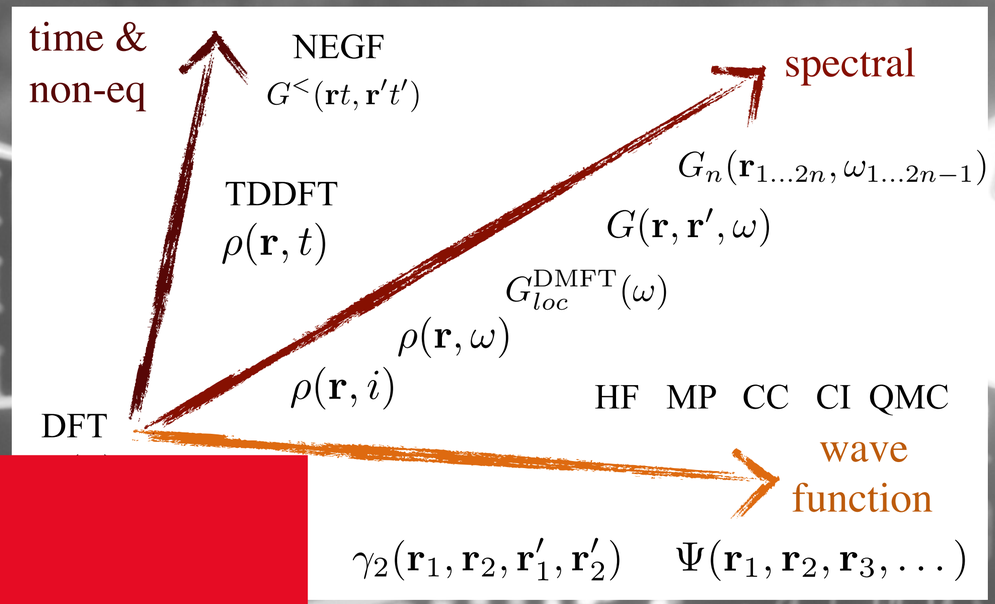
CNRNANO
ISTITUTO NANOSCIENZE CONSIGLIO NAZIONALE DELLE RICERCHE

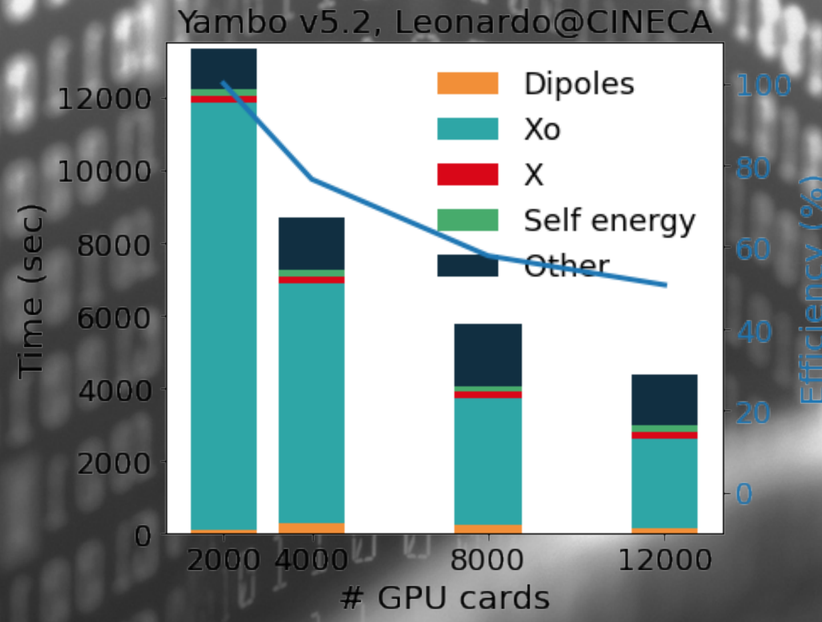
MAX



Benchmarking electronic structure codes on their way to exascale

Andrea Ferretti

[CNR-NANO, Modena, Italy]



outline

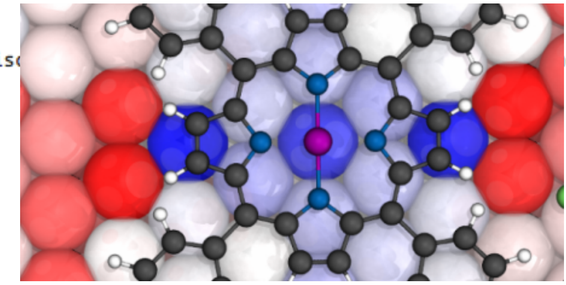
Materials Science &

- MaX: Materials design at the exascale
- Target workflows

Activities on MaX Codes

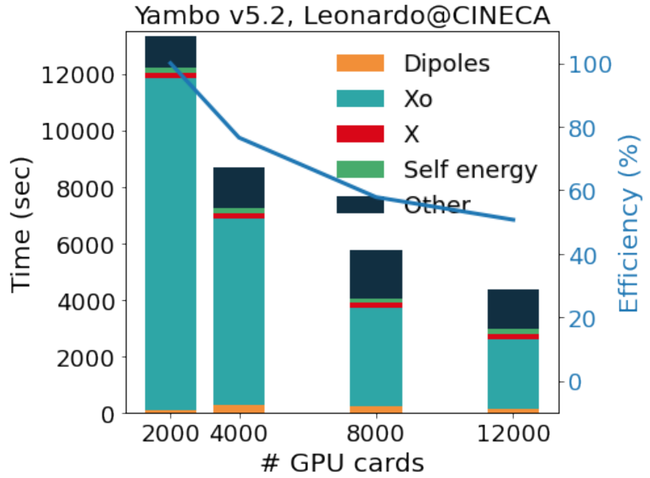
- Parallel performance
- Porting and performance portability

```
74    !$omp parallel default(shared), private(ir)
75    !$omp do
76    do ir = 1, fft_size
77      isc%rho_tw_rs(ir) = cmplx(isc
78    enddo
79    !$omp end do
```



More on benchmarking

- Metrics
- Hardware counters
- connection with co-design



ab initio materials modelling

quantum mechanics based
atomistic modelling of materials
+
interfacing with **multiscale** approaches

Electronic Structure Methods

- highly accurate (predictive)
- computationally demanding
- **a case for HPC**

the **exascale** opportunity:

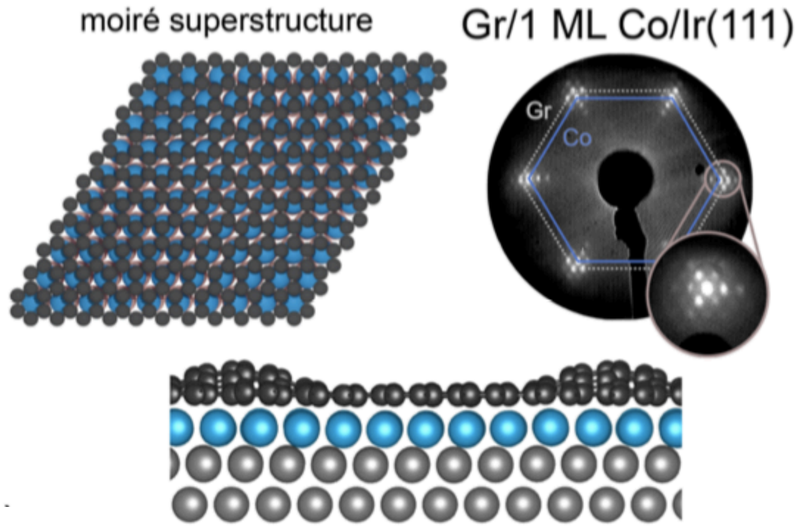
Higher accuracy

High throughput screening

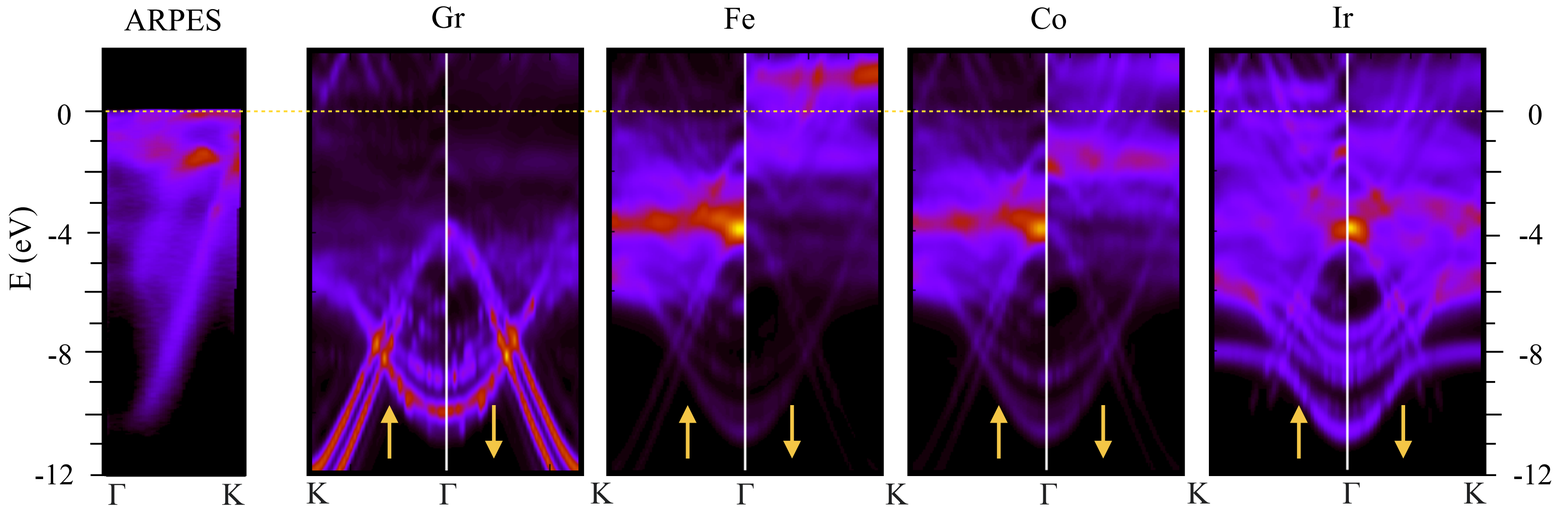
Improved modelling (complexity)

