Co-design for Green's-Function-Based Methods at the Exascale

Mini-apps Suite Development & Distribution

I.F. Mas Magre*, R. Grima Torres, J.M. Cela Espín, J.J. Gutiérrez Moreno

Barcelona Supercomputing Center (BSC)

* isidre.masmagre@bsc.es

Summer School Paphos

Towards exascale solutions in Green function methods and advanced DFT

7th of October 2023

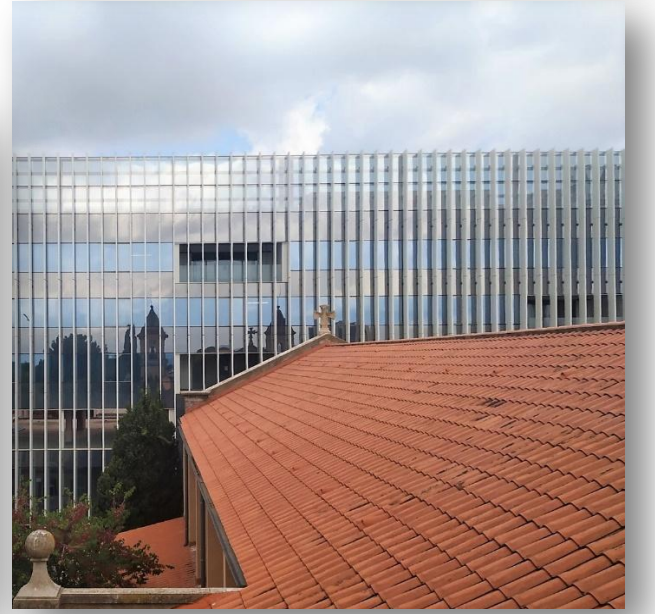
NOMAD CoE receives funding from the European Union's Horizon program under the grant agreement Nº 951786.

Outline

- BSC & Material Science Team
- HPC in the Exascale
- Co-Design for GW methods
- Summary & take aways

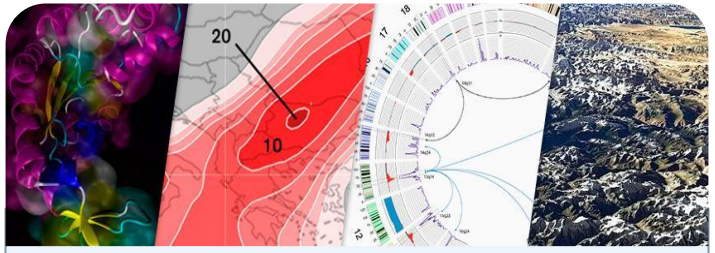
BSC & Material Science Team

The BSC site



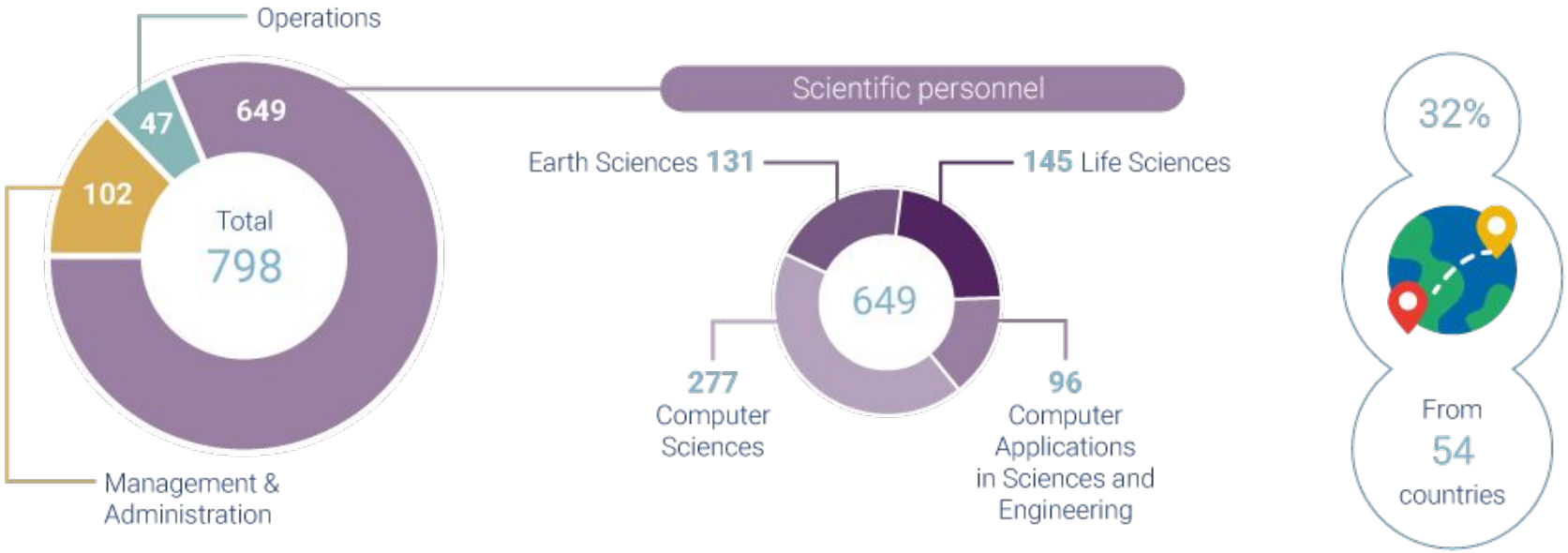
BSC-CNS Objectives

Supercomputing services
to Spanish and EU researchers



R&D in Computer, Life, Earth and
Engineering Sciences

PhD programme, technology
transfer, public engagement



MareNostrum 4

Total peak performance: **13,9** Pflops

General Purpose Cluster:	11.15 Pflops
MN4 CTE-Power:	1.57 Pflops
MN4 CTE-ARM:	0.65 Pflops
MN4 CTE-AMD:	0.52 Pflops

MareNostrum 5

Total peak performance: **314** Pflops

Coming soon!

