrho = 350,
    smearing='mp',
    occupations='smearing',
    degauss=0.01,
    nspin=2,
    starting_magnetization(1)=0.0,
    starting_magnetization(2)=0.5,
 /
 &electrons
    mixing_beta = 0.5,
    conv_thr =  1.0d-7,
    startingpot='atomic',
    startingwfc='atomic',
    electron_maxstep=4
 /

 &ions
 /

 &cell
    cell_dynamics = bfgs,
    press = 0
 /

ATOMIC_SPECIES
  O  1.0 o_pbe_v1.2.uspp.F.UPF
  Cu 1.0 cu_pbe_v1.2.uspp.F.UPF

CELL_PARAMETERS angstrom
        9.3660001755          0.0000000000          0.0000000000
        0.0000000000          6.8406000137          0.0000000000
       -1.7280819494          0.0000000000         10.1022633844

ATOMIC_POSITIONS crystal
    O  0.000000000          0.208500000          0.125000000
    O  0.000000000          0.791500000          0.875000000
    O  0.000000000          0.791500000          0.375000000
    O  0.250000000          0.458500000          0.125000000
    O  0.250000000          0.041500000          0.875000000
```

Periodicity, plane wave basis and 3D grids:

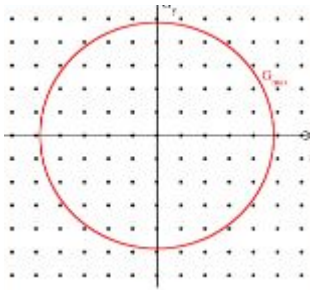
- Bravais and reciprocal lattices:

$$\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_3} \qquad \mathbf{b_i} = \frac{2\pi}{\Omega} \sum_{ijk} \epsilon_{ijk} \mathbf{a_j} \wedge \mathbf{a_k}$$

- real space grid: $\mathbf{r_n} = \sum_{i=1}^{3} \frac{n_i}{N_i} \mathbf{a}_i \quad with \quad n_i = 0, 1, 2, \ldots, N_i - 1$
- reciprocal space: $\mathbf{G_m} = \sum_{i=1}^{3} m_i \mathbf{b}_i \quad \text{with} \quad m_i = -\frac{N_i+1}{2}, \ldots, 0, \ldots, \frac{N_i}{2}$



$$m_i^{max} |\mathbf{b}_i| < 2\sqrt{E_{cut}}$$

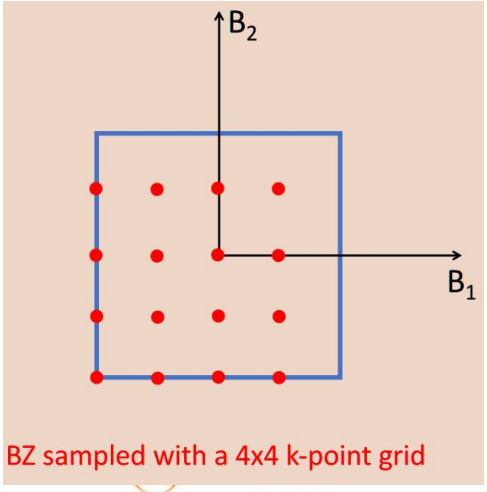


K points sampling:

```
K_POINTS automatic
2 2 2 0 0 0
```

Bloch Theorem: $\psi_{k,n}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{k,n}(\mathbf{r})$ → periodic function

We solve Kohn Sham equations only for a finite mesh or symmetric set of k-points and then sum approximating the integral.



BZ sampled with a 4x4 k-point grid

The Pseudopotentials:

- QE applications use pseudotentials in the UPF format.
- Possible to use Norm-Conserving and for most functionalities Ultra-Soft and PAW.
- Places where to find pseudos:
  - SSSP(https://www.materialscloud.org/discover/sssp/table/efficiency)
  - Pslibrary (https://dalcorso.github.io/pslibrary/)
  - PseudoDojo (http://www.pseudo-dojo.org/index.html)

Other links and Pslibrary tables on
http://pseudopotentials.quantum-espresso.org/

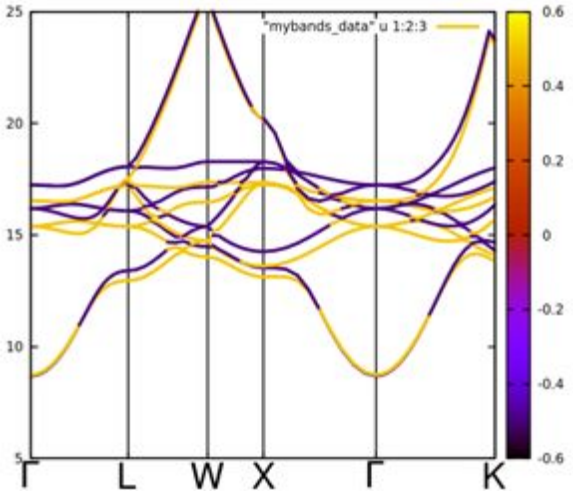
Output of PW

- Standard output: contains all results, information and warnings. No strict rule for the format.
- `prefix.save` directory contains:
  - XML data file with all results and information to reproduce/restart the run
  - charge density file (either `.dat` of `.hdf5`)
  - wave functions' files -one per k-point and spin- (either `.dat` or `.hdf5`)

Example of workflow: bands

For basic band structure calculations PW functionalities are integrated with those of post-processing utilities, for example for computing Kohn-Sham bands:

1. An SCF calculations with PW:
2. A non-SCF calculation with PW
3. The post-processing utility `bands.x` performs the symmetry analysis and save the data for the plot.



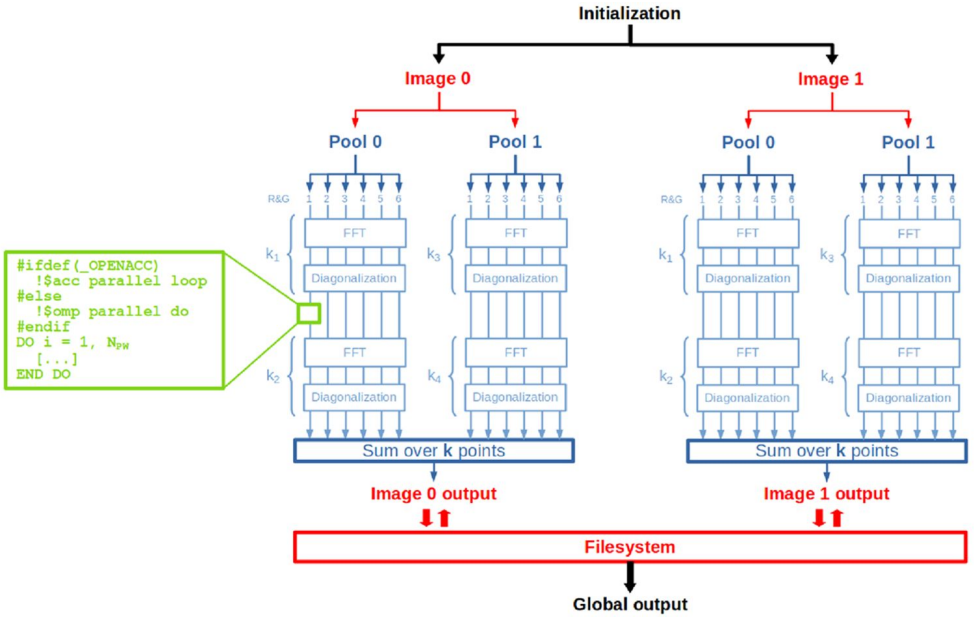
Parallelism

Throughput: low communication

- Images
- Pools
- Several kind of distribution of work on bands

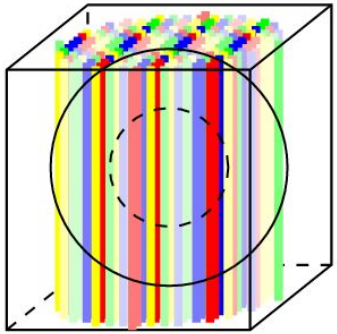
Latency: high communication

- R&G distribution
- GPU offload

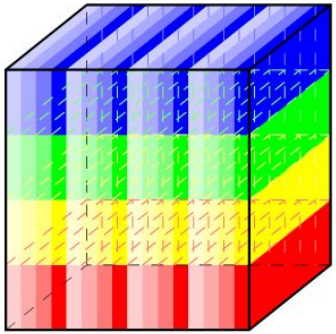


Data Distribution:

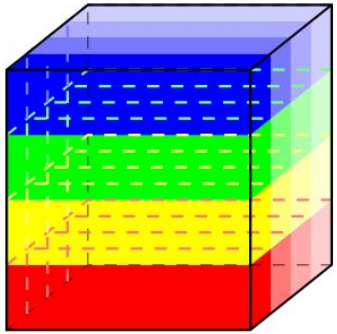
- Data in G space are distributed as sticks in the z directions
- Data in R space are distributed as z plane slices in the y direction
- At any moment a whole 1 dimensional FFT may be done in each process.



