Fast hybrid density-functional computations using plane-wave basis sets

note finali coverpage

(Article begins on next page)

12 May 2024

Fast hybrid density-functional computations using plane-wave basis sets

Ivan Carnimeo[a,b], Stefano Baroni[a,c], Paolo Giannozzi[b,c]

(a) SISSA – Scuola Internazionale di Studi Avanzati, via Bonomea 265, 34136 – Trieste, Italy
(b) Dipartimento di Scienze Matematiche, Informatiche e Fisiche, Università degli Studi di Udine, via delle Scienze 208, 33100 – Udine, Italy
(c) CNR-IOM DEMOCRITOS, SISSA, Trieste, Italy

Abstract. A new, very fast, implementation of the exact (Fock) exchange operator for electronic-structure calculations within the plane-wave pseudopotential method is described and carefully validated. Our method combines the recently proposed Adaptively Compressed Exchange approach, to reduce the number of times the exchange is evaluated in the self-consistent loop, with an orbital localization procedure that reduces the number of exchange integrals to be computed at each evaluation. The new implementation, already available in the QUANTUM ESPRESSO distribution, results in a speedup that is never smaller than 3-4× and that increases with the size of the system, according to various realistic benchmark calculations.

1. Introduction

Hybrid functionals, resulting from the introduction of a Fock exchange energy term into a density-functional framework, are very popular due to their accuracy and predictive capabilities.[1, 2, 3] Nowadays most if not all electronic-structure codes implement them. In quantum chemistry codes, and in general in codes using localized basis sets, exact-exchange terms do not significantly add to the computational complexity of electronic states calculations. When working with a plane-wave (PW) basis set, however, the calculation of the exact-exchange term, although straightforward,[4, 5, 6, 7, 8] is computationally heavy. The problem lies in the delocalized nature of orbitals and of PWs: all pairs of canonical orbitals in the system contribute to exact exchange. While each contribution is quickly and effectively computed using conventional Fast Fourier Transform (FFT) techniques, the overall computational workload scales unfavorably with the number of electrons in the system, being proportional to the square of the number of occupied electron states. In practice, the usage of modern and accurate hybrid functionals together with PW basis sets is limited to relatively small systems, described by units cells no larger than a few dozen atoms.

Over the years, several approaches have been proposed to extend the range of hybrid-functional calculations with PW-based methods. We mention, *e.g.*, the reduction of the dimension of the density matrix [9], or of the basis set, [10] and improvements in the parallelization strategies[11, 12]. Another promising approach is to leverage a localized representation of the occupied-state manifold, so as to reduce the number of significant exchange integrals to be computed. Several proposals[13, 14, 15, 16, 17, 18, 19] have been put forward, mostly using maximally localized Wannier functions.[20, 21] This approach is very effective and has proven to be reliable for large systems, showing that strategies based on molecular orbital localization are a robust route in order to cope with scaling problems in exact exchange calculations. Furthermore, another effective method named *Adaptively Compressed Exchange* (ACE) has been recently proposed by Lin Lin,[22] and integrated in QUANTUM ESPRESSO[23], achieving a speed-up factor of 4-5x by the integration of a projected exchange operator in the double-loop structure of the Self Consistent Field (SCF) algorithm. Notably such approach is very similar to other projection procedures previously proposed for PW calculations[24] and also to inner-projection methods,[25, 26] popular in Quantum Chemistry.[27, 28, 29, 30] In this work we combine for the first time an algorithm based on the molecular orbital localization with the ACE method. We show how the ACE approach can be extended to deal with approximate exchange potentials coming from the localization, and we show accurate benchmark calculations showing the computational performances on realistic systems. The new method, L-ACE, allows significant computational gains with respect to the previous implementation of exact exchange with PWs. As localization algorithm we use a modified version of the *Selected Columns of the Density Matrix* (SCDM) approach,[31, 32, 33] an algebraic technique recently introduced by Damle *et al.*.

The paper is organized as follows. In Sec. 2 we describe the L-ACE method. Sec. 3

briefly introduces computational details used in the extensive benchmarks of Sec. 4. Sec. 5 contains our conclusions.