exascale opportunity: **complexity**

Claudia Cardoso
CNR-NANO



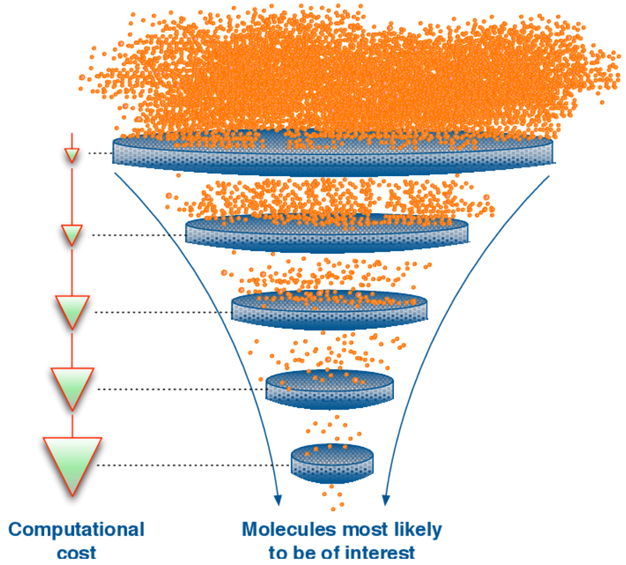
- Graphene / Transition Metal / Ir (111)
- **clear experimental evidence** for moire' pattern (**lattice mismatch**) and **Gr corrugation**
- 10x10 Graphene, 9x9 Iridium => 605 atoms / unit cell
- **Precise treatment of the structure** is important for modelling

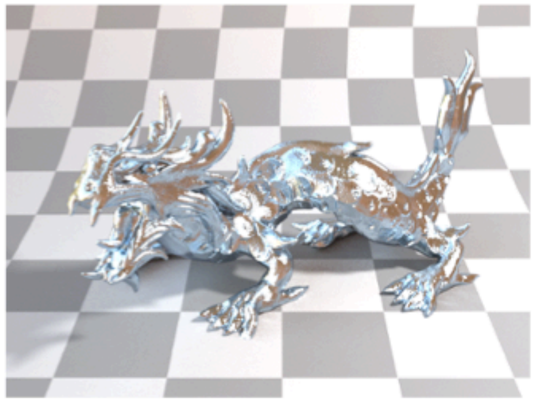


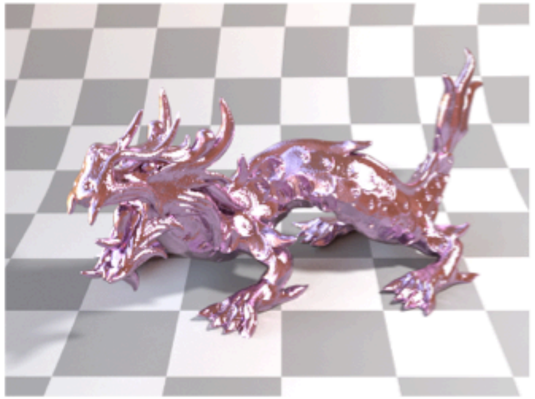
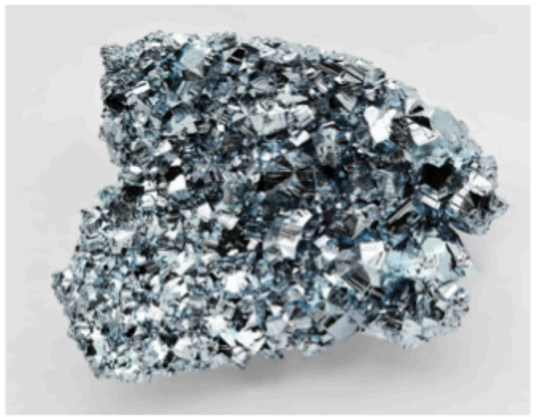
- Avvisati et al, J Phys. Chem. C **121**, 1639 (2017)

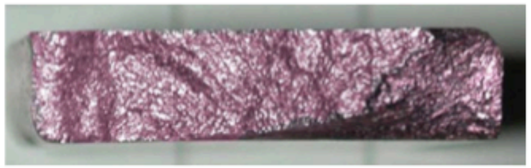
- Avvisati et al, Nano Lett. **18**, 2268 (2018)
- Calloni et al, J. Chem. Phys. **153**, 214703 (2020)
- Cardoso et al, Phys. Rev. Mat. **5**, 014405 (2021)
- Pacile' et al, Appl. Phys. Lett. **118**, 121602 (2021)

exascale opportunity: **high throughput screening**





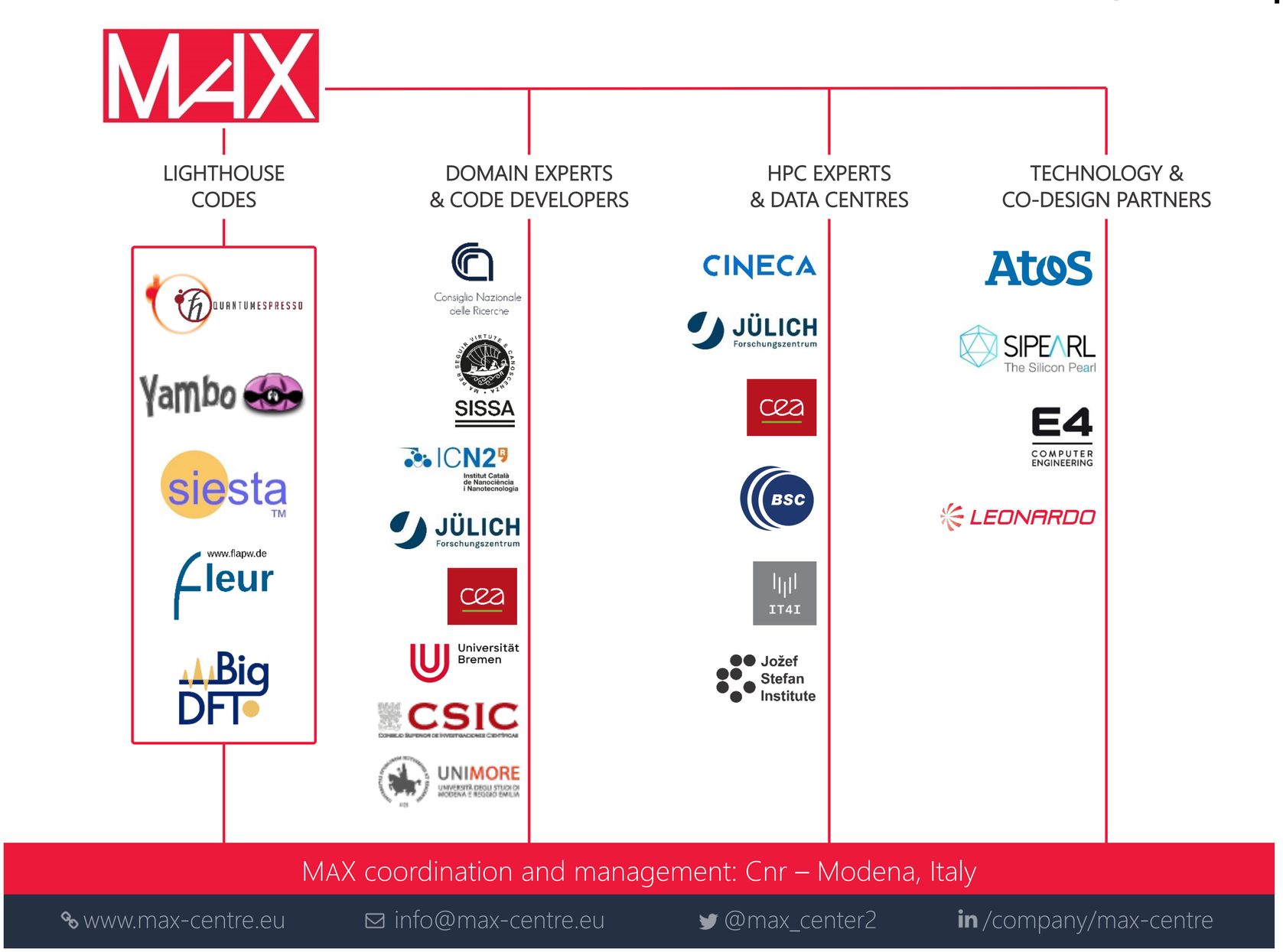


G. Prandini, G.M. Rignanese, N. Marzari, npj Comput. Materials **5**, 129 (2019)

Materials Design at the Exascale

- European centre of Excellence in HPC applications
- funded for 3 phases (2015-2026)
- head-quartered at CNR-NANO, Modena
- focused on **electronic structure codes**

