



HAL
open science

Affordable and accurate large-scale hybrid-functional calculations on GPU-accelerated supercomputers

Laura Ratcliff, A Degomme, A Flores-Livas, Stefan Goedecker, Luigi Genovese

► **To cite this version:**

Laura Ratcliff, A Degomme, A Flores-Livas, Stefan Goedecker, Luigi Genovese. Affordable and accurate large-scale hybrid-functional calculations on GPU-accelerated supercomputers. *Journal of Physics: Condensed Matter*, 2018, 30 (9), pp.095901. 10.1088/1361-648X/aaa8c9 . hal-02141904

HAL Id: hal-02141904

<https://hal.science/hal-02141904>

Submitted on 21 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accurate, Large-Scale and Affordable Hybrid-PBE0 Calculations with GPU-Accelerated Supercomputers

Laura E. Ratcliff

Argonne Leadership Computing Facility, Argonne National Laboratory, Illinois 60439, USA

Department of Materials, Imperial College London, London SW7 2AZ, UK

E-mail: laura.ratcliff08@imperial.ac.uk

A. Degomme

Department of Physics, Universität Basel, Klingelbergstr. 82, 4056 Basel, Switzerland

José A. Flores-Livas

Department of Physics, Universität Basel, Klingelbergstr. 82, 4056 Basel, Switzerland

Stefan Goedecker

Department of Physics, Universität Basel, Klingelbergstr. 82, 4056 Basel, Switzerland

Luigi Genovese

Univ. Grenoble Alpes, CEA, INAC-SP2M, L_Sim, F-38000, Grenoble, France

E-mail: luigi.genovese@cea.fr

December 2017

Abstract. Performing high accuracy hybrid functional calculations for condensed matter systems containing a large number of atoms is at present computationally very demanding – when not out of reach – if high quality basis sets are used. We present a highly efficient multiple GPU implementation of the exact exchange operator which allows hybrid functional density-functional theory calculations with systematic basis sets without additional approximations for up to a thousand atoms. This method is implemented in a portable real-space-based algorithm, released as an open-source package. With such a framework hybrid DFT calculations of high quality become accessible on state-of-the-art supercomputers within a time-to-solution of the same order of magnitude as traditional semilocal-GGA functionals.

1. Introduction

Density-functional theory (DFT) in principle is not an approximation, and should produce the exact ground-state energy and density, but in practice the crucial contribution, namely the exchange-correlation (XC) energy, is provided by an effective and therefore approximated, functional. Unsurprisingly, the quality of the result when compared to experiment depends on the quality of this approximation.

A bibliographic search (see [Appendix B](#)) reveals that some 80 to 90 percent of all density-functional calculations conducted over the last 5 years in the chemistry community with Gaussian-basis codes such as GAUSSIAN [1] or NWCHEM [2] use highly accurate hybrid functionals. Among the most popular and widely used are B3LYP [3, 4], PBE0 [5] and HSE06 [6] to name a few. These calculations use the best available functionals, however the total accuracy is limited by the size of the Gaussian basis set which cannot be taken to be very large in these calculations, because of the high computational cost. The most widely used basis set (in more than 90% of the publications analyzed) in such hybrid functional calculations is the **6-31G** basis set which gives for instance atomization energy errors of about 30 kcal/mol †, i.e. about 30 times worse than the desired chemical accuracy [7], see for instance Jensen *et al.* [8]. It is not surprising that one of the most accurate basis sets, AUG-cc-pV5Z, which provides in most cases chemical accuracy, is used in less than 2 percent of the hybrid functional calculations because it is computationally prohibitive.

In contrast to Gaussian-type basis sets, systematic basis sets such as plane waves or wavelets allow one to approximate the exact solution with arbitrarily high precision and with a moderate increase in cost within the adopted methodological framework. The framework is in this case the chosen exchange correlation functional together with the pseudopotential or PAW scheme [9, 10]. The latest generations of pseudopotentials and PAW schemes are able to deliver essentially chemical accuracy [7] such that they do not compromise the overall accuracy of the calculation [11]. Moreover the number of KS orbitals considered in all-electron quantum chemistry codes is much larger than in pseudopotential approaches, where one has to consider only valence and possibly semicore states. By far the most popular codes based on systematic basis sets are the plane wave codes used preferentially in materials science and solid state physics such as QUANTUM-ESPRESSO [12], VASP, [13] and ABINIT [14].

However, less than 10 percent of the production

†See NIST website for the enthalpy of atomization (experimental values), <http://cccbdb.nist.gov/ea1.asp>

results coming from these codes employ hybrid XC functionals. The main reason is that, for typical systems, hybrid functional calculations are *one to two orders of magnitude* more expensive than DFT-GGA calculations. It is therefore hardly possible, due to limited computer resources, to afford the computational power needed for such a treatment *without introducing* additional approximations of the treatment, which might spoil the precision of the result in an unpredictable way. In other terms, even though accuracies that are very close to the desired chemical accuracy are in principle possible – thanks to the availability of high quality basis sets as well as accurate exchange correlation functionals – calculations that really attain this accuracy are still rare because of the high numerical cost for such simulations. Unsurprisingly the great majority of studies therefore still rely on local and semilocal LDA/GGA functionals.

As anticipated, one way to circumvent the price that one has to pay for hybrid calculations is to exploit approximations or use a reduced representation of a set of KS orbitals. It has been shown by Gygi [15] that the algebraic decomposition of the matrix of wave-function coefficients provides a linear transformation which optimally localizes wave-functions on arbitrary parts of the basis set. This methodology drastically reduces the cost of hybrid-PBE0 calculations. However, the speed-up reached by deploying three-dimensional subspace bisection algorithms depends to a great extent on the degree of localization given by a tolerance ϵ that controls the quality of the results [16]. More recently, Lin [17] also used adaptive methods to compress the Fock-exchange operator, which also reduced the computational time associated with the calculation of this operator while not losing accuracy. However, these compressed methods are tailored for self-consistent field (SCF) optimizations based on a density-mixing scheme, on which the compression is performed for a given choice of the KS orbitals. This makes them unsuitable for use in less expensive SCF approaches like a direct minimization scheme, which are of utmost importance, for instance, in accurate and efficient *ab-initio* molecular dynamics trajectories. Thus, in order to study larger systems and more complex materials, we need to account for the requirements of a systematic basis set, computational affordability and algorithms without adjustable parameters.

In this work, we present a highly efficient GPU implementation of a real-space based algorithm for the evaluation of the exact exchange, which reduces the cost of hybrid functional calculations in systematic basis sets, *without any approximation*, by nearly one order of magnitude.

2. Algorithm and Implementation

We start this section with a brief introduction of the main quantities which will be the subject of analysis in the rest of the paper. In the (generalized) KS-DFT approach the one-body density matrix of the system is defined in terms of the occupied KS orbitals ψ_i :

$$F_\sigma(\mathbf{r}, \mathbf{r}') = \sum_i f_{i,\sigma} \psi_{i,\sigma}^*(\mathbf{r}') \psi_{i,\sigma}(\mathbf{r}), \quad (1)$$

where we explicitly specify the (collinear) spin degrees of freedom with the index $\sigma = \uparrow, \downarrow$, together with the occupation number $f_{i,\sigma}$. The system's electrostatic density is evidently the diagonal part of F , i.e. $\rho(\mathbf{r}) = \sum_\sigma F_\sigma(\mathbf{r}, \mathbf{r})$.

