Ref. Ares(2022)7221580 - 18/10/2022

D6.3

Report on the simulations performed and scientific outlook

Pablo Ordejón, Riccardo Bertossa, Pietro Delugas, Jose María Escartín, Andrea Ferretti, Alberto García, Pol Febrer, Ernane de Freitas, Luigi Genovese, Stefano de Gironcoli, Anton Kozhevnikov, Savio Laricchia, Elisa Molinari, Deborah Prezzi, Nicola Spallanzani, Daniele Varsano, Daniel Wortmann

Due date of deliverable:	30/09/2022
Actual submission date:	18/10/2022
Final version:	07/10/2022
Lead beneficiary:	ICN2 (participant number 3)
Dissemination level:	PU - Public

Document information

Project acronym:	MaX
Project full title:	Materials Design at the Exascale
Research Action Project type:	European Centre of Excellence in materials modelling, simulations and design
EC Grant agreement no.:	824143
Project starting / end date:	01/12/2018 (month 1) /30/09/2022 (month 46)
Website:	www.max-centre.eu
Deliverable No.:	D6.3

Authors: P. Ordejón, P. Delugas, J.M. Escartín, A. Ferretti, A. García, P. Febrer, E. de Freitas, L. Genovese, S. de Gironcoli, A. Kozhevnikov, S. Laricchia, E. Molinari, D. Prezzi, N. Spallanzani, D. Varsano, D. Wortmann

To be cited as: Ordejón, et al., (2022): Report on the simulations performed and scientific outlook. Deliverable D6.3 of the H2020 project MaX (final version as of 07/10/2022). EC grant agreement no: 824143, ICN2, Barcelona, Spain.

Disclaimer:
This document's contents are not intended to replace consultation of any applicable legal sources or the necessary advice of a legal expert, where appropriate. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user, therefore, uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors' view.

D6.3 Report on the simulations performed and scientific outlook

Content

1 Executive Summary 2

2 Introduction: the MaX Demonstrators 3

3 Report on demonstrators 4

3.1 On-the-fly computations of energy currents with CP and QEHeat (Demonstrator Dem.1) 4

3.2 Excited state electronic properties in GPU-accelerated infrastructures (Demonstrators Dem.1 and Dem.7) 8

3.3 Defected Titanium-dioxide surfaces for photocatalytic applications (Demonstrator Dem.1) 13

3.4 Strength of antibody/antigen complex interactions using QM/mMM (Demonstrators Dem.2 and Dem.6). 18

3.5 Full DFT simulation of new particle-like objects in chiral magnets (Demonstrator Dem.3) 27

3.6 RPA and double hybrid MD simulations of condensed phase systems with CP2K (Demonstrator Dem.4) 31

3.7 Exciton-Phonon coupling (Demonstrator Dem.5) 32

3.8 Electrified electrode-electrolyte interfaces: QM and QM/MM with SIESTA (Demonstrator Dem.7) 34

3.9 High-Throughput Machine-Learned DFT-quality Potentials (Demonstrator Dem.8) 42

Data-driven simulation and analysis of gold nanoparticle melting 42

4 Conclusions 43

5 References: 45

1 Executive Summary

MaX's WP6 is devoted to integrating the advances of the technical WPs and demonstrating the deployment of the codes and workflows to tackle specific, realistic applications in the framework of HPC infrastructures. For that purpose, specific demonstration cases (examples of new science that will be enabled by the availability of pre-exascale and future exascale environments and by the codes and workflows able to exploit them) were proposed in MaX. These include two sets of demonstrators: (1) five topics that were proposed originally in our initial work plan, as detailed in the description of the activities of WP6, and (2) three new topics that were proposed by MaX midterm, as detailed in Deliverable D6.2. This deliverable reports on the advance in these demonstration projects.

2 Introduction: the MaX Demonstrators

A critical goal of MaX is to enable computation of materials properties on current and future pre- and exascale machines. The MaX technical work-packages (WP1 to WP4) focused on making the MaX Flagship codes ready to run efficiently in large supercomputers, while WP5 provided tools to reach exascale via high-throughput calculations through the AiiDA platform. In turn, WP6's goal was to integrate all these advances to test and demonstrate the deployment of the codes and workflows to tackle specific, realistic applications in the framework of HPC infrastructures available during MaX, and to prepare for future pre-exascale and exascale environments. Also, to show, by means of specific demonstration cases, examples of new science that will be enabled by the availability of pre-exascale and future exascale environments and by the codes and workflows able to exploit them.

For that purpose, five specific demonstration projects were proposed originally in Task 6.1 of our initial work plan, as reflected in the description of activities of WP6:

Dem.1. "On-the-fly" evaluation of complex materials properties with Molecular Dynamics
Dem.2. Reduction of the complexity in very large scale systems
Dem.3. Full DFT simulation of new particle-like objects in chiral magnets.
Dem.4. RPA and double hybrid based MD simulations of condensed phase systems
Dem.5. Coupled electron-phonon dynamics within NEQ–MBPT

The advance on these demonstrators and results achieved by mid-term of the project were reported in deliverable D6.1.

Additionally, Task 6.1.7 in our work-plan intended to work further on the identification and implementation of new demonstrators during the development of MaX, taking into account

interesting, new scientific problems, but also applications of the technologies developed within the MaX codes and workflows during the first stages of the project. As a result of this work, also by MaX mid term, deliverable D6.2 proposed three new study cases:

Dem.6. Addressing SARS-CoV-2 molecular constituents from electronic structure: fragment analysis with BigDFT and SIESTA
Dem.7. Ab-initio investigation of electrode-electrolyte interfaces
Dem.8. High-Throughput Machine-Learned DFT-quality Potentials

This deliverable 6.3 reports on the advance made in these eight demonstrators, explaining both the scientific results achieved, and the aspects related to deployment in HPC machines, and the interplay with the technical work-packages WP1-5 of MaX.

This deliverable is structured in two sections, one reporting on the work done in the five initial demonstrators with an update of the advances made since the release of deliverable D6.1, and another on the deployment and results of the three new demonstrators proposed in deliverable D6.2. In some of the cases, the scientific topics extended well beyond those originally proposed. In other cases, more efforts were devoted to the study of certain aspects of the scientific case than others, as they were considered more interesting or relevant. This report encloses all these topics and activities, even if they were not explicitly considered initially. For the sake of clarity in the presentation, we structure this report in nine topics, in each of which we indicate the relation with the eight demonstrators projects listed above.

3 Report on demonstrators

3.1 On-the-fly computations of energy currents with CP and QEHeat (Demonstrator Dem.1)

Our work on the on-the-fly combination (see section 3.1.4 in MaX report 6.1) of ab initio molecular dynamics (AIMD) with the energy flux calculator (see Ref. [MUB2016]) continued along two main directions. On one side we focused on the testing of the workflow using the tools already provided by the Quantum ESPRESSO suite Ref. [GIAN2017, MARC2021] (cp.x for the molecular dynamics and QEHeat for the computation of the energy flux at each snapshot); on the other hand we also worked at the optimization on the flux-calculator, also implementing a more advanced workflow where at each snapshot the charge density and the wave-functions were passed from the AIMD quantum-engine directly to the flux calculator.

For what concerns the first point we present here a set of calculations of the thermal conductivity of water at different temperatures and phases that have also been recently published in Ref. [TISI2021]. We simulated liquid water at 521 K, 421 K, and 409 K. We also simulated the hexagonal ice-Ih phase of water at 260 K. For the simulation we used the PBE functional approximation of DFT, with norm-conserving pseudopotentials. We used a kinetic-energy cutoff of 85 Ry. The liquid phases were simulated with a cubic supercell containing 125 water molecules, the hexagonal phase was simulated using an orthorhombic supercell containing 128 water molecules. The dynamics was performed using a time-step of 0.073 fs, and for each of these systems we computed a 100 ps long trajectory, and computing the energy flux at snapshots separated by 3 fs.

The time series of heat fluxes was then treated with the velocity renormalization technique (see Ref. [TISI2021] and references therein for more details) and then the thermal conductivity was calculated using the cepstral analysis [ERCO2017, BERT2019] of the power spectrum, using the SPORTRAN code [ERCO2022]. We report the computed values in Table I.

phase	Temperature (K)	thermal transport coefficient (W/mK)
liquid	516	0.99 ± 0.05
liquid	423	1.03 ± 0.05
liquid	408	1.11 ± 0.05
ice Ih	270	1.08 ± 0.04

Table I: Thermal transport coefficients computed with the AIMD + QEHeat workflow. The ice Ih phase was simulated using an orthorhombic cell containing 128 H_2O molecules, with sides A= 18.084 Å, B = 15.664 Å, C = 14.724 Å. The liquid phases were simulated using a cubic cell containing 125 H_2O molecules, with side A = 15.52 Å.

These calculations were run in the A3 partition of Marconi@CINECA cluster. We used 4 nodes for running the AIMD quantum-engine (cp.x) while a team of flux calculators (QEHeat) were executed concurrently, each one on 1 node. We verified that with this system size and this set up at least 8 flux calculators were needed to process on the fly all the snapshots generated by the dynamics.

The main bottleneck of this on-the-fly workflow is evidently represented by the flux calculator, but this bottleneck can be straightforwardly resolved, tuning the throughput capabilities of the flux calculators' team with the rate of generated snapshots.

The simulation has also been facilitated by the extensions to QEHeat that were introduced following the first period of testing activity. These changes, already presented in the MaX deliverable D6.1, have unified the "self-consistent" phase for computing the density and the wave functions, with the actual computation of the energy current. This has reduced the amount of disk I/O, and simplified the workflow facilitating its management, thus allowing us to execute a larger number of QEHeat instances.

A further step in increasing the workflow's efficiency is represented by the reuse of the density and wave-functions generated during the dynamics. This would reduce by about 33% the computation time taken by a single QEHeat run and thus reduce the number of concurrent instances of the flux calculator. For an effective and efficient implementation of such a feature for HPC clusters, because of the large amount of data-exchange involved, it is crucial to avoid the usage of standard storage and it is in many cases unfeasible to use the conventional memory. More actions on the codesign side and the exploration of more efficient techniques for the intensive data-exchange are thus necessary, and are actually part of the work-plan of the MaX CoE in the next few years.

For the moment, we have thus made a proof-of-concept implementation. For the data-exchange and synchronisation we have first written a high level fortran interface to the library ZeroMQ [ZEROMQ]. This interface (available at https://github.com/rikigigi/fortran_cloud) implements a transparent data-exchange via the TCP/IP stack, avoiding the usage of the file system. The data-exchange and the synchronisation is done exploiting the MPI groups hierarchy, which must therefore be organised accordingly. The hierarchy should select a group of "producer" tasks that generate the data-sets, and one process is charged with the distribution and synchronisation between the "producers" and the "consumers" team that receive and process the data-sets. Using such an architecture allows a broad choice of parallelization schemes. It is in principle also possible to combine different types of nodes, with some tasks running on accelerated nodes with GPUs, and others like the distributing task running on nodes with more memory. Such a scheme also allows inserting more nodes during the calculation, with the proxy being able to automatically tune producers' and consumers' throughputs and collect and distribute the data-sets.

Thanks to the interface we have been able to rapidly adopt the ZeroMQ message passing paradigm. The modifications needed to the application that produces the data to be shared are minimal. In practice a call to initialise the ZeroMQ context and few calls to our API to send the data are required. On the other side, the code that consumes the atomic positions and the wavefunctions needs only to initialise the ZeroMQ context and to call our API for receiving the data. In https://gitlab.com/rikigigi/q-e/-/tree/on_the_fly we have made available our experimental implementation of this paradigm for passing charge density and wave-functions from cp.x to QEHeat.

The proof of concept is also allowing us to investigate what is the effective impact of the fictitious thermal fluctuations of the electronic state extracted directly from the Car-Parrinello molecular dynamics (CPMD) on the accuracy of the energy flux calculation (that in principle requires well converged ground state density and wave-functions). We are investigating this point both from the theoretical and numerical sides. For example, we show in Figure 1 the comparison between the energy current computed with well converged self consistent wave-functions, and the calculation done using those extracted from the CPMD.

It is also possible to damp the fictitious thermal fluctuations of CPMD performing a further electronic minimization inside QEHeat. This final minimization will be also necessary for using charge-densities from other AIMD quantum-engines, as for example SIESTA.

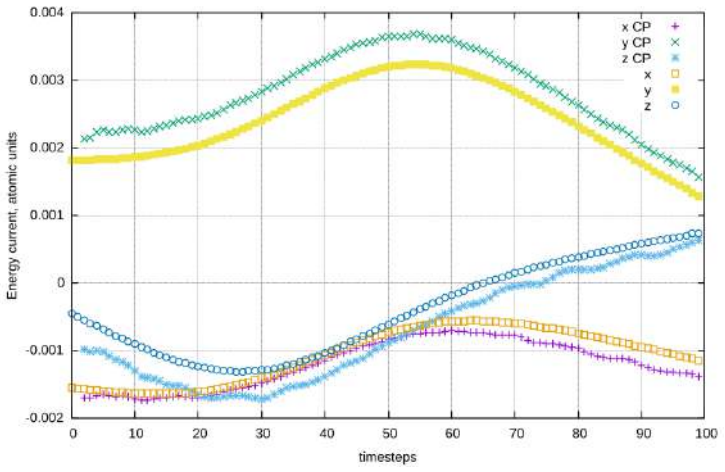


Figure 1. Comparison of the components of the energy currents computed using density and wave functions generated by the CP code and those generated by QEHeat

In conclusion, we have shown the feasibility of the "on-the-fly" calculations, and we are providing a valid test-bed for a general, minimally invasive, framework for all the workflows that require an intensive data exchange between different concurrent tasks. The framework can be easily used for other calculations, in every situation where passing of information between different codes is needed, but merging or glueing the codes together can be very difficult.

3.2 Excited state electronic properties in GPU-accelerated infrastructures (Demonstrators Dem.1 and Dem.7)

This task focuses on the calculation of quasi-particle (QP) electronic properties of materials and interfaces as investigated within the GW approximation by using the Yambo code. As mentioned in deliverable D6.1, one of the most critical issues for these types of simulations is the memory footprint. To partially overcome this issue and afford the calculation of QP properties of layered materials and slab configurations of interest for battery applications (related to Demonstrator Dem.7), we have prioritised the developed of a computational algorithm able to accelerate the convergence of QP properties of 2D semiconductors with respect to the number of k-points in the BZ sampling. By combining Monte Carlo integration techniques with an interpolation scheme of the screened potential, this method proves successful for efficiently evaluating the band gap of prototypical layered materials, with a reduction of the k-points grids by a factor that ranges between 20 and 80 (see Figure 2a).

