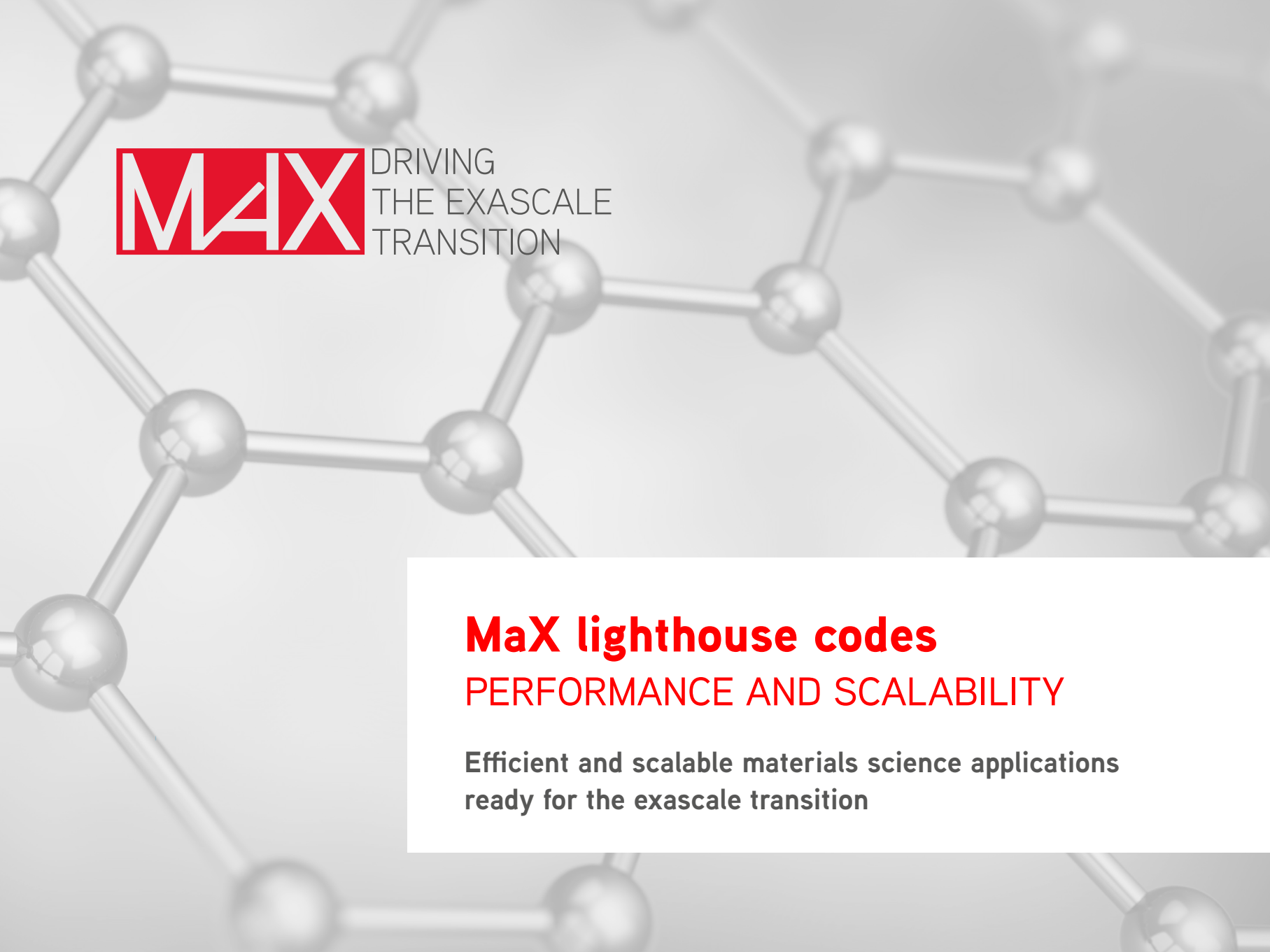