A partnership with the required skills



QUANTUMESPRESSO

www.flapw.de
fleur

BigDFT

siesta

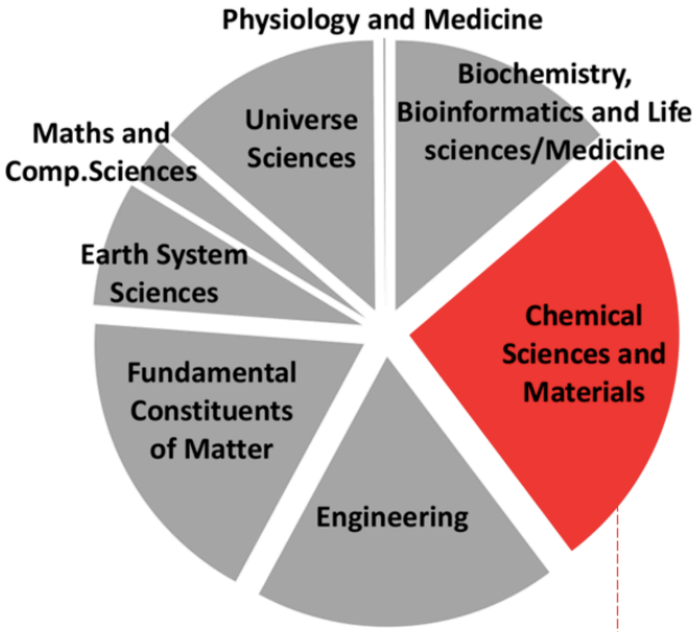
Yambo

AiiDA

MaX

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

- widely used **open source, community codes** in electronic structure



SELECTED ACTIVITIES

- **parallel optimization and performance portability** are key to keep exploiting HPC resources
- All MaX flagship codes released for **production with GPU support**

- hardware-software **codesign vehicles**
- **energy-efficiency** of codes

- large effort on **education and training**: hands-on schools and hackathons

today: HPC at the exascale

the exascale challenge
in high performance computing

- 10^{18} Flops/s
- 10^{18} Bytes
- abrupt technology changes
- **action is needed** for full exploitation
- **heterogeneous** machines (multiple HW and SW stacks)

Jupiter: > 1 ExaFlops

MareNostrum V: Atos + NVIDIA H100 => 208 PFlops (estimated)

US DOE

Frontier (@ORNL): HDE+AMD
=> 1194 PFlops

Leonardo: Atos + NVIDIA A100 (CUDA backend) => 239 PFlops

LUMI: CRAY + AMD cards (ROCm, HIP) => 309 PFlops

Codes

GOAL: turn MaX flagship codes into exascale-enabled applications

- **large scale MPI** parallelism (order of 10000 tasks)
- combined with **GPU awareness**

Frontier (@ORNL): 1100 PFlops
=> **37888 GPUs** (AMD MI250X)

Nicola Spallanzani
CNR-NANO

```
229       do jb=Sx_lower_band,Sx_upper_band
230         !
231         if (.not.PAR_IND_G_b%element_1D(jb)) cycle
232         !
233         isc%os(1)=jb
234         iscp%os=isc%os
235         !
236         call DEV_SUB(scatter_Bamp)(isc)
237         !
238         ! Normal case, the density matrix is diagonal
239         !
240         if (isc%is(1)/=iscp%is(1)) then
241           call DEV_SUB(scatter_Bamp)(iscp)
242         else
243           ! dev2dev, iscp%rhotw = isc%rhotw
244           call dev_memcpy(DEV_VAR(iscp%rhotw),DEV_VAR(isc
245         endif
246         !
247         DP_Sx_l=DEV_SUB(Vstar_dot_VV)(isc%ngrho,DEV_VAR(i
248     &                                 DEV_VAR(isc%rhotw),
249         DP_Sx=DP_Sx + DP_Sx_l * ( -4._SP/spin_occ*pi*E%f(
250         !
251         if (master_thread.and.is_ibz==1.and.n_lt_steps>0)
252         !
253       enddo
```

https://gitlab.com/max-centre/components/devicexlib

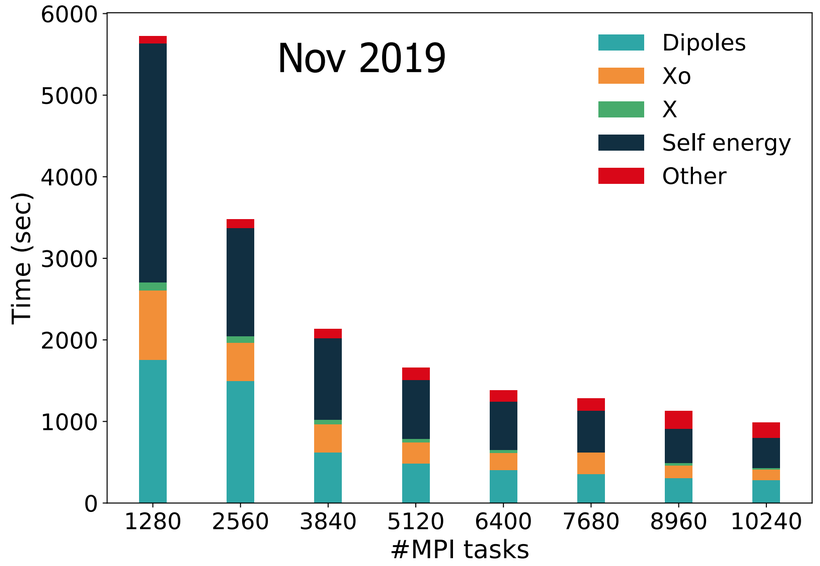
Codes

GOAL: turn MaX flagship codes into exascale-enabled applications

- **large scale MPI** parallelism (order of 10000 tasks)
- combined with **GPU awareness**

Frontier (@ORNL): 1100 PFlops
=> **37888 GPUs** (AMD MI250X)

yambo @
Marconi-KNL (CPU-only)

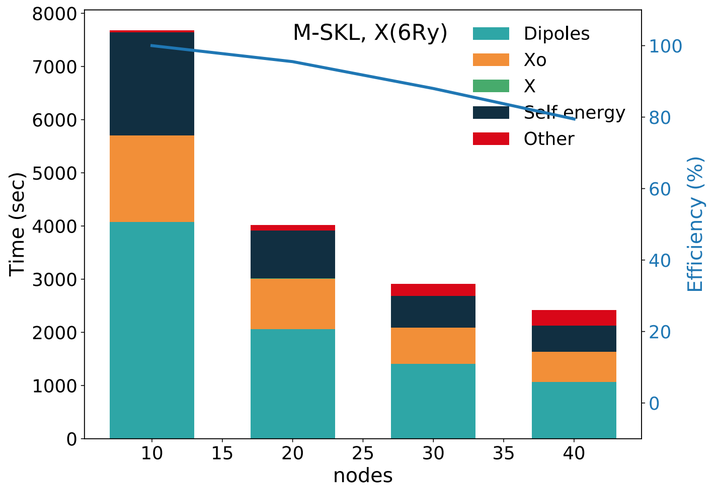



```
229         do jb=Sx_lower_band,Sx_upper_band
230           !
231           if (.not.PAR_IND_G_b%element_1D(jb)) cycle
232           !
233           isc%os(1)=jb
234           iscp%os=isc%os
235           !
236           call DEV_SUB(scatter_Bamp)(isc)
237           !
238           ! Normal case, the density matrix is diagonal
239           !
240           if (isc%is(1)/=iscp%is(1)) then
241             call DEV_SUB(scatter_Bamp)(iscp)
242           else
243             ! dev2dev, iscp%rhotw = isc%rhotw
244             call dev_memcpy(DEV_VAR(iscp%rhotw),DEV_VAR(isc
245           endif
246           !
247           DP_Sx_l=DEV_SUB(Vstar_dot_VV)(isc%ngrho,DEV_VAR(i
248      &                                  DEV_VAR(isc%rhotw),
249           DP_Sx=DP_Sx + DP_Sx_l * ( -4._SP/spin_occ*pi*E%f(
250           !
251           if (master_thread.and.is_ibz==1.and.n_lt_steps>0)
252           !
253         enddo
```

Yambo: performance (GPU)

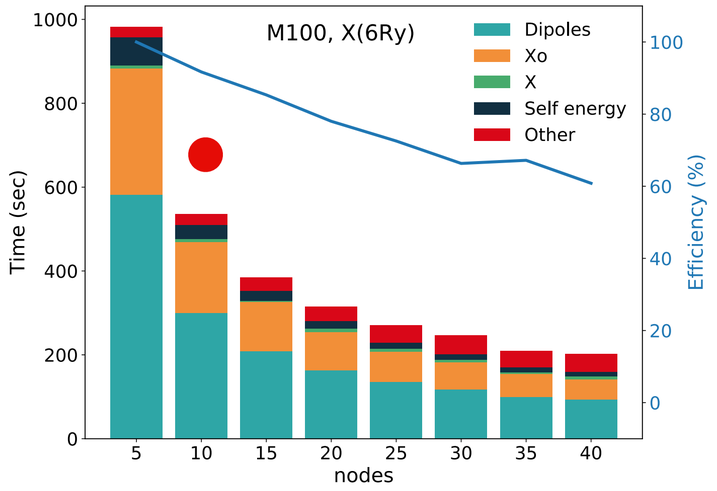
heterogeneous architectures: **MPI** + **OpenMP** + **CUDA**