2. The L-ACE approach

In the framework of density-functional theory (DFT), the energy for a hybrid functional[34] in a system with N_P Kohn-Sham (KS) molecular orbitals (or pseudo-orbitals if pseudo-potentials are used), $\{\psi_k(\mathbf{r})\}_{k=1}^{N_P}$, can be written in terms of the charge density, $n(\mathbf{r}) = \sum_i^N |\psi_i(\mathbf{r})|^2$, and of the density matrix, $\gamma(\mathbf{r}, \mathbf{r}') = \sum_i^N \psi_i(\mathbf{r})\psi_i^*(\mathbf{r}')$, where N is the number of occupied orbitals. In the following the index i will be used to label the occupied orbitals ($i = 1, .., N$), the indexes j and k will be used for the KS orbitals ($j, k = 1, ..., N_P$), also including virtual orbitals if $N_P > N$. The spin part, omitted for simplicity of notations, should be assumed wherever relevant. In order to determine the molecular orbitals, a set of single-particle KS equations

$$\left(-\frac{1}{2}\nabla^2 + \hat{V}_{ext} + \hat{V}_{Hxc}[n, \alpha] - \alpha\hat{V}_X[\gamma] \right)\psi_k(\mathbf{r}) = \varepsilon_k\psi_k(\mathbf{r}) \tag{1}$$

must be solved. The first term of the last equation is the kinetic energy operator, the second term is the external potential, $\hat{V}_{Hxc}$ includes the Hartree and the DFT exchange-correlation contributions, and $\hat{V}_X$ is the exact (Fock) exchange operator (with a minus sign with respect to the usual definition, to make it positive definite). The empirical parameter α is used to weight the Fock and DFT exchange contributions, and takes different values according to the particular functional parametrization (e.g. see Refs. [35, 36, 3]). The Fock exchange operator can be defined through its action on a generic function ψ_k

$$\hat{V}_X\psi_k(\mathbf{r}) = e^2 \sum_{i=1}^{N} \psi_i(\mathbf{r}) \int d\mathbf{r}' \frac{\psi_i^*(\mathbf{r}')\psi_k(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \tag{2}$$

e being the electronic charge. When a PW basis set is used, Eq. 2 is usually represented in reciprocal space as a matrix $\mathbf{V}_X$ of dimension $N_P \times N_{PW}$, where N_{PW} is the number of plane waves, and the integrals are numerically solved using FFT algorithms. The latter is the bottleneck of the calculation especially when large molecular systems are involved, because many FFTs must be performed over large grids. A significant speed-up can be achieved by employing the ACE method.[22] In the ACE method $\mathbf{V}_X$ is used to build the ACE operator, which can be written in matrix form as

$$\mathbf{W}_X = \mathbf{V}_X \cdot \mathbf{M}^{-1} \cdot \mathbf{V}_X^{\dagger} \tag{3}$$

where $\mathbf{M}_{jk} = \langle\psi_j|\hat{V}_X|\psi_k\rangle$ is the exchange matrix. In practice the $\mathbf{M}^{-1}$ matrix is factorized via the Cholesky decomposition and the ACE operator takes the form

$$\mathbf{W}_X = \boldsymbol{\xi} \cdot \boldsymbol{\xi}^{\dagger} \tag{4}$$

where

$$\boldsymbol{\xi} = \mathbf{V}_X \cdot \mathbf{L}^{-T} \tag{5}$$

and $\mathbf{L}^{-T}$ is the Cholesky factor of $\mathbf{M}^{-1}$. The application of the ACE operator to a generic function costs as little as N_P scalar products and it is much cheaper than the application of $\hat{V}_X$. When a double-loop SCF algorithm is employed, as in the `pw.x` code of QUANTUM ESPRESSO,[8, 23] it is much more convenient to compute $\mathbf{V}_X$ only to build the ACE operator, using the latter to solve the KS equations iteratively, instead of computing $\mathbf{V}_X$ at every SCF iteration. Once convergence is achieved, $\mathbf{V}_X$ is evaluated again in order to update the ACE operator, the KS equations are solved with the new operator, and this process is repeated until the total energy reaches the minimum. As a result, the exact exchange potential needs not to be evaluated inside every SCF iteration, but only once per SCF loop, resulting in a significant speed-up. It is important to notice that the ACE operator is perfectly equivalent to $\hat{V}_X$ only at convergence or whenever it is applied to the same set of functions used to build it. In all other cases, it is an approximation, whose accuracy improves as more functions are used for the projection, and it can been shown that its expectation values are greater than the expectation values of $\hat{V}_X$.[37, 25, 26] The ACE method has been implemented in the QUANTUM ESPRESSO code[23] with minor modifications, so that we refer to the original work[22] for a detailed description of the method.