MareNostrum 1
2004 – 42,3 Tflops
1st Europe / 4th World
New technologies

MareNostrum 2
2006 – 94,2 Tflops
1st Europe / 5th World
New technologies

MareNostrum 3
2012 – 1,1 Pflops
12th Europe / 36th World

MareNostrum 4
2017 – 11,1 Pflops
2nd Europe / 13th World
New technologies

Material Science Team

People

Julio Gutiérrez Moreno

Rogeli Grima Torres

Isidre Mas Magre

Projects

NOMAD
NOVEL MATERIALS DISCOVERY

Co-design for ab initio computational materials

MAX
DRIVING THE EXASCALE TRANSITION

Optimization of SIESTA ab initio code for exascale computing

Lines of Research

Atomic scale simulation of materials based on **quantum mechanical** ab initio calculations & **classical molecular dynamics**

- **Materials:** metals, alloys, oxides, ceramics…
- **Applications and uses:** energy materials (catalysis, thermoelectrics, fusion power), biomaterials, sensors, surface processes and growth techniques, atomic-level characterization...
- **HPC:** Performance analyses, code optimization, co-design...

HPC in the Exascale

Exascale is (almost) here

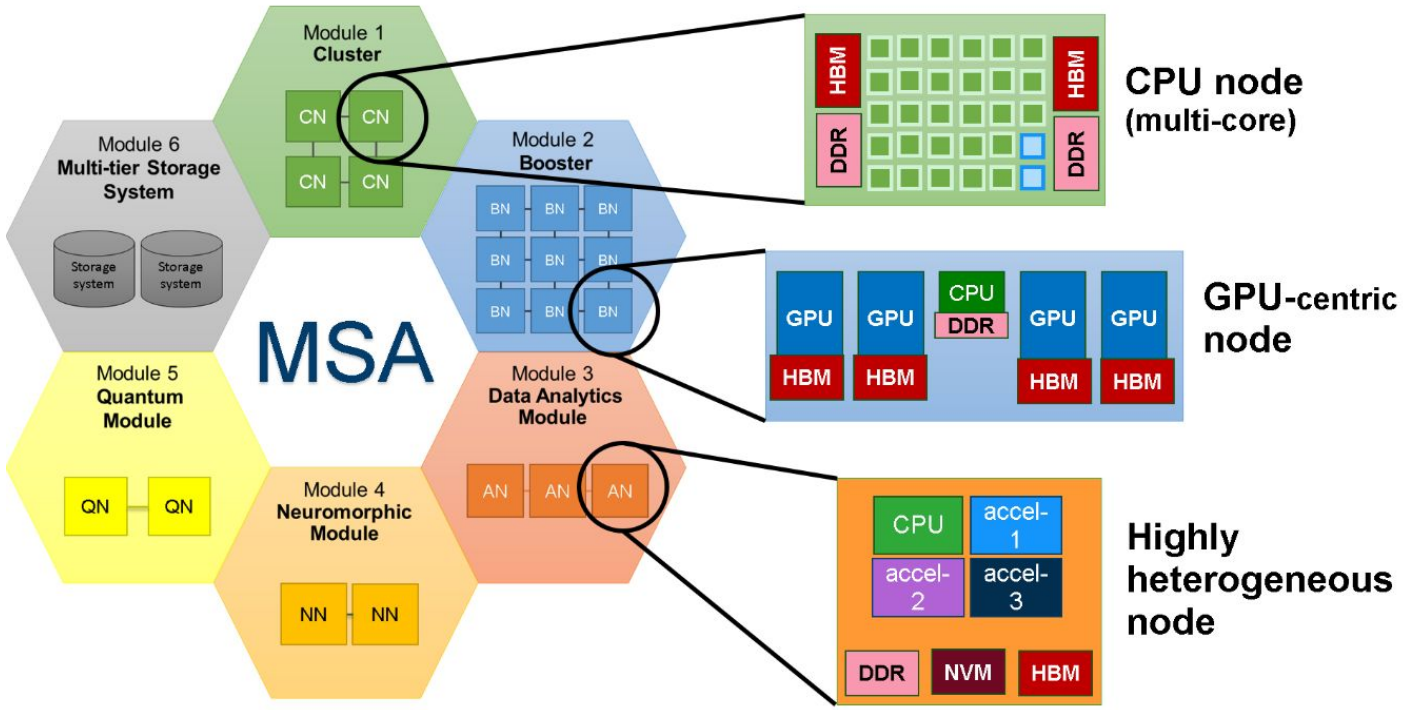
- Frontier (HPE/AMD) @ ORNL
- 1.102.000.000.000.000.000 FLOP/S
- 8.730.112 cores (8.138.240 accelerated)
- 21,1 MW

EuroHPC pre-exascale supercomputers are the current targets to focus on:
LUMI, LEONARDO and MARENOSTRUM 5 …