CPU: **Skylake**



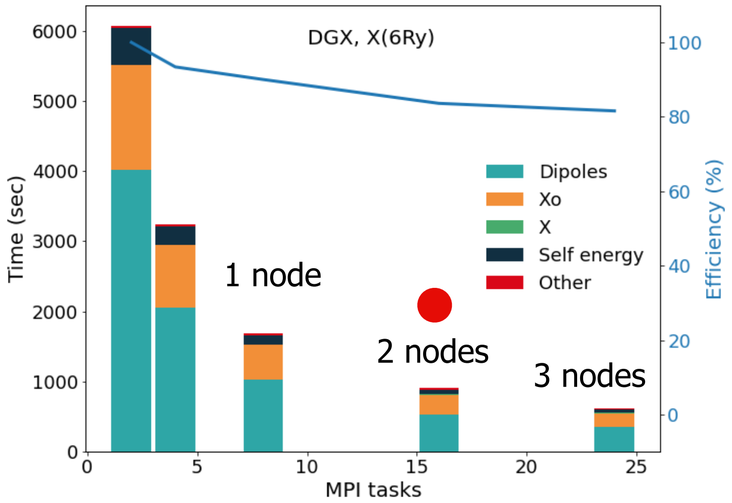
x14 wrt SKL (10 nodes)

GPU: P9+**V100**

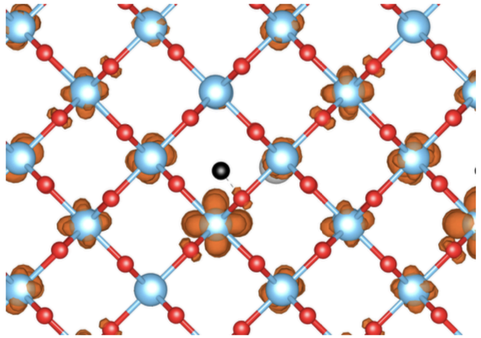


x1.4 wrt V100 (2 nodes)

GPU: AMD-Rome + **A100** (DGX)



- complete **GW workflow** for defected TiO2 (rutile)
- small system, **stress test**
- 1 MPI task/GPU
- data obtained on Marconi100, 4 V100 GPUs/node
- and DGX arch, 8 A100 GPUs/node



system size: 72+1 atoms, 2000 bands, 6 Ry for Xo repr (N=1317); ~290 occ states, 8 kpts.

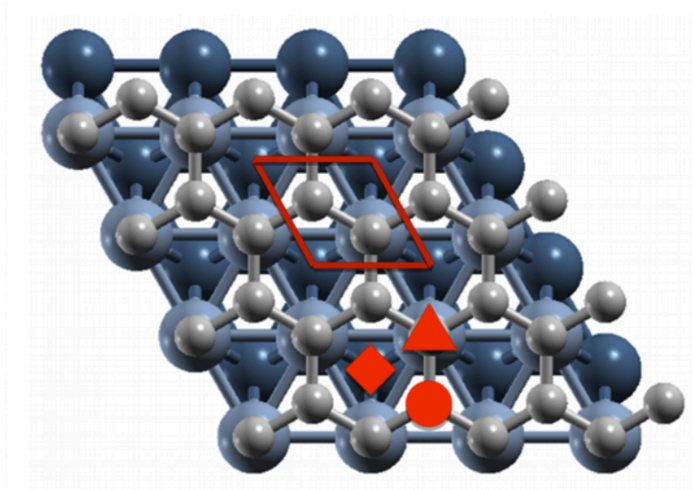
data available at: http://www.gitlab.com/max-centre/Benchmarks

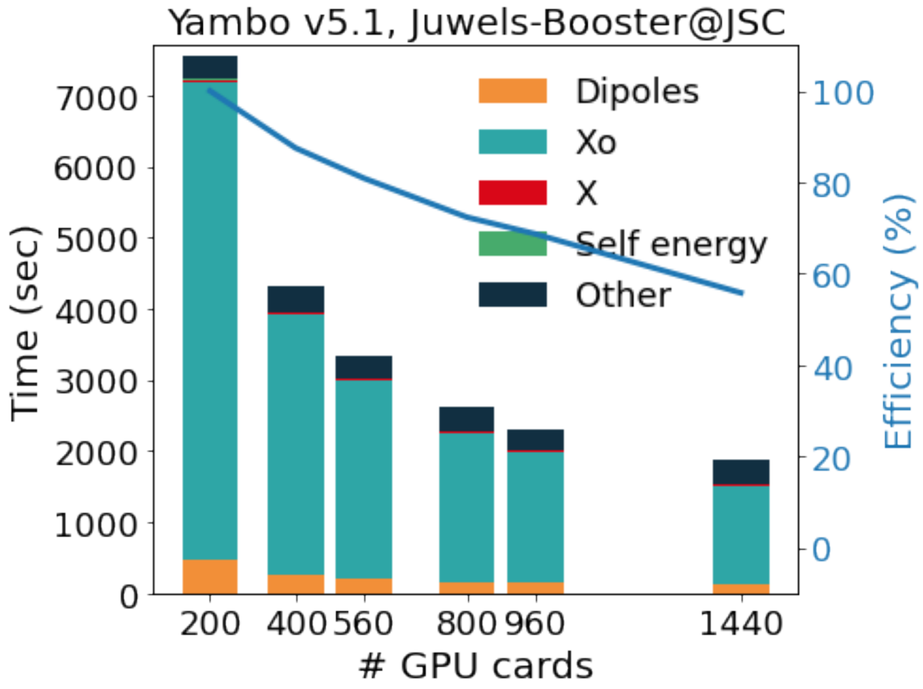
Yambo: performance (GPU)

Juwels-Booster: 4 Nvidia A100 / node

runs: up to 360 J-B nodes about 40% of the whole machine (960 nodes)

GW study of
Graphene @ Co(0001) interface





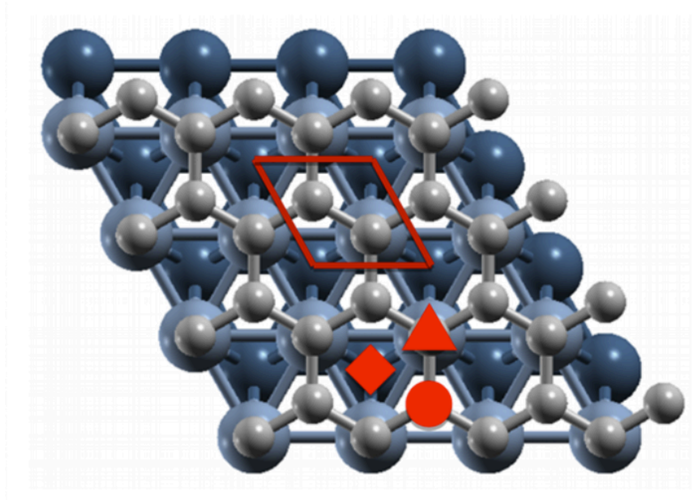
data available at: http://www.gitlab.com/max-centre/Benchmarks

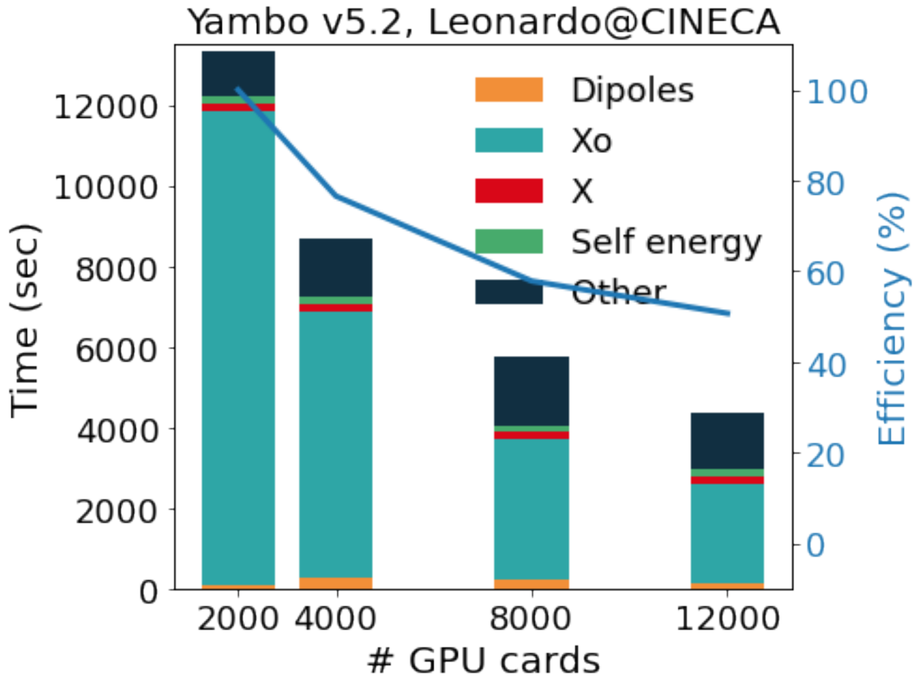
Yambo: performance (GPU)

Leonardo @ CINECA: 4 Nvidia A100 next / node

runs: up to 3000 nodes, about 90% of the whole machine (3456 nodes)