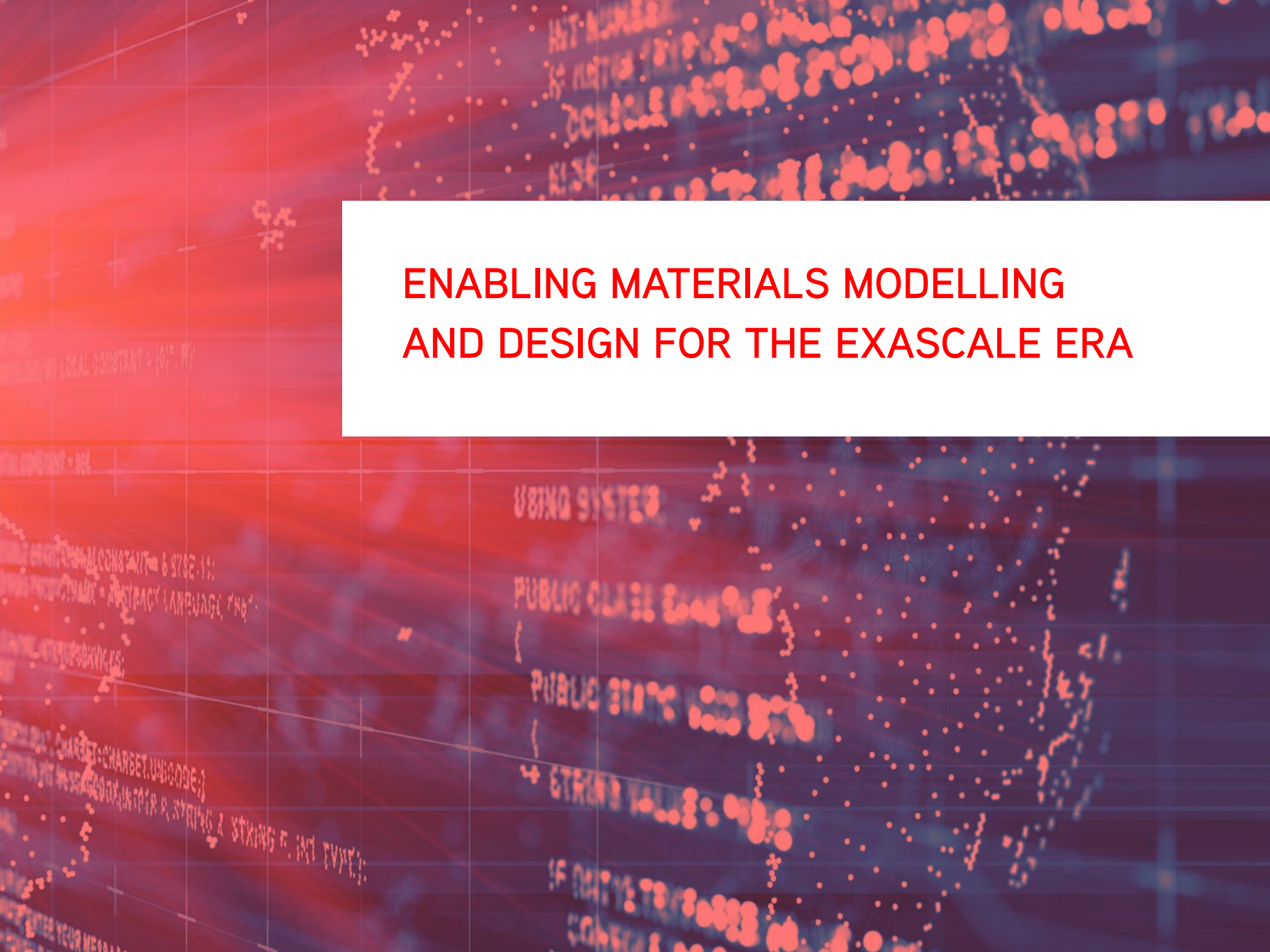