… with **petascale** EuroHPC supercomputers: VEGA, MELUXINA, KAROLINA, DISCOVERER and DEUCALION

JUPITER: The Arrival of Exascale in Europe

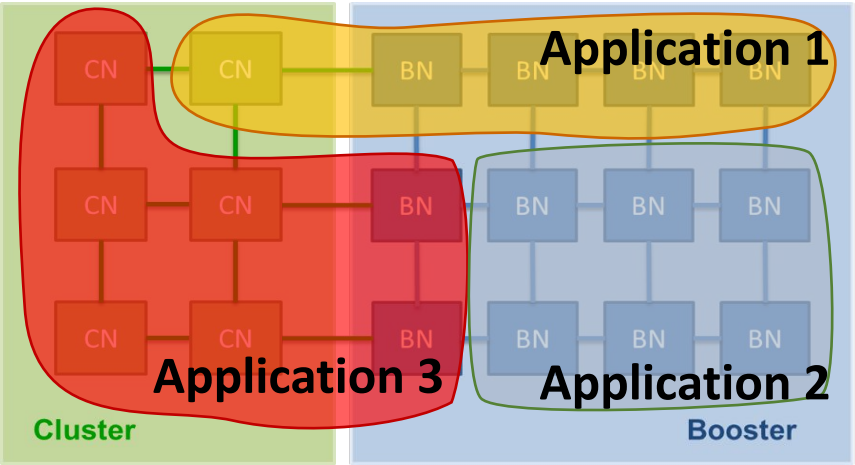
Modular Supercomputer Architecture



In MSA each module is tuned to best match the needs of a certain class of algorithms

(Some) Challenges at the Exascale

- Abrupt technology changes
- Hardware heterogeneity
  - CPUs, GPUs, APUs…
  - HBM, SSD…
- Level of parallelism
  - O(10^{18}) flops/s, Bytes



N. Eicker et al. Concurrency and comp. 28 (2016) 2394

- Novel numerical approaches
  - Low scaling algorithms, highly scalable computations, mixed precision
- Modularity
  - Single simulation → integrated workflows
  - Simulation + AI + Data Analytics + …

- Resilience
  - Technical support for codes on exascale systems
  - Deployment and tuning codes and workflows
  - Checkpointing long runs
- Programming – Performance Portability
  - MPI, OpenMP, CUDA, OpenACC, HIP, OneAPI …

European Processor Initiative (EPI)

- BSC leads the RISC-V HPC accelerator development in EPI
- EPI is a 100% funded EuroHPC project (120 M€) to develop European processor technology
- BSC was the original initiator of EPI and most active proponent in the scientific and technical community
- EPI is led by Atos/Bull with 28 partners from leading HPC industrial and academic centres

European Processor Initiative

Strategic importance to attain European chip sovereignty

The EPI Chip Roadmap

EPI 1

2019-2021

2022

Rhea platform

EPAC V1.5 sent for manufacturing

KVX RISC-V based accelerator architecture definition

Menta FPGA chiplet architecture definition

Rhea Family - Gen1 GPP

EPI Common Platform
Arm & Risc-V (STX, VRP, ..)
Arm Neoverse V1 Core – N6
External IP's

EPAC V1.5 multi-node demonstration cluster
EPAC V2.0 sent for manufacturing

Menta FPGA chiplet sent for maufacturing

Kalray RISC-V SDK release

2023

Rhea2 Family - Gen2 GPP

Dual chiplet implementation
First EU Exascale system with Rhea processors
FPGA Die-to-Die Demonstrator of the HW Common Platform

RISC-V KVX FPGA emulator

EPAC V2.0 Platform

2024

Cronos Family - Gen3 GPP

EU Exascale system with Rhea2 processors

Hurricane PCIe Acceleration card based on EPAC technology

2025

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Co-Design for GW methods

Motivation & Objectives of Co-Design

- **Prepare aiCMS codes for exascale and modular HPC computing**
  - Processor level (CPU): memory characteristics (heterogeneity, bandwidth), vector length, etc.
  - Node level: # sockets/accelerators per node.
- **Provide technology developers with realistic data**
  - optimal # cores & accelerators, memory, etc.
- **Get ready for advanced hardware platforms**
  - EU RISC-V ecosystem (co-design at the Instruction Set Architecture level).
- **Foster collaboration between software and hardware developers**

Main challenges of Co-Design

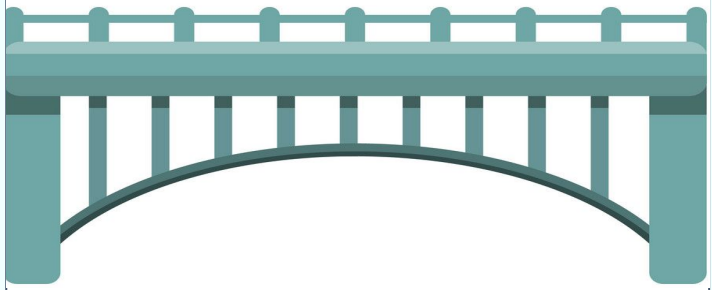
- Performance analysis of relevant kernels on experimental HPC hardware due to code complexity.
- Effective communication between software and hardware developers

Mini-apps as a bridge for Co-Design

Software developers

- Work on complex production codes
- Want to benchmark real-world systems
- Need insights on how to take advantages of the available hardware

Mini-apps



Hardware engineers

- Design algorithm-specific computing architectures
- Want to benchmark simple non-redundant algorithms
- Need fast insights on the performance of specific algorithms