The proposed algorithm, available in the GPL releases of the code, has been used for production runs of QP properties of graphene-based materials (Figure 2b, hydrogenated few-layer graphene [BETT2022]), which are among the most common materials for anodes in Li-ion technologies. Along the same direction, we developed the Multipole Approximation [VALI2022], which allows for computing accurate QP quantities beyond the generalised plasmon-pole approximation without resorting to the more computationally demanding full-frequency integration approach. This allowed us to benchmark GW calculations on prototypical graphene/metal interfaces (121 electrons) in slab configuration. The calculations included 2000 bands, a Xo cutoff of 20Ry, a 24×24×1 k-points grid, and 2 frequency values. The scalability on Juwels-Booster (Juelich) and relative efficiency are reported in Figure 3a as a function of the number of GPU cards, showing a very good scalability up to 1440 GPUs (360 nodes, 4x NVIDIA A100 per node).

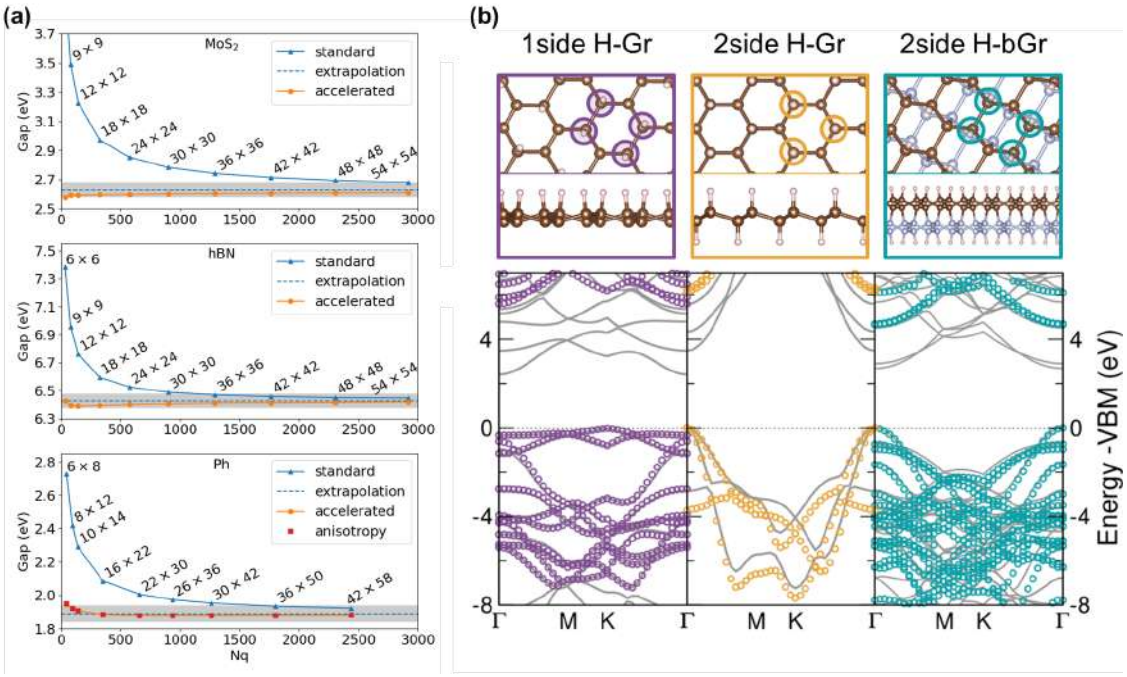


*Figure 2: Efficient GW calculations in two dimensional materials through a stochastic integration of the screening potential [*GUANArXiV*]. (a) From top to bottom, tests on a wide band gap insulator (hBN), a transition metal dichalcogenide (MoS_2), and an anisotropic semiconductor (phosphorene) are shown, with a reduction of the k-points grids of 80, 40 and 20, respectively. (b) Application to the calculation of QP properties of graphene - based materials, after Ref. [BETT2022].*

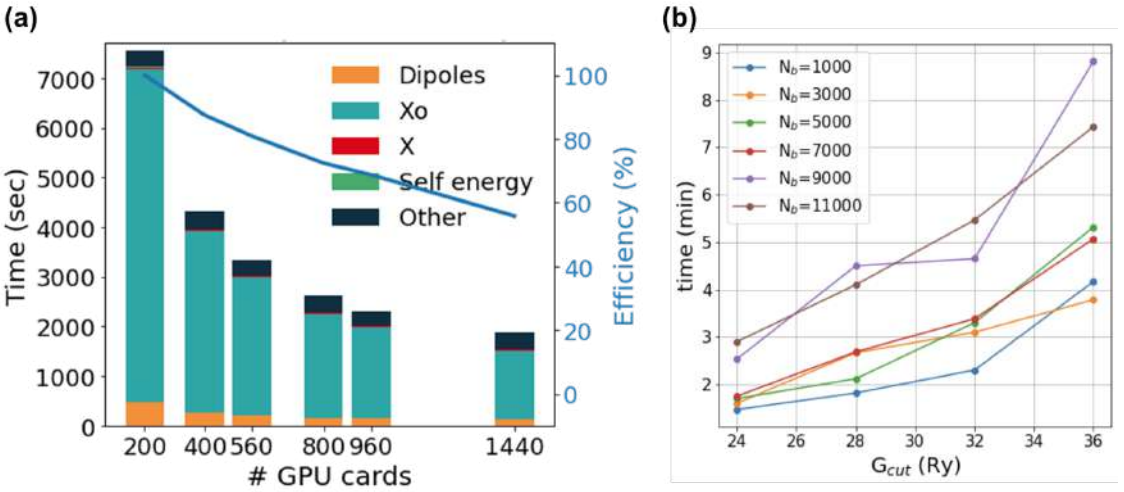


Figure 3: Scalability and time-to-solution on Juwels-Booster (Juelich). (a) Scalability test for a graphene/metal slab, up to 1440 GPUs (360 nodes, 4 NVIDIA A100 per node),

corresponding to a partition of 14 PFlops. (b) Time-to-solution for the LiF molecule. A calculation with Kohn-Sham (KS) states Nb=1000 and PW cutoff Gcut=24 Ry done with the same number of MPI tasks on a CPU-only machine takes around 20 minutes instead of 1 minute.

As a further important issue, we have tackled the well-known problem of interdependent convergence parameters in GW calculations. Indeed, the convergence procedure is a critical step in these types of calculations to obtain accurate and reliable results. However, parameters like those determining the summation over empty states and the size of the response matrix are inherently coupled, and ignoring this interdependence can result in a very tedious as much as time-consuming convergence procedure, which requires the cyclic repetition of multiple univariate convergences. An efficient algorithm taking care of the convergence of GW calculations has been implemented in the framework of the AiiDA infrastructure, within a GLP release of the aiida-yambo plugin (https://github.com/yambo-code/aiida-yambo). This plugin contains a set of python workflows – including the newly developed YamboConvergence workflow – that enable automated and high-throughput (HT) oriented simulations of excited state properties of materials. Thanks to these novel implementations, the least possible number of convergence steps are performed with the least possible human intervention, while ensuring reproducibility and robustness (in terms of error handling). An example of YamboConvergence provenance graph is shown in Figure 4.

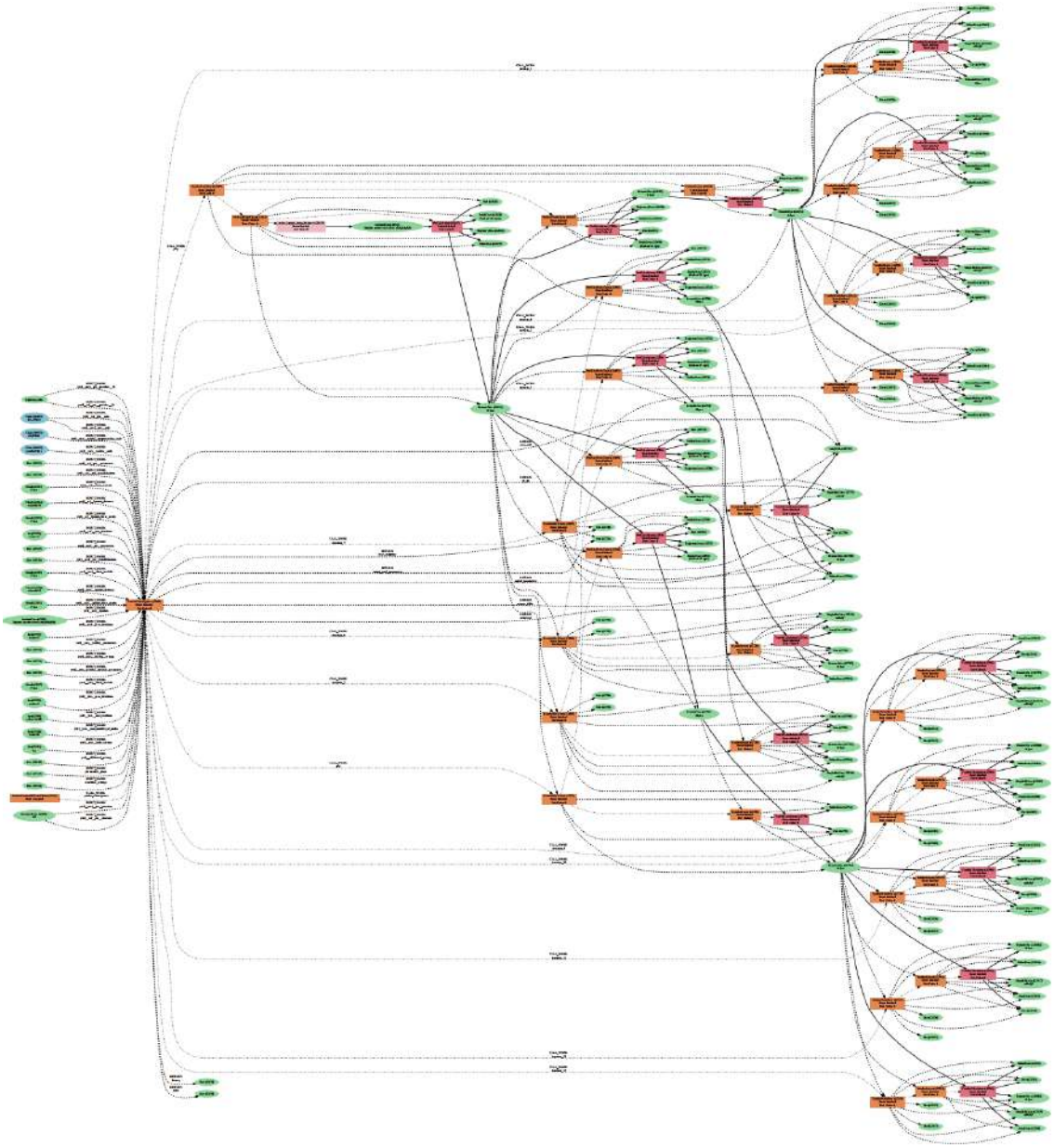


Figure 4: An example of YamboConvergence provenance graph.

This implementation has been validated for the calculation of QP gaps of well-known systems, like Silicon, bulk TiO_2 and 2D hexagonal Boron Nitride (hBN), and readily applied to production calculations, as for the graphene-based materials cited above [BETT2022]. A further application concerns the evaluation of IP and EA for the GW100 benchmark dataset, (a set of 100 molecules for which QP properties are computed within the GW approximation). By considering a 2D parameters grid of 24 points (simulations), we performed an automatised HT run of 2400 calculations. The predicted

wall-clock time required to run the whole GW100 set was estimated at about 150 hours. Calculations were performed by exploiting the speedup provided by GPU-accelerated machines, enabled both in Yambo and Quantum ESPRESSO packages. By using 20-64 nodes of Juwels-Booster@JSC, each of them equipped with 4 A100 GPUs (40 GB of memory), we were able to use from 80 to 256 GPUs together for each single calculation. The timing of a typical full molecule study is shown in Figure 3b, showing that all calculations, even the largest ones, take less than 10 minutes to be completed, for a total of 1.5 hours per molecule (including all 24 simulations). We remark that such a demonstration opens the way to HT screening of materials, such as novel electrolytes with desired EA/IP characteristics.

The possibility of taking advantage of both the novel stochastic integration and convergence algorithm has also opened the way for the unforeseen investigation of the excited-state properties of a subset of recently discovered 2D systems (contained in the Materials Cloud 2D database, MC2D), in order to identify potential excitonic insulators (EI) in an HT fashion. Here, very accurate estimation of QP band gap and exciton binding energies are fundamental in order to identify excitonic instabilities. GW-BSE calculations were performed on a subset of 35/1800 candidate materials, chosen by means of a DFT pre-screening. Most of these calculations target >100 electrons and are at the verge of the possibilities of modern machines. Calculations were performed by using a large part of the CPU partition of the LUMI machine (Kajaani, Finland), i.e., from 80 up to 200 nodes (128 cores/node, 256 GB of maximum RAM memory per node) with massive usage of the OpenMP (shared memory) parallelism. The average number of CPUs per single MPI (distributed memory) task was between 16 and 128, due to memory reasons.

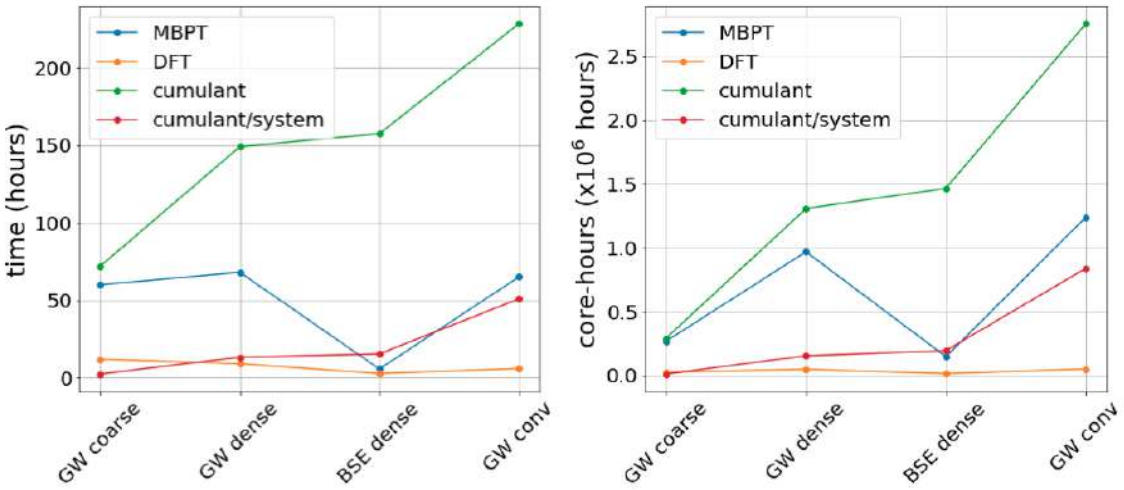


Figure 5: Timings for a subset of 254 GW-BSE simulations on LUMI (CPU partition). Left: wall-clock time needed to run the simulations. The cumulant per system, i.e. the cumulant for all systems (green curve) divided by their number, shows that the full set required only 50 hours. Right: computational time, in terms of core-hours. The work was performed in ~2 running days consuming ~2.5 million core-hours.