# MaX lighthouse codes

## PERFORMANCE AND SCALABILITY

**Efficient and scalable materials science applications ready for the exascale transition**



# ENABLING MATERIALS MODELLING AND DESIGN FOR THE EXASCALE ERA

## MaX—Materials design at the eXascale is a European Centre of Excellence that enables materials modelling, simulations, discovery, and design at the frontiers of current High-Performance Computing (HPC).

Our mission is to develop and optimise **materials science applications** by implementing state-of-the-art algorithms for quantum mechanical materials simulations based on density functional theory and many-body perturbations theory.

**MaX encompasses five lighthouse codes: Quantum ESPRESSO, YAMBO, SIESTA, BigDFT, and FLEUR.**

All of our codes are **open-source.** They are tested, tuned, and supported by many experts, including materials and computational scientists. Our codes are efficient, robust, and flexible. They are regularly upgraded with new features (scientific capabilities, performance, and robustness algorithms) to progressively address the *grand research challenges* emerging in computational and materials science. **The codes are interrelated and complementary; they are ready for interoperability**, allowing users to integrate them into complex workflows to calculate advanced materials' properties.

Today, MaX codes can exploit the most recent technologies and the computational power expressed by accelerated systems, leveraging NVIDIA and AMD GPUs. **The codes are running on most of the heterogeneous EuroHPC systems with high parallel efficiency**, offering new opportunities for HPC research and development in materials science.

The brochure covers the main features of the MaX lighthouse codes. Their **outstanding results in achieving performance and scaling on different types of EuroHPC architectures** are of particular interest for both computational and materials scientists. The brochure comes with illustrations and descriptions of results.

Thanks to the increased performance of its lighthouse codes and their orchestration in exascale-oriented workflows, **MaX promotes a new paradigm to solve outstanding scientific problems at unprecedented scales and degrees of accuracy**, looking forward to the exascale transition in Europe.

We invite you to explore the main features of the MaX lighthouse codes and to compare their performance and scalability on different types of EuroHPC architectures.

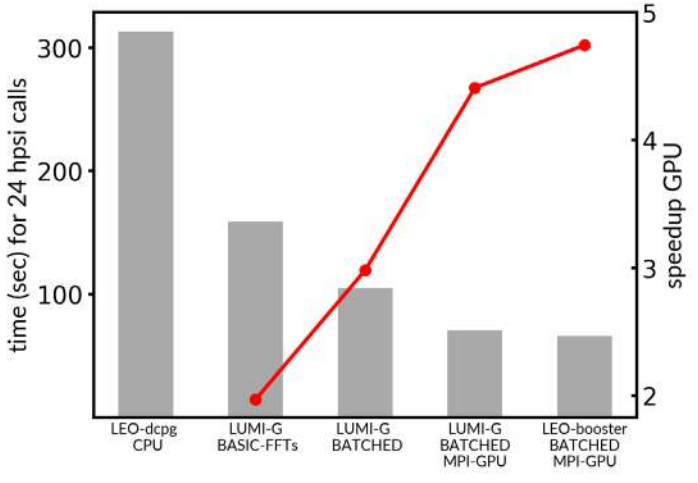

**QUANTUM ESPRESSO is a widely used software for atomistic and electronic structure calculations. It is an open-source collection of codes based on density-functional theory, plane waves, and pseudopotentials, offering portability, modularity, and high performance. The software comprises core components, plug-ins, and applications for energy and force computations, phonons with linear response, excitation spectra with TDDFT, and many others. It encourages usage and contributions from researchers worldwide.**

## PERFORMANCE

QUANTUM ESPRESSO suite supports many types of processors, accelerators, and architectures. It is designed for use on laptops and exascale supercomputers. Regarding HPC clusters, the suite's components are organised with multiple parallelisation schemes and levels, allowing optimal usage of homogeneous and heterogeneous parallel clusters. This strength leads to increasingly improved times-to-solutions and energy-efficient usage of computational resources.