Co-Design Framework for GW implementations

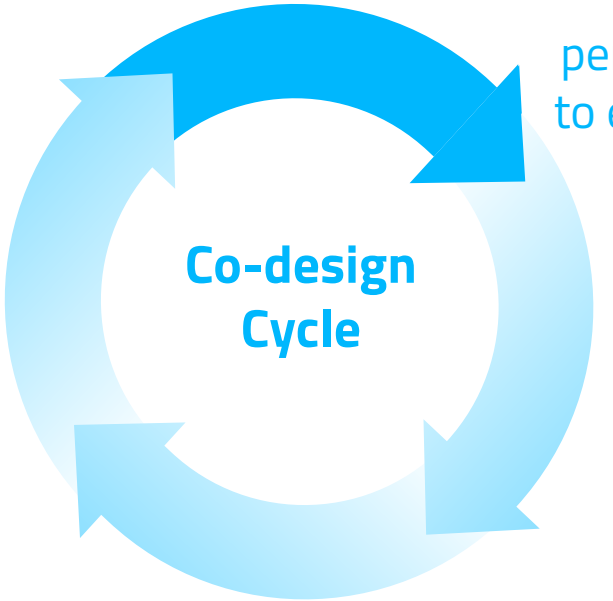
Mini-Apps Facilitate Co-Design

- High portability to experimental HPC systems
  - **Proxies for relevant computational kernels**
  - **Minimum** volume of **profiling data** & **execution time**
  - **Simple compilation**
- Rapid testing of new versions of the kernels
- Communication vehicle between software and hardware developers

Distributing mini-apps to different HPC architectures

Isolating relevant computation kernels into mini-apps

Providing performance insights to engineers and HPC manufacturers



Profiling and analysing performance metrics of production codes

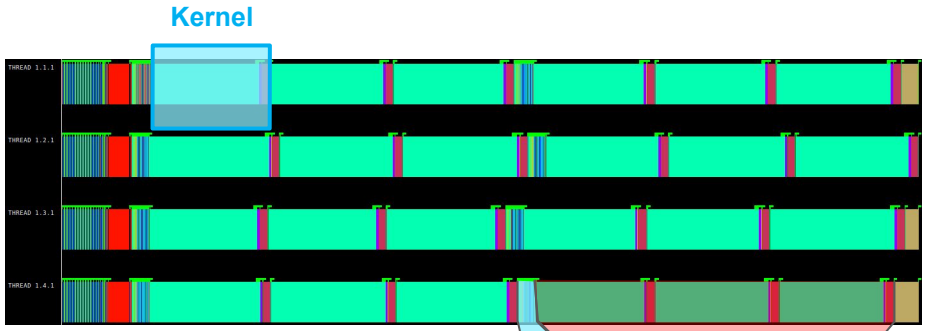
Modifying kernels code to run efficiently on target architectures

Mini-App Development

4 nodes (4x48 threads) at MN4

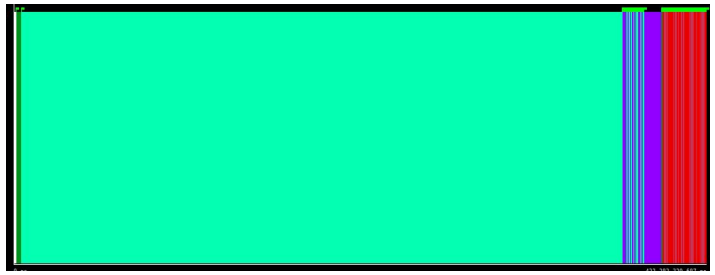
1 node (48 threads) at MN4

ZrO_2 all bands (54 min)

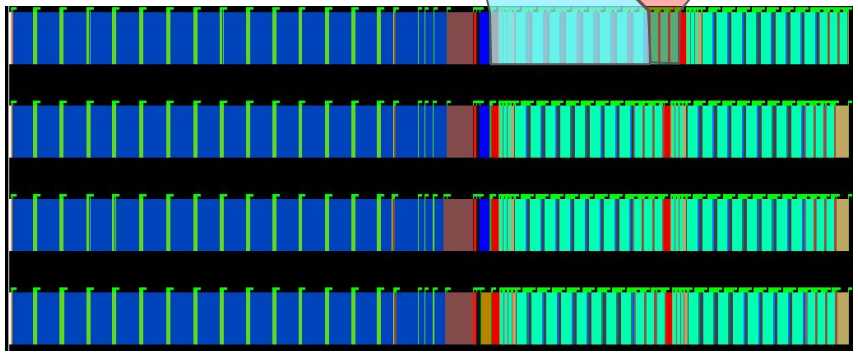


Kernel isolation

~7 min

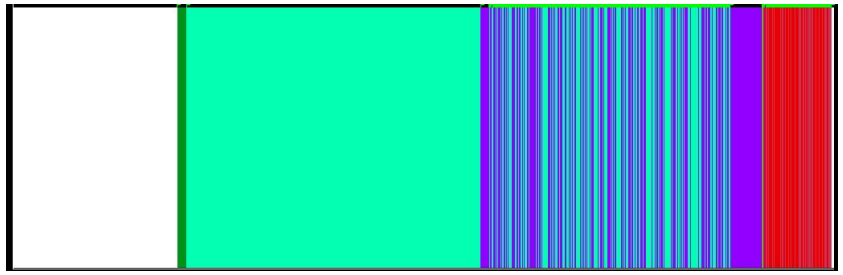


	THREAD 1.1.1
expand_products_	3,951,239,197 ns
calcminm2_	387,657,700,308 ns
mod_mini..emm_tmp_	12,111,175,962 ns
calcminc_	268,859,195 ns
calcmwm_	76,521,476 ns
mod_mini..emv_tmp_	26,950,958,673 ns
Total	431,016,454,811 ns