GW study of
Graphene @ Co(0001) interface

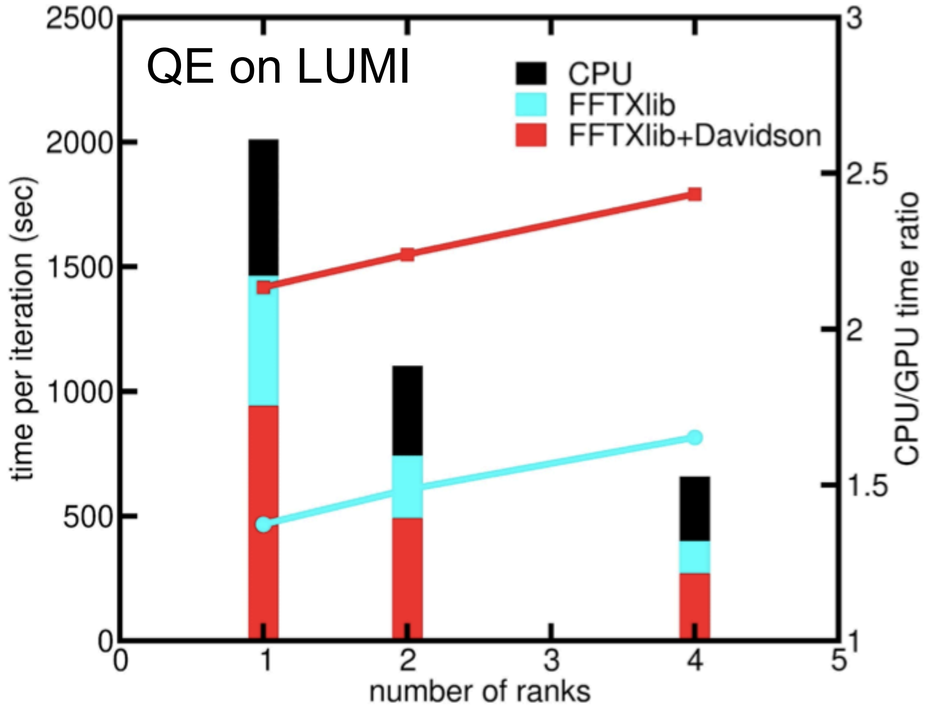


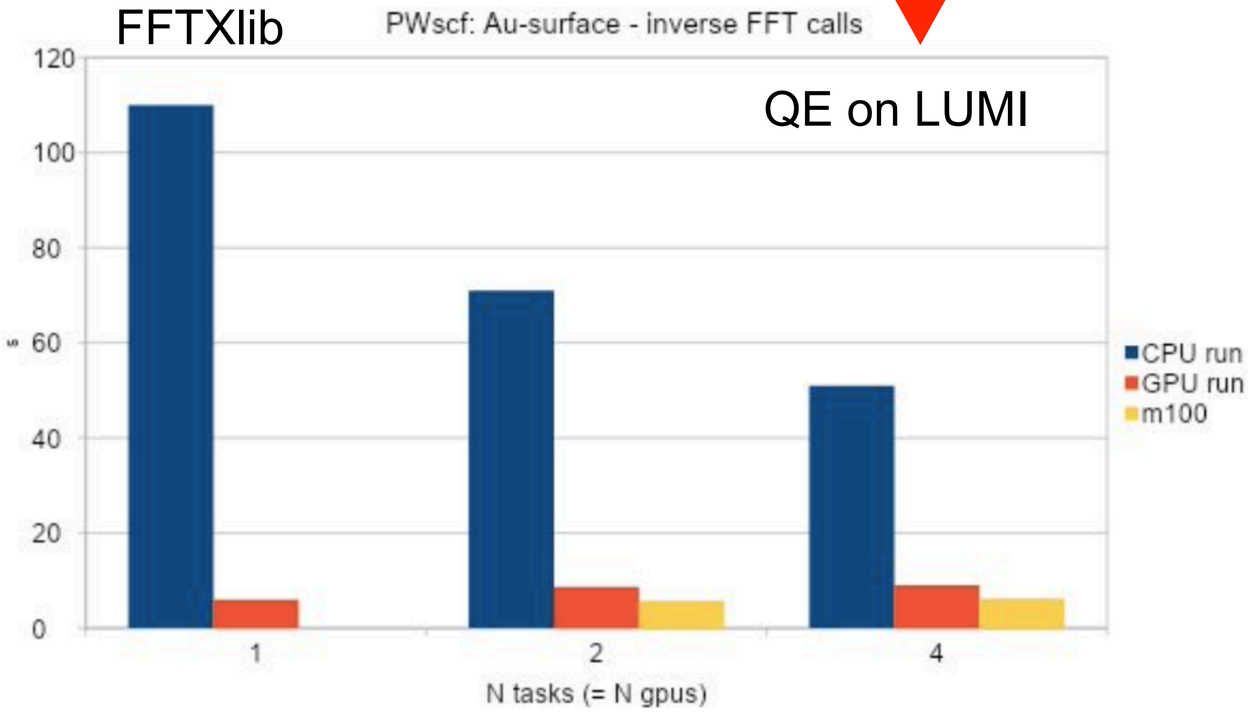


data available at: http://www.gitlab.com/max-centre/Benchmarks

Quantum ESPRESSO on pre-exascale

- currently running on **NVIDIA, AMD, and INTEL accelerated machines**
- using **Cuda-Fortran, OpenACC, and OpenMP** programming models
- DATA: courtesy of **I. Carnimeo & F. Ferrari Ruffino**





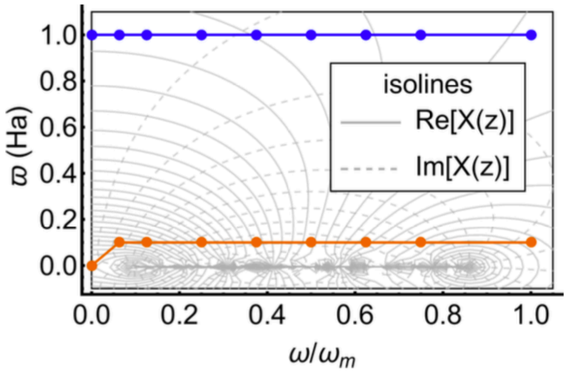
Full frequency GW using MPA

PHYSICAL REVIEW B **104**, 115157 (2021)

Frequency dependence in *GW* made simple using a multipole approximation

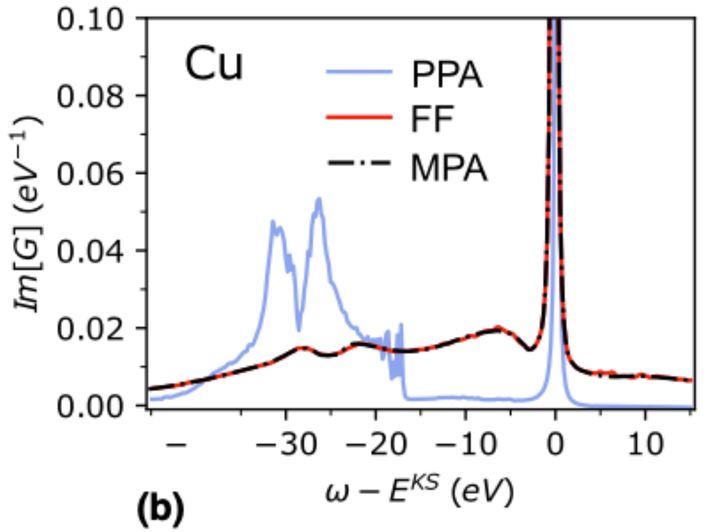
Dario A. Leon [1,2,*] Claudia Cardoso [2] Tommaso Chiarotti [3] Daniele Varsano [2]
Elisa Molinari [1,2] and Andrea Ferretti [2]

[1] *FIM Department, University of Modena & Reggio Emilia, Via Campi 213/a, Modena, Italy*
[2] *S3 Centre, Istituto Nanoscienze, CNR, Via Campi 213/a, Modena, Italy*
[3] *Theory and Simulation of Materials (THEOS), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*



- exploits an **advanced frequency sampling** of the response in the complex plane
- bridges from PPA to **full-frequency accuracy** (at a fraction of the cost)
- demonstrated for semiconductors and metals
- also: D.A. Leon, et al PRB **107**, 155130 (2023)