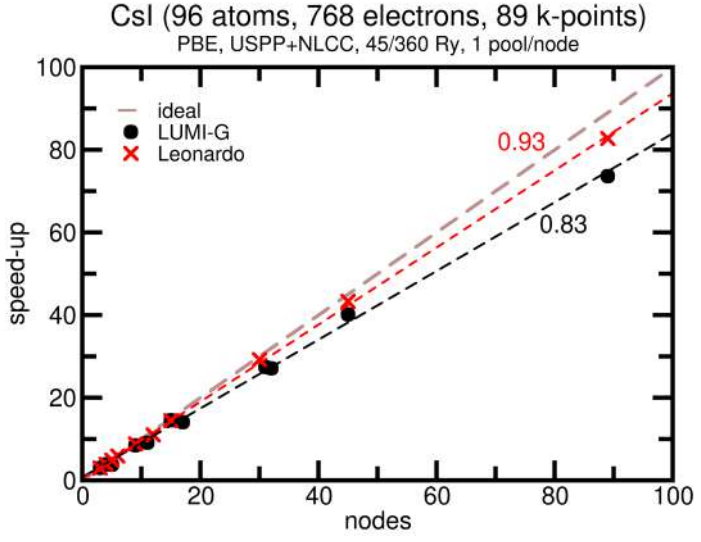
## PERFORMANCE EVOLUTION

Comparison of a middle-sized benchmark calculation on various parallel systems. The plot shows the performance of 2 nodes on a homogeneous parallel cluster (LEONARDO-DCPG@CINECA) compared to two nodes of two Tier0 heterogeneous parallel clusters (LUMI-G and LEONARDO-BOOSTER@CINECA). On LUMI-G, we show the results of several implementations of the communication scheme to demonstrate the incremental improvements of AMD-Hip-based parallel accelerated machine support over the last two years.




## STRONG SCALING SPEEDUP

Benchmark of pool parallelism scaling efficiency of Quantum ESPRESSO kernels in heterogeneous parallel machines with NVIDIA A100 GPUs (LEONARDO-BOOSTER@CINECA) and AMD MI250x GPUs (LUMI-G@LUMI). For small-size cases, the Quantum ESPRESSO kernels in heterogeneous machines are already very efficient with one node. The suite provides several throughput parallelisation schemes to speed up these calculations.



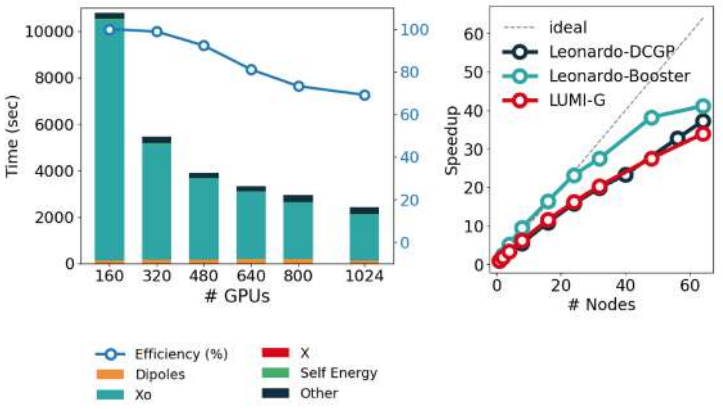

YAMBO is an open-source code released within the GPL licence implementing first-principles methods based on Green's function theory to describe excited-state properties of realistic materials. These methods include the GW approximation, the Bethe-Salpeter equation (BSE), electron-phonon interaction and non-equilibrium Green's function theory (NEGF), allowing the user to compute a wealth of physical properties, including reliable band gaps, band alignments, defect quasiparticle energies, optical and non–equilibrium properties.

## PERFORMANCE

YAMBO is parallelised using an hybrid MPI-OpenMP approach, integrated with support for GPU-accelerated hardware. This allows the code to distribute the workload to a number of parallel levels, both in terms of computation and memory usage. The GPU porting exploits different programming models (CUDA-Fortran, OpenACC, or OpenMP5) and targets different GPU-accelerated hardware (including NVIDIA, AMD, and INTEL GPUs).

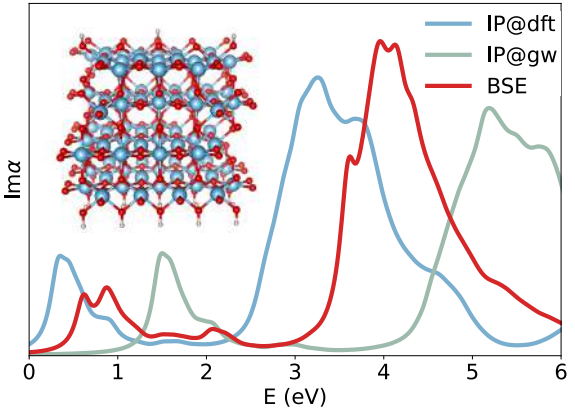
## STRONG SCALING AND SPEEDUP

Calculation of quasi-particle corrections on a graphene/Co interface (GrCo). The test involves the evaluation of the response function, Hartree-Fock, and correlation self-energy, treated at the GW level. On the left, the scalability test on LEONARDO-BOOSTER (CINECA), while on the right a speedup comparison between LEONARDO-DCGP, LEONARDO-BOOSTER and LUMI-G (CSC) for the same system but with reduced parameters.