Kernel isolation

~4 s



	THREAD 1.1.1
expand_products_	56,406,736 ns
calcminm2_	2,779,636,167 ns
mod_mini..emm_tmp_	432,441,987 ns
calcminc_	4,661,088 ns
calcmwm_	11,014,899 ns
mod_mini..emv_tmp_	360,135,327 ns
Total	3,644,296,204 ns

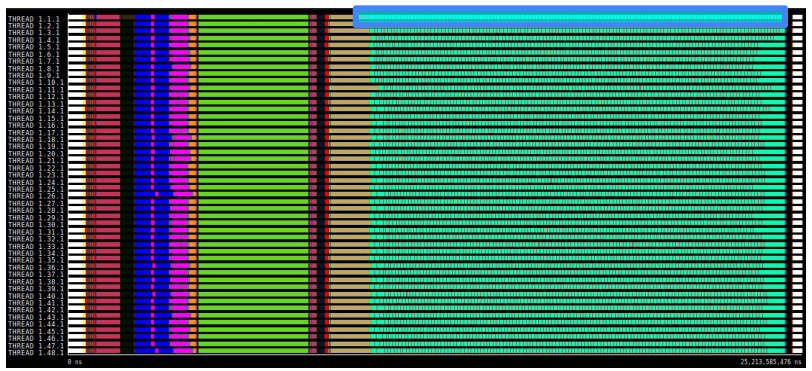
ZrO_2 reduced bands (5 min)

Source file: calcselfc.f90

Mini-App Development

ZrO2 benchmark (1 q-point)

Kernel

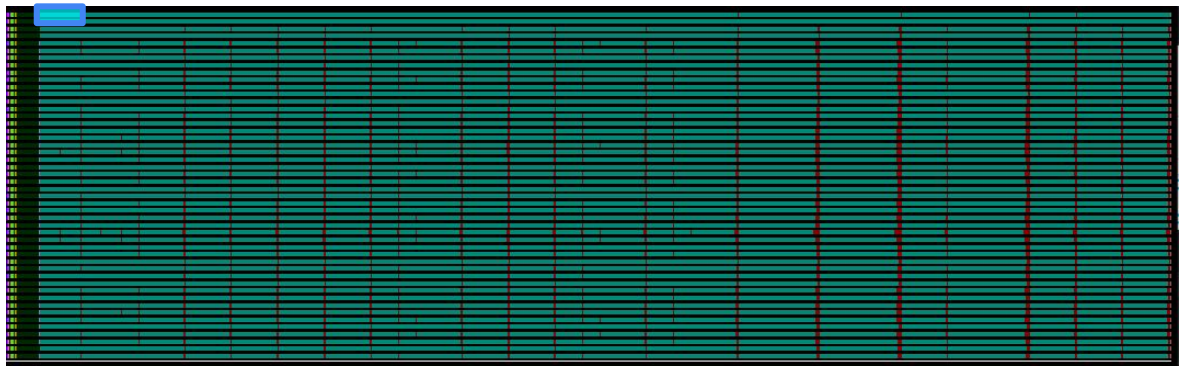


q-point reduction

ZrO2 benchmark (35 q-point) running in 48 MPI processes in MN4

ABINIT full-code execution **(15 min)**

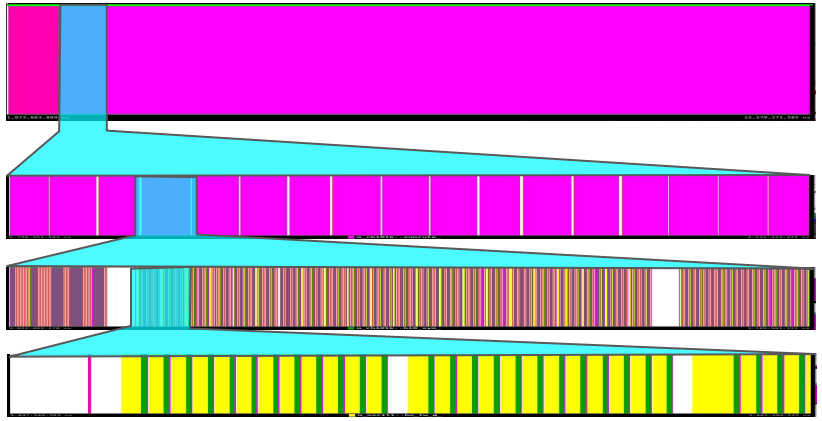
Kernel



Source code: **chi0.F90**

Kernel isolation

ABINIT mini-app



14 s

k-points and bands iterations to calculate χ_0

ZrO2 mini-app running in 1 MPI process in MN4

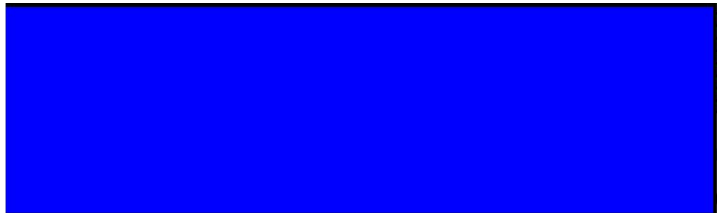
m_bz_mes..bz_item_	188,331 ns
m_bz_mes..bz_diff_	10,172,252 ns
m_chi0tk..sumrule_	211,467,435 ns
m_chi0tk..hi0_sym_	2,930,783,486 ns
m_oscill..ho_tw_g_	7,831,687,333 ns
Total	10,984,298,837 ns