The calculation of the exact exchange energy E_X requires a double summation over all the N occupied orbitals

$$\begin{aligned} E_X[\hat{F}] &= -\frac{1}{2} \sum_\sigma \int d\mathbf{r} d\mathbf{r}' \frac{F_\sigma(\mathbf{r}, \mathbf{r}') F_\sigma(\mathbf{r}', \mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} \\ &= -\frac{1}{2} \sum_{i,j,\sigma} f_{i,\sigma} f_{j,\sigma} \int d\mathbf{r} d\mathbf{r}' \frac{\rho_{ij}^\sigma(\mathbf{r}) \rho_{ji}^\sigma(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \end{aligned} \quad (2)$$

where we have defined $\rho_{ij}^\sigma(\mathbf{r}) = \psi_{j,\sigma}^*(\mathbf{r}) \psi_{i,\sigma}(\mathbf{r})$. The diagonal ($i = j$) contribution to E_X exactly cancels out the Hartree electrostatic energy $E_H[\rho]$. The action of the Fock operator \hat{D}_X to be added to the KS Hamiltonian directly stems from the E_X definition:

$$\begin{aligned} \hat{D}_X |\psi_{i,\sigma}\rangle &= \int d\mathbf{r} d\mathbf{r}' \frac{\delta E_X[\hat{F}]}{\delta F_\sigma(\mathbf{r}, \mathbf{r}')} \psi_{i,\sigma}(\mathbf{r}') |\mathbf{r}\rangle \\ &= - \sum_j \int d\mathbf{r}' f_{j,\sigma} V_{ij}^\sigma(\mathbf{r}) \psi_{j,\sigma}(\mathbf{r}') |\mathbf{r}\rangle, \end{aligned} \quad (3)$$

where we have defined

$$V_{ij}^\sigma(\mathbf{r}) = \int d\mathbf{r}' \frac{\rho_{ji}^\sigma(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad (4)$$

that is the solution of the Poisson's equation $\nabla^2 V_{ij}^\sigma = -4\pi \rho_{ij}^\sigma$. In a KS-DFT code which searches for the ground state orbitals, one has to repeatedly evaluate, during the SCF procedure, for a given set of $\psi_{i,\sigma}(\mathbf{r})$, the value of E_X as well as the action of the corresponding Fock operator \hat{D}_X on the entire set of occupied orbitals.

If each orbital $\psi_{j,\sigma}(\mathbf{r})$ is written as a linear combination of M Gaussian basis functions, a straightforward evaluation of the exact exchange energy has an M^4 scaling. Such a scaling might constitute a severe limitation for calculations with highly precise basis sets, where diffuse functions are needed to approach completeness. As a consequence the exact exchange becomes extremely expensive to calculate when very high quality atomic basis sets are used. Plane wave and wavelet basis set density-functional codes evaluate the exact exchange in a different way. They form all the $N(N+1)/2$ charge

densities $\rho_{ij}^\sigma(\mathbf{r})$ and then solve the Poisson's equation for each of them. Then the operator of Eq. 3 can be evaluated as well as the value of

$$E_X = -\frac{1}{2} \sum_{i,j,\sigma} f_{i,\sigma} f_{j,\sigma} \int d\mathbf{r} V_{ij}^\sigma(\mathbf{r}) \rho_{ji}^\sigma(\mathbf{r}). \quad (5)$$

The scaling of these operations is therefore $\mathcal{O}(N^2)$ multiplied by the scaling of the Poisson's equation, which is generally of $\mathcal{O}(N \log N)$ for typical Poisson solvers used in the community.

For large systems which exhibit the nearsightedness principle, exponentially [18] localized orbitals $\phi_\alpha(\mathbf{r})$ can be constructed [19] to represent the density matrix in terms of the matrix \mathbf{K} :

$$F(\mathbf{r}, \mathbf{r}') = \sum_{\alpha\beta} K^{\alpha\beta} \phi_\alpha(\mathbf{r}) \phi_\beta(\mathbf{r}'), \quad (6)$$

where we omit for simplicity the spin index and consider all the KS orbitals as real functions. After truncation within a localization region they can be used for the calculation of the Hartree exchange term. In this case the matrix \mathbf{K} becomes sparse so that the scaling reduces to a scaling proportional to N with a very large prefactor:

$$E_X = \sum_{\alpha\beta\gamma\delta} K^{\alpha\gamma} K^{\beta\delta} \int d\mathbf{r} V_{\alpha\beta}(\mathbf{r}) \rho_{\gamma\delta}(\mathbf{r}) \quad (7)$$

More precisely the scaling is given by $N \log(N) m^3$ where m is the number of local orbitals for which the factor $K^{\alpha\gamma} K^{\beta\delta}$ is nonzero. The method we have developed could in principle also exploit orbitals constrained to localization regions, however we do not present such results since we are instead concentrating on optimization aspects for a given fixed simulation region.

2.1. Calculation Steps

The starting point for evaluating the exact exchange terms (Eqs. 2, 3) is the construction of the electrostatic potentials arising from the co-densities,

$$\rho_{i,j}(\mathbf{r}) = \psi_i^*(\mathbf{r}) \psi_j(\mathbf{r}), \quad (8)$$

for each pair of orbitals i and j . In the following we omit for simplicity the spin index σ . We here describe the communication mechanism implemented to calculate the entire set of the $\rho_{i,j}(\mathbf{r})$.

For a pool of P MPI processes each labelled by $p = 0, \dots, P-1$ we assume that each process owns a subset of orbitals $\psi_{i_p}(\mathbf{r})$. Such a process is clearly able to calculate the ρ_{i_p, j_p} codensities and the corresponding V_{i_p, j_p} . While performing such calculations, each process also sends (receives) the orbitals to (from) the $p \pm 1 \pmod P$ process. Once this communication is terminated, the process p is also capable of calculating $\rho_{i_p, j_{p-1}}$ and $V_{i_p, j_{p-1}}$. The latter transition potential is

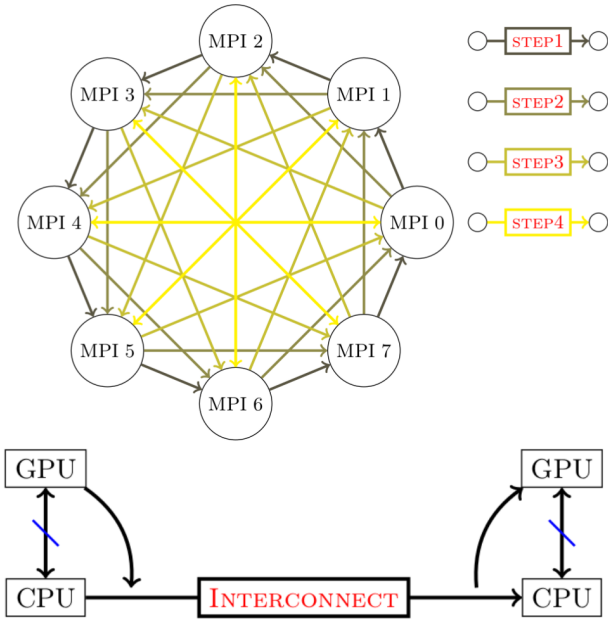


Figure 1. Sketch of the round-robin parallelization scheme used to calculate E_X from Eq. 2. Each communication step is overlapped with the calculation of the Poisson’s equation for the data that was communicated in the previous step. The advantage of the GPU-Direct scheme is illustrated in the bottom panel. GPU buffers are explicitly passed to MPI routines instead of sending data back and forth to the GPU.

of interest at the same time for calculating the value of E_X and the action of \hat{D}_X on *both* the orbitals ψ_{i_p} and $\psi_{j_{p-1}}$. For this reason, once calculated, all the potentials $V_{i_p, j_{p-1}}$ are sent back to the $p-1$ process for the partial calculation of $\hat{D}_X |\psi_{i_{p-1}}\rangle$. Such a procedure is repeated $[P/2] + 1$ times, where at each step s each process communicates with the $p \pm s \bmod P$ MPI processes of the same pool. A schematic of this communication procedure is illustrated in Fig. 1.