3.3 Defected Titanium-dioxide surfaces for photocatalytic applications (Demonstrator Dem.1)

Titanium dioxide (TiO_2) is one of the most thoroughly investigated metal oxides, due to its broad range of applications in several key technologies including photocatalysis and solar energy conversion. Despite the broad interest and the efforts made so far, an accurate theoretical description able to capture the complexity of the photoexcitation process is still lacking. Here, to be able to directly compare with the experimental outcomes [WANG2015, TANN2021, WANG2021], it is particularly important to resort to highly accurate methods to treat the excited-state properties, such as the GW-BSE approach. A further complication in such a study concerns the modelling of surface-localised band-gap states (SLBGSs), which requires (i) large supercells in slab configuration in order to properly describe the electron self-trapping mechanism underlying the localization of a polaronic charge in the semi-infinite TiO_2 surface, and (ii) large amounts of vacuum in order to prevent spurious interaction between slab replicas (resulting in a large number of plane waves and bands). The complexity of the required methodologies, combined with the size of the systems, makes these calculations a remarkable challenge to be tackled on the current (pre)exascale HPC architectures.

The main goal of this demonstrator was to achieve an accurate characterisation of the optical response of defected TiO_2 surfaces within the GW-BSE framework by exploiting the recent GPU porting of the Yambo code. To this end, a preliminary scalability test on the Yambo parallelization has been carried out on the Juwels-Booster GPU-accelerated supercomputer (936 nodes equipped with 2 AMD EPYC 7402 CPUs and 4 NVIDIA A100 GPUs with 40 Gb, connected via NVLink 3.0 to each other), by considering a reliable reference able to verify both MPI parallelization and GPU acceleration, i.e., a 2×2×3 TiO_2 bulk supercell with an interstitial H impurity [ATAM2019]. As it can be seen in Figure 6, the dipoles and X_0 polarizability runlevels are, for this test, the most computationally expensive parts of a typical GW run. Remarkably, the calculation of nonlocal commutators [r,Vnl] in the “dipole” runlevel was originally unaffordable for systems containing a large number of atoms, like the ones studied here, because the atomic projectors were all stored on each GPU memory. Recent developments have

addressed this issue by enabling the distribution of the projectors over all the GPUs available, thus allowing us to perform this type of calculations.

In order to investigate the *d-d* transitions from polaronic SLBGSs in defective TiO_2 surfaces, we considered here the simplest case study, i.e., the adsorption of atomic H on a bridging O atom for a TiO2(110) unreconstructed surface. The adopted 4x2-TiO_2(110) supercell (see Figure 7) is made up of 209 atoms, corresponding to 1553 electrons.

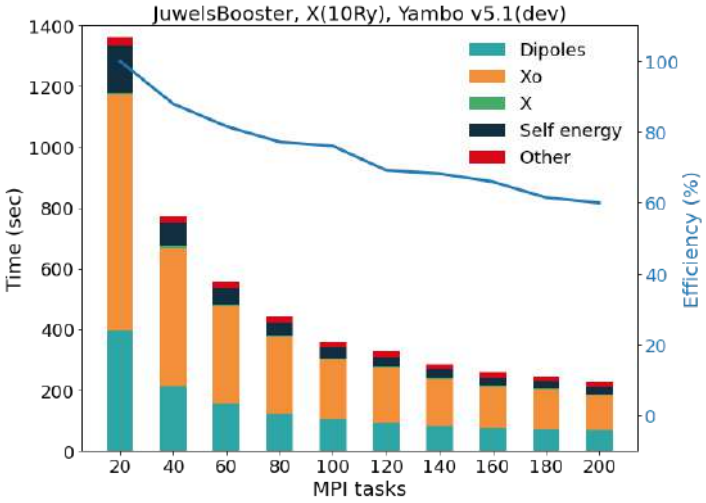


Figure 6: Bulk TiO_2 rutile (72 atoms/cell) with interstitial H. Strong scaling test for a GW calculation on Juwels Booster (Juelich) whose nodes are equipped with 4 NVIDIA A100 GPUs. In these calculations we have used one MPI task for each GPU available, which corresponds to an upper bound of 200 GPUs, i.e. 50 nodes.

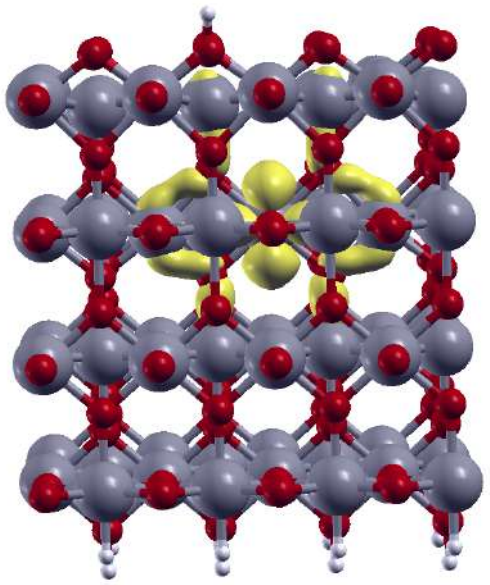


Figure 7: Side view of the ball and stick model of the relaxed mono-hydroxylated four-trilayers 4x2 (110) surface supercell of rutile TiO_2 made up of 209 atoms. Grey, red, and white spheres represent Ti, O, and H atoms, respectively. The polaron charge density originating from the hydroxyl defect is also shown in yellow.

First, we have employed the GPU porting of Quantum ESPRESSO (QE) to analyse the *charge localization* induced by the defect at the DFT level, by looking for the most stable polaron configuration. To properly describe the charge localization on strongly correlated *d*-states, we have adopted the PBE+U method, with U=3.5 eV, and the most stable configuration, with the corresponding polaron charge density, is shown in Fig. 7. Starting from the PBE+U optimised geometry, we have computed the G_0W_0 QP corrections to the DFT-PBE eigenenergies and the BSE optical spectra by employing the Yambo code. All the G_0W_0-BSE calculations required for our project have been performed on two different systems:

- Marconi100 GPU-accelerated supercomputer (Cineca, Bologna), which consists of 980 nodes equipped with 2 IBM P9 CPUs and 4 NVIDIA V100 GPUs with 16 Gb (NVLink 2.0);

- Perlmutter GPU-accelerated supercomputer (NERSC, Berkeley), which consists of 1536 nodes equipped with 1 AMD EPYC 7763 CPU and 4 NVIDIA A100 GPUs with 40 Gb (NVLink 3.0), similarly to Juwels Booster (Juelich);

where the choice of the class of GPUs to be used was mainly determined by the memory requirements in the polarizability computation.

The QP-corrected gap and defect binding energies for the system in Figure 7 have been computed on Marconi100 by using 400 MPI tasks (one MPI task for each GPU available, 100 nodes). In particular, for the calculation of the irreducible polarizability X_0, we adopted a cutoff in the Coulomb potential in the direction perpendicular to the surface in order to eliminate spurious interactions with periodic replicas while reducing the computational effort. Considering a ~15 Å vacuum layer in the non-periodic direction (corresponding to a supercell length of ~29.45 Å), we used a 2×2×1 sampling of the Brillouin Zone (BZ), two frequency values within the Godby-Needs plasmon-pole approximation (PPA), 4000 states (including 777 valence states), a X_0 plane wave cut off of 8 Ry (corresponding to 11809 G vectors).

Given the size of the matrices (1.08 Gb for each frequency) and the large number of wavefunctions to deal with, the main bottleneck of the calculations turned out to be the memory usage on the V100 GPUs, which prevents calculations on k-grids denser than the 2×2×1 one. This can be traced back to the lack of parallel distributed linear algebra support on GPUs (required to invert the dielectric matrix) and to the porting strategy in Yambo. The latter is based on reading and copying DFT wave functions on the GPU card memory, making them available for heavy computational kernels. Similarly, the response function is also calculated and temporarily stored on the card memory. However, as a design principle, in order to improve the performance of the porting, the number of data transfers between host and device is minimised.

All the G_0W_0-BSE calculations requiring denser BZ samplings have been performed on the Perlmutter system (equivalent to Juwels Booster tested above, see Figure 6) by exploiting the wider memory and larger computational power of the A100 GPU cards. In addition, we have taken advantage of a novel stochastic integration technique, described in Sec. 3.2 and Ref. [GUANArXiV] for GW calculations, and recently extended to BSE. This introduces a remarkable speed up in the convergence of both G_0W_0 and BSE calculations with respect to the k-mesh, as shown in Figure 8, enabling calculations otherwise unfeasible even for systems equipped with A100 GPU cards.

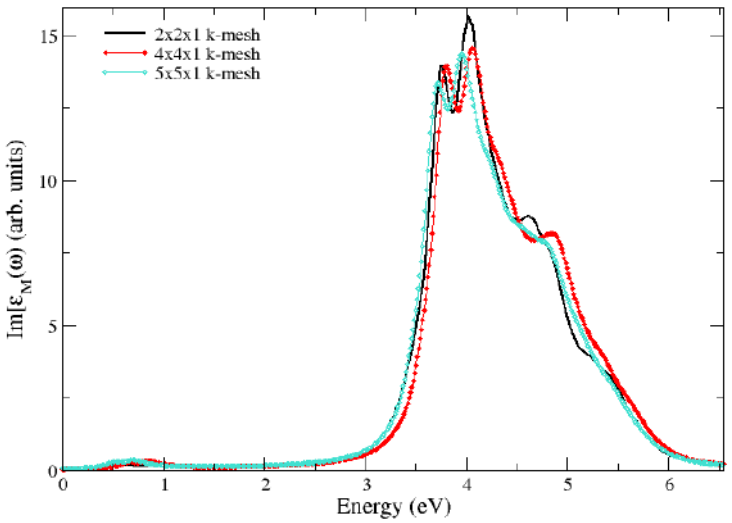


Figure 8: Optical spectra of the four-trilayers mono-hydroxylated 4x2 TiO_2(110) supercell made up of 209 atoms at the BSE level for light polarised parallel to the surface and reported for different samplings of the BZ.

Furthermore, we have found that, in order to characterise the optical transitions from the polaronic SLBGSs to the *d*-like excited states identified by 2PPE experiments [WANG2015, TANN2021, WANG2021], a considerably large energy range needs to be sampled, i.e. the required two-particle Hamiltonian matrix contains at least 231 conduction states (up to 5.31 eV above the VBM), and 62 valence states (up to 3.80 eV below the VBM), corresponding to a size n=113896. These calculations have been computed on M100 distributing the kernel evaluation over 484 GPUs. The diagonalization of such a large matrix by using the Scalapack routines on the hosts (i.e. without exploiting the GPU acceleration) would be totally inefficient (more details are reported in D6.4). To perform these calculations we have thus employed a subspace iterative algorithm implemented in the SLEPc library, which has been fully interfaced with the Yambo code. The corresponding spectra are reported in Figure 9.

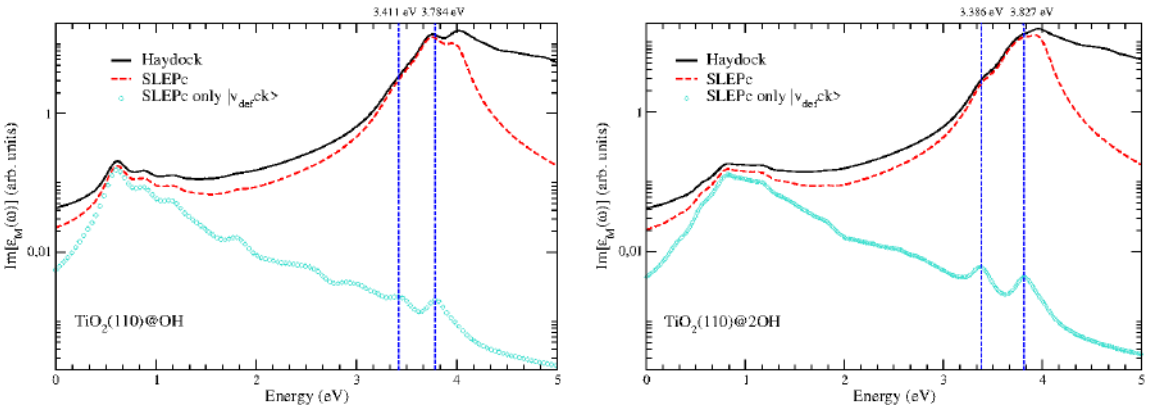


Figure 9: On the left (right) the optical spectra of the four-trilayers mono-hydroxylated (di-hydroxylated) 4×2 TiO_2(110) supercell at the BSE level for light polarised parallel to the surface. The Lanczos-Haydock and SLEPc solvers are compared with the SLEPc spectra obtained by including only the defect states in the computation of the macroscopic dielectric function. To highlight the peaks obtained with this approach we reported the results on a semi-logarithmic scale.

3.4 Strength of antibody/antigen complex interactions using QM/mMM (Demonstrators Dem.2 and Dem.6).

In our previous work described in deliverable D6.1 on the demonstrator Dem.2 on "Reduction of the complexity in very large scale systems", we demonstrated the possibility of analysing very large biological molecules in terms of coarse grain fragments, obtaining very useful information stemming from the (quantum) electronic structure. This method has allowed us to quantify the strength of the chemical interactions between different fragments of the biomolecule.

With this methodology developed during the first phase of MaX, we proposed a new demonstrator (Dem 6), focused on its application to two different systems: proteins related to the SARS-CoV-2 virus, and antibody/antigen (AA) assemblies. Both lines of application of the fragment analysis machinery have progressed quickly, and have already produced results relevant to the biochemistry community and the pharma industry

The work on SARS-CoV-2 has focused on the main protease M^{pro} and the spike protein. We have investigated, for instance, the microscopic factors that modulate the interactions between M^{pro} and α−Ketoamide inhibitors [GENO2020], relevant for the search of drugs to fight COVID-19, and the effect of mutations in the different virus

variants on the binding efficacy to the cellular receptors (and therefore, the infectious potential of each variant) [ZACC2021, GENO2021, ZACC2022].

In this report, we focus on the detailed description of the results of our work on antibody/antigen assemblies, which illustrates the power of our techniques in the understanding of biomolecular complexes. In particular, we focus on monoclonal antibodies targeting specific antigens, which are promising biochemical macromolecules for therapeutic purposes. However, whereas targeting specific monoclonal antibodies can be routinely achieved, increasing the antigen affinity as much as desired is still a challenging task. Most of the available theoretical methods mostly focus on investigating close contact antibody/antigen (AA) local regions and usually ignore the effect of more distant domains on affinity. Because of the atomic size of AA complexes (from 500 to 1,000 proteinic residues and above) only standard pairwise molecular modelling force fields and more likely empirical cost functions are used to quantify the strength of the interactions within such complexes. However, these theoretical approaches are known to be based on crude approximations (see among others Ref. [PIA14]), which prevents reaching a high level of accuracy to estimate the stability of AA assemblies. Despite the recent huge progress in predicting protein 3D structures at the atomic level using AI techniques [TUNY2021], the accurate investigation of assemblies like AAs is still an ambitious goal.