## OPTICAL ABSORPTION SPECTRA

Optical absorption spectra computed at different levels of theory using YAMBO. IP@dft, IP@gw: independent particle response, evaluated on top of Kohn-Sham DFT and GW electronic structure. BSE: response function from the Bethe-Salpeter equation. The system studied is a defected (hydroxylated) TiO2 slab, relevant for photocatalysis, counting 210 atoms in the unit cell.



siesta

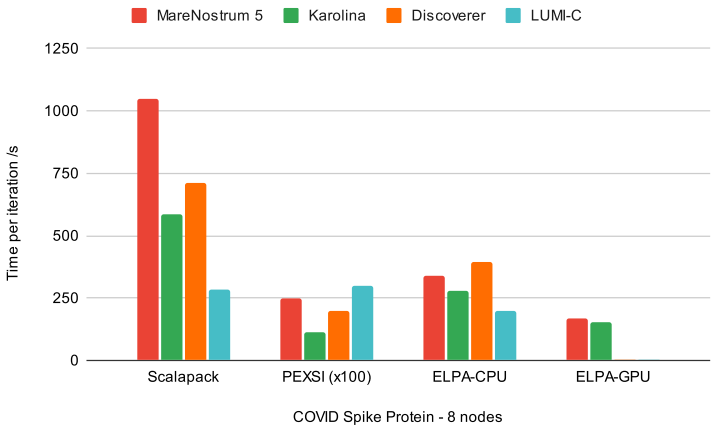
SIESTA is a DFT code designed from the outset to perform efficient simulations for large systems. SIESTA's efficiency stems from the use of a basis set of strictly-localised atomic orbitals, which leads to moderate matrix sizes in diagonalization and enables the use of low-complexity methods (e.g., linear scaling and the PEXSI scheme) that exploit sparsity. A very important feature of the code is that its accuracy and cost can be tuned in a wide range, from very fast exploratory calculations to highly accurate simulations matching the quality of other approaches, such as plane-wave methods.

## PERFORMANCE

SIESTA supports very diverse hardware architectures, including GPU acceleration via the ELPA library and massive parallelization with the PEXSI solver. SIESTA's cost-effectiveness enables DFT calculations of hundreds of atoms even in personal computers. In HPC infrastructures, SIESTA can deal easily with calculations for thousands of atoms, giving it a distinct competitive advantage.

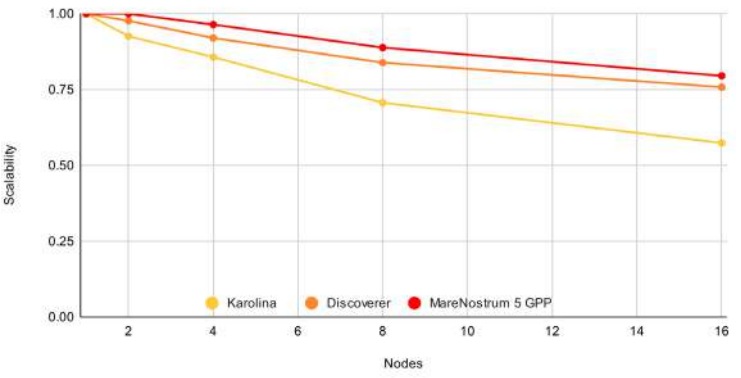
## SOLVERS AND ARCHITECTURES

Time per solver step for a system consisting of a COVID protein in water with 8,783 atoms and 57,647 orbitals. The calculation was performed using 8 nodes in each of the EuroHPC JU partitions evaluated, with various solvers available in SIESTA.




## STRONG SCALING

Strong scaling in different EuroHPC JU supercomputer CPU partitions using the ELPA solver. The scalability measures the efficiency of the use of computational time when compared to a single node on the same partition. Work on fine-tuning performance on large numbers of GPU nodes is ongoing. The system is the same COVID protein in water described earlier.