2.1.1. The Calculation Kernels on Graphics Processing Units Since the orbitals represent substantial data packets the communication cost of this step is significant. Therefore, the communication is overlapped with the calculation using the data communicated at the previous step, thereby hiding the cost of the communication. The entire calculation procedure is decomposed into the following subprograms (kernels):

- Kernel 1: Calculate the charge density $\rho_{i,j}(\mathbf{r}) = \psi_i^*(\mathbf{r}) \psi_j(\mathbf{r})$ on a real space grid;
- Kernel 2: Solve Poisson’s equation $\nabla^2 V_{ij} = -4\pi \rho_{ij}$;
- Kernel 3: Obtain the energy density and gradient by multiplying $V_{i,j}(\mathbf{r})$ with $\rho_{i,j}(\mathbf{r})$ or $\psi_j(\mathbf{r})$ and $\psi_i(\mathbf{r})$ respectively.

When a GPU accelerator is available, the above

operations can be accelerated considerably with respect to the equivalent CPU kernels.

For what concerns the communication scheduling, best performance was obtained using non-blocking Isend/Irecv MPI communications and by taking advantage of the streaming capacities of GPUs, allowing the MPI library to perform the necessary tasks in the background. One-sided remote memory access (RMA) communications were evaluated and found to be slower on the tested machine and MPI implementations. This may however change depending on the communication libraries, the network or the tested cluster, and so this communication scheme can be turned on through an option in the input file if required. The Isend/Irecv scheme also has the advantage that it is applicable if using localized orbitals, which lead to unstructured communication patterns.

2.1.2. Poisson Solver Most of the numerical work is done in kernel 2. For this step we use the established approach based on interpolating scaling functions [20, 21], which has also been ported to GPU [22]. The basic advantage of this approach is that the real-space values of the potential V_{ij} are obtained with very high accuracy on the uniform mesh of the simulation domain, via a direct solution of Poisson’s equation by convolving the density with the appropriate Green’s function of the Laplacian. The Green’s function can be discretized for the most common types of boundary conditions encountered in electronic structure calculations, namely free, wire, slab and periodic. This approach can therefore be straightforwardly used in all DFT codes that are able to express the densities ρ_{ij} on uniform real-space grids. This is very common because the XC correlation potential is usually calculated on such grids. This approach has already turned out in its parallel CPU version to be fastest under most circumstances [23] and is therefore integrated in various DFT codes such as ABINIT [14], CP2K [24], OCTOPUS [25] and CONQUEST [26]. Recently this Poisson solver has been extended to also include electrostatic environments [27, 28].

The convolution is performed exactly with $N \log(N)$ operations using Fourier techniques, in which appropriate padding is employed in order to avoid any supercell effects. In our GPU implementation we use the cuFFT library from NVIDIA. This leads to an efficient implementation of the Poisson solver, as described in Ref. [22]. Such a GPU-based implementation of the Poisson solver already gives a gain in speed over the CPU version, but the speed-up is limited by the large amount of host-GPU communication in such a simple scheme where the data are stored on the host

memory.

2.1.3. GPU Direct Communication Since large scale density-functional calculations are typically done on massively parallel computers, one clearly also has to exploit such architectures for hybrid functional calculations. Even if all the floating point operations are performed on a GPU, part of the efficiency would be lost in a parallel run if the data have to be transferred to the CPU for the MPI communication calls. For this reason we use the GPUDirect communication scheme which allows the MPI calls to be performed on the GPU.

Host-GPU communications have always been one of the main issues with GPU acceleration. However, the Nvidia GPUDirect technology allows the acceleration of communication between GPUs by reducing or eliminating transfers between device memory, host memory and network buffers. It also provides an extension for MPI libraries, to use GPU memory directly in MPI communications, thus hiding the retrieving of the buffer from the user. The third version of GPUDirect, which is the most recent, allows the complete bypass of copies to the host memory, by connecting the GPU directly to the network card. This technology is currently available in most supercomputer centers, including the CSCS cluster PIZ DAINT. Moreover, Cray’s MPI implementation allows the use of this copy-free method without changing a CUDA-aware MPI code.

Using this capability we have an almost perfect overlap of communication and computation, hiding data transfer involving the GPU. The GPU occupancy during the exact exchange step is over 80%, with cuFFT kernels of the Poisson solver accounting for around 65% of this GPU time. Using this optimized GPU to GPU MPI communication gives a speedup of three compared to the standard CPU based MPI communication used in the previous GPU implementation.

3. Tests and Case Studies

We tested our approach on two different materials, to consider two opposing scenarios: a standard benchmark system with few electrons that allows us to put our results into context by comparing with other publications, and a highly challenging system with many electrons. For the former, we used H_2O , with sizes ranging from 64 to 512 molecules. For the latter, we investigated uranium dioxide (UO_2), which plays a central role in nuclear fuel, and is both a technologically important and scientifically interesting material, with simulation challenges that go beyond the need for a hybrid functional, as discussed below.

Fig. 2 shows the structures used for UO_2 and H_2O .

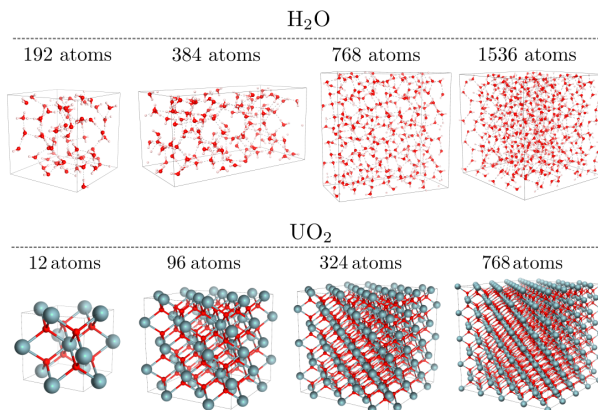


Figure 2. Systems used for benchmarking our implementation: water ranging from 64 to 512 molecules [top] and bulk UO_2 for system sizes of 12 atoms up to 768 atoms [bottom].

All DFT calculations were performed with the BIGDFT code [20, 29, 30], which uses a systematic wavelet basis set. As the basis functions used in our calculation do not depend on the atomic positions of the systems, no Pulay terms have to be computed for the forces on the atoms. In other terms the calculation with hybrid functionals does not generate additional terms for the evaluation of the atomic forces other than the conventional Hellmann-Feynman terms. The exchange-correlation functionals used were GGA-PBE [31] and hybrid-PBE0 [5]. The pseudopotentials used were of the HGH form in the Krack variant [32].

In order to compare performance across different architectures, including both CPU and CPU-GPU, we employed three different parallel supercomputers. The machines used were the XC-30 and XC-50 PIZ DAINT supercomputers hosted by the Swiss national supercomputing center (CSCS), Lugano, together with the MIRA Blue Gene/Q cluster at the Argonne Leadership Computing Facility (ALCF). For further details about the machines and calculation setups see Appendix A.

To give an idea of the computational workload of our calculations we have written in Table 1 the number of evaluations of the Poisson solver that the calculation of E_X and \hat{D}_X (for a *single* wavefunction iteration) requires for the systems considered in these benchmarks.