In this work we have achieved a further speed-up in the exact exchange evaluation, by projecting the ACE operator on a set of localized orbitals, $\{w_i\}_{i=1}^N$ and neglecting the integrals of Eq. 2 involving two orbitals localized in different regions of space. Among the many known localization schemes,[13, 14, 15, 20, 21, 38, 39, 31, 32, 33] we have chosen the *Selected Columns of the Density Matrix* (SCDM) approach,[31, 32] which has the advantage of being fast, non-iterative, and to be estensible to periodic systems using k-points[33]. The method proposed in the following is however completely general and can be applied independently of the particular choice of the localization algorithm.

In order to decide whether a given exchange integral can be discarded or must be retained, a reliable criterion needs to be found. In the limit case of orbitals shaped like Gaussian functions, the exchange integrals depend linearly upon the overlap of the two functions, and can be thus neglected when the overlap is small enough. Extending the idea to more general functions which can take positive and negative values, a safe criterion is that whenever the overlap of the moduli

$$S_{ik} = \int d\mathbf{r} \quad |w_i(\mathbf{r})| \cdot |w_k(\mathbf{r})| \tag{6}$$

is smaller than a predefined threshold (S_{thr}), i.e. $S_{ik} < S_{thr}$, the corresponding exchange integral can be set to zero.

The (approximate) matrix of truncated exchange potential, $\mathbf{V}'_X$, can be formally written as

$$\mathbf{V}'_X = \mathbf{V}_X - \sum_{\substack{\alpha\beta \\ S_{\alpha\beta} < S_{thr}}} \mathbf{u}^{(\alpha\beta)} \tag{7}$$

where α and β run over the discarded pairs of orbitals, and $\mathbf{u}^{(\alpha\beta)}$ have all columns set

to zero except two:

$$\mathbf{u}_\alpha^{(\alpha\beta)}(\mathbf{r}) = e^2 w_\beta(\mathbf{r}) \int d\mathbf{r}' \frac{w_\beta^*(\mathbf{r}')w_\alpha(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{8}$$

and $\mathbf{u}_\beta^{(\beta\alpha)}(\mathbf{r})$, with switched α and β indexes. The truncated exchange matrix $\mathbf{M}'$ can then be written as

$$\mathbf{M}' = \mathbf{M} - \sum_{\substack{\alpha\beta \\ S_{\alpha\beta} < S_{thr}}} \mathbf{m}^{(\alpha\beta)}, \tag{9}$$

where $\mathbf{m}^{(\alpha\beta)}$ contains only column vector corrections. The matrix $\mathbf{M}'$ is still invertible and the matrix inversion lemma (Woodbury identity)[40, 41] allows to write $\mathbf{M}'^{-1} = \mathbf{M}^{-1} + \Delta\mathbf{M}$. Furthermore, recalling that the determinant is a linear function of the columns (rows) and that the α and β columns of $\mathbf{m}^{(\alpha\beta)}$ are small by construction, we can safely assume that $\det(\mathbf{M}') \simeq \det(\mathbf{M})$, where $\det(\mathbf{M}) > 0$ is the product of positive definite eigenvalues. This suggests that the truncated exchange matrix $\mathbf{M}'$ remains positive definite and invertible, provided that the $\mathbf{m}^{(\alpha\beta)}$ matrices are small compared with $\mathbf{M}$. The approximate ACE operator then reads

$$\mathbf{W}'_X = \mathbf{V}'_X \cdot \mathbf{M}'^{-1} \cdot \mathbf{V}'^\dagger_X \simeq \mathbf{W}_X + \Delta\mathbf{W}. \tag{10}$$

The last equality just formalizes the fact that the approximate ACE operator can be written as the sum of the exact ACE operator plus a truncation error. The ACE method exploiting the localization algorithm will be referred to as the Localized ACE (L-ACE) method in the following. Unlike $\mathbf{M}$, the $\mathbf{M}'$ matrix is not symmetric and it has to be symmetrized in order to perform the Cholesky decomposition. We have tested three different symmetrizations: copying the upper (lower) triangle into the lower (upper) triangle, and replacing the matrix with $(\mathbf{M}' + \mathbf{M}'^T)/2$. In all cases, also after the symmetrization the matrix can be decomposed as in Eq. 9, into a sum of column-wise (or row-wise) corrections and the method is still valid.