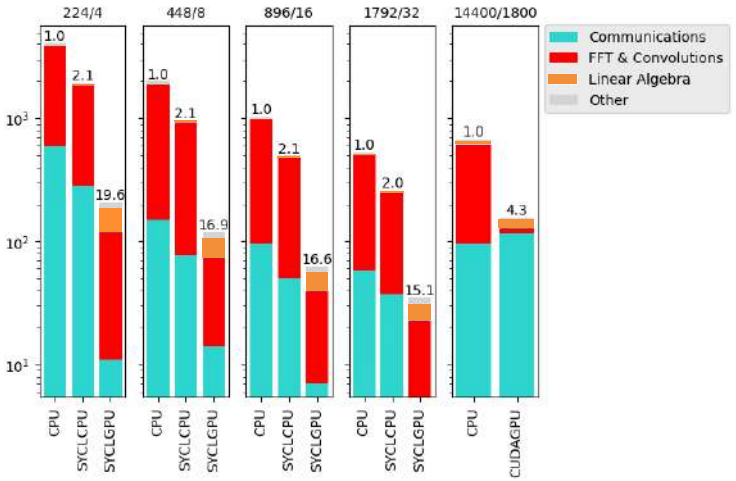

BigDFT suite offers a variety of features, ranging from various ground-state DFT algorithms up to potential energy surface exploration techniques. BigDFT uses dual space Gaussian type norm-conserving pseudpotentials including those with non-linear core corrections, which deliver all-electron precision on a number of quantities. Its flexible poisson solver can handle a number of different boundary conditions including free, wire, surface, and periodic. The code simulates implicit solvents and external electric fields. The code possesses a Linear-Scaling algorithm enabling unbiased DFT simulations of systems up to many thousand atoms.

## PERFORMANCE

BigDFT employs a mathematical formalism based on Daubechies Wavelets, suitable for efficient usage of modern HPC platforms. The multi-level parallelization of BigDFT (shared, distributed, accelerated parallelization) allows the code to scale to thousands of cores and to run efficiently on the most powerful supercomputers.

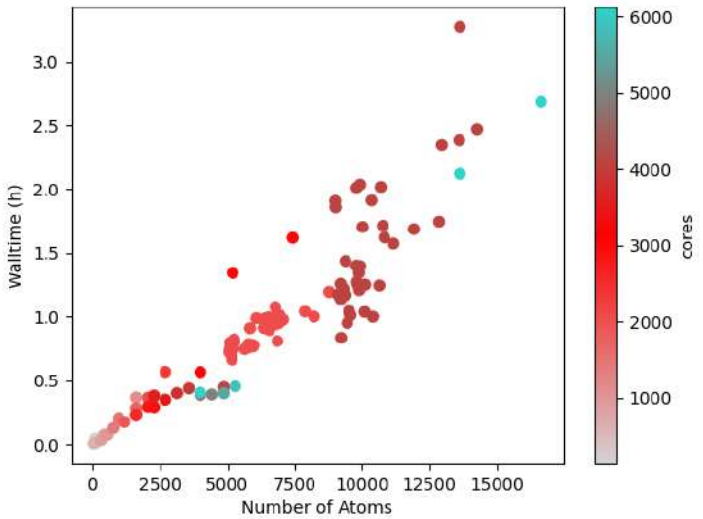
## GPU ACCELERATION EFFICIENCY

Capability of BigDFT code to efficiently employ GPU acceleration. The acceleration of BigDFT thanks to the SYCL and CUDA programming paradigms of UO2 bulk systems (PBE0 functional) is presented (CPU cores/GPU in the panel title, speedup above bars). The first four panels on the left show the recent results on PVC Intel GPU cards, with a system of 1432 electrons. The last panel on the right illustrates previously obtained results on NVidia architectures, with a system of 5400 electrons, presenting a computational workload which is 14 times larger. CPU SYCL acceleration is also presented.