To ease the understanding of our results, we use in the following the factor γ , defined as the ratio of walltime between a ground-state, self-consistent evaluation between a hybrid PBE0 and a semilocal PBE ground state energy and forces evaluation. Since in our benchmarks the calculations are performed with the same number of wavefunction iterations such a factor is a reliable evaluation of the relative wall-time

increase for PBE0 with respect to a PBE calculation. The reported timings in the following were obtained from BIGDFT’s internal profiling routines, which links to the standard library with real-time functions `librt`. These timings are in agreement with those obtained from Cray’s Craypat tool and Nvidia profilers.

3.1. H₂O

Internal coordinates of H₂O molecules were relaxed for different supercells and with both PBE and PBE0 functionals. Calculations were Γ -centered single k -point, using a grid spacing of 0.35 Å to reach convergence in the ionic forces of less than 2 meV/Å. Pseudopotentials were constructed using the PBE approximation to the exchange and correlation with a single valence electron for hydrogen and 6 valence electrons for oxygen.

3.1.1. MD Energy Conservation The first check was to verify the energy conservation in molecular dynamics (MD) trajectory calculations. This benchmark was conducted on 64 H₂O molecules (192 atoms) with a starting temperature of 300 K and was evolved for more than 1000 steps on which wavefunction optimization was converged. These calculations represent a real production run with tight parameters and were conducted in Piz Daint (XC50). In Fig. 3 the top panel shows the kinetic energy and the bottom panel shows the normalized total energy per MD-step. The energy is conserved for a large number of steps for both PBE and hybrid PBE0. The absence of energy drifting in the normalized energy is a clear evidence that energy and forces are accurately determined.

While the use of PBE does not represent a challenge for any *ab-initio* molecular dynamics code, the use of hybrid PBE0 has until now been impossible for such large systems that require too many MD steps. This is now feasible thanks to the great reduction in time for optimized MD steps. The PBE and PBE0 timings obtained for H₂O as a function of the number of calculated molecules are summarized in Table 2. It is worth mentioning that the hybrid-PBE0 energies are accurately determined for a large number of atoms, while the absence of any adjustable parameters makes our approach suitable for tackling biological systems and nanostructures.

3.2. UO₂

It is well known that standard Kohn-Sham DFT at the LDA or GGA level fails to handle self-interactions present in transition metal oxides, Mott Hubbard insulators and rare-earth materials. This situation is particularly bad for systems with partially occupied d or f shells, such as UO₂, and may lead to incorrect

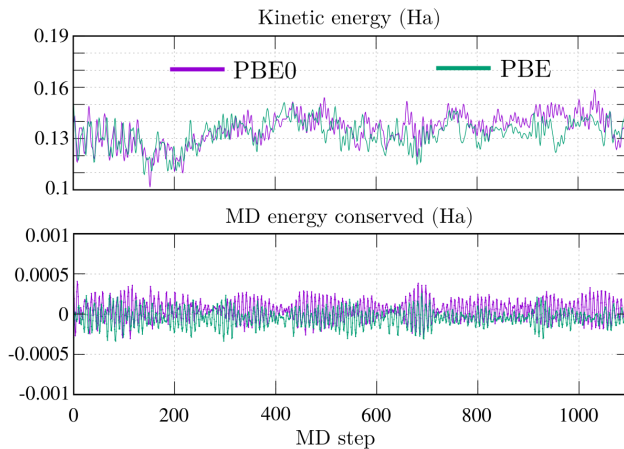


Figure 3. Kinetic energy [top] and normalised total energy [bottom] per MD step for PBE and PBE0 MD calculations of 64 H₂O molecules. Energy is conserved for a large number of steps for both PBE and hybrid PBE0.

(metallic) ground states for insulating systems. Indeed, the self interaction or delocalization error reduces the attractive potential, resulting in a rather extended orbital that increases its hybridization, therefore failing to capture the correct physics of the system. The problem is frequently circumvented by adding *ad-hoc* terms to the KS Hamiltonian, modelling the on-site electronic interactions, whose strengths are usually called U and J . However, these on-site interactions are adjustable parameters and it becomes more challenging to use this approach for predicting properties of materials. Hybrid functionals, on the other hand, which incorporate exact-exchange to the Kohn-Sham energy, also correct this self-interaction error without the need for additional, system-dependent parameters.

Because of the relative computational costs, the vast majority of studies on UO₂ have thus far focused on DFT+U, although there have been some hybrid calculations either for a full hybrid approach [33, 34, 35] or using the “exact exchange for correlated electrons” (EECE) approach [36, 37]. However these were restricted to small cells.

3.2.1. Metastable states and Occupancy Control

Aside from the need for a DFT+U or hybrid functional treatment for UO₂, another major complication arises, namely the existence of metastable states. These result from the many different combinations of orbitals that the two f -electrons of each uranium atom can occupy, which act as local minima. Both DFT+U and hybrid functional calculations easily become “trapped” in these local minima and therefore end up converging to different solutions for different initial guesses. These metastable states can have widely varying energies and band gaps, so that convergence to a metastable state

H ₂ O					UO ₂				
N	192	384	768	1536	12	96	324	768	1029
$\# \psi_i^\sigma$	256	512	1024	2048	164	1432	5400	12800	17150
$\# \rho_{ij}^\sigma$	32 896	131 328	524 800	2 098 176	6 658	513 372	7 292 700	81 926 400	73 539 200

Table 1. Number of Poisson solver evaluations per self-consistent iteration required to calculate the exact exchange energy and operator on the different systems used in the study. The number of atoms as well as the number of KS orbitals is indicated.

H ₂ O	Atoms	Nodes	PBE	PBE0	γ
64	192	32	11	24	2.18
128	384	64	23	74	3.2
256	768	128	49	244	5.6
512	1536	256	115	1030	8.95

Table 2. PBE and PBE0 timings (s) obtained for complete SCF calculations of H₂O systems as a function of the number of calculated molecules (first column) with Piz Daint (XC50).

can significantly affect the accuracy and reliability of the simulation.

A number of schemes for countering the problem of metastable states have been proposed, as summarized in Ref. [38]. One of the most popular is the occupation matrix control (OMC) scheme [39, 40, 41], which is used to explore different metastable states by imposing various occupancy matrices. The systematic approach of the OMC forms the basis of its appeal, however it can require calculations on a rather large number of potential occupancy combinations, especially if not all uranium atoms are considered to be identically occupied or if non-zero off-diagonal elements of the occupation matrix are also considered. When combined with the use of large cells which are required when e.g. a defect is inserted, this provides a further strong motivation for the need to have a hybrid functional implementation with a low overhead.

Previously, such an approach has been out of reach due to the excessive computational cost, however such calculations are possible using our new hybrid functional implementation. To this end, in a similar spirit to the occupancy control scheme of DFT+U, we have implemented an occupancy biasing scheme in BIGDFT, which we use to ensure convergence to a particular state.

3.2.2. Computational Details We used a cubic cell, taking the structure from the Materials Project [42], keeping the lattice constant fixed at 5.42 Å. In lieu of k -point sampling, we investigated increasing supercell sizes at the Γ -point only. We considered four different periodic cells covering a wide range of sizes: a small cell (12 atoms), medium sized cell (96 atoms) and large cells (324 and 768 atoms). Calculations employed a moderate grid spacing of 0.23 Å. The pseudopotential used for uranium had 14 electrons, while that for

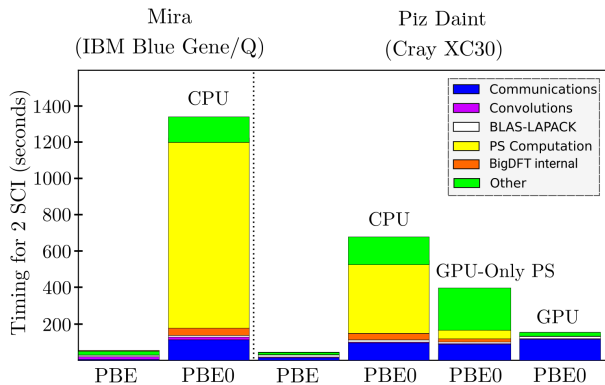


Figure 4. BIGDFT time profiling for 324 atoms of UO₂ (5,400 KS orbitals) on a Blue Gene/Q (MIRA) and a Cray XC30 (Piz Daint), both using 1,800 MPI processes. The calculations on Mira were performed using 4 MPI processes per node, with 16 OpenMP threads per MPI process. The solution of Poisson’s equation in the exact exchange part consumes most of the time on the CPU architecture, whereas it is drastically reduced on the GPU accelerated architecture.

oxygen had 6 electrons.