$F(G_x,G_y,G_z)$
$F(G_x,G_y,R_z)$



$F(G_x,G_y,G_z)$
$F(G_x,R_y,R_z)$



$F(G_x,G_y,G_z)$
$F(R_x,R_y,R_z)$

More info:

- about QUANTUM ESPRESSO
  - Web: www.quantum-espresso.org
  - Downloads: https://www.quantum-espresso.org/download-page/
  - papers:
    - P. Giannozzi *et al.*, J.Phys.:Condens.Matter **21**, 395502 (2009) http://dx.doi.org/10.1088/0953-8984/21/39/395502
    - P. Giannozzi *et al.*, J.Phys.:Condens.Matter **29**, 465901 (2017) http://iopscience.iop.org/article/10.1088/1361-648X/aa8f79
    - P. Giannozzi et al. J. Chem. Phys. **152**, 154105 (2020); https://doi.org/10.1063/5.0005082
    - I. Carnimeo et al. J. Chem. Theory Comput. 2023, **19**, 6992-7006; https://doi.org/10.1021/acs.jctc.3c00249
- about QEF foundation.quantum-espresso.org
- Development on gitlab.com/QEF/q-e

Thanks !!!

Outline

1. QUANTUM ESPRESSO
   a. The project
   b. The Suite
   c. What QE do
2. Porting to heterogeneous architecture
3. Overall performance of the subprograms

Why exascale

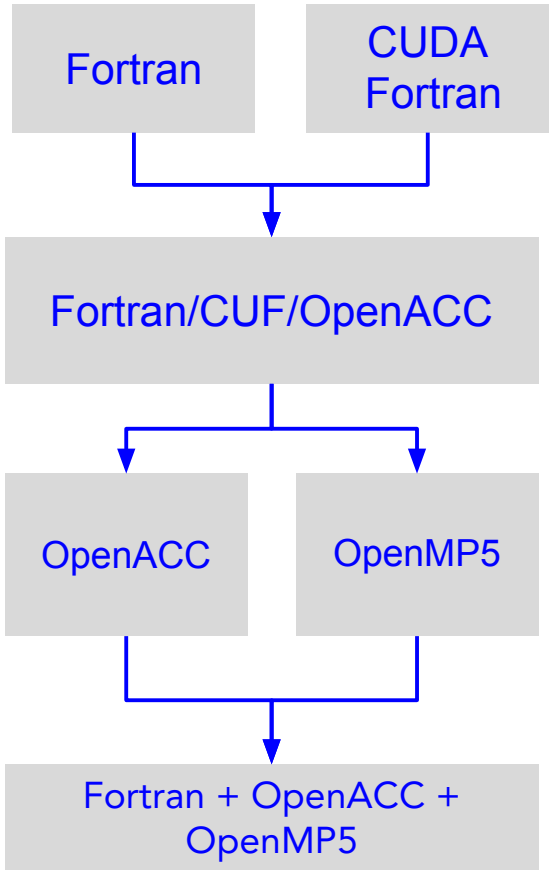
The current strategy for performance portability is to specialize the code to different hardware configurations by using directive based approaches.

Porting to heterogeneous architectures

J. Chem. Phys. **152,** 154105 (2020)

J. Chem. Theory Comput. **19**, 6992 (2023)

QUANTUM ESPRESSO is being used for validation and co-design by major hardware vendors



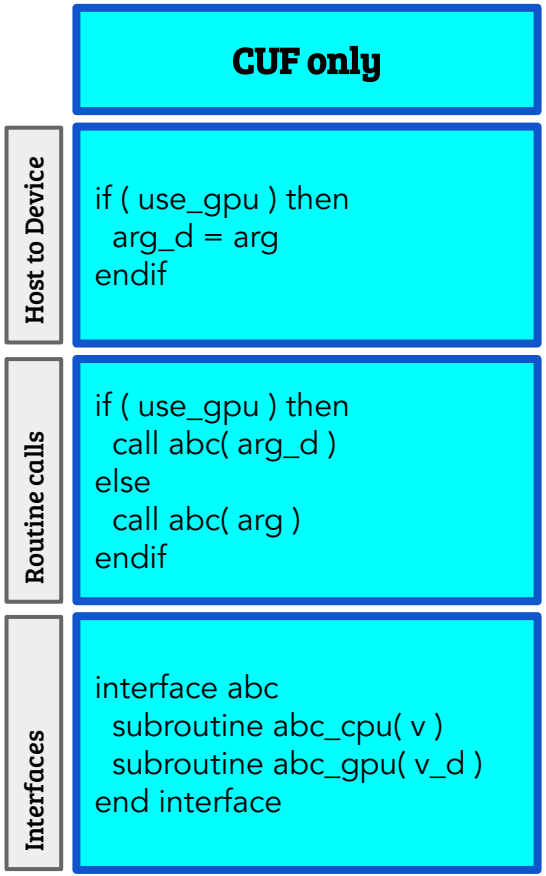