Dario A. Leon
now: NULS Norway

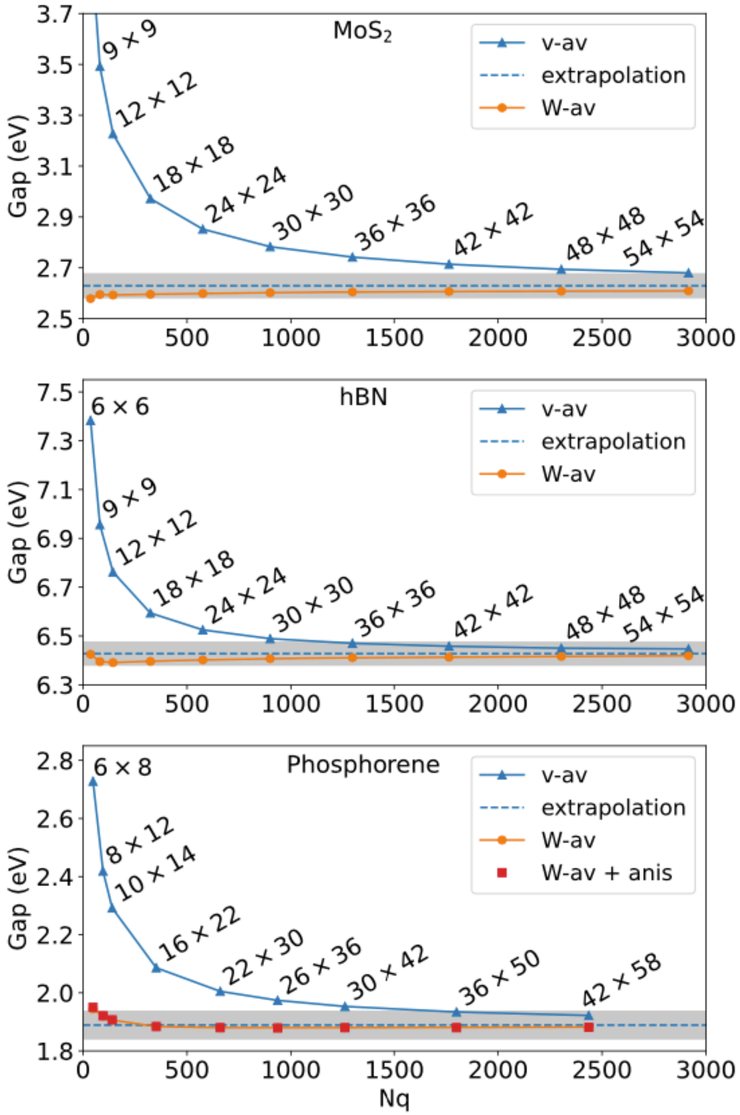


charged excitations in 2D materials

npj Comput Materials **9**, 44 (2023)

Efficient GW calculations in two dimensional materials through a stochastic integration of the screened potential

Alberto Guandalini[1,2]✉, Pino D'Amico[1]✉, Andrea Ferretti[1]✉ and Daniele Varsano[1]✉



- Accurate integration of the momentum dependence of the response function, critical in **2D materials**
- devise a **convergence accelerator** for the GW method applied to 2D systems
- 2 orders of magnitude speed up
- the approach can be generalised to other dimensionalities (see G. Sesti's work)

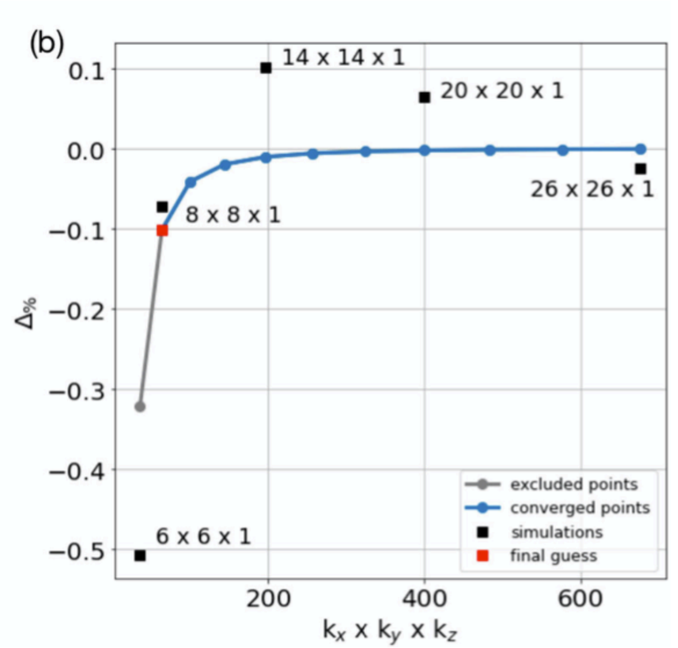
Alberto Guandalini
now: Sapienza Uni

automated workflows for MBPT

npj Comput Materials **9**, 74 (2023)

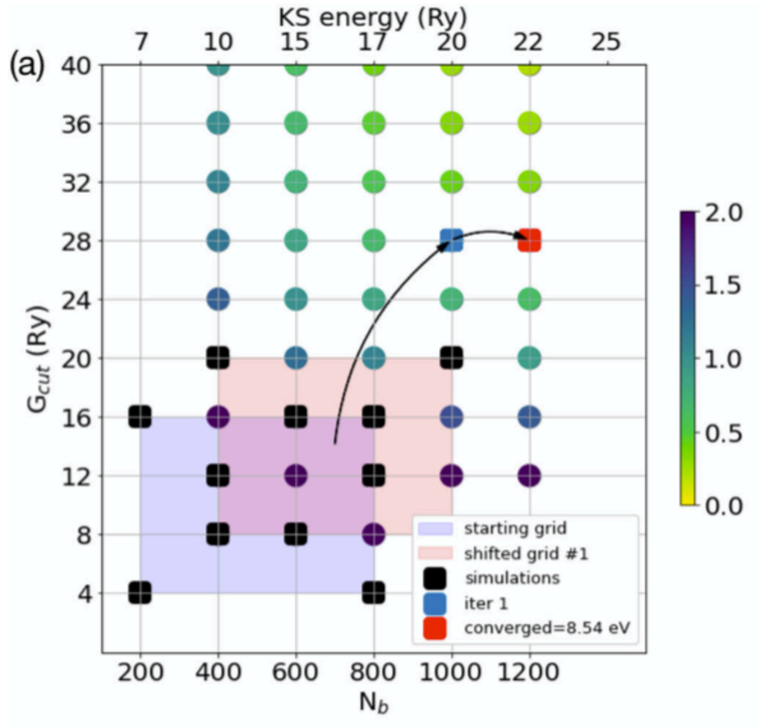
Towards high-throughput many-body perturbation theory: efficient algorithms and automated workflows

Miki Bonacci[1,2]✉, Junfeng Qiao[3], Nicola Spallanzani[2], Antimo Marrazzo[4], Giovanni Pizzi[3,5], Elisa Molinari[1,2], Daniele Varsano[2], Andrea Ferretti[2] and Deborah Prezzi[2]



- Algorithms for **automatic convergence** of many-body perturbation theory methods
- **efficient sampling** in multi-dimensional parameter space
- combines yambo workflows with the AiiDA automation engine