Mini-App Development

Sitest (48 MPI processes) **(6 min)**

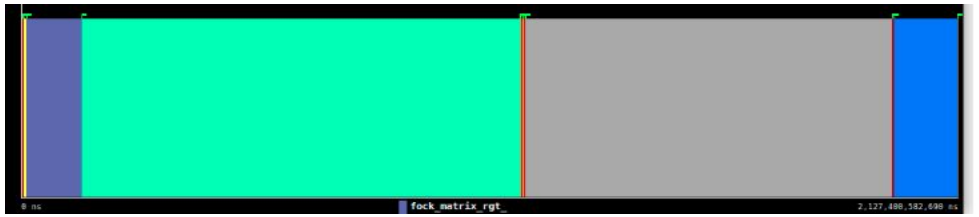
Kernel isolation

Sitest Mini-App **(1.5 min)**



read_checkpoint	990.26 us
sgw_init_lvl_tricoeff_recip	89,013,828.56 us
Total	89,014,818.81 us
Average	44,507,409.41 us
Maximum	89,013,828.56 us
Minimum	990.26 us
StDev	44,506,419.15 us
Avg/Max	0.50

ZrO_2 (384 MPI processes) **(36 min)**

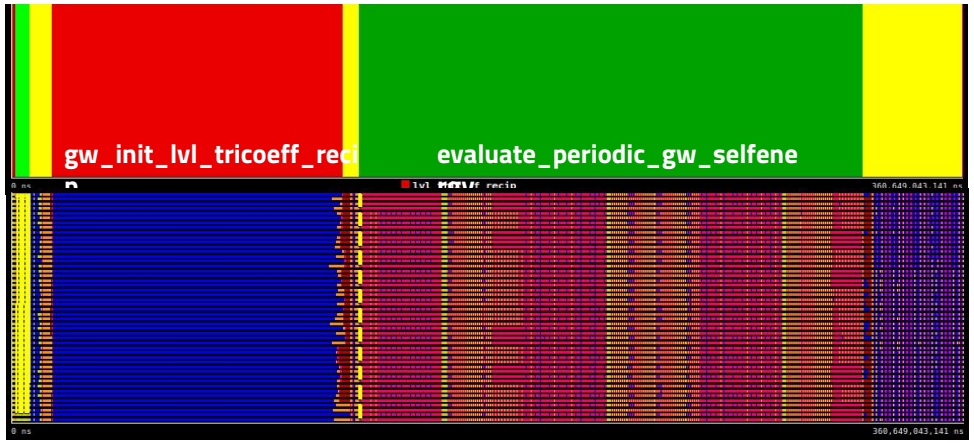


Construction of the Auxiliarity Basis Functions (ABF)

$$\psi^*_{m,\sigma}(\mathbf{r})\psi_{n,\sigma}(\mathbf{r}) \approx \sum_{\mu} C^{\mu}_{m,n,\sigma} P_{\mu}(\mathbf{r}), \qquad (6)$$

It is relevant for other methods

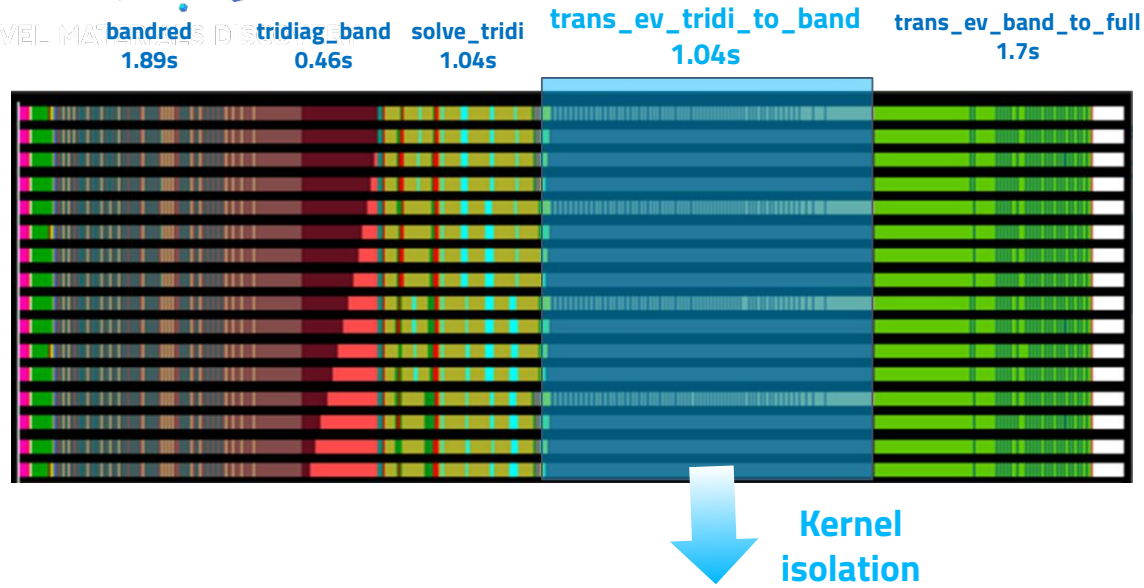
Sitest benchmark (48 MPI processes)