Until qe-v6.8

From qe-v7.0

Under development

Goal

Porting to heterogeneous architectures



Porting to heterogeneous architectures

	CUF only	CUF interfaces OpenACC parent code

Host to Device	if (use_gpu) then
	  arg_d = arg
	endif
		!$acc update device(arg)

Routine calls	if (use_gpu) then
	  call abc(arg_d)
	else
	  call abc(arg)
	endif
		!$acc host_data use_device(arg)
		call abc(arg)
		!$acc end host_data

Interfaces	interface abc
	  subroutine abc_cpu(v)
	  subroutine abc_gpu(v_d)
	end interface	

Porting to heterogeneous architectures

	CUF only	CUF interfaces OpenACC parent code	**OpenACC only**

Host to Device	if (use_gpu) then
	  arg_d = arg
	endif
		!$acc update device(arg)	

Routine calls	if (use_gpu) then
	  call abc(arg_d)
	else
	  call abc(arg)
	endif
		!$acc host_data use_device(arg)
		call abc(arg)
		!$acc end host_data
			call abc_acc(arg)

Interfaces	interface abc
	  subroutine abc_cpu(v)
	  subroutine abc_gpu(v_d)
	end interface
			subroutine abc_acc(v)

Porting to heterogeneous architectures

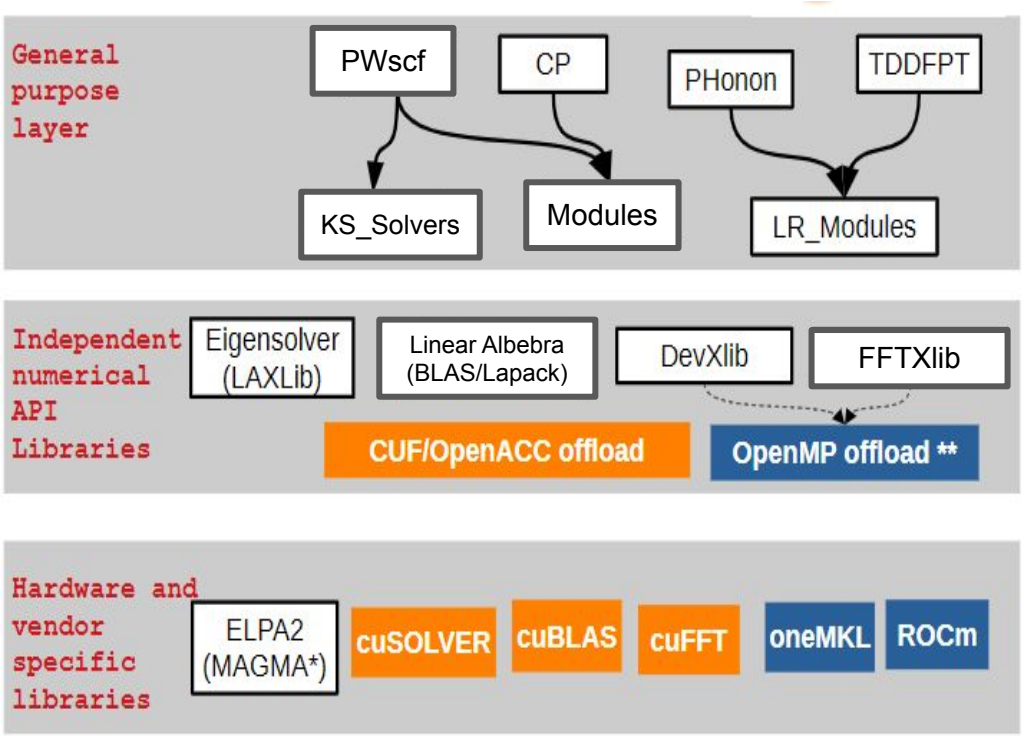
	CUF only	CUF interfaces OpenACC parent code	OpenACC only	**OpenACC + OpenMP5**

Host to Device	if (use_gpu) then
	  arg_d = arg
	endif
		!$acc update device(arg)		!$acc update device(arg)
				!$omp target update to (arg)

Routine calls	if (use_gpu) then
	  call abc(arg_d)
	else
	  call abc(arg)
	endif
		!$acc host_data use_device(arg)
		call abc(arg)