Restricting to the case of all atoms having the same f -orbital occupancy, we took the lowest energy configuration found by Krack [43] subject to this constraint, namely $[f_1 f_3]$. The occupancy was imposed for the first 12 self-consistent iterations (SCI), after which it was allowed to evolve freely. Spin polarized calculations with non identical spin up and down orbitals were set for this material. Following previous works, we used the 1- \mathbf{k} anti-ferromagnetic ordering for the magnetic structure with no spin-orbit coupling, since the effects of spin-orbit coupling have been shown to be rather small [35]. No symmetry was imposed on the density.

A convergence threshold of 10^{-4} was used for the wavefunction gradient, except where only a fixed number of iterations were used for benchmarking purposes, as described when relevant. However calculations were not considered fully converged until the energy difference between two successive diagonalizations was less than 0.1 meV/atom, with 12 SCI between each diagonalization.

3.2.3. Performance and Scaling In order to analyze the performance of our approach, we first compared

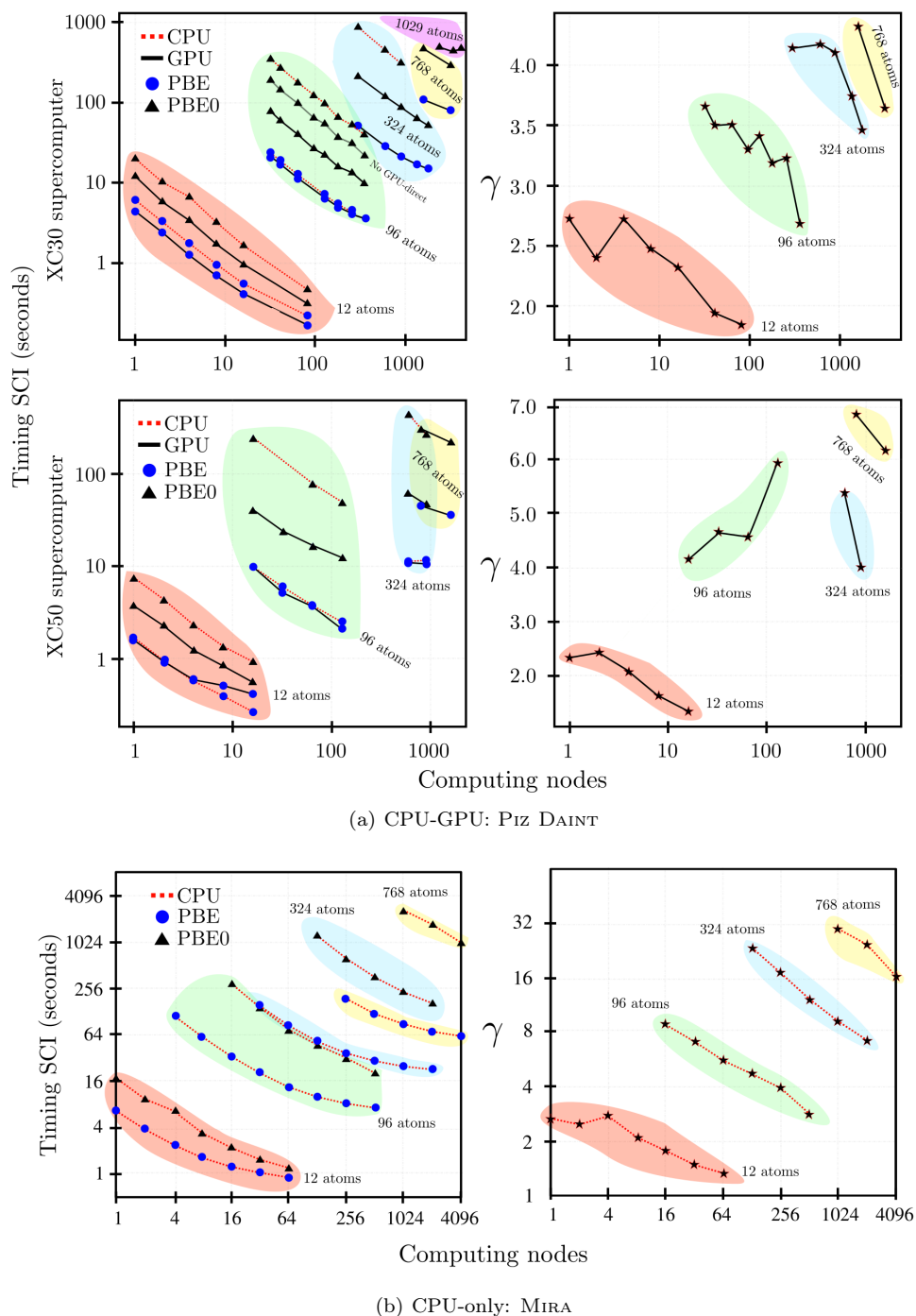


Figure 5. Timings for PBE and PBE0 with and without GPU acceleration per iteration for different cells of UO_2 as a function of computing nodes on XC30 [top] and XC50 [middle]. For comparison, CPU-only results are also shown for MIRA [bottom]. Right panels represent the ratio PBE0/PBE for GPU runs [top and middle] and CPU runs [bottom].

the breakdown in the timings for two SCIs on different architectures: CPU (MIRA) and CPU-GPU (PIZ DAINT XC30), as shown in Fig. 4. In this plot the y -axis represents wall-time in seconds and the color code categorizes the different computing operations necessary per iteration. This profiling was conducted

for a system of 324 atoms of UO_2 (5,400 KS orbitals). The Nvidia profiler `nvprof` was used to profile the GPU kernels, and compute the efficiency of the implementation. We took timings for all the sections of BIGDFT that are related to the wavefunction optimization, not considering the time needed to set

up the input wavefunctions and to evaluate the atomic forces as these operations do not involve calculations from our library. It is unsurprising that most of the time goes to the solution of Poisson’s equation (PS computation) particularly on CPU architectures (yellow).

On XC30 systems exploiting only the CPU brings this down to 15 times the cost of PBE. Whilst the usage of GPU already brings the PBE0 computation within the same order of magnitude of the PBE runs, further reduction of the time required for the computation of PBE0 can be achieved by allowing GPU-Direct communication, in this case the cost to fully compute PBE0 on a systematic basis set for 324 atoms of UO_2 is only three times more expensive than for PBE.

We performed calculations for the different system sizes over a range of numbers of compute nodes. Fig. 5 shows the time per iteration for these calculations, for both the PBE and PBE0 functionals. For PIZ DAINTE, for 12 and 96 atoms, timings were obtained for runs where the wavefunction was fully converged. For the larger runs, as discussed above, we considered only two complete self-consistent iterations and the time is normalized per wavefunction iteration. For MIRA five SCI were considered for each system size. Each SCI still consists of a full calculation of both the exact exchange energy and gradient. Table 3 shows the ratios of PBE0/PBE for CPU and GPU architectures. The timings demonstrate not only the high scalability of our approach, but also show that in our GPU implementation the hybrid PBE0 calculation is just 2 times (for about 200 orbitals) and up to 4 times (for more than ten thousand orbitals) more time consuming than the GGA PBE calculation. In the CPU version we still have a relatively low ratio, i.e. a factor of 15 on PIZ DAINTE.