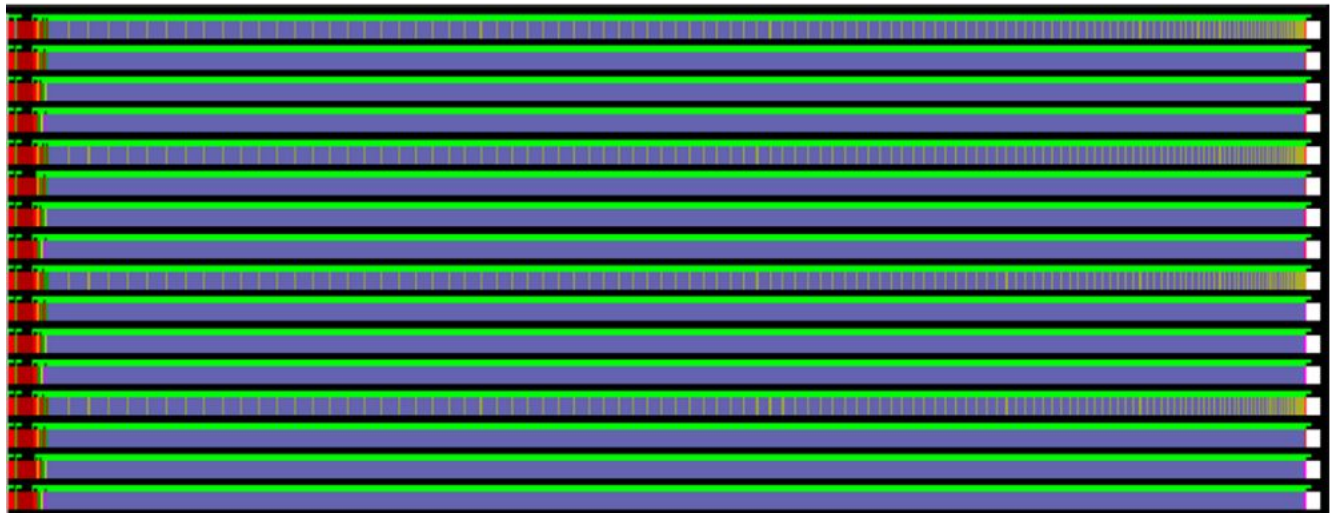
XINGUO REN *et al.*

Our ABFs are also atom-centered, and they are constructed in a way similar to the mixed product basis in the LAPW framework, but without the plane-wave component in the interstitial region. In Ref. [59] we have described how to make use of the expansion in Eq. (6) to achieve efficient implementations of Hartree-Fock, second-order Møller-Plesset perturbation theory (MP2), the RPA, and *GW* within the NAO basis set framework. Approximations based on Eq. (6) are known as density fitting [75,76], or RI [77–79] in the context of evaluating two-electron Coulomb repulsion integrals,

Mini-App Development



ELPA mini-app: isolated **trans_ev_tridi_to_band** subroutine



ELPA run: matrix size 8000x8000.
Runtime: (8.444 s (16 one-threaded MPI processes)

Not vectorized mini-app / portable

my_mpi_send_	20,234,349 ns
my_mpi_recv_	45,852,380 ns
miniapp_..and_mat_	133,947,949 ns
my_mpi_wait_	167,022,498 ns
MAIN__	231,619,272 ns
compute_.._double_	9,053,366,388 ns
Total	9,697,508,245 ns
Average	881,591,659 ns
Maximum	9,053,366,388 ns
Minimum	4,103 ns
StDev	2,585,230,313 ns
Avg/Max	0

Mini-Apps General Features

- Checkpoint size ~ **90-360 MB** (binary files)
- Minimal number of modules (**BLAS** & MPI*)
  * Only for ELPA Mini-App
- 1:1 reproduction of the original kernels
- Matching numerical results to original executions

Mini-App Suite Distribution

Features

- CMake build system

```
cmake -B build && make -C build
```

- Easy compilation and execution of the tests.
- Already tested compatibility for different platforms
- Release paper in preparation
- **Available on GitHub**

https://github.com/nomad-coe/co-design-WP8

master / nomad-mini-apps /

README | Add LICENSE | Add CHANGELOG | Add CONTRIBUTING | Add Kubernetes cluster | Se

Name	Last commit
benchmarks	Changed checkpoints from HDF5 to Binary
results	Added results directory
src	Added some informative prints to exciting and abinit mi...
.gitignore	We have moved find_package to the root CMAKE file
CMakeLists.txt	This projects does not depend on HDF5 anymore
README.md	Added instructions to dowload the checkpoints

Already tested HPC systems

Machine \ Compiler	INTEL	GNU	XLF	PGI	NVIDIA
MN4	✓	✓			
CTE-POWER		✓	✓	✓	
Karolina IT4I	✓				✓
Leonardo	✓				

Collaborations are welcome!

Summary & Take aways

Current work focused on testing

- Testing on EuroHPC machines as they become available
  - MN5
  - Leonardo
  - …
- Testing on prototypes: RISC-V (European Processor Initiative)
- In touch with companies
  - Nvidia
  - ATOS

Promote usage of the Mini-App Suite for Co-Design

- **Paper in preparation**

We are open to contributions

NOMAD

NOVEL MATERIALS DISCOVERY

Barcelona Supercomputing Center

Centro Nacional de Supercomputación

Thank you!