		!$acc end host_data
			call abc_acc(arg)	call abc(arg, offload)

Interfaces	interface abc
	  subroutine abc_cpu(v)
	  subroutine abc_gpu(v_d)
	end interface
			subroutine abc_acc(v)	interface abc
				  subroutine abc_cpu(v,off)
				  subroutine abc_acc(v,off)
				  subroutine abc_omp(v,off)
				end interface

Porting to heterogeneous architectures



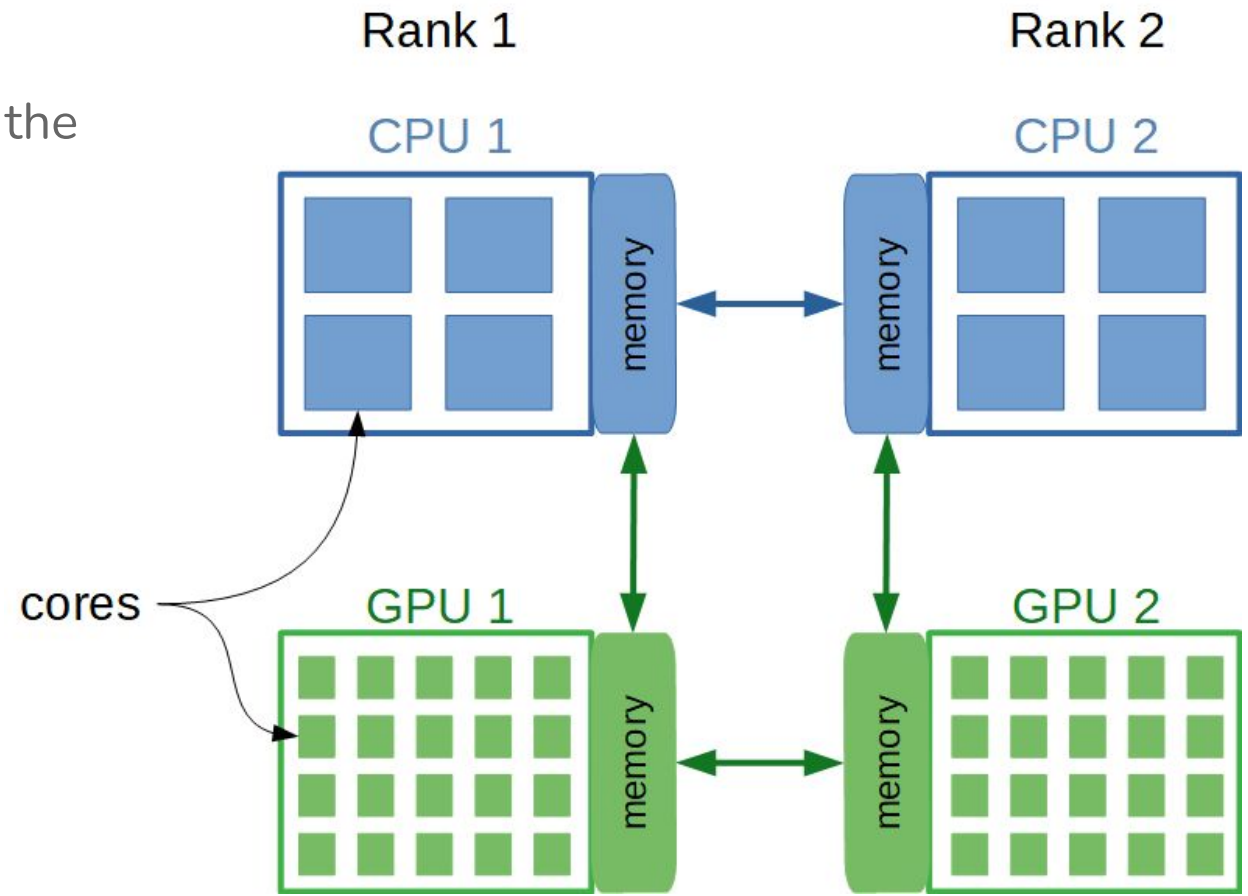
directives → MAINTAINABILITY, PORTABILITY

multiple back-ends → FLEXIBILITY, PERFORMANCE

Modularity enables interoperability and new programming models

Porting to heterogeneous architectures

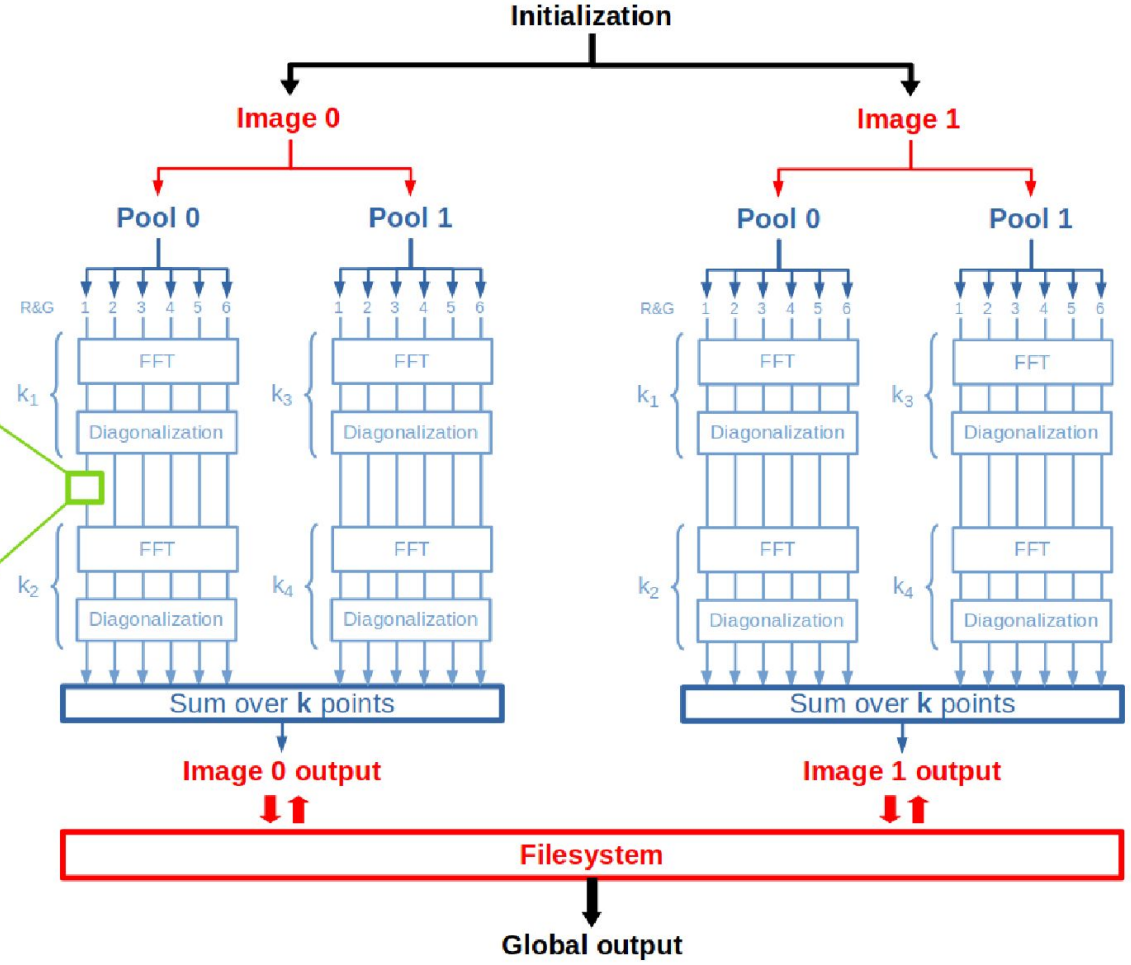
Each MPI process off-loads the calculation workload to the device



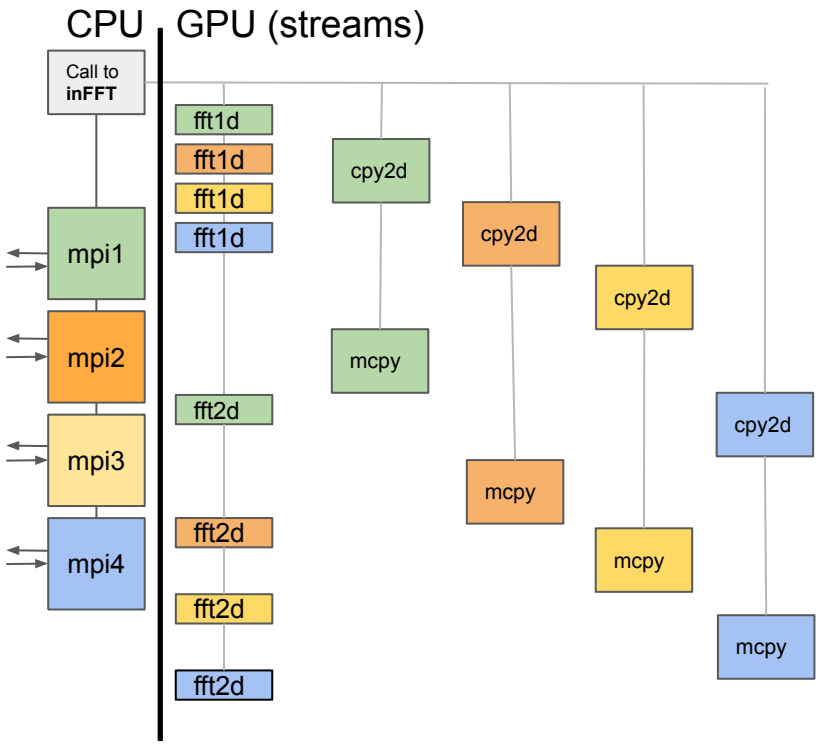
Porting to heterogeneous architectures

Several parallelization schemes are implemented and integrated with GPU offload

```
#ifdef(_OPENACC)
  !$acc parallel loop
#else
  !$omp parallel do
#endif
DO i = 1, N_PW
  [...]
END DO
```



Porting to heterogeneous architectures

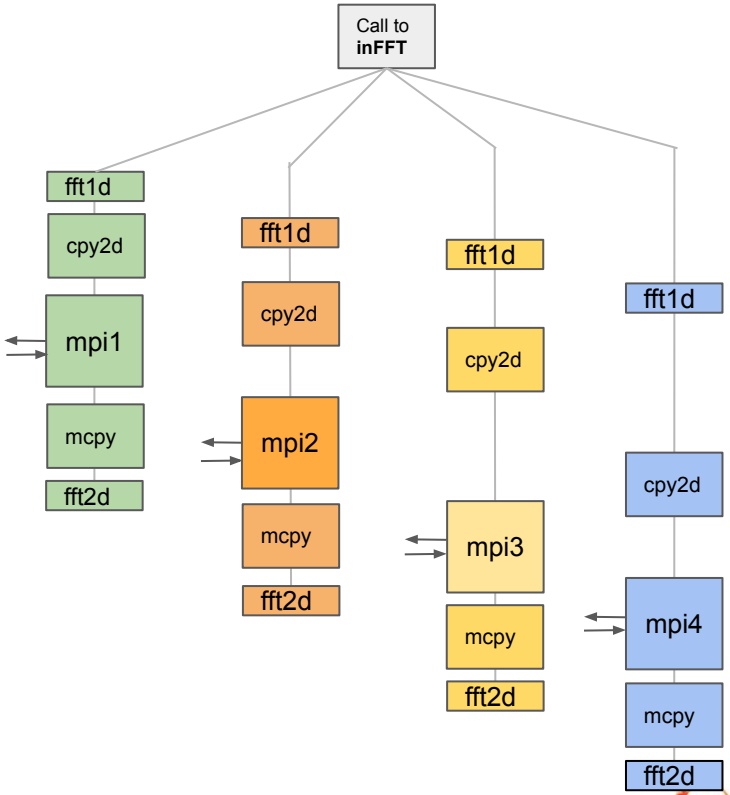


- **Batched** 3d-FFT of the **wave-function**;
- the input array divided in **4 batches** (on bands);
- 1 stream for **FFTs**, 4 streams for **data movements**;
- 4 **async mpi** communications (ISEND, IRECV).

Notes:

- **non-asynchronous memcpy**;
- memcpy operations **D2H/H2D** much more time consuming than FFT calls;
- memcpy operations **D2D** same order of FFT calls.