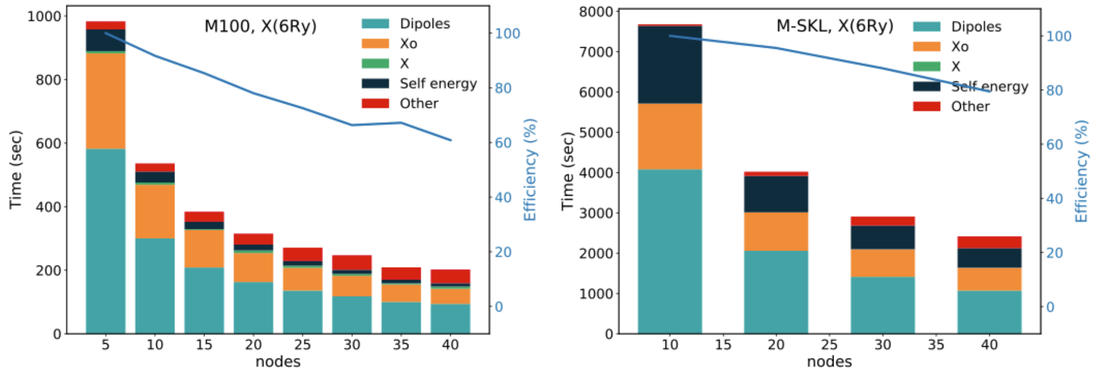
Miki Bonacci
now: PSI Zurich, CH

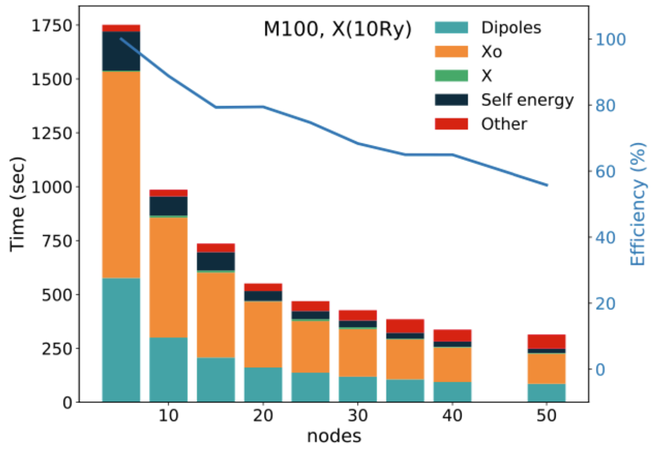


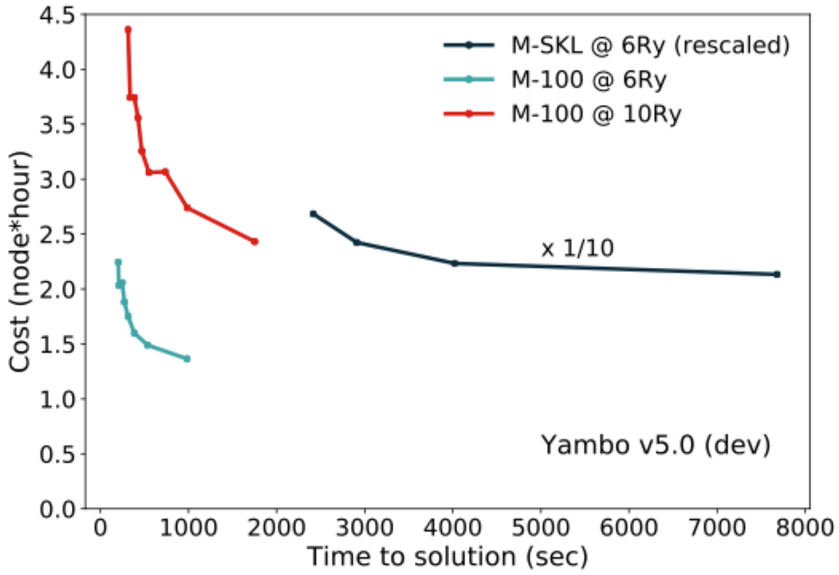
… more about benchmarking

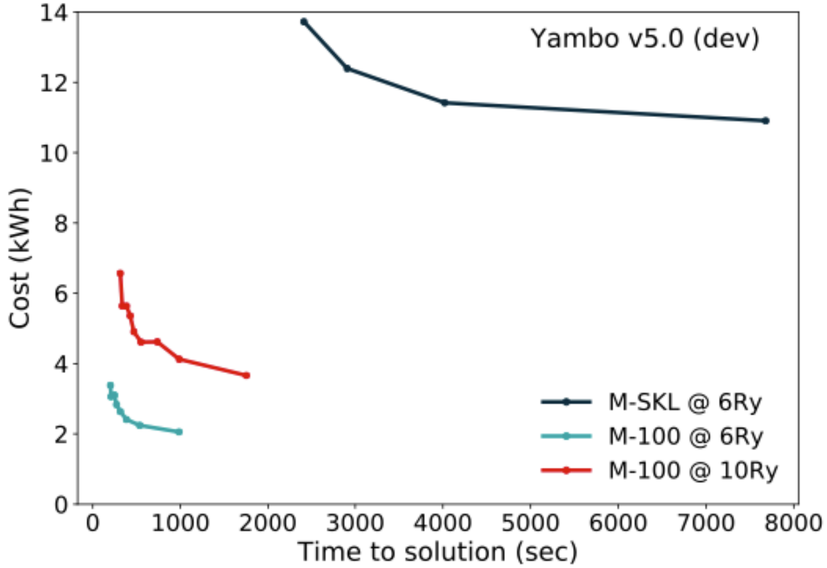
cost vs time

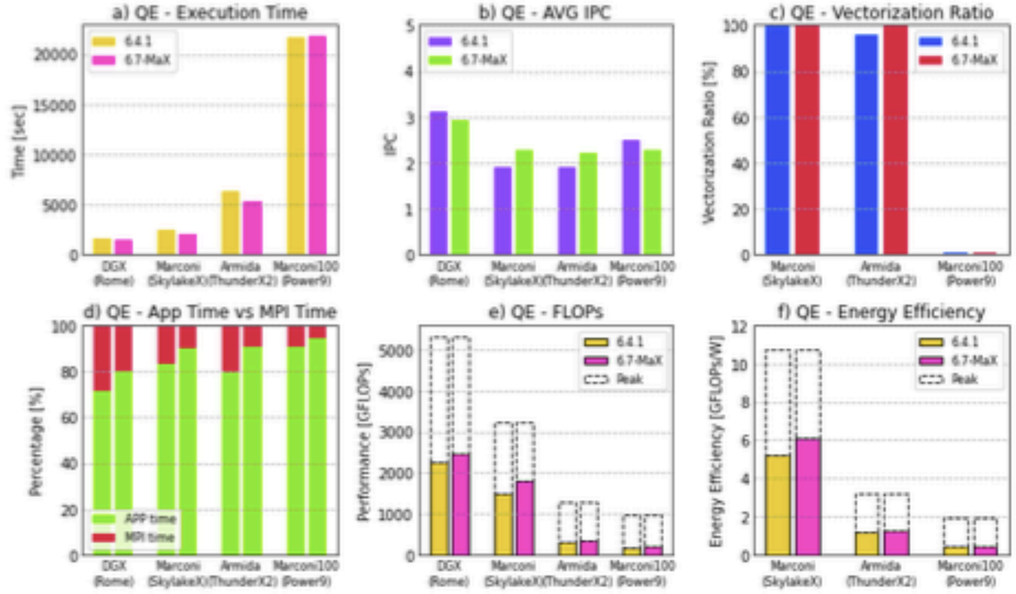
- usage of COST-VS-TIME
- same data as for time-to-solution
- practically meaningful (eg to users)
- simple proxy for advanced metrics (eg energy-to-solution)