One drawback of the L-ACE method is that the truncation introduces a small error in the exchange energy, resulting in an oscillatory behaviour of the total energy during the self-consistency cycle which may plague the convergence, as observed in other linear-scaling approaches based on truncated potentials.[17, 16] However, in our case the truncation is done only on the exact exchange energy, which is a small fraction of the total energy, so that such fluctuations are very small. Furthermore, such fluctuations depend on the threshold S_{thr}, which can be tuned to be large enough to allow significant computational savings but still small enough to allow smooth SCF convergence. For the molecular systems reported in Section 4 we have verified that values of the threshold in the interval $0 < S_{thr} \leq 0.005$ allow to achieve a significant speed-up and yet converge smoothly the SCF to an estimated accuracy of 10^{-6} Ry. When very tight convergences are required, a good strategy to improve SCF convergence is to align the orbitals at a given iteration to the ones at the previous step, thus ensuring convergence to arbitrary precision.

3. Computational methods

All calculations have been performed using a development version ‡ of the QUANTUM ESPRESSO `pw.x` code.[8, 23] Pure DFT calculations have been performed using the PBE[42] functional, while PBE0,[36] B3LYP,[35] and HSE[3] functionals have been used for hybrid calculations. Three FFT grids are defined for hybrid functional calculations, one for KS orbitals, one for the charge density and one for the exchange potential. Such grids are uniquely identified by the cell parameters and the cutoff values, which will be indicated in the following for each test system. Norm-conserving pseudopotentials have been used throughout the work. We will refer to the old method for the exact exchange, not using either ACE or localization, as the "Full" method; to the ACE algorithm without localization as "ACE"; to ACE plus localization as "L-ACE(S_{thr})", with the value of S_{thr} always specified (e.g., L-ACE(0.004) means that we are neglecting all exchange integrals in the potential with $S_{ik} < 0.004$).

The calculations have been run on the Galileo supercomputer hosted at CINECA. In the current configuration, the nodes of the system are based on a IBM NeXtScale architecture equipped with 2 x 18-cores Intel Xeon E5-2697 v4 (Broadwell) running at 2.30 GHz nominal clock speed and 128 GB of RAM memory. The nodes are connected through a QDR InfiniBand network with 40 Gbps links.

4. Results and discussion

We used three different typologies of molecular systems (shown in Figure 1) to validate the truncation scheme of the L-ACE method, ranging from amorphous materials (Silica), to organic molecules (Anthocyanine in water) and nanoparticles (TiO$_2$ cluster).
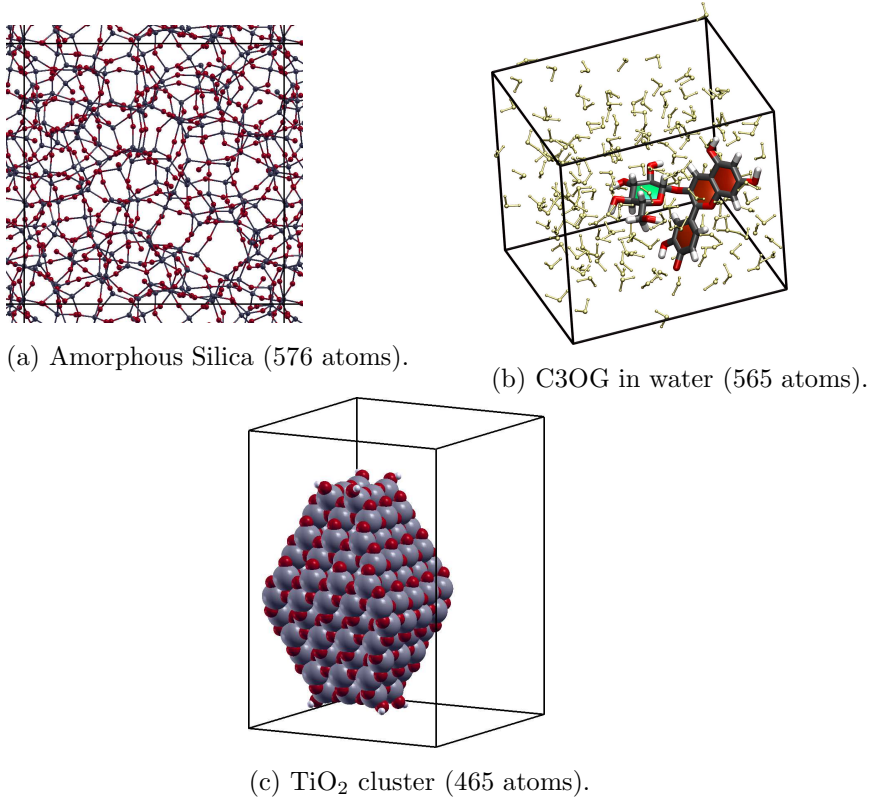
4.1. Amorphous Silica