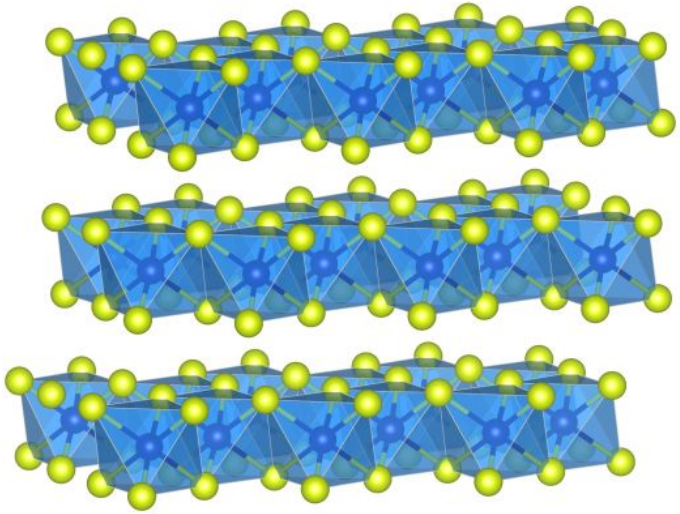
Porting to heterogeneous architectures

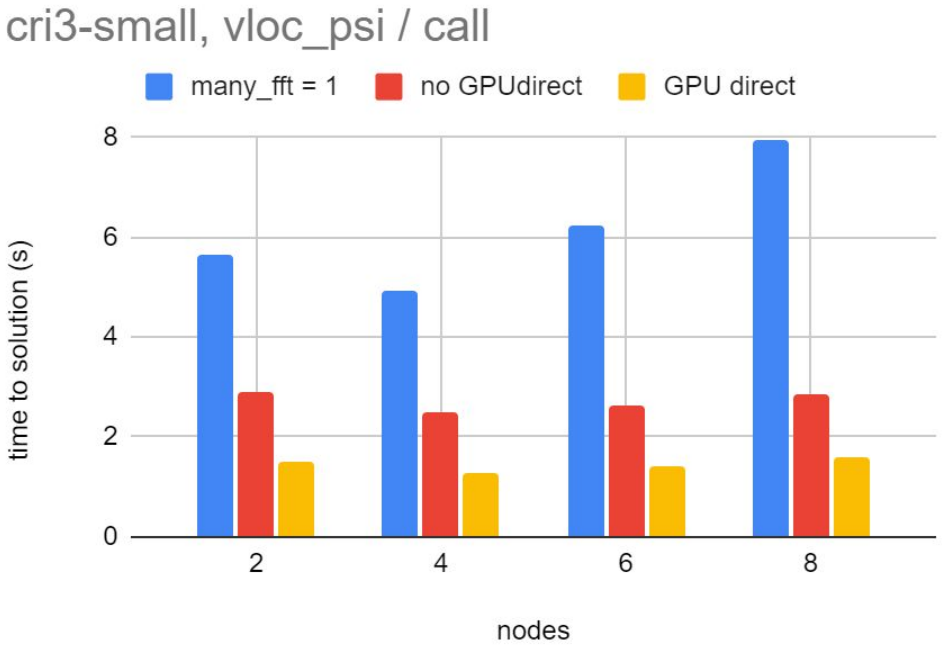


- ntasks associated to nbatches
- work in progress...

Execution on LUMI

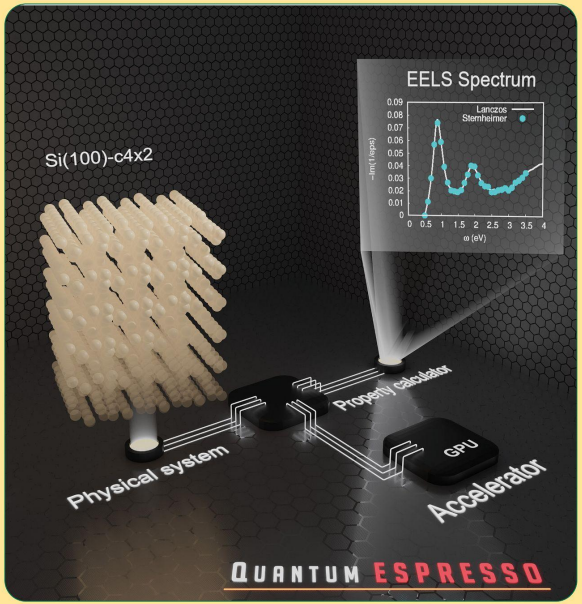
Chromium Iodide
~2700 electrons
480 atoms





CUF/OpenACC offload

JCTC Journal of Chemical Theory and Computation
Xxxxx XX, XXXX Volume XX Number XX pubs.acs.org/JCTC



ACS Publications
Most Trusted. Most Cited. Most Read.
www.acs.org

JCTC Journal of Chemical Theory and Computation
Open Access

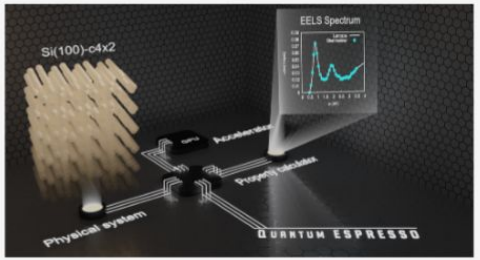
pubs.acs.org/JCTC
Article

QUANTUM ESPRESSO: One Further Step toward the Exascale

Ivan Carnimeo,* Fabio Affinito, Stefano Baroni, Oscar Baseggio, Laura Bellentani, Riccardo Bertossa, Pietro Davide Delugas, Fabrizio Ferrari Ruffino, Sergio Orlandini, Filippo Spiga, and Paolo Giannozzi

Cite This: J. Chem. Theory Comput. 2023, 19, 6992–7006
Read Online

ABSTRACT: We review the status of the QUANTUM ESPRESSO software suite for electronic-structure calculations based on plane waves, pseudopotentials, and density-functional theory. We highlight the recent developments in the porting to GPUs of the main codes, using an approach based on OpenACC and CUDA FORTRAN offloading. We describe, in particular, the results achieved on linear-response codes, which are one of the distinctive features of the QUANTUM ESPRESSO suite. We also present extensive performance benchmarks on different GPU-accelerated architectures for the main codes of the suite.