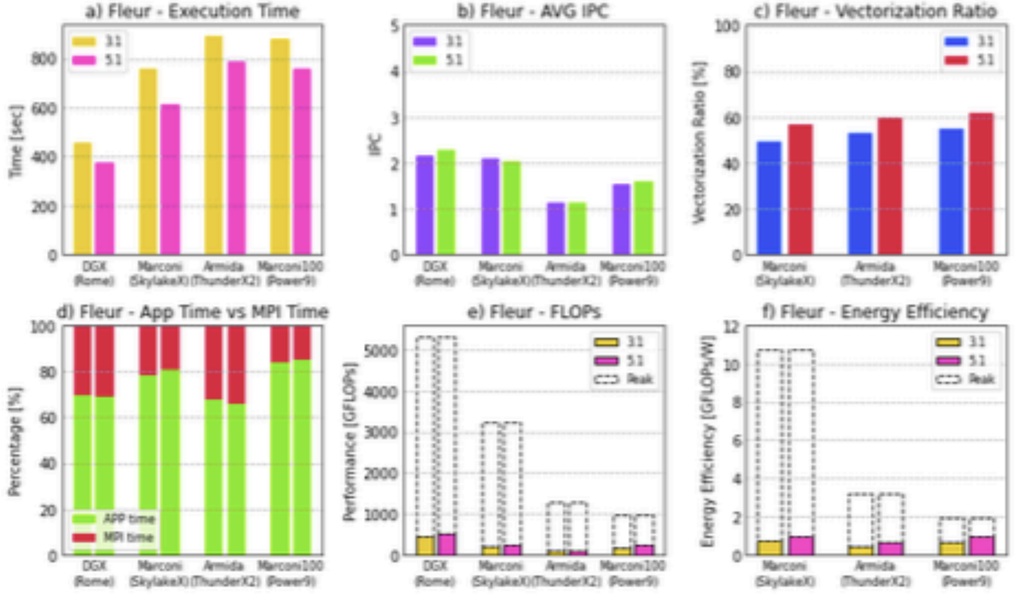


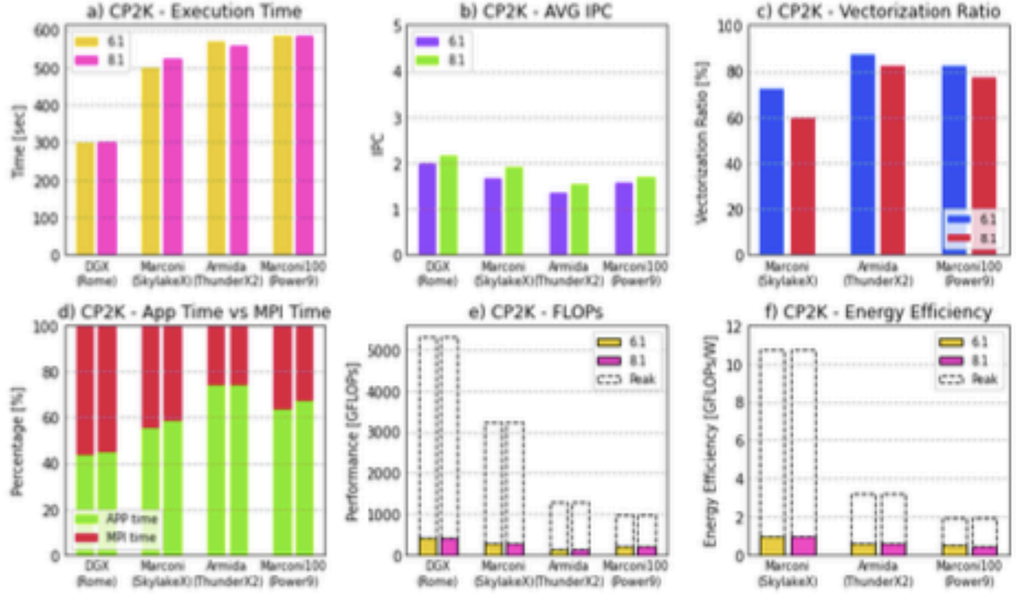










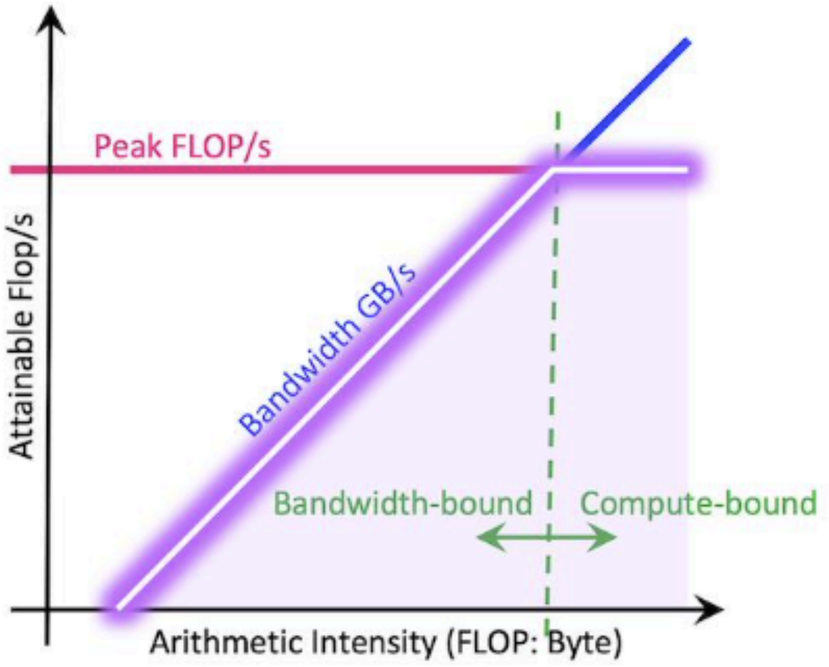


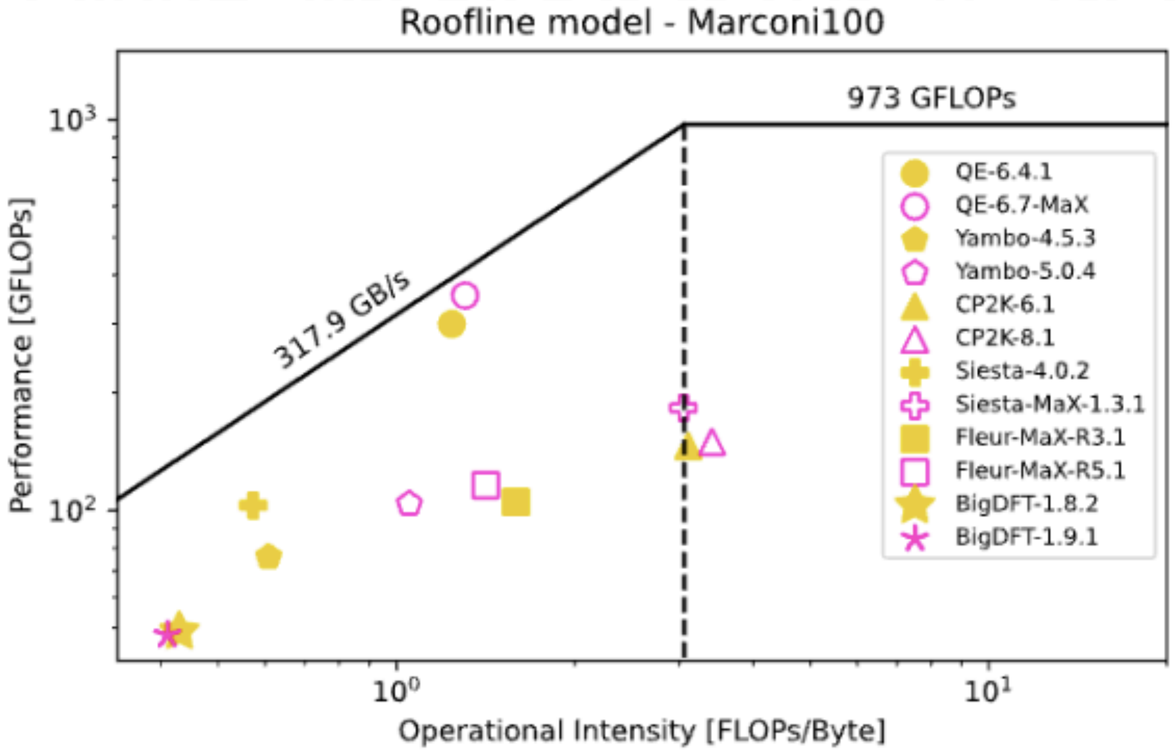
Multiple Metrics

For each code:

- execution time
- instruction per cycle
- vectorization ratio
- app time vs MPI time
- FLOP/s
- energy efficiency

Roofline model





adoption of JUBE

The **JUBE benchmarking environment**:

Script based framework to easily

- create benchmark sets
- run benchmarks on different computer systems
- evaluate and display the results

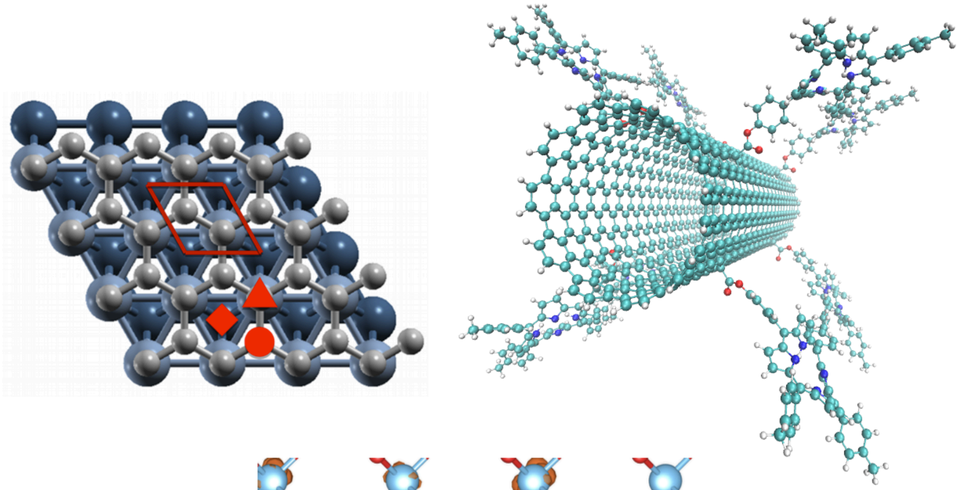
Developed at JSC

nspallan@login07:/leonardo_scratch/large/userexternal/nspallan/benchmaks

```
(jube) [nspallan@login07 benchmaks]$ jube analyse ./run_bench_grco -i 3 -t leonardo yambo grco
################################################################################
# Analyse benchmark "yambo" id: 3
################################################################################
>>> Start analyse
>>> Analyse finished
>>> Analyse data storage: ./run_bench_grco/000003/analyse.xml
################################################################################
(jube) [nspallan@login07 benchmaks]$ jube result ./run_bench_grco -i 3 -t leonardo yambo grco
result:
| Nodes | Tasks/Node | Threads/Task | k-points | nbands | X-cutoff | Time-profile | Dipoles |      Xo |    X | Self energy | Other |
|-------|------------|--------------|----------|--------|----------|--------------|---------|---------|------|-------------|-------|
|     1 |          4 |            8 |        7 |   2000 |       20 |      2940.00 |   40.40 | 2683.00 | 0.01 |      177.79 | 35.60 |
|     4 |          4 |            8 |        7 |   2000 |       20 |       753.00 |   15.28 |  654.00 | 0.01 |       51.60 | 23.61 |
|     8 |          4 |            8 |        7 |   2000 |       20 |       399.00 |   11.31 |  341.06 | 0.01 |       27.41 | 14.81 |
|    12 |          4 |            8 |        7 |   2000 |       20 |       287.00 |   11.82 |  234.06 | 0.01 |       19.50 | 14.96 |
|    16 |          4 |            8 |        7 |   2000 |       20 |       234.00 |   14.01 |  182.43 | 0.01 |       16.40 | 17.69 |

(jube) [nspallan@login07 benchmaks]$
```

Conclusions



Benchmarking rationale

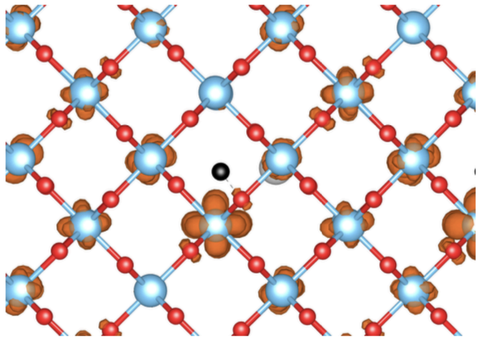
- **Test cases identified** early on and **regularly updated**
- **data available** at: http://www.gitlab.com/max-centre/Benchmarks

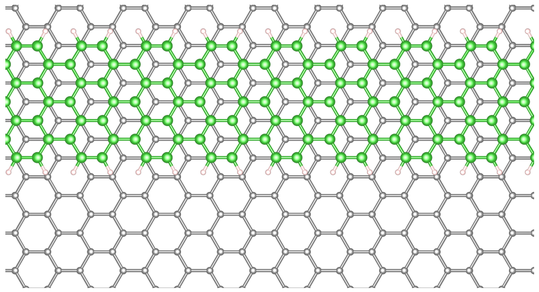
- in MaX-3, adoption of **JUBE**

- **Mostly TIME-TO-SOLUTION** so far, but also **COST-VS-TIME**
- Access to **hardware counters** and different metrics
- More on the analysis of **ENERGY-TO-SOLUTION** ongoing

- Used to validate the **parallel performance** and **performance portability** on the codes across different architectures
- **Non-maturity** and **inhomogeneity of the software stacks** strong limitation

- Even more difficult to find a **suitable metric(s)** when dealing with **workflows**





Special thanks to:

Nicola Spallanzani
CNR-NANO

Fabio Affinito
Cineca

Enjoy !

Follow us on:

@max_center2

company/max-centre/

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

youtube/channel/MaX Centre eXascale