Our project, awarded by a SANOFI iTech AWARDS this year and supported by the pharmaceutical group SANOFI, aims at identifying the microscopic interactions that favour/disfavour the stability of AA assemblies by means of a hybrid (sequential) Quantum Mechanical/multi-scale Molecular Modeling (QM/mMM) simulation protocol that we developed recently [GEN20] to investigate the microscopic factors at the origin of the stability of SARS-COV-2 main protease/-ketoamide inhibitor complexes [ZHAN2020].

Our QM/mMM scheme relies on a QM method (an O(*N*) method based on localised Daubechies wavelets, see http://bigdft.org/) and a mMM approach (based on polarizable force fields coupled with an O(*N*) polarizable multi-scale coarse grained water model, see http://biodev.cea.fr/polaris/) that are able to efficiently simulate microscopic systems with up to tens of thousands atoms (QM) and at the million atom scale (mMM), on modern supercomputers. The mMM approach is used to sample the microscopic system potential energy surface, PES, whereas the QM one is used to assess/refine the quality of the MM results as well as to identify AA domains (that can be close and/or far from the main AA contact regions) pivotal to understand the AA complex stability. More specifically, our goal is to add two new steps to standard computational protocols used

to model AA complexes (like the popular one based on the Rosetta package of programs [WEIZ2017]), to further evaluate and refine their solutions. Both these steps rely on our own simulation codes, POLARIS(MD) [POL] and BIGDFT [BIG], that are stable and mature enough to be routinely used to investigate very large AA assemblies.

As proposed in the last deliverable, our main goals were to assess our mMM/QM scheme by investigating (for starters) five standard benchmark AA assemblies whose PDB labelings are 1IQD, 2JEL, 1BVK, 1MLC and 2VIS and that were investigated both experimentally and theoretically using standard docking approaches. In particular their experimental dissociation constants, Kds, are < .0014, 2.8, 14, 91 and 4000 nM, respectively.

The mMM numerical scheme and the HRE simulations

From the experimental crystallographic structures of the AA assemblies, we first investigated the Potential Energy Surface (PES) of the AA interaction region (the epitope and paratope domains of the antibody and antigen constituents) in aqueous phase by means of our multi-scale mMM approach coupled to a Hamiltonian Replica Exchange (HRE) scheme in order to provide statistically meaningful sets of 3D structures of the assemblies. To preserve the global assembly conformations, we systematically constrained the backbone carbon alpha's of the antibody (antigen) residues that are distant by more than 6 Å from any antigen (antibody) non hydrogen atoms in experimental crystallographic structures. We simulate the assembly as immersed in a large enough solvent cubic box (whose dimensions are set to the longest inter atomic distance within the AA assembly to which 15 Å is added). That box is then filled by coarse grained pseudo particles to model the solvent water [MASE2013]: the number of pseudo-particles added ranges from about 40k (1BVK) to about 250k (2VIS). Examples of AA assembly starting structures are shown in Figure 10. For the present purpose we don't account for counter ions to neutralise the assembly total charges.

The hydrated AA assembly structures are first relaxed by minimising the norm of the force gradient using a standard conjugate gradient algorithm. Then we perform a 2ns scale simulation in the NPT ensemble (using a Nosé-Hoover barostat, targeting ambient temperature and pressure: 300 K and 1 atm) to relax the box dimensions. The final structures are the starting ones for HRE simulations.

The HRE scheme consists of 24 independent MD simulations. Along each simulation the Hamiltonian describing all the microscopic interactions between the atoms of the epitope/paratope regions is scaled by a factor ranging from 1.00 to 0.36 (typically spaced by 0.04). We didn't consider values of < 0.36 as preparatory tests show that they

can yield to numerical instabilities arising from steric clashes among the epitope/paratope atoms as a replica exchange occurs.

Epitope/paratope regions are defined as the antibody (antigen) residues that are distant from any antigen (antibody) non hydrogen atoms by less than 6 Å. The simulation duration is 20 ns, the integration time steps of the Newtonian equations of motion are set to 0.25 fs (standard harmonic potentials modelling chemical bonds, angles between chemical bonds and the so-called improper torsions) and to 2 fs (all the other interactions). The temperature is monitored according to a Langevin thermostat [CANC2007] with a target temperature of 300 K. X-H bonds and H-X-H angles are restrained to their ideal values by means of the iterative RATTLE procedure (the convergence criterion is set to 10^{-5} Å). Along the simulations the centre of mass of the AA assembly is periodically reset to the solvent box centre and the pseudo particle coordinates are shifted accordingly. Replica exchanges among the simulations are checked each 250 fs and AA assembly structures along the reference simulation (=1.00) are stored each 10 ps once an equilibrium starting period of 2 ns is achieved. That yields sets of AA assembly snapshots comprising 1,800 conformations.

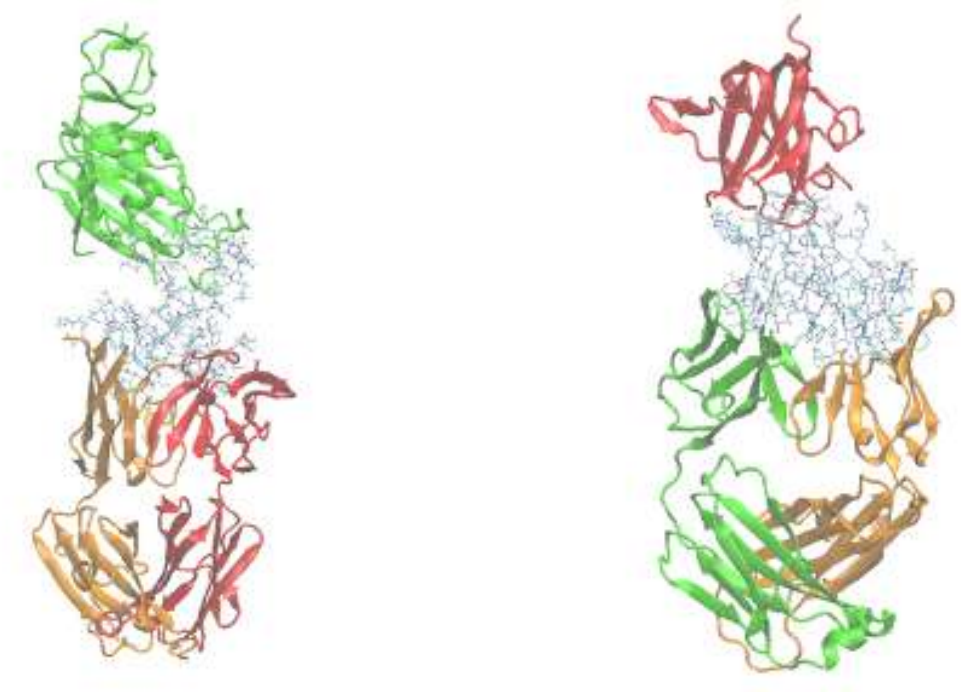


Figure 10. Starting simulation structures of the AA assemblies 2VIS (left) and 1IQD (right). The paratope/epitope regions are shown in an all atom representation. The antigen is shown in red (1IQD) and green (2VIS). Solvent pseudo particles are not shown.

Each HRE simulation is performed on a full 64-cores AMD processor of the TGCC-IRENE ROME partition (we thus only use the OPENMP scheme of POLARIS(MD) to perform each simulation). When two replicas are exchanged, the AA assembly (including the solvent) coordinates are permuted between the two corresponding simulations, i.e. between the two MPI processes handling these simulations. We note the simulations to be from 2 to 3 times slower as compared to fully independent simulations. They typically run at rate ranging from 2 ns per day (for the largest AA assembly 2VIS comprising 1,122 residues) up to about 5 ns per day for the smallest assemblies (1BVK, comprising about 520 residues). All the 20 ns simulations are thus performed within less than 10 days.

Clustering the simulation snapshot sets

To provide a reduced but statistically reliable set of AA snapshots for the QM post processing step, we performed a clustering of the AA snapshot set extracted along the reference HRE simulation (that comprises 1,800 snapshots), based on the vector of the coordinate deviations from the crystallographic structure of all the non-hydrogen atoms belonging to the AA epitope/paratope regions (see above for definition). Then we randomly choose 1 every ten snapshots belonging to each of the main clusters identified (see Figure 11, for instance). We thus post process 180 representative snapshots per AA assembly using our QM approach. Note the solvent particles are ignored during the QM post processing.

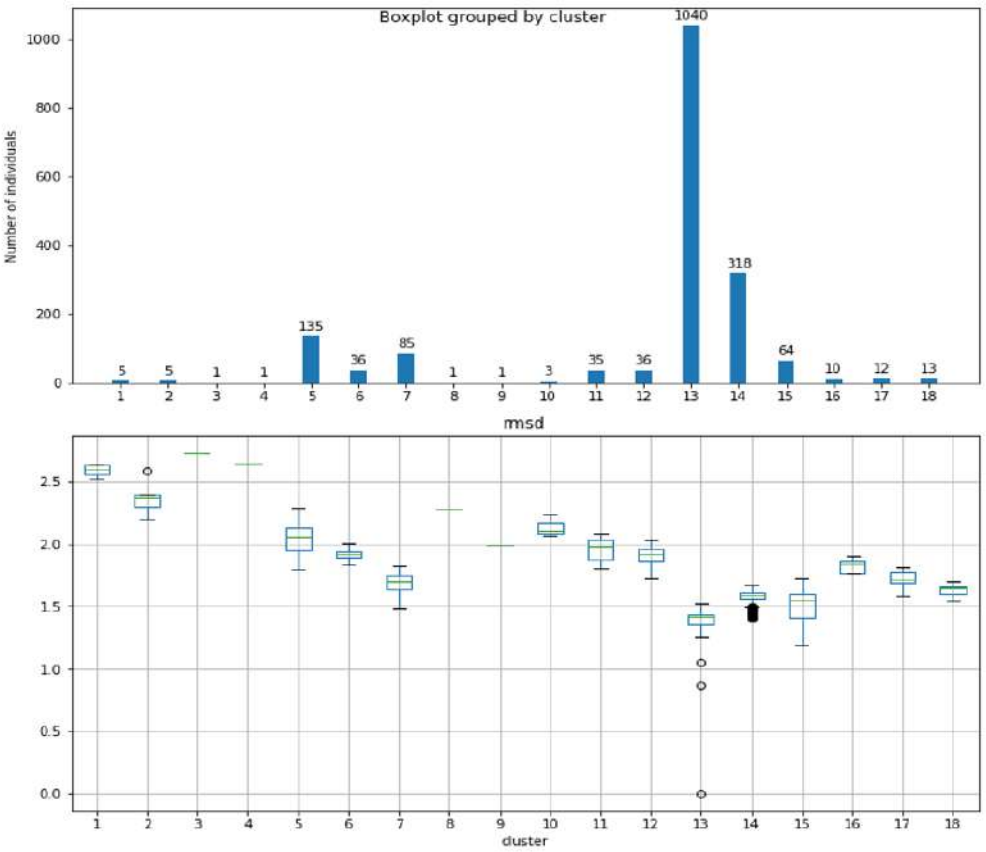
The 1,800 snapshots were decimated as follows. First, the epitope/paratope regions of all the snapshots were rototranslated with the Kabsch algorithm in order to minimise the Root Mean Square Deviation, RMSD, of their backbone atoms from the initial configuration. Then, the distance matrix of all the atoms of those regions was considered as a descriptor. Agglomerative clustering of the structures was then performed following this descriptor. We obtain histograms like the one depicted below, where each snapshot is associated with a cluster which shares a similar value of RMSD with respect to the initial structure. We then choose one out of ten snapshots per cluster, in order to select a representative ensemble of the population sampled by HRE simulations.

All the largest clusters correspond to sets of atomic coordinates that are close to the experimental crystallographic ones (see Figure 12). As these atomic conformations are predicted to be the most probable by our simulation code POLARIS(MD), that supports the accuracy of our mMM approach, in particular to investigate the effects of antibody

mutations on the stability of AA assemblies ahead of experiments and for which experimental data are not available.

The QM scheme and the AA interaction patterns

We initially proposed to systematically estimate the interaction energy between the antibody and antigen moieties of each AA assembly for 100 snapshots extracted along the reference HRE simulation using a standard triple point computation scheme (i.e.: we compute the difference between the total energy of the assembly and the energies of each of its elements). However an interesting property of the localised Daubechies wavelets as implemented in the code BIGDFT is to enable a fragmentation of an AA assembly in a set of interacting 'fragments' from which an estimate of the antibody/antigen interaction energy can be readily computed. We have recently demonstrated that the values of these approximate interaction energies linearly correlate with accurate interaction energies computed from the triple point scheme. Hence, instead of performing systematic triple point computations, the approximate estimate allowed us to perform quantum post-processing of about twice as many snapshots (180 structures) than those initially planned, for each AA assembly.



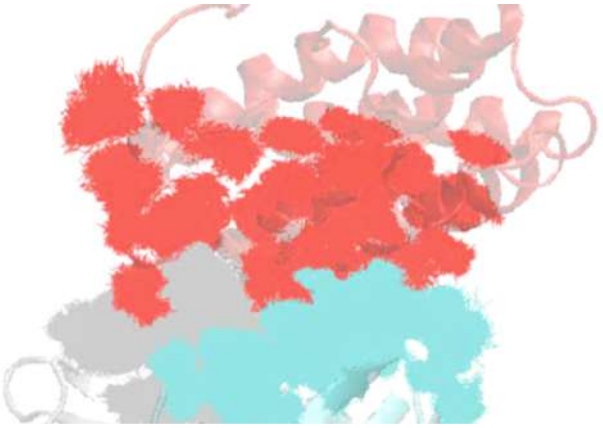
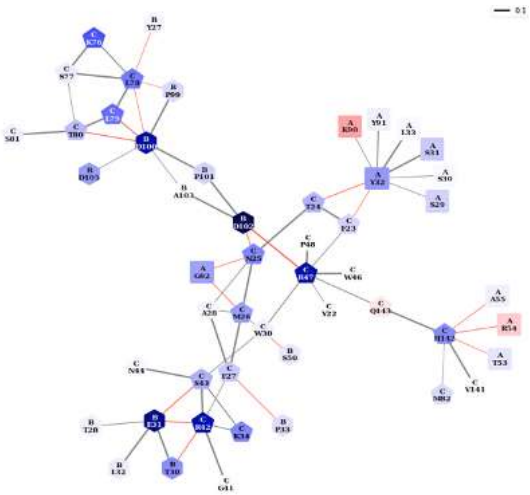


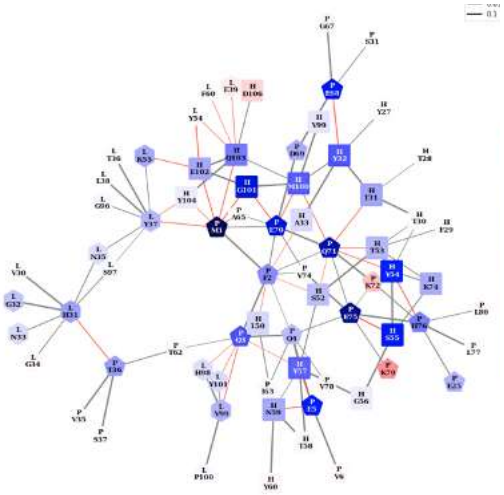
Figure 11 Top: size of the main clusters identified of the AA assembly 1IQD along its reference HRE simulation. Middle: Root Mean Square Deviation (RMSD) values (and their standard deviation) for all the atoms of each cluster. The reference conformation for the RMSD computations corresponds to the experimental crystallographic one. Note that the largest cluster '13' turns out to have the smallest mean RMSD, i.e.: the atom conformations within cluster '13' are the closest of the crystallographic structure.