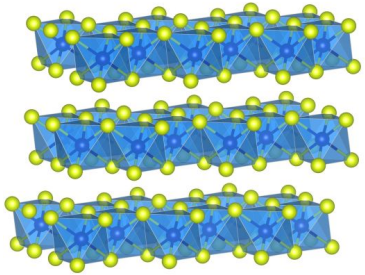


Outline

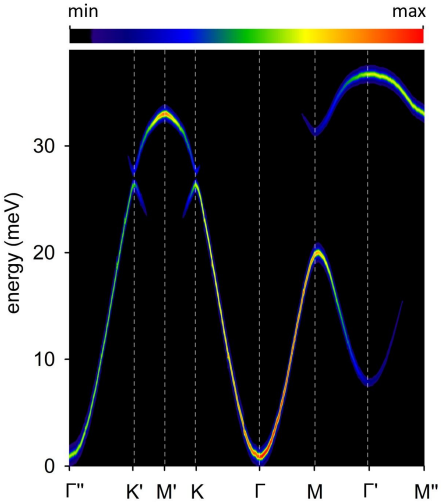
1. QUANTUM ESPRESSO
   a. The project
   b. The Suite
   c. What QE do
2. Porting to heterogeneous architecture
3. Overall performance of the subprograms

PWSCF: Ground state properties

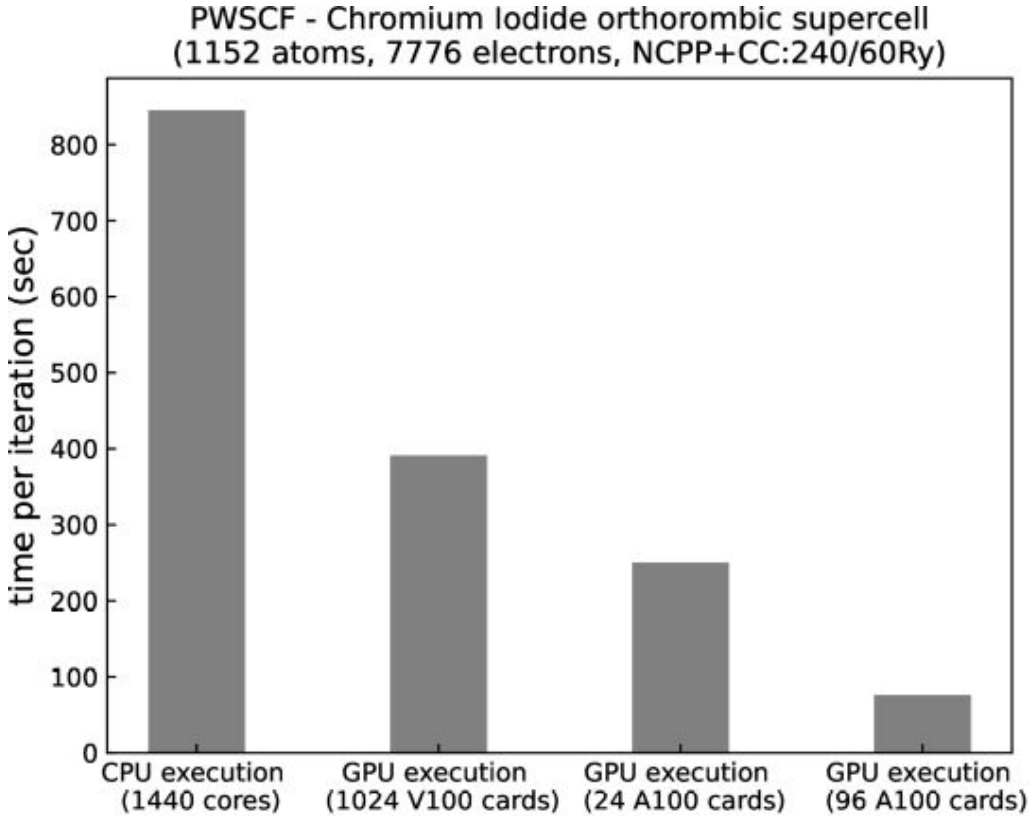
Chromium Iodide, 7776 electrons, 1152 atoms



Magnon dispersions along the high-symmetry directions of the BZ



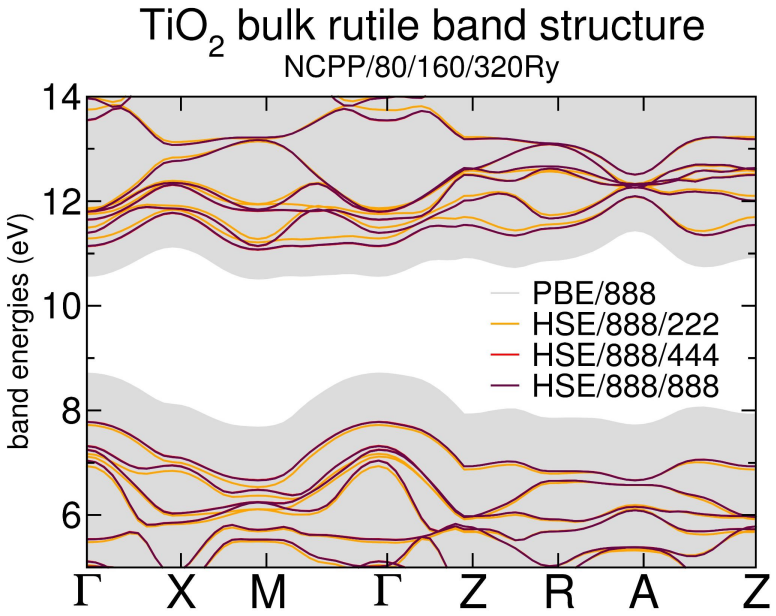
Delugas et al., Phys. Rev. B,, 107, 214452 (2023)
Gorni et al., *Phys. Rev. B,, 107, L220410 (2023)*



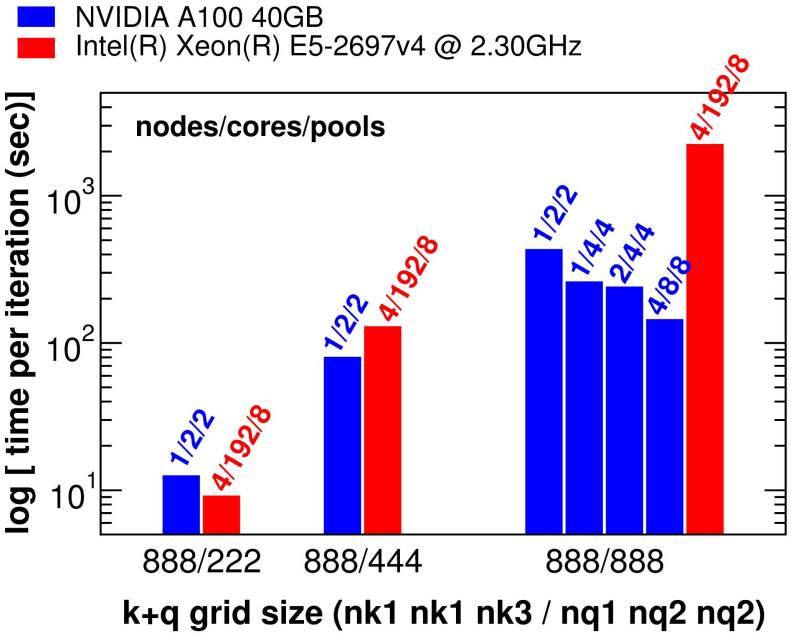
Quantum ESPRESSO: one further step towards the exascale, I. Carnimeo et al., JCTC, **19**, 20, 6992-7006 (2023)

PWSCF: band structures

TiO_2 bulk band structure



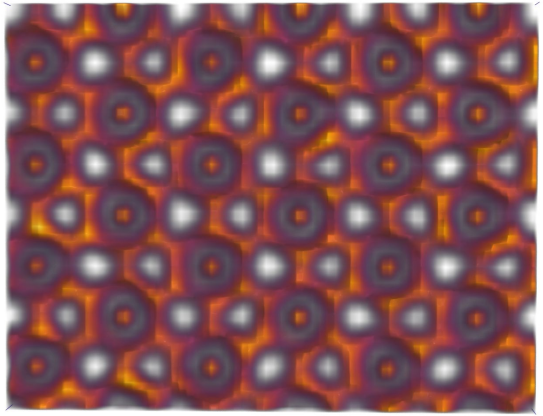
GPU acceleration provide further speed-up to the overall computational time

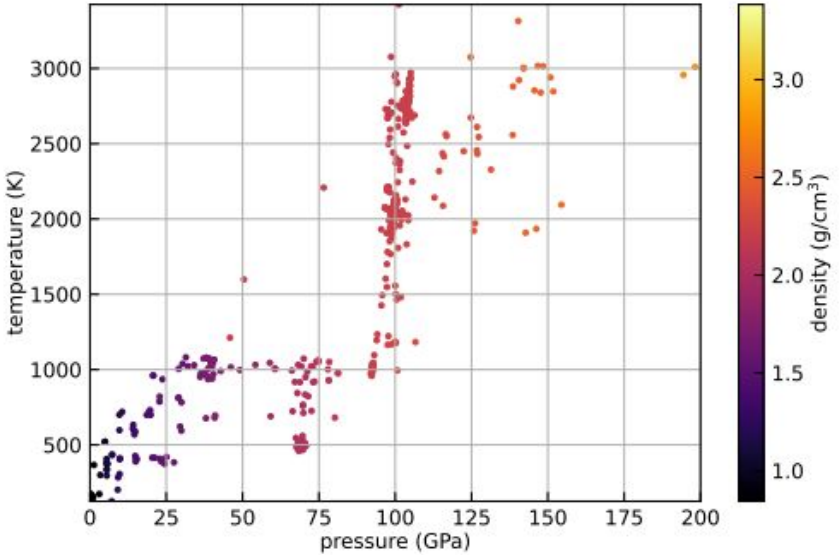


Quantum ESPRESSO: one further step towards the exascale, I. Carnimeo et al., *JCTC*, **19**, 20, 6992-7006 (2023)

CP: Car-Parrinello Molecular Dynamics

Equation of state diagram of ammonia, computed with the Car–Parrinello method. Each point is a CP run, and the color represents the density of the system for each particular combination of pressure and temperature.