## LINEAR SCALING BEHAVIOUR

Linear Scaling behaviour of BigDFT code. The total walltime needed for a converged ground state calculation of a variety of systems (OLED morphologies, protein dimers) is represented as a function of the number of atoms. The total number of CPU cores employed for the simulation is represented as a colour code. For those calculations the computational resources require, on average, 20 CPU minutes per atom.



www.flapw.de

# fleur

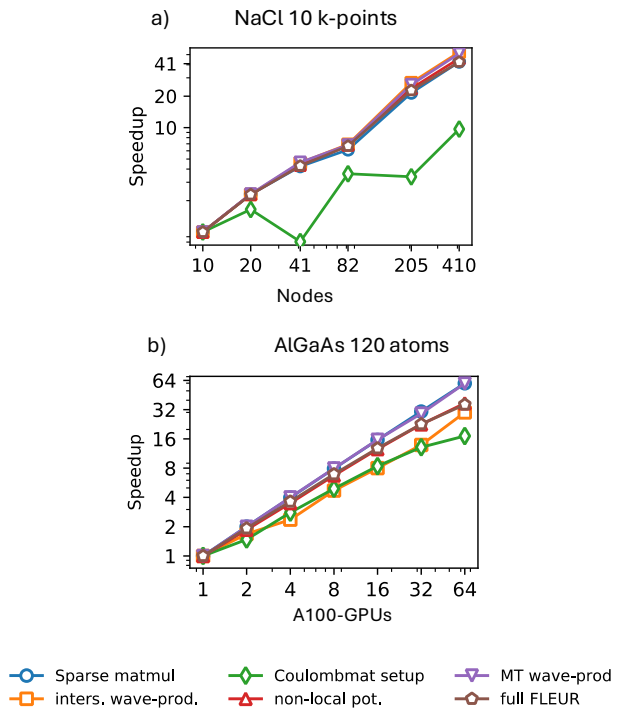
FLEUR is an all-electron density functional theory code based on the full-potential linearized augmented plane wave (FLAPW) method. It can be applied to crystalline solids of all chemical compositions in three-dimensional bulk setups as well as in two-dimensional films for the simulation of surface properties. Due to the highly accurate approach, the code is particularly well positioned to describe phenomena on small energy scales, like those arising in complex, non-collinear magnetism or due to spin-orbit physics.

## PERFORMANCE

FLEUR is able to utilise various CPU architectures as well as NVIDIA GPUs. Multiple parallelization levels are implemented to achieve optimal scaling, to be able to adjust the memory and computing demands to different hardware, and to enable scaling up to thousands of GPUs.

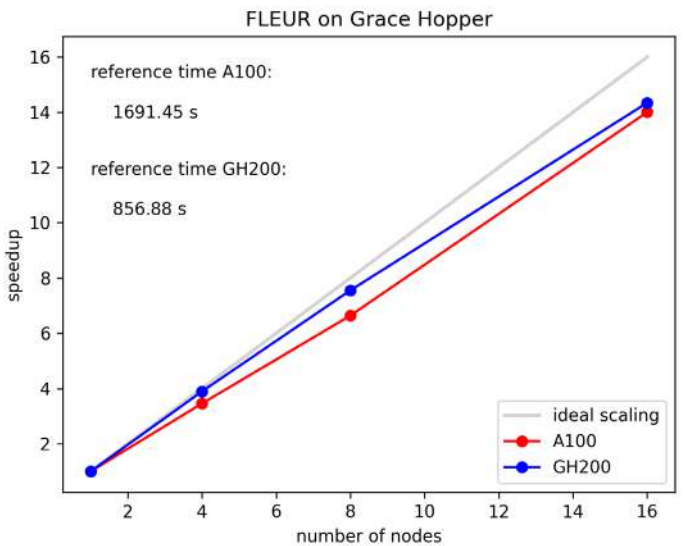
## STRONG SCALING SPEEDUP

Scaling of the hybrid functionals' implementation of FLEUR. a) shows the parallelization over independent kq-points, b) demonstrates the scaling of a single kq-point over multiple GPU (4 per node). Most parts of the code show excellent scaling, the Coulomb matrix setup requires only little computational time.




## PERFORMANCE COMPARISON

Comparison of the performance and the scaling of FLEUR on different NVIDIA GPU architectures. The measurements have been performed on JURECA-DC and the JEDI test system for the upcoming JUPITER exascale machine. Each node features 4 GPUs of either A100 or GH200 kind. While the scaling on both machines is similarly good, the higher single node performance of the GH200 system leads to a lower reference time.