*Bottom: superposition of all the local conf*ormations of the epitope/paratope atoms corresponding to the most probable cluster of the AA assembly 1BVK.

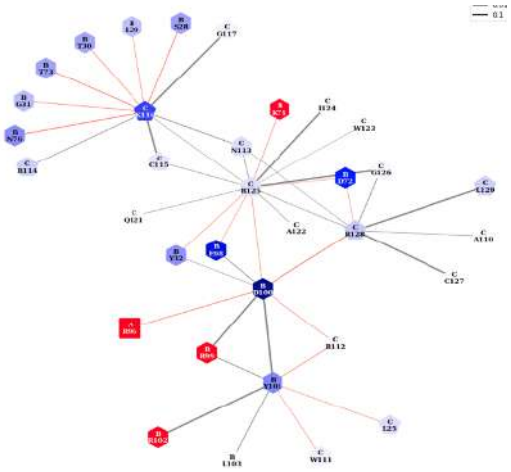
Each set of 180 snapshots were post-processed using the BIGDFT code. Most of the computations for each snapshots were achieved within about 2h by using 2,048 cores (small AA assemblies 1BVK and 2JEL) and 3h using 4096 cores (large assemblies 1IQD and 2VIS).



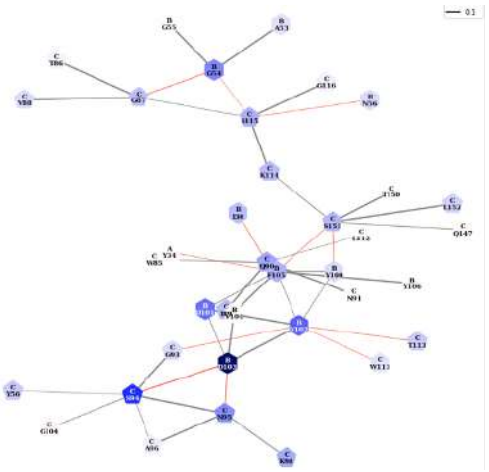
1IQD (Kd = 0.0014)



2JEL (Kd = 2.8)



1BVK (Kd = 14)



2VIS (Kd=91)

*Figure 12. Local AA assembly patterns for the largest clusters. In blue and red: residues favouring/disfavoring the assembly stability, respectively. The darker the colour, the stronger the stabilising/destabilising effect of the residue. Black lines interconnect residues exclusively of the antibody (or of the antigen) that interact at short range. Red lines interconnect pairs of residues belonging to both the antibody and the antigen and which interact at short range. The assembly dissociation constants Kd are expressed in nM. The antigens are denoted **A** (1IQD, 1BVK and 2VIS) and **P**(2JEL).*

In Figure 12 we summarise the details of the QM fragmentation computations for the four assemblies 1BVK, 2VIS, 1IQD and 2JEL by means of a schematic representation of the resulting 'local AA interaction patterns' allowing one to readily identify the epitope/paratope residues that contribute to globally stabilise/destabilise an AA assembly, as well as to identify pairs of residues interacting at short range within the largest snapshot clusters (i.e. the set of most probable atomic conformations of epitope/paratope regions). The residues of the latter pairs can belong either to both the AA entities or only to a single one: a set of interconnected pairs correspond thus to a local interaction network of residues at the interface of an AA assembly.

Interestingly, the data of Figure 12 show the interaction patterns of the 'strongest' AA assemblies (those whose dissociation constant is the smallest, namely 1IQD and 2JEL) to be more complex and to involve more residues than for the 'weakest' assemblies (1BVK and 2VIS). With the exception of the assembly 1BVK, the interaction patterns do not exhibit large sets of residues that reduce the assembly stability. For 1IQD and 2JEL, no more than three residues are destabilising; however they are outnumbered by extended sets of stabilising residues.

The assembly Kd value increases as the number of strongly stabilising residues decreases (i.e. residues with dark blue colours corresponding to interaction energies smaller than -80 kcal mol^{-1} in Figure 12). That suggests that the strength of the four AA assemblies is modulated by the magnitude of enthalpic effects arising from local interactions among epitope and paratope residues. We note also a dissymmetry regarding the weight of the antibody and of the antigen residues on the interactions stabilising the assemblies: residues of both the assembly moieties are present in the interaction patterns for 1IQD and 2JEL (the strongest assemblies), whereas mainly antibody residues stabilise the 1BVK and 2VIS assemblies (the weakest ones).

The largest complexity of the 2JEL pattern as compared to 1IQD suggests the weakest Kd value for 1IQD to arise from entropic effects: the 1IQD pattern involves less residues, the structural constraints are thus weaker than for a larger pattern such as 2JEL. Lastly the complex interaction patterns of 1IQD and 2JEL suggest their stability to not be highly sensitive to mutations.

The schematic representation of our results (see Figure 12) allows one to readily interpret the simulations and to estimate potential enthalpic and entropic effects on a AA assembly. All the results of our hybrid QM/mMM simulation approach are in line with the experiments.

To complement the data provided by the local interaction patterns, we are presently looking to define a schematic representation that allows one to readily identify residues distant from the epitope/paratope regions and that also noticeably affect the assembly stability (a priori the effects of these residues on the AA stability correspond to long range electrostatic interactions). Currently, we plan to investigate at least 20 AA and protease/receptor assemblies using our hybrid approach for the forthcoming years. These assemblies are all at the 600-1000 residues scale.

3.5 Full DFT simulation of new particle-like objects in chiral magnets (Demonstrator Dem.3)

Our simulations on particle-like magnetic objects, on which we reported in deliverable D6.1, continued with a strong focus on the properties of a Bloch-point localised at the end of a chiral bobber structure. This point, being a singular point in a continuum model of magnetism, constitutes a high energy structure at which we expected significant modification of the local magnetic structure.

Our simulations were performed in several steps: in the initial stage (already reported in D6.1) we used an idealised model for the magnetic structure in which we constrained the magnetization directions. This allowed us to simulate the Bloch point in a series of different unit-cells starting from a 2×2×4 supercell containing 128 atoms up to larger cells with over 1000 atoms.

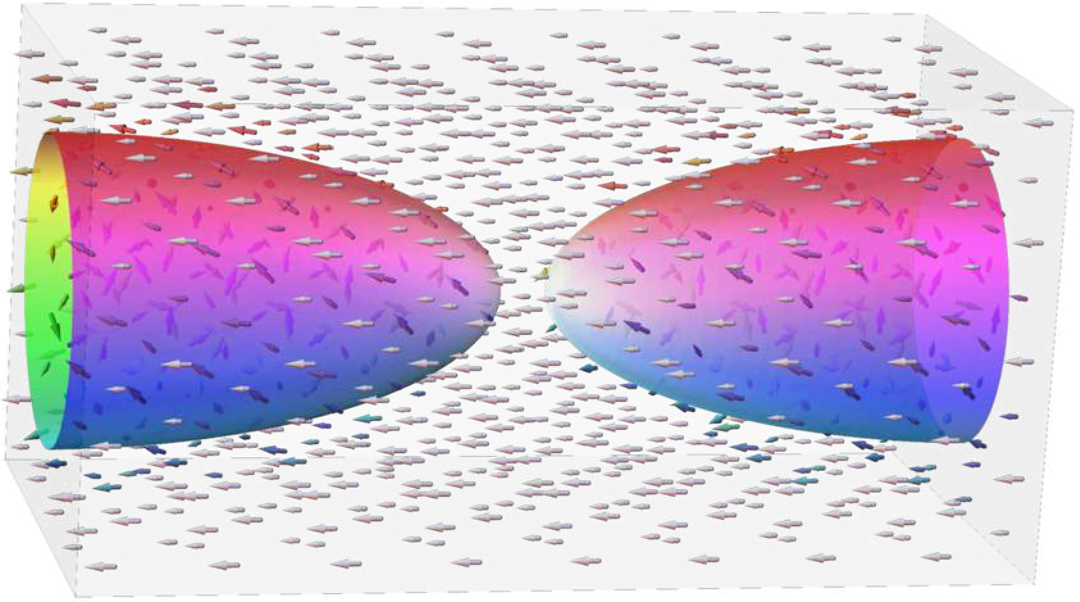


Figure 13: Idealised magnetic structure in which the Bloch point is constructed at the tip of the chiral bobber structure.

These simulations used typically some 256 nodes with 128 CPU cores and about a million core hours for a single self-consistency, and they already allowed us to obtain two important results:

1. The magnetic moment of the atom located closest to the position of the Bloch point (as defined in a continuum model) is significantly reduced. While we usually find a magnetic moment of up to 2.2 mBohr per magnetic Mn atom, this is reduced to about 1.6 mBohr close to the Bloch point. This is of course in line with the picture of this point as a point of vanishing moment and its nature of a high energy magnetic arrangement. We could also verify that this result is independent of the treatment of spin-orbit interaction.

2. The output magnetization in our simulations shows a significant tilting with respect to our constraining directions indicating that our structure is not free of magnetic torque and thus not the ground state configuration.

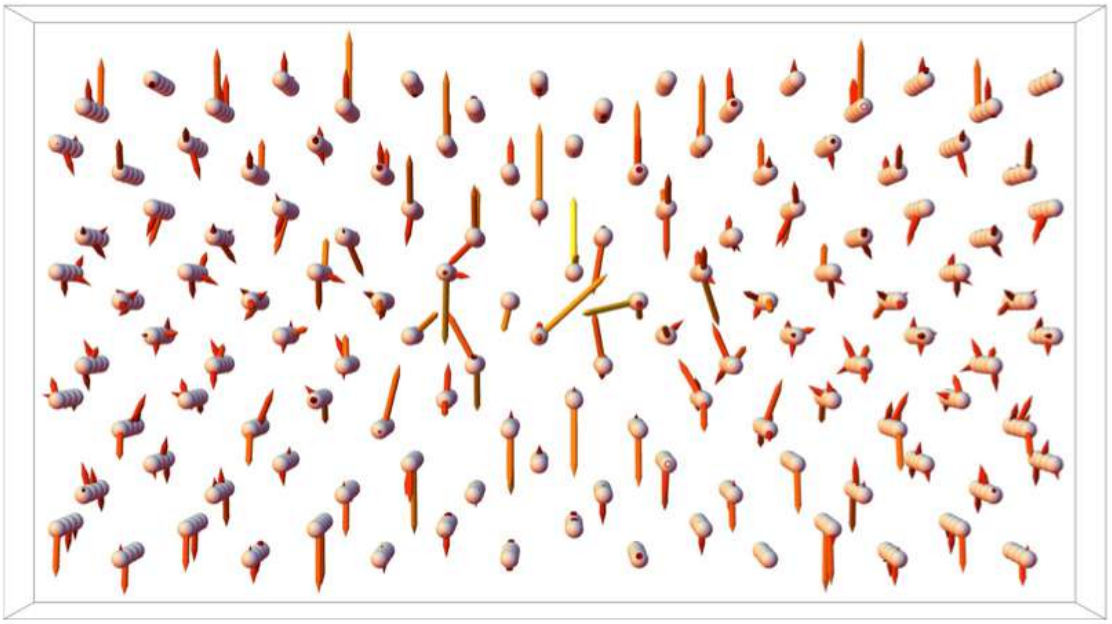


Figure 14: Visualisation of the magnetic torque acting on the atomic moments. While the torque is largest in the vicinity of the Boch point, the structure obviously differs from the magnetic ground state in all places and a non-vanishing torque can be seen on practically all sites.

To improve on this second point and to confirm that our simulation, while not performed directly at the minimum energy magnetic structure, still has some significance, we started to improve on our magnetic setup. This was done by replacing the idealised structure by a configuration obtained as the ground state in atomistic spin simulations. The structures obtained in this scheme are considerably larger than those considered previously, now containing 1,600 atoms in a 5×5×8 unit cell. Nevertheless, the improvements on the performance and the scalability of the FLEUR code enabled us to again achieve self-consistency for these setups using approximately a million core-hours with 128 nodes of the Intel Skylake architecture.

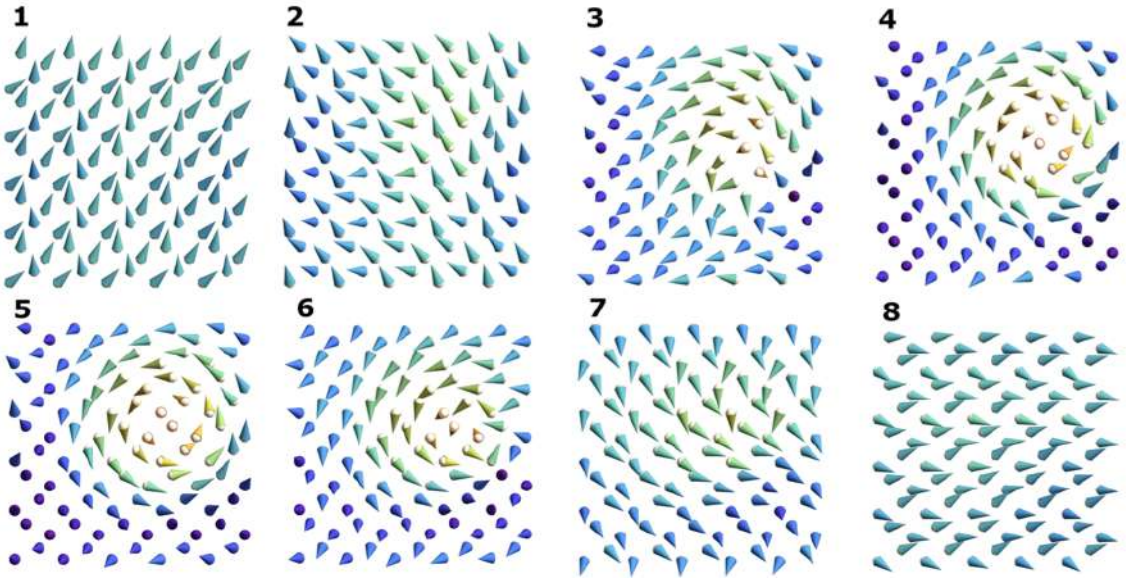


Figure 15: Layer resolved magnetic directions as obtained from atomistic spin dynamics calculations. The Bloch point is located again at the tip of chiral structure close to layers 2 and 7.

We could confirm that, in this structure, the remaining magnetic torque was indeed significantly reduced indicating a more stable state. Hence we could confirm:

1. That the atomistic spin dynamics is able to capture further relevant aspects of the magnetic interaction and thus provide a realistic picture of the structure. We could also verify that the spin-orbit coupling only has a minor effect on the torque still on an energy scale smaller than the errors inherent in the approximate setup.