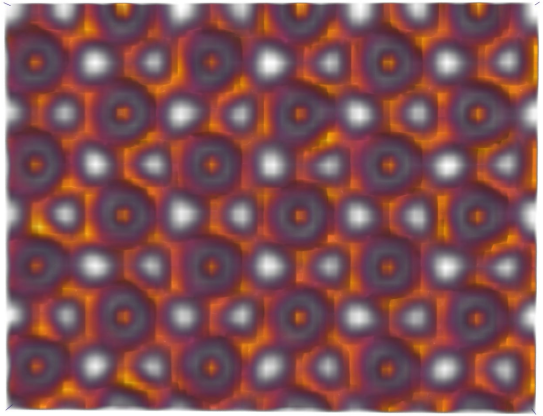
Quantum ESPRESSO: one further step towards the exascale, I. Carnimeo et al., *JCTC*, **19**, 20, 6992-7006 (2023)

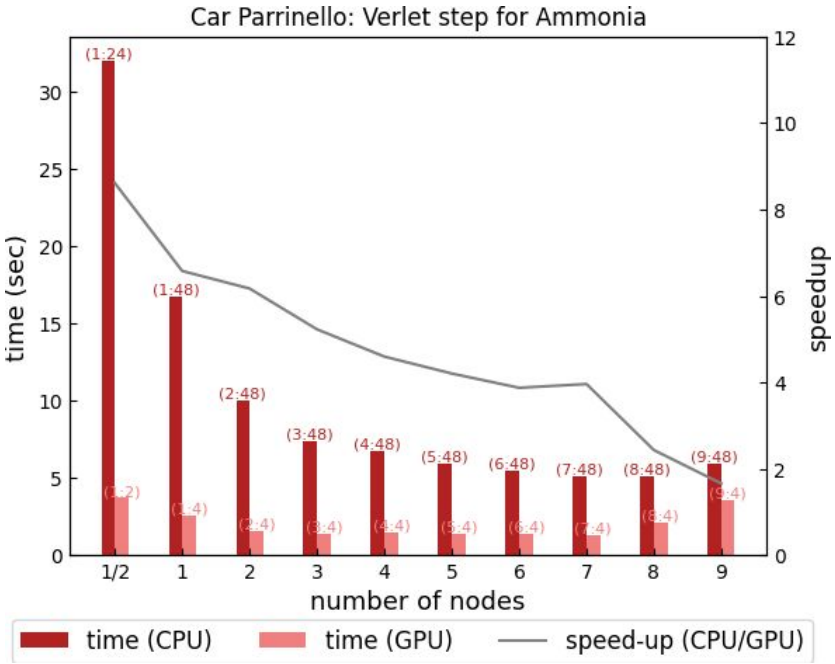
QUANTUMESPRESSO

MAX DRIVING THE EXASCALE TRANSITION

CP: Car-Parrinello Molecular Dynamics

Equation of state diagram of ammonia, computed with the Car–Parrinello method. Each point is a CP run, and the color represents the density of the system for each particular combination of pressure and temperature.

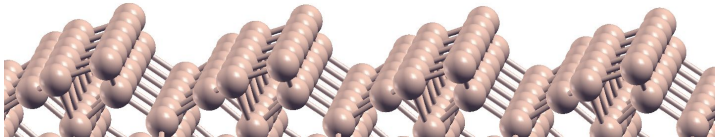




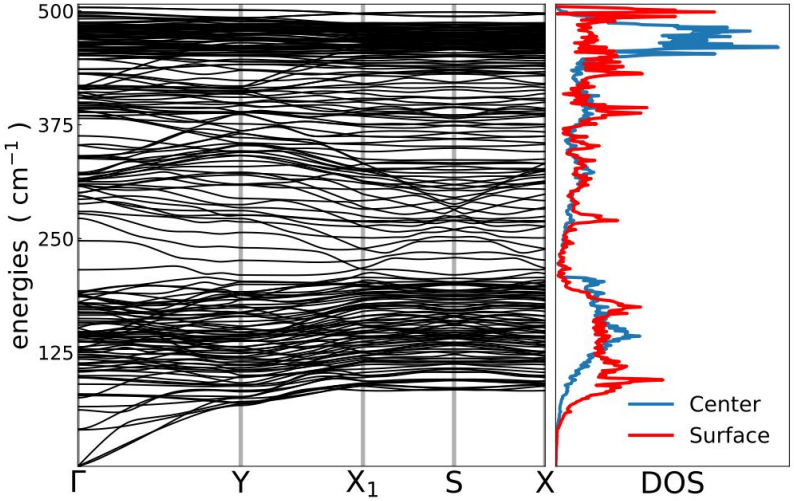
Quantum ESPRESSO: one further step towards the exascale, I. Carnimeo et al., JCTC, **19**, 20, 6992-7006 (2023)

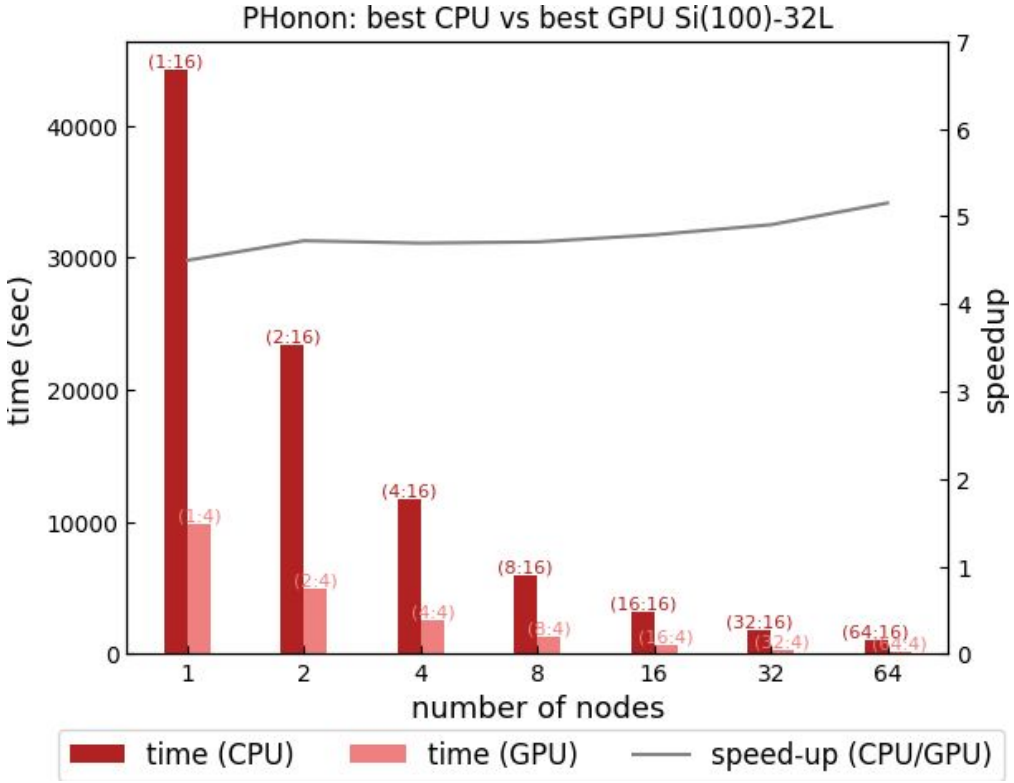
PHonon: vibrational properties

Silicon 100 surface, 512 electrons, 128 atoms



Phonon dispersions along the high-symmetry directions of the BZ

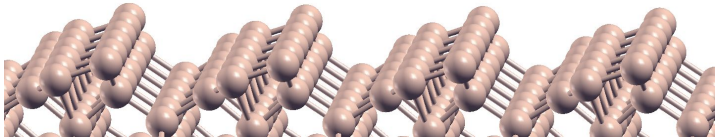




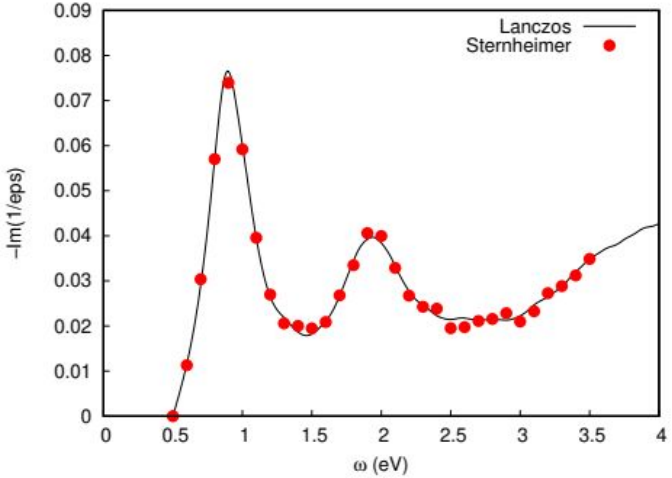
Quantum ESPRESSO: one further step towards the exascale, I. Carnimeo et al., *JCTC*, **19**, 20, 6992-7006 (2023)

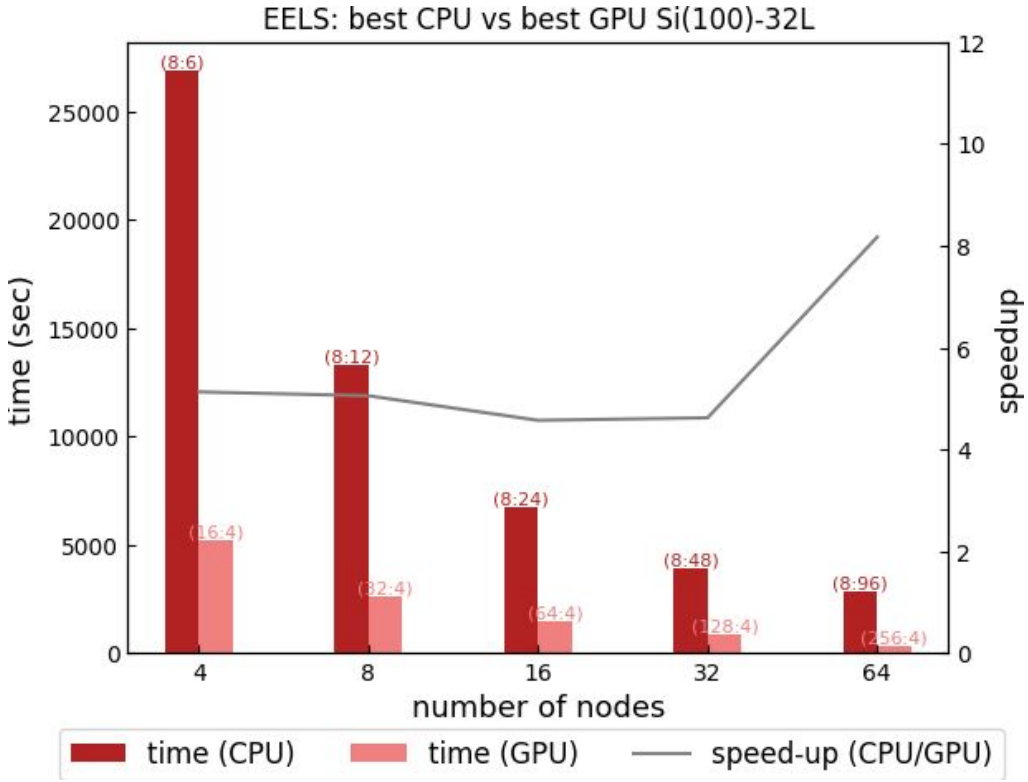
TD-DFPT: EELS spectroscopy

Silicon 100 surface, 512 electrons, 128 atoms



The turbo_EELS code can employ Lanczos or Sternheimer (CPHF) based approaches

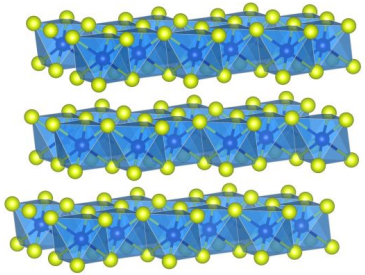