When going to two or even one orbital per node the degradation of the scalability is due to the fact that computation time is not high enough to overlap the communications. For one of the large systems used in this study (768 atoms, i.e. 12,800 orbitals), the 3,200 node run (4 orbitals/node) was 75% quicker than the 1,600 nodes run. This amounts to a 110 TFlops/s performance for the exact exchange computation. Even larger sizes (1,029 atoms, resulting in 17,150 orbitals, on up to 4,288 nodes) were reached during this study, although network issues arose and prevented a good scalability. These limits, although only reached on sizes that are currently considered as uncommon for scientific production work, could be overcome by adding a wavelet based compression/uncompression step before each communication. This transformation is already implemented on the GPU in BIGDFT, and currently used only before entering the exact exchange step. On some systems this could reduce

communication sizes by a factor of four. The new version of PIZ DAINTE provides a huge bump in computing power on the GPU, while the network was only slightly improved. This means that this optimization is even more important.

During the exact exchange computation, when communication is properly overlapped, each of the GPUs spends 80% of the time computing, and reaches 40 GFlops of sustained double precision performance. Memory throughput on the GPU is on average more than 165 GBps, over 65% of the peak theoretical bandwidth of the GPU, which is the limiting factor here. The CPU performance for the same run without acceleration on PIZ DAINTE is 6.4 GFlops per CPU (using 8 OpenMP processes per CPU).

We have reported the performance results for the PBE0 hybrid functional. However quasi-identical timings for other hybrid functionals are expected to be of the same order of magnitude. We would also obtain identical timings for screened hybrid functionals such as HSE06, since in this case only the filter F of kernel 2 would be different. In all cases the calculation of the Hartree exchange on the real-space grid would be done without any approximation.

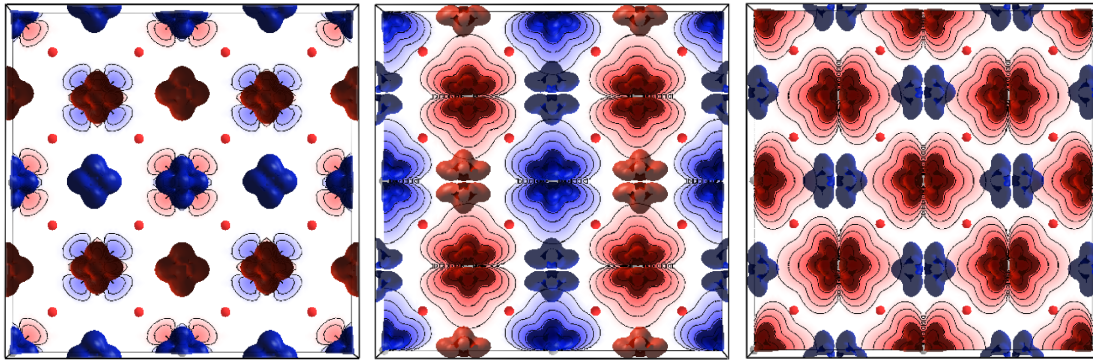
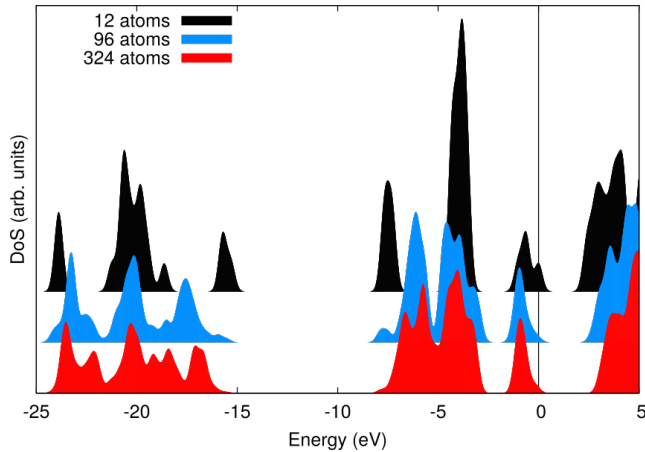
3.2.4. Electronic Structure To confirm that the occupancy biasing scheme is working as anticipated, we first printed the density matrix in the space of the f -orbital pseudopotential projectors at the end of the calculation, and verified that the correct states had non-zero values. As a further visual confirmation, we plotted the spin density difference, as shown in Fig. 6, for the 96 atom cell. Aside from the expected rotation, the spin densities closely resemble those shown by Krack [43] for the $[f_{-3}f_{-1}]$ state (which is degenerate with $[f_1f_3]$), confirming the ability of the occupancy biasing scheme to find the desired metastable state.

The density of states (DoS) is plotted for the different supercells in Fig. 7, while the corresponding band gaps are given in Table 4. Only the DoS for up electrons is plotted, as results for spin down are indistinguishable. Since no k -point sampling is used, it is unsurprising that the electronic structure and band gap are poorly converged for the smallest cell. The band gap is particularly affected by the supercell size, in line with previous observations where the level of k -point sampling was noted to affect the band gap accuracy [37]. Nonetheless, the differences in the DoS between the 96 atom and 324 atom supercells are much smaller. Comparing the results with DFT+U, the calculated band gap for the largest supercell is very close to that according to Krack for an effective U value ($U_{\text{eff}} = U - J$) of $U_{\text{eff}} = 3.96$, which is the value most commonly used in the literature.

These results serve to demonstrate the ability

Table 3. The ratio γ obtained on PIZ DAINTE for CPU and GPU architectures, and on MIRA for CPU only. The number of orbitals considered varies due to a different number of unoccupied states used in the calculation.

System (atoms)	PIZ DAINTE (XC30)					MIRA			
	Orbitals	Nodes	Orbitals/Node	γ CPU	γ GPU	Orbitals	Nodes	Orbitals/Process	γ CPU
12	164	16	10	3.31	2.06	104	64	1-2	1.33
96	1,432	128	11	11.69	3.09	832	512	1-2	2.83
324	5,400	600	9	14.4	4.1	2,808	2,048	1-2	7.23
768	12,800	1,600	8	-	4.33	6,656	4,096	1-2	16.34
PIZ DAINTE (XC50)									
12	164	16	10	3.47	1.33				
96	1,432	128	11	19.76	5.91				
324	5,400	600	9	40	5.41				
768	12,800	1,600	8	-	6.14				


Figure 6. Final spin density for an occupancy of $[f_1 f_3]$ in a $2 \times 2 \times 2$ supercell as viewed along the different axes. Oxygen (uranium) atoms are indicated in red (grey), while blue (red) density corresponds to spin up (down). The contour plots are shown for a single plane.

Figure 7. Density of states of UO_2 for spin up, for three different cell sizes, with an imposed occupancy of $[f_1 f_3]$. A Gaussian smearing of 0.2 eV has been applied and the curves have been shifted so that the highest occupied molecular orbital (HOMO) energy of each curve is at zero. The curves have also been normalized with respect to the number of atoms in the cell.