DFT & beyond: The NOMAD CoE

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

CHALLENGES IN MATERIALS SIMULATIONS

So far, Density Functional Theory has been the workhorse of ab initio computational materials

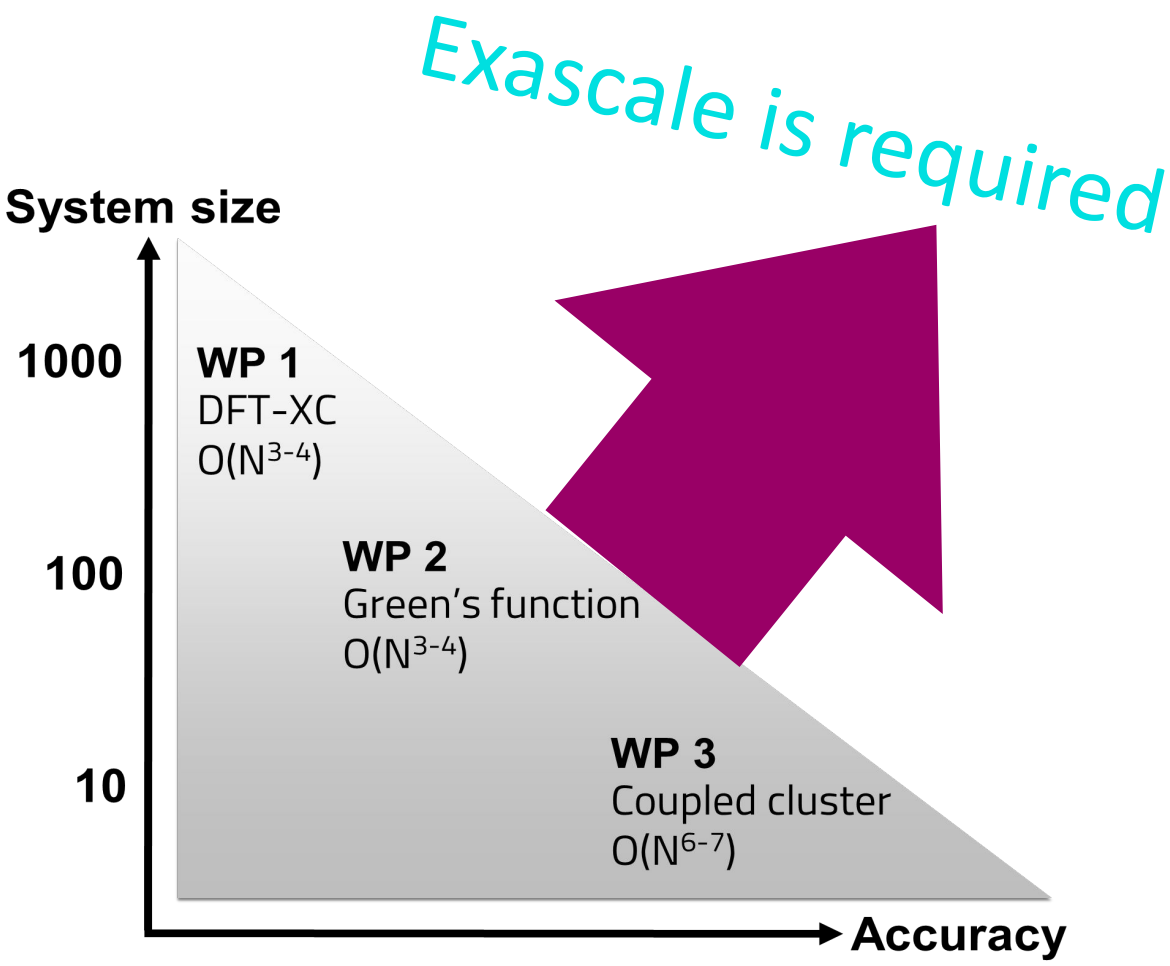
Many materials and/or properties require methods better than DFT

- Energy research, optics...

Complex systems require larger simulation cells

… and/or better methodologies

- Green´s function (GW), Coupled cluster theory
- Larger computational cost than DFT



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Source: Matthias Scheffler (FHI) NOMAD-CoE

Co-design: applications for materials science

- **Materials science is a very strong and relevant use-case**
- **Ab initio materials science codes should influence system design/procurements**
- **Influencing hardware design is difficult, especially for the HPC community**
  - Examples of good practice: EPI, RISC-V, Fugaku
- **Potential as a co-design vehicle for next-gen hardware**
  - Market is more malleable
  - Interaction with EU developed exascale prototypes
  - Prepare codes for future hardware

WP8 Structure & Collaborations

EU HPC Centres

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

MPCDF

MAX PLANCK COMPUTING & DATA FACILITY

WP1

CSC

ab-initio Computational Materials Science

Abinit

FHI-aims
The ab initio materials simulation package

WP2

exciting

GPAW!

ELPA: eigensolver library

guide development of future exascale systems

GW implementations

exciting: all-electron LAPW
FHI-AIMS: all-electron, numeric atom-centred orbitals
ABINIT: PAW, planewaves
GPAW: PAW, LCAO/planewave basis

NOMAD CoE receives funding from the European Union's Horizon program under the grant agreement Nº 951786.