Quantum ESPRESSO: one further step towards the exascale, I. Carnimeo et al., JCTC, **19**, 20, 6992-7006 (2023)

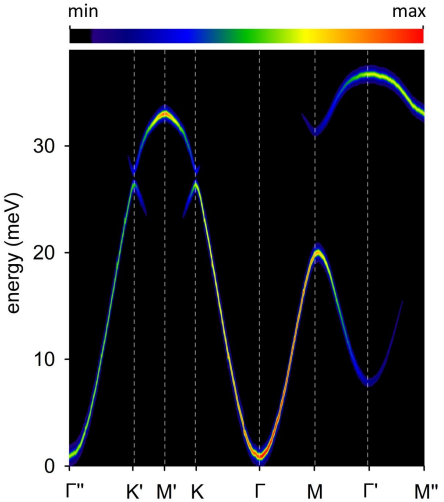
TD-DFPT: magnetic perturbation

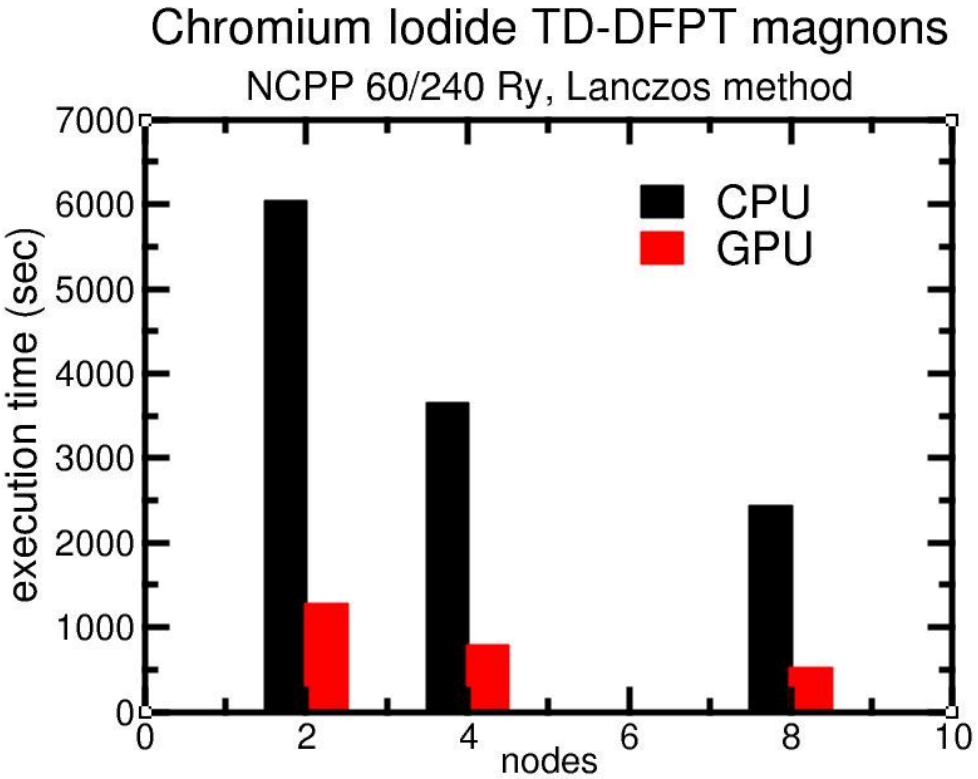
Chromium Iodide, 54 electrons, 8 atoms



64 k-points for each spin component + noncollinear + Spin-orbit + 30k Lanczos iterations

Magnon dispersions along the high-symmetry directions of the BZ





Delugas et al., Phys. Rev. B,, 107, 214452 (2023)
Gorni et al., *Phys. Rev. B,, 107, L220410 (2023)*

What's next

- **Complete the OpenMP porting** of PWscf minor routines and other codes (PHonon, TD-DFPT, CP)
- FFT optimization with OpenMP
- incorporate **DevXlib**.
- **Exascale workflows**

Acknowledgments

QUANTUM ESPRESSO developers

- Pietro Delugas, SISSA
- Ivan Carnimeo, SISSA
- Fabrizio Ferrari Ruffino, CNR-IOM
- Oscar Baseggio, SISSA
- Riccardo Bertossa, SISSA
- Aurora Ponzi, CNR-IOM
- Stefano Baroni, SISSA, CNR-IOM
- Paolo Giannozzi, UniUD, CNR-IOM

CINECA

- Laura Bellentani
- Sergio Orlandini
- Fabio Affinito

QUANTUM ESPRESSO Foundation

- Francesca Garofalo

Other collaborators and vendor technical support (chronological order)

- Ye Luo (Argonne)
- Filippo Spiga (NVIDIA)
- Louis Stuber (NVIDIA)

- Giacomo Rossi (Intel)
- Ossian O'Reilly (AMD)
- Jakub Kurzak (AMD)