Num. atoms	PBE0	DFT+U	
		$U_{\text{eff}}=3.96$	$U_{\text{eff}}=2.00$
12	2.35	-	-
96	2.65	-	-
324	3.01	-	-
768	-	2.99	1.90

Table 4. Band gaps in eV for different supercell sizes of UO_2 for an imposed occupancy of $[f_1 f_3]$, calculated with PBE0. For comparison, DFT+U results are also shown for two different values of U_{eff} , using the values according to Krack [43].

of our approach to explore different f -electron occupancies in UO_2 using accurate hybrid functionals. Furthermore, the computational cost is low enough to allow calculations on cell sizes which are big enough to treat point defects while keeping spurious interactions between defects in neighbouring cells to a minimum. A detailed investigation of convergence with respect to basis size, the equilibrium lattice constant and exploration of the occupancy space using this approach will be published in the near future.

4. Conclusions

Thanks to a highly efficient GPU implementation of a novel wavelet based algorithm for the evaluation of the exact exchange, we have demonstrated a reduction in the cost of hybrid functional calculations in systematic basis sets by nearly one order of magnitude. As a consequence, hybrid functional calculations with our method are only about three times more expensive than a GGA functional calculation. This is a price that the community is ready to pay for the significantly improved accuracy offered by such functionals.

This methodology is available as a stand-alone library and can therefore be coupled to virtually any code that uses a systematic basis set and which calculates at some point the electronic orbitals on a Cartesian real space grid. Our implementation is completely general and available as an open-source package which can treat isolated and mixed boundary conditions as well as non-orthorhombic cells. As GPU acceleration is now also available on commodity computing clusters, our method will also allow researchers to do hybrid functional calculations for medium sized systems (hundreds of atoms) on affordable local computers. This advantageous price/performance aspect could also lead to an increased use of this technology in industry. Moreover, our developments will have implications both for future systems as well as for scientific applications. The availability of a code that gives a huge performance gain on accelerated architectures will further accelerate the spread of GPU accelerated systems. Finally, from a HPC viewpoint, the usage of such methods will enable extensive use of petaflop machines for electronic structure calculation communities, on the brink of the exascale era.

Acknowledgments

This work was done within the MARVEL and PASC programs. We acknowledge computational resources from the Swiss National Supercomputing Center (CSCS) in Lugano (project s707 and 752). Computer time was also provided by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

Appendix A. Supercomputer Details

Performance was measured for runs conducted on three different parallel supercomputers. The system used to test our GPU accelerated runs is the PIZ

DAINT supercomputer hosted by the Swiss national supercomputing center (CSCS), Lugano. Benchmarks were performed on both the previous version of the cluster, and the upgraded version, deployed in December 2016. The original system machine was a 5,272 node Cray XC-30 system, which had 8-core Intel SandyBridge CPUs (Intel Xeon E5-2670), with 32 GB of RAM and an Nvidia Tesla K20X GPU with 6 GB of RAM per node. It used an Aries proprietary interconnect from Cray with a dragonfly network topology, and achieved a peak performance of 6,271 Teraflops. The compiler used for these runs was Intel 15.0.1.133, with CUDA toolkit 7.0. The MPI Library was Cray’s MPICH 7.2.6. Each run was performed using 1 MPI process per node, and 8 OpenMP threads per MPI process. The GPU portion of the new iteration of PIZ DAIN is a 4,800 nodes Cray XC-50 system, with a 12-core (24 virtual cores with HyperThreading) Intel Haswell processor (Intel Xeon E5-2690 v3). Each node has 64 GB of RAM and is accelerated by an Nvidia Tesla P100 GPU with 16 GB of on board memory. This new system reached the peak performance of 9,779 Teraflops, due to the high increase in bandwidth at the GPU level. The network topology and interconnect is the same as on the former iteration of PIZ DAIN. The compiler for these new runs was Intel 17.0.1, with CUDA toolkit 8.0. The MPI Library was Cray’s MPICH 7.5.0.

The MIRA Blue Gene/Q cluster at the Argonne Leadership Computing Facility (ALCF) features 49,152 nodes and has a peak performance of 10 Petaflops. Each node consists of a 16 core PowerPC A2 1600 MHz processor, and 16GB of DDR3 memory. The network is an IBM 5D Torus Proprietary Network. The compiler used was IBM’s xlc, version 14.1, with the provided MPI library, based on MPICH itself. Except where stated otherwise, runs were performed with 1 MPI process and 32 OpenMP threads per node, in order to maximize the range of node counts over which scaling tests could be performed.

Appendix B. Bibliographic Search of DFT Calculations

The bibliographic search was conducted using the GOOGLE SCHOLAR engine in 2017. The procedure was as follows: the main papers suggested for citation in the GAUSSIAN [1] 2009 and NWCHEM [2] 2010 papers were used. Key words matching either in the abstract, title or block of the paper were accounted for. These results are shown in Table B1. About 80 to 90 percent of all density-functional calculations conducted with those two codes employed hybrid functionals. Not surprisingly these calculations use the best available functionals. The most widely used basis set in such

hybrid functional calculations is the **6-31G** basis set which gives for instance atomization energy errors of about 30 kcal/mol. The more accurate AUG-cc-pV5Z basis set, which give in most cases chemical accuracy, is used in less than 2 percent of the hybrid functional calculations. Clearly, while one problem is alleviated by using hybrid functionals, another problem is raised by employing poor basis sets.