2. The reduction of the magnetic moment close to the Bloch point was now less pronounced. The closest atom showed a reduced moment of about 1.9 mBohr. This effect can be due to the different alignment of the atoms in the surrounding of the Bloch point and needs further investigation.

We will now continue our investigations by trying to relax the constraint on the magnetic directions. While this sounds like a rather straight-forward idea we already experienced the difficulty to drive such a relaxation of the magnetic directions into convergence as the associated energy scale is very small and thus the driving forces of the dynamics are weak. Despite these difficulties, the advances made during MaX in the scalability and performance of the FLEUR code in state-of-the-art computational infrastructures will allow us to tackle these important but challenging physical systems in the future EU pre- and exascale machines.

3.6 RPA and double hybrid MD simulations of condensed phase systems with CP2K (Demonstrator Dem.4)

The CP2K team reached the goal of implementing forces for the RPA and associated methods. The work is reflected in the github merge request #1647 (MP2-based Double-Hybrid Gradients and Stress Tensors) and published in [STEI2022]. Double-Hybrid (DH) functionals try to combine the best of both worlds: DFT is augmented with the description of long-ranged interactions from a correlated method like MP2 or RPA. DH functionals can be routinely applied to molecular systems, but large computational resources are necessary for condensed phase calculations. Therefore, the number of applications to condensed phase systems is rather limited. Nevertheless, in these applications, the increased accuracy of DH functionals over (meta)-GGA functionals and hybrid functionals for the description of condensed phase systems has been established. For practical applications, the availability of analytical gradients for these methods is important, enabling geometry optimizations and molecular dynamics simulations. The items below make a summary of the accomplishments:

- RPA calculations make a full use of the COSMA library developed by CSCS:
  COSMA library provides a communication-optimal GPU accelerated distributed matrix-matrix multiplication implementation. It is now fully integrated into CP2K release and tested in production. It provides both NVIDIA and ROCm backends and enables CP2K to run on the majority of GPU-accelerated supercomputers.

- RPA was used as acceptance benchmark on the LUMI-G system, being on target to reach the performance requirements:
  CSCS and AMD software engineers worked in collaboration to fix the remaining performance issues and small bugs both in ROCm programming environment and COSMA AMD GPU backed in order to achieve the highest possible performance in the acceptance tests for LUMI-G. RPA calculations of the 128-water molecule system with CP2K are part of the acceptance benchmark.

- RPA+forces has started as LUMI-G Pilot project, delivering first results on the pre-Exascale EuroHPC system:
  The exponential increase in computational power of modern computers makes calculations with DH functionals for large sys- tems more and more feasible. We will demonstrate the current state with calculations carried out for large benchmark

systems performed efficiently on compute nodes with 1024 GPUs. We present two different benchmark systems for which we calculated energies and forces: the benzene crystal and the anatase crystal. The geometrical parameters, the number of basis functions, and the number of floating point operations of the quintically scaling steps for each test system are summarised in Table II. The computational timings and some performance related parameters are summarised in Table III.

System	Super cell size	a; b; c (Å)	α; β; γ (deg)	Reference
Benzene	2 × 2 × 2	7.398; 9.435; 6.778	90; 90; 90	125
Anatase	1 × 1 × 1	9.674; 9.826; 15.125	74.9; 75.9; 78.3	

Table II. Geometrical parameters of the benchmark systems (see Ref. [STEI2022])

System	N	o	n	v	n_{ADMM}	n_{RI}	*mem* (GB)	*FLOPs*	t
Benzene	384	480	8256	7776	3456	20 352	1703	$1.7 \cdot 10^{18}$	515
Anatase	144	576	5424	4848	2928	15 360	961	$7.2 \cdot 10^{17}$	354

Table III. Information about the performance of the benchmark systems on the Piz Daint supercomputer. N, o, n, v, n_{ADMN} and n_{RI} refer to the number of atoms, the number of occupied orbitals, unoccupied orbitals, basis functions, ADMN auxiliary basis functions, and RI basis functions, respectively. mem provides the minimal required memory for the MP2 gradient calculation. FLOPs refer to the number of floating point operations in the quintical scaling steps. t provides the computational time in node hours. (See Ref. [STEI2022] for details).

3.7 Exciton-Phonon coupling (Demonstrator Dem.5)

Demonstrator Dem.5 was originally intended for the study of electron-phonon interactions and coupled dynamics, and important advances in this direction were already reported in deliverable D6.1. However, in the last period, we have focused our attention on a harder problem, facilitated by the advances made in our codes (Quantum Espresso and Yambo) during this project: the description and understanding of exciton-phonon interactions in materials. We report our results on this area here.

The design of novel optoelectronic and photovoltaic devices requires an in-depth understanding of the microscopic physical mechanisms governing light-matter interaction. In the last few years, many experimental and theoretical breakthroughs have demonstrated the importance of excitonic effects in the interpretation of the optical properties of many promising materials.

The theoretical modelling of excitons - i.e., electron-hole pairs bound by the Coulomb interaction - requires a microscopic, quantum many-body description, formally accessible through the numerical solution of the Bethe-Salpeter equation (BSE). Yet, it completely neglects an important aspect of exciton physics, namely their interaction with the degrees of freedom of the ionic lattice - the phonons or lattice vibrations.

We aim at a description of the exciton-phonon coupling mechanism through a full ab initio approach, in which both excitonic and vibrational properties of realistic materials are calculated from first principles, i.e., without the use of system-specific models and parametrizations. We have then conceived the scheme represented in Figure 16 and described below. The procedure involves the interfacing of Quantum ESPRESSO and Yambo, in order to allow for data transfer from the former to the latter.

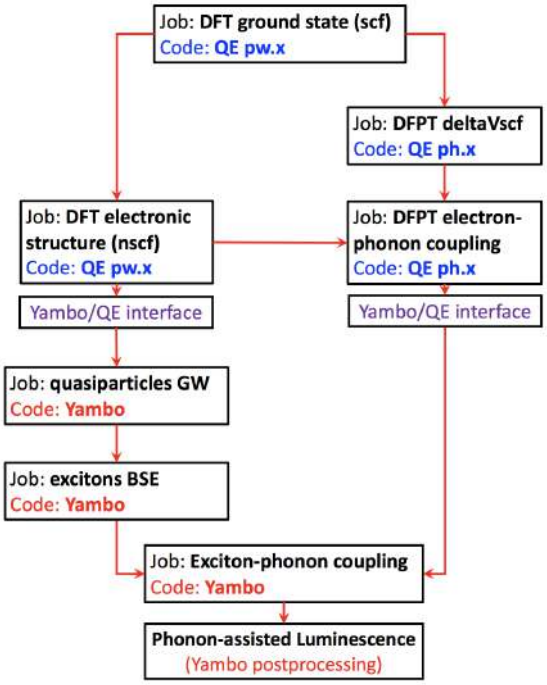


Figure 16. Scheme of workflow with the codes and interfaces for the calculation of exciton-phonon coupling mechanisms.

As a first step, we calculate structural optimization, ground state and electronic structure at the DFT level, using the Quantum Espresso code. These data are used as the starting point for: (i) phonon frequency and electron-phonon coupling calculations (QE, ph.x) and (ii) MBPT calculation first in the GW approximation, followed by the solution of the BSE to determine the optical spectrum and excitonic dispersion (Yambo). Finally, BSE eigenstates, phonon frequencies and electron-phonon coupling elements are used to compute exciton-phonon coupling and photoluminescence spectra.

All the workflow can be run on GPU-accelerated HPC architectures and here we show scalability tests taking advantage of the numerical resources allocated by the PRACE CoE Allocation MaX on Juwels-Booster. The cluster featured 2xAMD EPYC Rome CPUs per node with 512 GB DDR4-3200 RAM and accelerated with 4xNvidia Ampere A100-40 GPUs per node.

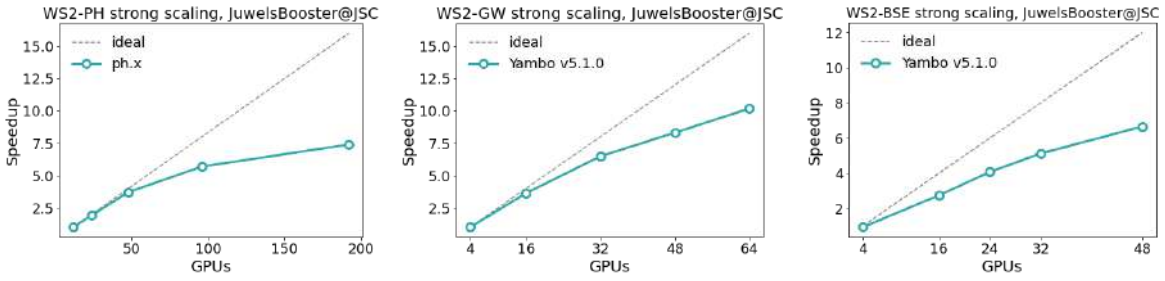


Figure 17. Scaling of Quantum ESPRESSO (ph.x, left) and Yambo (GW in the centre, Bethe-Salpeter in the right) vs. number of GPUs in JewelsBooster@JSC.

The plots in Figure 17 show good scalability for all the codes and run-levels considered, namely Quantum ESPRESSO (ph.x), Yambo-GW, and YAMBO-BSE. All the steps show good scalability and performance, also considering that the tests were performed reducing the input parameters. For the Quantum ESPRESSO ph.x executable, we used a development version that will be available in an upcoming stable release.

3.8 Electrified electrode-electrolyte interfaces: QM and QM/MM with SIESTA (Demonstrator Dem.7)

The structure of water interacting with metallic surfaces is still not fully understood. Although many efforts and progress are being made in experimental techniques, including some with an amazing espacial resolution, the level of detail, space and time resolution that atomistic simulations can provide make them extremely attractive for

achieving a deeper understanding of the processes happening at these interfaces. Numerous applications are reported in the literature on the interaction between water molecules and many kinds of surfaces, using quantum mechanics (QM) and molecular mechanics (MM) simulations. However, an important question, crucial in many fields and especially in electrochemistry, is yet to be answered: what is the effect of the potential applied to the electrodes on water's structure? This question extends to other electrolytes including different solvents and ionic concentrations, which are even less understood than pure water, but are much more relevant for many technologies including batteries, supercapacitors, electro and photocatalysis, to mention just a few.

Answering that question using atomistic simulation is very challenging. One needs a fully QM description to correctly account for the electron's flow on the electrodes and in the possible chemical reactions taking place at or near the interface (in electrochemical processes). However, the typical system sizes and time scales required to describe the structure and dynamics of the electrolyte are computationally prohibitive.

A second difficulty is that many of the relevant processes occur when the electrode is subject to an external voltage, and in the presence of electric current. This implies a non-equilibrium situation which is usually out of the scope of usual DFT technologies. Although some schemes have been devised to treat at least the situation in which an external voltage is applied and induces an electrified electrode-electrolyte interface (EEI), they have not been sufficiently tested and do not provide robust and general solutions to address these non-equilibrium situations. SIESTA has the capability to perform calculations in systems where a voltage bias is imposed between two electrodes, using Non-Equilibrium Green's Functions, as implemented in the TranSIESTA method [BRAN2002, PAPI2017]. We have advanced in using this capability to study electrified EEI's between noble metals and aqueous solutions. In deliverable D6.2 we already presented preliminary results for proof of concept runs in small systems of water confined between two Au electrodes, showing how the dipoles of the H_2O molecules orient according to the voltage applied to the electrodes [ORDE2021]. We also demonstrated the diffusion and drift of Na^+ and Cl^- ions in solution, with and without the presence of a bias potential between the electrodes.

In this period, we advanced in these complex simulations, in two different directions:

Very large QM simulations of the water/gold interface

First, we have made systematic fully QM runs, increasing considerably the size of the systems considered, both in the width of the water layer contained between the electrodes (to provide an accurate account of the evolution of the structure of water from

the interface to the bulk), and in the periodicity of the system along the surface directions (to minimise the spurious effects of the periodic boundary conditions along these directions). Some of the cells used in our simulations (together with those of our first proof-of-concept calculations) are shown in Figure 18. These simulations have confirmed the basic results obtained in our preliminary proof-of-concept runs, but provide a much better account for the water structure, especially the layering near the surface, and the evolution towards the bulk farther from the electrode. They were also run for extended simulation times, reaching some 30 ps and therefore allowing for better statistics. Figure 19 shows a profile of the density of water molecules obtained in an MD simulation of 30 ps in the system shown in the central panel of Figure 18. The layered structure of the liquid is clearly visible at the interfaces with the gold electrodes, with three clearly defined layers, and then going towards the bulk density farther from the interface. Figure 20 shows the evolution of the position of the water molecules along the simulation, also indicating a clear layered structure, but with interdiffusion of molecules between layers at a time scale of picoseconds. A careful analysis of these results indicate that the molecules close to the surface (green and red curves in Figure 20) are less mobile than those farther from the interface with gold, indicating a reduction of their mobility near the surface, but not reaching a frozen structure and conserving its liquid character.

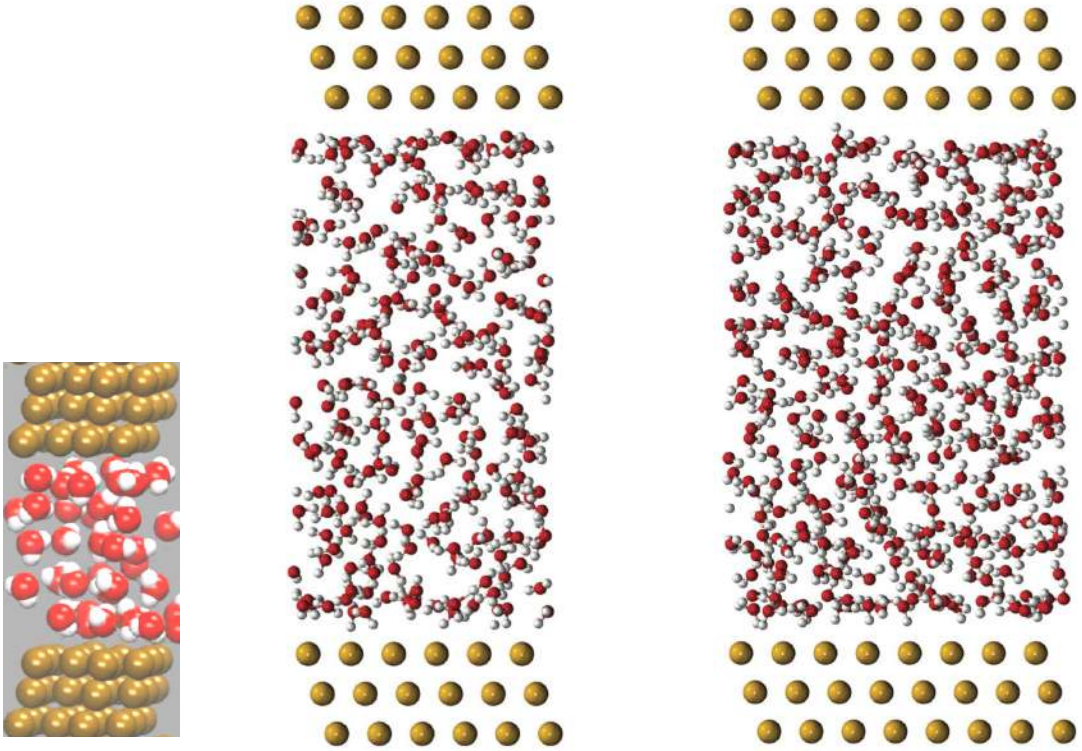


Figure 18: Side view of three models used to study the water-Au interface. The golden, white and red spheres represent Au, H, and O atoms, respectively. Left: model used in the proof-of-concept simulations, containing 40 water molecules and a supercell of the Au(111) surface with 12 Au atoms per layer, reported in deliverable D6.2. Centre and right: models used during this period, with 236 and 378 water molecules and 30 and 46 Au atoms per layer, respectively.