## CONTACT


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

@MaXCentreeXascale

@max_center2

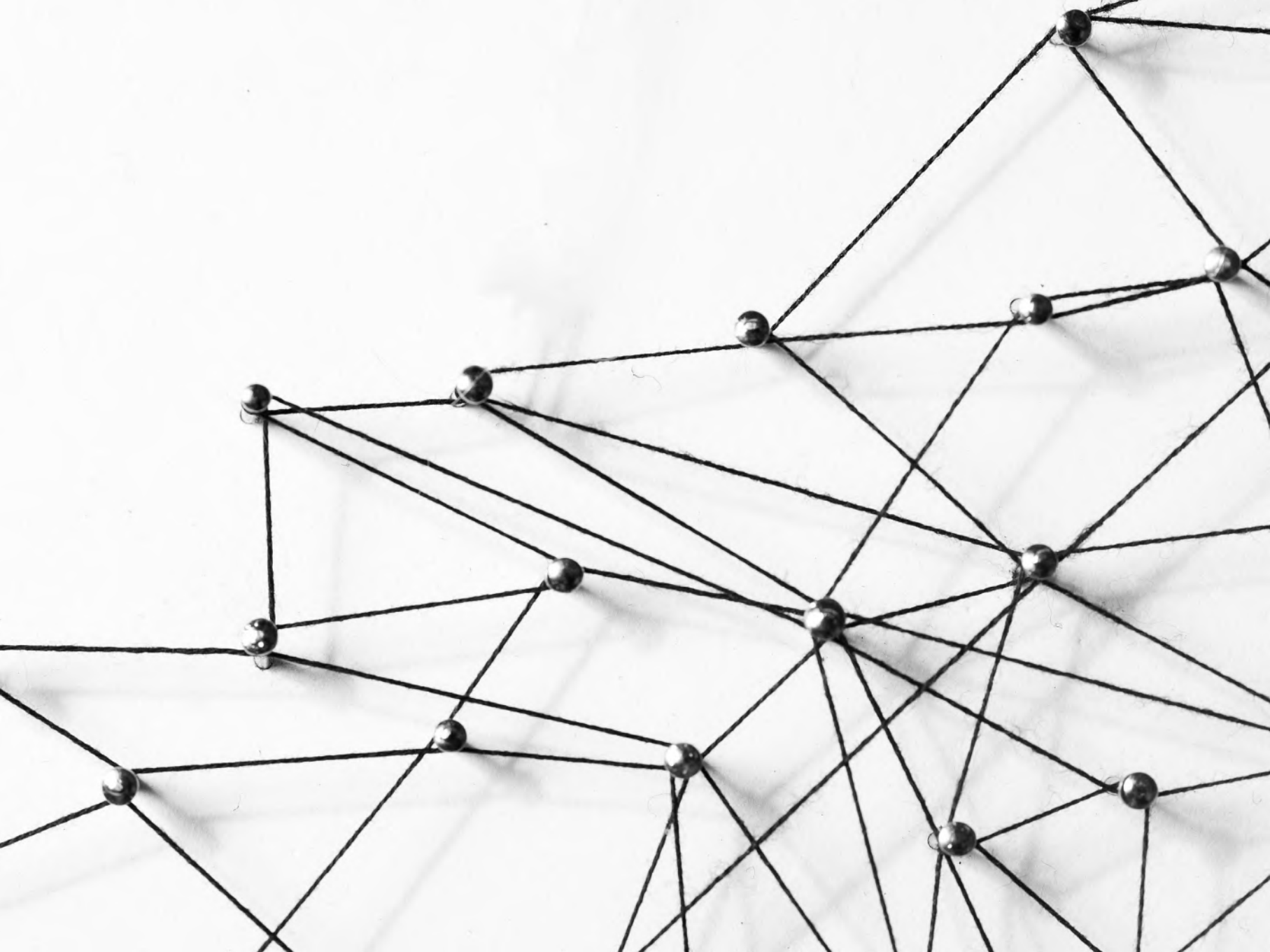
/company/max-centre


## ACKNOWLEDGEMENTS


MaX - Materials design at the Exascale has received funding from the European High Performance Computing Joint Undertaking and Participating Countries in Project (Czechia, France, Germany, Italy, Slovenia, and Spain) under grant agreement no. 101093374.

Co-funded by the European Union

EuroHPC Joint Undertaking




DRIVING THE EXASCALE TRANSITION