References

- [1] Frisch J, Trucks G W, Schlegel H B, Scuseria G E, Robb M A, Cheeseman J R, Scalmani G, Barone V, Mennucci B, Petersson G A, Nakatsuji H, Caricato M, Li X, Hratchian H P, Izmaylov A F, Bloino J, Zheng G, Sonnenberg J L, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery Jr J A, Peralta J E, Ogliaro F, Bearpark M, Heyd J J, Brothers E, Kudin K N, Staroverov V N, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant J C, Iyengar S S, Tomasi J, Cossi M, Rega N, Millam J M, Klene M, Knox J E, Cross J B, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann R E, Yazyev O, Austin A J, Cammi R, Pomelli C, Ochterski J W, Martin R L, Morokuma K, Zakrzewski V G, Voth G A, Salvador P, Dannenberg J J, Dapprich S, Daniels A D, Farkas Ö, Foresman J B, Ortiz J V, Cioslowski J and Fox D J 2009 Gaussian 09, revision a.02 URL <http://gaussian.com/g09citation/>
- [2] Valiev M, Bylaska E, Govind N, Kowalski K, Straatsma T, Dam H V, Wang D, Nieplocha J, Apra E, Windus T and de Jong W 2010 *Comput. Phys. Commun.* **181** 1477–1489
- [3] Becke A D 1993 *J. Chem. Phys.* **98** 5648–5652
- [4] Lee C, Yang W and Parr R G 1988 *Phys. Rev. B* **37** 785
- [5] Adamo C and Barone V 1999 *J. Chem. Phys.* **110** 6158–6170
- [6] Heyd J, Scuseria G E and Ernzerhof M 2006 *J. Chem. Phys.* **124** 219906
- [7] Willand A, Kvashnin Y O, Genovese L, Vázquez-Mayagoitia Á, Deb A K, Sadeghi A, Deutsch T and Goedecker S 2013 *J. Chem. Phys.* **138** 104109
- [8] Jensen S R, Saha S, Flores-Livas J A, Huhn W, Blum V, Goedecker S and Frediani L 2017 *J. Phys. Chem. Lett.* **8** 1449–1457
- [9] Singh D J and Nordstrom L 2006 "Planewaves, Pseudopotentials, and the LAPW method" (Springer)
- [10] Rangel T, Caliste D, Genovese L and Torrent M 2016 *Comput. Phys. Commun.* **208** 1 – 8 ISSN 0010-4655
- [11] Lejaeghere K, Bihlmayer G, Björkman T, Blaha P, Blügel S, Blum V, Caliste D, Castelli I E, Clark S J, Dal Corso A, de Gironcoli S, Deutsch T, Dewhurst J K, Di Marco I, Draxl C, Dulak M, Eriksson O, Flores-Livas J A, Garrity K F, Genovese L, Giannozzi P, Giantomassi M, Goedecker S, Gonze X, Grånäs O, Gross E K U, Gulans A, Gygi F, Hamann D R, Hasnip P J, Holzwarth N A W, Iuşan D, Jochym D B, Jollet F, Jones D, Kresse G, Koepnick K, Küçükbenli E, Kvashnin Y O, Locht I L M, Lubeck S, Marsman M, Marzari N, Nitzsche U, Nordström L, Ozaki T, Paulatto L, Pickard C J, Poelmans W, Probert M I J, Refson K, Richter M, Rignanese G M, Saha S, Scheffler M, Schlipf M, Schwarz K, Sharma S, Tavazza F, Thunström P, Tkatchenko A, Torrent M, Vanderbilt D, van Setten M J, Van Speybroeck V, Wills J M, Yates J R, Zhang G X and Cottenier S 2016 *Science* **351** ISSN 0036-8075
- [12] Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, Ceresoli D, Chiarotti G L, Cococcioni M, Dabo I, Corso A D, de Gironcoli S, Fabris S, Fratesi G, Gebauer R, Gerstmann U, Gougoussis C, Kokalj A, Lazzeri M, Martin-Samos L, Marzari N, Mauri F, Mazzarello R, Paolini S, Pasquarello A, Paulatto L, Sbraccia C, Scandolo S, Sclauzero G, Seitsonen A P, Smogunov A, Umari P and Wentzcovitch R M 2009 *J. Phys.: Condens. Matter* **21** 395502
- [13] Kresse G and Furthmüller J 1996 *Phys. Rev. B* **54** 169
- [14] Gonze X, Amadon B, Anglade P, Beuken J, Bottin F, Boulanger P, Bruneval F, Caliste D, Caracas R, Côté M, Deutsch T, Genovese L, Ghosez P, Giantomassi M, Goedecker S, Hamann D, Hermet P, Jollet F, Jomard G, Leroux S, Mancini M, Mazevet S, Oliveira M, Onida G, Pouillon Y, Rangel T, Rignanese G, Sangalli D, Shaltaf R, Torrent M, Verstraete M, Zerah G and Zwanziger J 2009 *Comput. Phys. Commun.* **180** 2582–2615
- [15] Gygi F 2009 *Phys. Rev. Lett.* **102** 166406
- [16] Gygi F and Duchemin I 2012 *J. Chem. Theory Comput.* **9** 582–587
- [17] Lin L 2016 *J. Chem. Theory Comput.* **12** 2242–2249
- [18] Ratcliff L E, Mohr S, Huhs G, Deutsch T, Masella M and Genovese L 2017 *Wiley Interdisciplinary Reviews: Computational Molecular Science* **7** e1290–n/a ISSN 1759-0884 e1290
- [19] Damle A, Lin L and Ying L 2015 *J. Chem. Theory Comput.* **11** 1463–1469
- [20] Genovese L, Deutsch T, Neelov A, Goedecker S and Beylkin G 2006 *J. Chem. Phys.* **125** 074105
- [21] Cerioni A, Genovese L, Mirone A and Sole V A 2012 *J. Chem. Phys.* **137** 134108
- [22] Dugan N, Genovese L and Goedecker S 2013 *Comput. Phys. Commun.* **184** 1815–1820
- [23] García-Risueño P, Alberdi-Rodríguez J, Oliveira M J, Andrade X, Pippig M, Muguerza J, Arruabarrena A and Rubio A 2014 *J. Comp. Chem.* **35** 427–444
- [24] Hutter J, Iannuzzi M, Schiffmann F and VandeVondele J 2014 *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4** 15–25
- [25] Andrade X, Alberdi-Rodríguez J, Strubbe D A, Oliveira M J T, Nogueira F, Castro A, Muguerza J, Arruabarrena A, Louie S G, Aspuru-Guzik A, Rubio A and Marques M A L 2012 *J. Phys.: Condens. Matter* **24** 233202
- [26] Gillan M J, Bowler D R, Torralba A S and Miyazaki T 2007 *Comput. Phys. Commun.* **177** 14–18
- [27] Fiscaro G, Genovese L, Andreussi O, Marzari N and Goedecker S 2016 *J. Chem. Phys.* **144** 014103
- [28] Fiscaro G, Genovese L, Andreussi O, Mandal S, Nair N N, Marzari N and Goedecker S 2017 *J. Chem. Theory Comput.* **13** 3829–3845 PMID: 28628316
- [29] Genovese L, Neelov A, Goedecker S, Deutsch T, Ghasemi S A, Willand A, Caliste D, Zilberberg O, Rayson M, Bergman A and Schneider R 2008 *J. Chem. Phys.* **129** 014109
- [30] Genovese L, Videau B, Ospici M, Deutsch T, Goedecker S and Méhaut J F 2011 *C. R. Mécanique* **339** 149–164
- [31] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77**(18) 3865–3868
- [32] Krack M 2005 *Theor. Chem. Acc.* **114** 145–152
- [33] Kudin K N, Scuseria G E and Martin R L 2002 *Phys. Rev. Lett.* **89**(26) 266402
- [34] Prodan I D, Scuseria G E and Martin R L 2006 *Phys. Rev. B* **73**(4) 045104
- [35] Roy L E, Durakiewicz T, Martin R L, Peralta J E, Scuseria G E, Olson C G, Joyce J J and Gziewicz E 2008 *J. Comp. Chem.* **29** 2288–2294 ISSN 1096-987X
- [36] Novák P, Kuneš J, Chaput L and Pickett W E 2006 *Phys. Status Solidi B* **243** 563–572 ISSN 1521-3951
- [37] Jollet F, Jomard G, Amadon B, Crocombette J P and Torumba D 2009 *Phys. Rev. B* **80**(23) 235109

Table B1. Approximate number of entries counted with the GOOGLE SCHOLAR engine for two widely used DFT codes in the chemistry community.

GAUSSIAN 09 Rev. A.02 (2009)		NWCHEM (paper 2010)	
Total calls: 22,835		Total calls: 1,455	
Keywords	Citations	Keywords	Citations
"B3LYP" "6-31G"	11,000	"B3LYP"	882
"B3LYP" "6-311G"	3,840	"Hybrid functional"	274
"B3LYP" "AUG-cc-pVDZ"	1,950	"PBE0"	270
"B3LYP" "AUG-cc-pV5Z"	269	"HSE06"	19
"PBE" "6-31G"	1,970		
"PBE" "6-311G"	855		
"PBE" "AUG-cc-pVDZ"	552		
"PBE" "AUG-cc-pV5Z"	80		
"PBE0" "6-31G"	1,540		
"PBE0" "6-311G":	607		
"PBE0" "AUG-cc-pvdz"	425		
"PBE0" "AUG-cc-pV5Z"	79		

- [38] Dorado B, Freyss M, Amadon B, Bertolus M, Jomard G and Garcia P 2013 *J. Phys.: Condens. Matter* **25** 333201
- [39] Amadon B, Jollet F and Torrent M 2008 *Phys. Rev. B* **77**(15) 155104
- [40] Jomard G, Amadon B, Bottin F and Torrent M 2008 *Phys. Rev. B* **78**(7) 075125
- [41] Dorado B, Amadon B, Freyss M and Bertolus M 2009 *Phys. Rev. B* **79**(23) 235125
- [42] Jain A, Ong S P, Hautier G, Chen W, Richards W D, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G and Persson K a 2013 *APL Materials* **1** 011002 ISSN 2166532X
- [43] Krack M 2015 *Phys. Scripta* **90** 094014