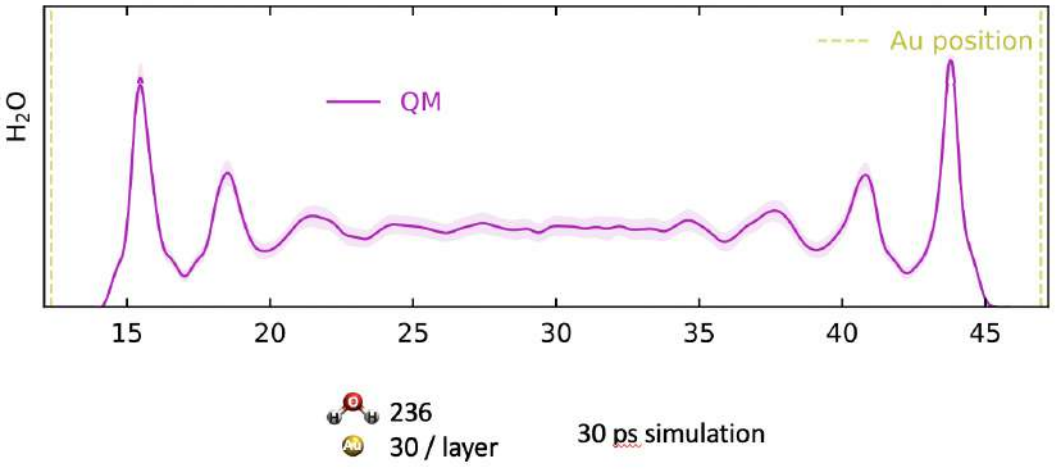


Figure 19. Density profile of water molecules as a function of position along the water slab between Au electrodes. The positions of the two Au surfaces are indicated by the broken lines. The system contains 236 water molecules, and the gold surfaces are periodic, containing 30 gold atoms per layer of the electrode.

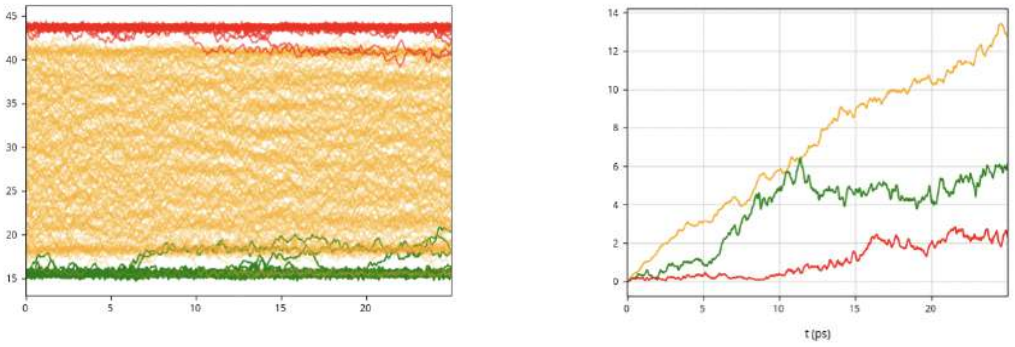


Figure 20. Left: Evolution of the position of each water molecule along the simulation cell, (in Ang) as a function of simulation time (in ps). Right: mean squared displacement of the water molecules along the slab direction, as a function of time. Red and green lines represent water molecules located in the layers in contact with the top and the bottom interfaces, respectively, in the initial steps of the simulation. Yellow lines represent molecules located initially in inner layers.

These calculations were possible due to the work done to improve the parallel performance of SIESTA, and in particular, of its TranSIESTA module that solves the Nonequilibrium Green's Functions (NEGF) which allow us to include the effect of the external potential at the electrodes. In particular, we have greatly improved the OpenMP threading of the code, which allows us to reach considerably larger systems and simulation times. Figure 19 shows the scaling with respect to the number of OpenMP threads in the TranSIESTA run for each MPI process for the system shown in the central panel of Figure 18 (with 236 water molecules). Our implementation profits from speed-ups even when the number of threads reaches the number of cores in one node of MareNostrum 4 (although optimal scaling reaches about half of the cores in one node).

In turn, our MPI implementation currently only includes parallelization over the energy points in the complex plane where the Green's function is calculated. Each of these energy points involves the inversion of a square matrix with the dimension of the number of orbitals in the simulation cell (nearly 10.000 in the system of Figure 21). Each of these inversions is currently assigned to a MPI process. As the number of

energy points is typically of the order of 100, our current code can efficiently use some 1,000 cores (~100 energy points × ~10 OpenMPI processes). We are now working on an MPI implementation where each of these inversions is further distributed among a number of MPI processes, using standard distributed linear algebra libraries (currently SCALAPACK). Results will be presented in the near future, which will allow scaling to at least 50,000 cores (with the inversions distributed over 50 cores or more). Further, GPUs will be enabled by the use of implementation of the distributed linear algebra libraries (ELPA and ELSI) in these architectures, as we successfully did for the diagonalization in SIESTA in WP1 (see D1.1, D2.1, D2.2, D4.3).

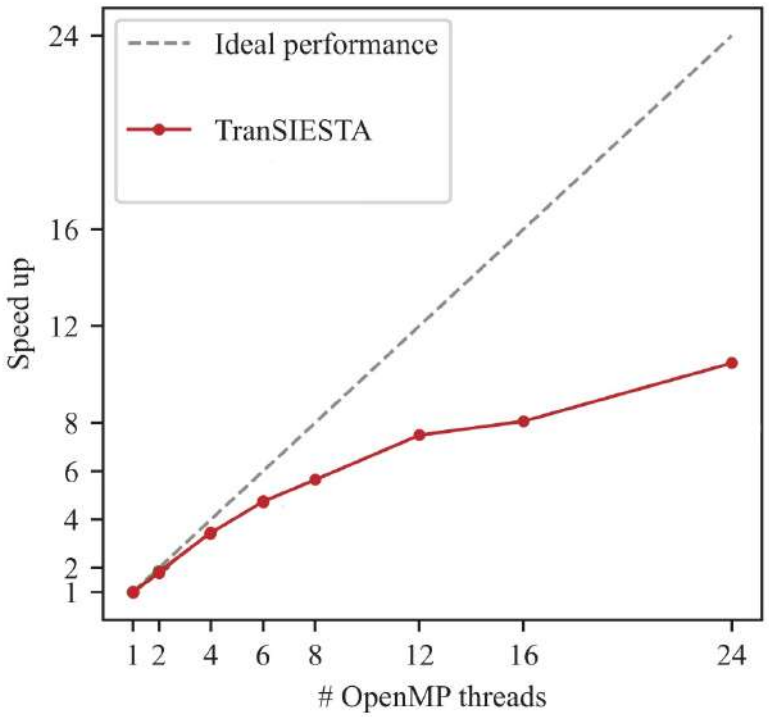


Figure 21. Speedup as a function of OpenMP threads for TranSIESTA in MareNostrum 4. This performance is obtained for the system shown in the central panel of Figure X1 (236 water molecules, and a supercell with 30 atoms on each Au atomic layer in the electrodes). MPI parallelization was done using one MPI process for each of the 96 energy points of the integration contour for the Green's functions. Optimal efficiency is obtained up to 12 OpenMP threads, although substantial speedup is achieved for up to the 24 threads, corresponding to the 24 cores of each of the two sockets of each Intel Xeon Platinum 8160 node of MareNostrum 4.

Extending the size and time scales: QM/MM simulations of the water/gold interface

Although our implementation of NEGF in TranSIESTA is very efficient and allows us to deal with systems of unprecedented size as those shown above, these are still far from

the sizes required to include ions in realistic concentrations (involving one ion per 100 water molecules or less). Dealing with these concentrations is key to describe structure of the double layer formed at the EEI, and therefore, for an appropriate account of all the physical and chemical processes occurring at that interface. Moreover, studying liquid water requires long time scale simulations, and even longer for the diffusion of ions. To address precisely that, we are using a quantum mechanics/molecular mechanics (QM/MM) approach [SANZ2011] coupled to the NEGF method, as implemented in the SIESTA [GARC2020] package combining the best of each approach: the accuracy of ab-initio simulations (which correctly models the electron response in the electrode, from quantum mechanics) and the affordable cost of classical molecular dynamics for the description of the electrolyte.

We have recently revamped the QM/MM interface with SIESTA allowing for very efficient calculations with a notably increased parallel performance, especially for calculations under a bias using the TranSIESTA module. This is enabling us to investigate even larger systems, including counterions in our simulations, fully accounting for the electrode's potential in molecular dynamics (MD) calculations.

Figure 22 shows a comparison of the structure of the water system between Au electrodes, comparing the results of our QM/MM implementation with those of the fully QM simulations presented in Figure 19. The QM/MM approach correctly reproduces the layering of water near the interface with gold, although the specific MM parametrization used results in a less pronounced layering structure from the second layer towards the bulk than the QM simulations. Nevertheless, the main features of the structure and dynamics of water are well reproduced by the QM/MM approach. For this particular system (shown in the centre of Figure 18), the QM/MM runs are approximately 5 times faster than the fully QM runs (as shown in Figure 23), allowing to simulate considerably longer simulations with the same computational cost.

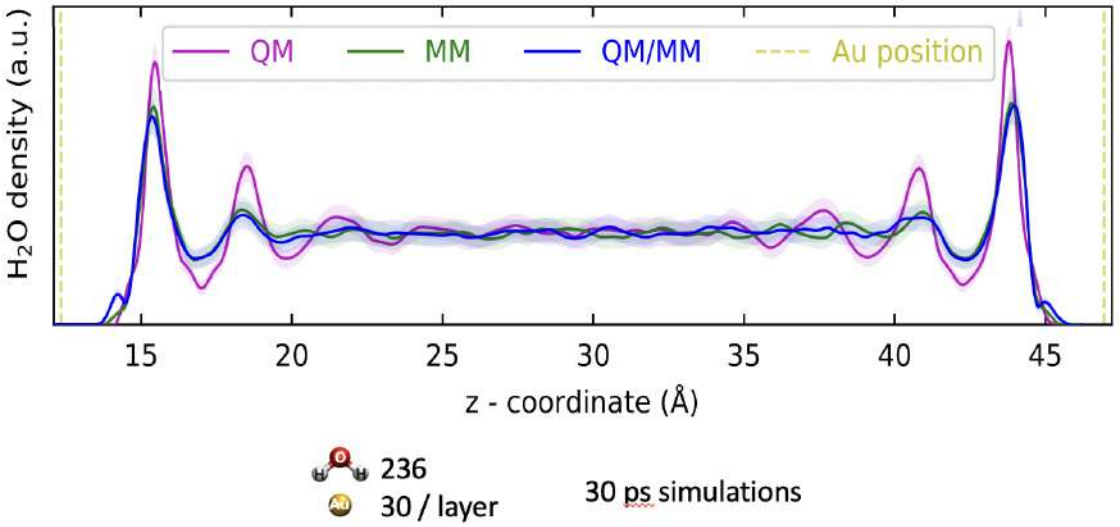


Figure 22. Same as in Figure 19, but comparing the results of a fully QM simulation with those of the QM/MM approach.

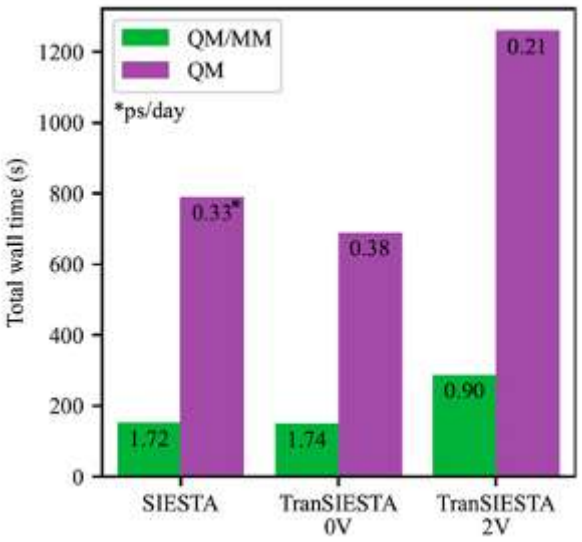


Figure 23. Total wall time for MD runs of the system shown in the central panel of Figure 18, for each 3fs of simulation time. Runs done in 384 cores (16 nodes) of MareNostrum 4. Results labelled SIESTA show the CPU time for SIESTA calculations in periodic supercells (not using Green's functions, but standard diagonalization). TranSIESTA results are those using the NEGF technique, with two different values of the potential applied between the two Au electrodes: 0 Volt (centre) and 2 Volt (right). Green bars are results for the QM/MM runs, and magenta for the fully QM ones. The numbers on each bar indicate the number of picoseconds that can be run per day in

each case, in the 384 cores of MareNostrum 4. QM/MM allows for nearly 1ps per day in this number of nodes.

3.9 High-Throughput Machine-Learned DFT-quality Potentials (Demonstrator Dem.8)

Development of efficient atomic descriptors

Atomistic descriptors are numerical representations of the geometry and chemistry of atomic structures, and are often employed for regression tasks in materials science. Because of their role in building machine learning force fields, and predicting materials' properties, a lot of research and effort has gone into the development of efficient and accurate atomistic descriptors in recent years. Of particular interest, due to their computational efficiency, completeness and generality, are the so-called "atomic cluster expansion" (ACE) descriptors, introduced by Drautz in 2019 [DRAU2019]. Within MaX, we worked on a Python implementation of such descriptors, which is now publicly available as a Python package, named "Raffy", available at https://github.com/ClaudioZeni/Raffy. The development of such implementation also led to the writing of Ref. [ZENI2021], where we explored the accuracy of ridge regression force fields employing ACE descriptors when predicting formation energies for small molecules, and forces and total energies in crystalline solids. In addition, in that work we also investigated principal component analysis (PCA) and least absolute shrinkage selection operator (LASSO) as dimensionality reduction techniques to obtain a more compact, yet fully informative, form of the atomic descriptor. We find that, for data sets that span six elemental crystals, the features selected by PCA share a high fraction of their space, hinting at the possibility of devising a very low-dimensional atomic descriptor that yields prediction accuracies on energies on-par with the ones of the original ACE descriptors.

Furthermore, the so-designed ACE descriptors have been added to the PANNA [LOT2020] python package, improving its functionalities by adding explicit 3-body descriptors as options.

Data-driven simulation and analysis of gold nanoparticle melting

The simulation and analysis of the thermal stability of nanoparticles, a stepping stone towards their application in technological devices, require fast and accurate force fields, in conjunction with effective characterisation methods. We worked in collaboration with groups at the Ecole Polytechnique Federale de Lausanne (EPFL, CH), King's College London (KCL, UK), Swansea University (UK), and Aristotle University of

Thessaloniki (Greece) on the development of efficient, transferable, and interpretable machine learning force fields for gold nanoparticles based on data gathered from Density Functional Theory calculations. We then applied them to the investigation of the thermodynamic stability of gold nanoparticles of different sizes (1 to 6 nm), containing up to 6266 atoms, concerning a solid- liquid phase change through molecular dynamics simulations. The methods employed predict melting temperatures in good agreement with available experimental data. Furthermore, the solid-liquid phase change mechanism is observed thanks to the use of an unsupervised learning scheme to classify local atomic environments. We were therefore able to provide a data-driven definition of liquid atomic arrangements in the inner and surface regions of a nanoparticle and employ it to show that melting initiates at the outer layers. The above results have been published in Nature Communications [ZENI2021-2].

On the robust extrapolation of machine learning potentials and Ensemble of linear potentials

In collaboration with researchers at EPFL (lausanne, CH), Roche Farma Research (Basel, CH) we worked on the extrapolation and interpolation of high-dimensional machine learning potentials. In a paper appeared on Physical Reviews B [ZENI2022] we showed that, contrary to widespread assumptions, predictions from machine learning potentials built upon high-dimensional atom-density representations almost exclusively occur in regions of the representation space which lie outside the convex hull defined by the training set points. We then propose a perspective to rationalise the domain of robust extrapolation and accurate prediction of atomistic machine learning potentials.

In collaboration with the same team we worked on the theory and implementation of locally linear machine learning potentials. The method allows one to fit one fast ridge regression linear potential based on atomic cluster expansion (ACE) descriptors for each “class” of local atomic environment. Such classes of local atomic environment are identified via clustering of the training set. In prediction, atoms are assigned one weight per class, and energy/force/stress predictions are obtained as a weighted average of the predictions yielded by each linear potential. This method is as fast as linear potentials, while being typically more accurate, especially on large databases. A manuscript on the topic is under preparation.

4 Conclusions

The Demonstrators in MaX have been extremely useful in understanding the barriers challenging the deployment of our codes in extreme HPC infrastructures. The availability of

many advanced computational installations in Europe, including the recent pre-exascale efforts (those already partially available like LUMI), have allowed us to experiment challenging and realistic environments in the way towards the exascale.

Most of the Demonstrator projects have obliged us to modify, refine, or develop new numerical algorithms, convergence procedures, computing schemes and approaches, to adapt them to the massive calculations required. Tuning compilation and run-time parameters has also been a key aspect of our advance.

Many workflows for automatising the multiple calculations and code executions required in several Demonstrators have been developed. Some of these have been focused on the interconnection between different codes involved in the calculation of complex materials properties, in which the output of one code is needed as input for another. In those cases where this interconnection requires the transfer of massive amounts of data from one code to the other (such as the wave functions obtained for each MD time step), the workflows have been focused on handling these communications without resourcing to standard (disk) I/O, while the codes are being executed concurrently. This approach has demonstrated to be possible, and successful, in several of our Demonstrators, and opens an interesting way of interfacing different codes, possibly replacing steps in the calculation with the most adequate code for each task (e.g., picking the quantum engine to compute the wave functions, depending on the type of system and the accuracy required).

The large amount of work invested in MaX technical work-packages in adapting our codes to GPU-based machines has paid off in their deployment on the most recent HPC infrastructures (e.g.: Marconi100, Jewels Booster, LUMI in Europe; Perlmutter in the US). All the MaX codes are not only CPU, but GPU-ready, and have shown good potential for future deployment in exascale machines. Some of them are indeed included among the standard benchmarks of the HPC centres.
The work has also produced interesting results in the development of ideas, methods and algorithms for the analysis of the Quantum Mechanical results from our codes, to extract physical properties which were previously unavailable or difficult to obtain from these complex simulations. This allows us to predict new properties like the thermal conductivity, or analyse complex results in simple terms, like the fragment interaction analysis developed for the study of biomolecules. Processes like those happening at the electrified electrolyte-electrode interface are also now accessible due to the new techniques based on non-equilibrium Green's functions.

We have also explored new computational schemes, including the interaction of our calculations with materials databases (like toe Materials Cloud 2D Database), or incorporating

Artificial Intelligence methods in the workflows. These techniques have been explored in many of our Demonstrators, and are a future avenue of disruptive improvement of our computational technologies.

In conclusion, the Demonstrators have allowed us to be ready for the use of pre-exascale machines, and prepared to tackle the challenges of the coming exascale machines. Much work is needed to improve performance, automation, data analysis, and many other aspects of the massive calculations (as described in more detail in deliverable D6.4), but the results of MaX represent an optimal starting point in this read towards exascale.

5 References:

[ATAM2019] M. O. Atambo, D. Varsano, A. Ferretti, S. S. Ataei, M. J. Caldas, E. Molinari, and A. Selloni, "Electronic and optical properties of doped TiO2 by many-body perturbation theory", Phys. Rev. Mater. **3**, 045401 (2019)
https://doi.org/10.1103/PhysRevMaterials.3.045401

[BERT2019] R. Bertossa, F. Grasselli, L. Ercole, and S. Baroni, "Theory and Numerical Simulation of Heat Transport in Multicomponent Systems", Phys. Rev. Lett. **122**, 255901 (2019).
https://doi.org/10.1103/PhysRevLett.122.255901

[BETT2022] M. G. Betti et al., "Gap Opening in Double-Sided Highly Hydrogenated Free-Standing Graphene", *Nano Lett.* **22**, 7, 2971 (2022)
https://doi.org/10.1021/acs.nanolett.2c00162

[BIG] http://bigdft.org/

[BRAN2002] Mads Brandbyge, José-Luis Mozos, Pablo Ordejón, Jeremy Taylor, and Kurt Stokbro, "Density-functional method for nonequilibrium electron transport", Phys. Rev. B **65**, 165401 (2002).
https://doi.org/10.1103/PhysRevB.65.165401

[CANC2007] Cancès, E.; Legoll, F.; Stoltz, G. "Theoretical and Numerical Comparison of Some Sampling Methods for Molecular Dynamics". ESAIM: M2AN 2007, **41**, 351–389.

[DRAU2019] Ralf Drautz, "Atomic cluster expansion for accurate and transferable interatomic potentials", Phys. Rev. B **99**, 014104 (2019); Erratum Phys. Rev. B **100**, 249901 (2019).

https://doi.org/10.1103/PhysRevB.99.014104

[ERCO2017] L. Ercole, A. Marcolongo, and S. Baroni, “Accurate thermal conductivities from optimally short molecular dynamics simulations”, Sci. Rep. **7**, 15835 (2017)
https://doi.org/10.1038/s41598-017-15843-2

[ERCO2022] Loris Ercole, Riccardo Bertossa, Sebastiano Bisacchi, Stefano Baroni, “SporTran: A code to estimate transport coefficients from the cepstral analysis of (multivariate) current time series”, Computer Physics Communications, **280**, 108470 (2022)
https://doi.org/10.1016/j.cpc.2022.108470

[GARC2020] A. García et al., "Siesta: Recent developments and applications", J. Chem. Phys. 152, 204108 (2020)
https://doi.org/10.1063/5.0005077

[GENO2020] L. Genovese, W. Dawson, T. Nakajima, V. Cristiglio, V. Vallet and M. Masella, "Microscopic Factors Modulating the Interactions Between the SARS-CoV-2 Main Protease and α−Ketoamide Inhibitors", ChemRxiv. Cambridge: Cambridge Open Engage; 2020; This content is a preprint and has not been peer-reviewed.
https://doi.org/10.26434/chemrxiv.12924974.v2

[GENO2021] Luigi Genovese, Marco Zaccaria, Michael Farzan, Welkin Johnson, and Babak Momeni, "Investigating the mutational landscape of the SARS-CoV-2 Omicron variant via ab initio quantum mechanical modeling", bioRxiv preprint 2021
https://doi.org/10.1101/2021.12.01.470748

[GIAN2017] P Giannozzi et al, “Advanced capabilities for materials modelling with Quantum ESPRESSO”, J. Phys.: Condens. Matter **29,** 465901 (2017)
https://doi.org/10.1088/1361-648x/aa8f79

[GUANArXiV] A. Guandalini, P. D’Amico, A. Ferretti and D. Varsano, “Efficient GW calculations in two dimensional materials through a stochastic integration of the screened potential”.
https://arxiv.org/abs/2205.11946

[LOT2020] Ruggero Lot, Franco Pellegrini, Yusuf Shaidu, and Emine Küçükbenli, “Panna: Properties from artificial neural network architectures”, Computer Physics Communications, **256**, 107402 (2020)
https://doi.org/10.1016/j.cpc.2020.107402

[MARC2021] Aris Marcolongo, Riccardo Bertossa, Davide Tisi, Stefano Baroni, "QEHeat: An open-source energy flux calculator for the computation of heat-transport coefficients from first principles", Computer Physics Communications, **269**, 108090 (2021)
https://doi.org/10.1016/j.cpc.2021.108090

[MASE2013] Masella M and Borgis D and Cuniasse P, "A multiscale coarse-grained polarizable solvent model for handling long tail bulk electrostatics", J Comput Chem, **34,** 1112 (2013)
https://doi.org/10.1002/jcc.23237

[MUB2016] Marcolongo, A., Umari, P. & Baroni, S. "Microscopic theory and quantum simulation of atomic heat transport", *Nature Phys* **12**, 80–84 (2016).
https://doi.org/10.1038/nphys3509

[ORDE2021] P. Ordejón and P. Febrer, "Addressing electrified metal-electrolyte interfaces with Non-Equilibrium Green's Functions", Bulletin of the American Physical Society, 2021
https://meetings.aps.org/Meeting/MAR21/Session/X19.1

[PAPI2017] Nick Papior, Nicolás Lorente, Thomas Frederiksen, Alberto García and MadsBrandbyge, "Improvements on non-equilibrium and transport Green function techniques: The next-generation transiesta", Computer Physics Communication **212**, 8-24 (2017)
https://doi.org/10.1016/j.cpc.2016.09.022

[POL] http://biodev.cea.fr/polaris/

[SANZ2011] Carlos F. Sanz-Navarro, Rogeli Grima, Alberto García, Edgar A. Bea, Alejandro Soba, José M. Cela and Pablo Ordejón, "An efficient implementation of a QM–MM method in SIESTA", Theor Chem Acc **128**, 825 (2011).
https://doi.org/10.1007/s00214-010-0816-5

[STEI2022] Frederick Stein and Jürg Hutter, "Double-hybrid density functionals for the condensed phase: Gradients, stress tensor, and auxiliary-density matrix method acceleration", J. Chem. Phys. **156**, 074107 (2022)
https://doi.org/10.1063/5.0082327

[TANN2021] A. Tanner, B. Wen, Y. Zhang, et al., "Photoexcitation of bulk polarons in rutile TiO2", Phys. Rev. B **103**, L121402 (2021)
https://doi.org/10.1103/PhysRevB.103.L121402

[TISI2021] D. Tisi, L. Zhang, R. Bertossa. H. Wang, R. Car, S. Baroni, " Heat transport in liquid water from first principles and deep neural network simulations", Phys. Rev. B **104,** 224202 (2021)
https://doi.org/10.1103/PhysRevB.104.224202

[TUNY2021] Tunyasuvunakool, K., Adler, J., Wu, Z. *et al.* "Highly accurate protein structure prediction for the human proteome". *Nature* **596,** 590–596 (2021).
https://doi.org/10.1038/s41586-021-03828-1

[VALI2022] D. L. Valido, C. Cardoso, D. Varsano, E. Molinari and A. Ferretti, "A multipole approach for dielectric screening in metallic systems", Bulletin of the American Physical Society, 2022
https://meetings.aps.org/Meeting/MAR22/Session/A48.7

[WANG2015] Z. Wang, B. Wen, Q. Hao, et al., "Localized Excitation of Ti3+ Ions in the Photoabsorption and Photocatalytic Activity of Reduced Rutile TiO2", J. Am. Chem. Soc. **137**, 9146 (2015)
https://doi.org/10.1021/jacs.5b04483

[WANG2021] T. Wang, W. Chen, S. Xia, *et al.*, "Anisotropic d–d Transition in Rutile TiO2", J. Phys. Chem. Lett. **12**, 10515 (2021)
https://doi.org/10.1021/acs.jpclett.1c02931

[WEIZ2017] Weitzner, B D and Jeliazkov, J R and Lyskov, S and Marze, N and Kuroda, D and Frick, R and Adolf-Bryfogle, J and Biswas, N and Dunbrack Jr, R L and Gray, J J, "Modeling and docking of antibody structures with Rosetta", Nat. Protocols., **12**, 401 (2017).
https://doi.org/10.1038/nprot.2016.180

[ZACC2021] M. Zaccaria, L. Genovese, M. Farzan, W. Dawson, T. Nakajima, W. Johnson and B. Momeni, "Anticipating future SARS-CoV-2 variants of concern through ab initio quantum mechanical modeling", bioRxiv preprint 2021
https://doi.org/10.1101/2021.11.25.470044

[ZACC2022]Marco Zaccaria, Luigi Genovese, William Dawson, Viviana Cristiglio, Takahito Nakajima, Welkin Johnson, Michael Farzan, Babak Momeni, "Probing the mutational

landscape of the SARS-CoV-2 spike protein via quantum mechanical modeling of crystallographic structures", PNAS Nexus 2022, pgac180,
https://doi.org/10.1093/pnasnexus/pgac180

[ZENI2021] Claudio Zeni, Kevin Rossi, Aldo Glielmo, and Stefano de Gironcoli, "Compact atomic descriptors enable accurate predictions via linear models", Journal of Chemical Physics **154**, 224112 (2021)
https://doi.org/10.1063/5.0052961

[ZENI2021-2] Claudio Zeni, Kevin Rossi, Theodore Pavloudis, Joseph Kioseoglou, Stefano de Gironcoli, Richard E Palmer, and Francesca Baletto, "Data-driven simulation and characterisation of gold nanoparticle melting", *Nat Commun* **12**, 6056 (2021)
https://doi.org/10.1038/s41467-021-26199-7

[ZENI2022] Claudio Zeni, Andrea Anelli, Aldo Glielmo, and Kevin Rossi, "Exploring the robust extrapolation of high-dimensional machine learning potentials", Phys. Rev. B **105**, 165141 (2022)
https://doi.org/10.1103/PhysRevB.105.165141

[ZEROMQ] https://zeromq.org

[ZHAN2020] Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. "Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved -Ketoamide Inhibitors". Science, **368,** 409-412 (2020).
https://doi.org/10.1126/science.abb3405