Five different amorphous Silica (SiO$_2$) geometries, containing 72, 144, 288, 432 and 576 atoms, have been generated by classical MD simulations using the BKS force field,[43, 44] quenching from the melt at 6500K down to 500K at a constant quenching rate, as described in Ref.[45] The geometries are in cubic cells of dimension 18.6826, 24.1996, 30.3976, 35.1143 and 38.6458 Bohr, respectively. In Figure 1a the largest system of 576 atoms is shown. We have used such geometries for hybrid SCF calculations, using the PBE0 functional, and cutoffs of 80 Ry for the orbital grid and 320 Ry for the density and the exchange grid. For pure DFT calculations the PBE functional has been used.

First of all we have numerically checked the consistency of the L-ACE approximations. In practice the L-ACE method is based on the assumption that whenever the absolute overlap S_{ik} is small, all the exchange integrals§ of the type $(jkki)$

---

‡ to be released by the time this article is published; work is currently under way to port the same methodology to the GPU version of QUANTUM ESPRESSO.
§ In this notation[46] for the electrostatic integrals, the first (last) two indexes refer to functions integrated over $\mathbf{r}$ ($\mathbf{r}'$)

(a) Amorphous Silica (576 atoms).

(b) C3OG in water (565 atoms).

(c) TiO$_2$ cluster (465 atoms).

**Figure 1.**   Unit cells of three representative systems.

and $(jiik)$ are also small for any j. In Figure 2, the values of $(jkki)$ and $(jiik)$ integrals are plotted versus the corresponding values of S_{ik}, for the 72-atoms system. We observe that the largest integrals – those of the type $(kkki)$ and $(iiik)$ – show a linear dependence upon S_{ik} in the range between 0 and 0.005, confirming that the absolute overlap is a good quantity for the prescreening. Following Figure 2, values of S_{thr} in the range between 0 and 0.005 have been used for all the calculations in this work.
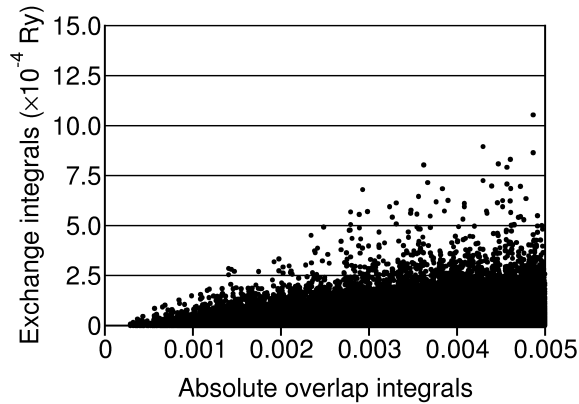
Figure 2.   Exchange integrals of type $(jiik)$ and $(jkki)$ plotted versus the corresponding absolute overlap integrals S_{ik}, for any j.

In Figure 3 the WALL times for one SCF with standard convergence (10^{-6} Ry) are plotted as a function of the number of atoms of the Silica cells in a logarithmic scale, in order to visualize the scaling with the system size. The same numbers are shown in the last column of Table 1. When the Full algorithm is used, we could perform the SCF only for the three smallest systems, and the SCF for the system of 288 atom took nearly three days. The use of the ACE method improves the performance, and we could perform calculations also for the system composed by 432 atoms, in a reasonable time. With the L-ACE approach, the WALL times are strongly reduced especially for the largest systems, and the calculation of the system composed by 576 atoms becomes feasible. Noteworthly, in all cases PBE0 calculations are still much slower than pure PBE ones. The lines in Figure 3 are fits of the computational times to a function of the type $y = \alpha x^\beta$, where y is the SCF time and x the number of atoms composing the system, and the values of β are reported in the Figure. We observe that although the Full method is much more expensive than ACE, they scale pretty much in the same way with increasing system size: $\beta \simeq 2.8$ in both cases. For comparison, L-ACE scales with an exponent of about 2, closer to the pure DFT value of about 1.7. The behavior of the L-ACE(0.002) and L-ACE(0.004) cases can be explained by observing (Table 1) that the percentage of integrals included in the calculation is 91% and 71% respectively for the 72-atom system, 12% and 8% respectively for the largest system.
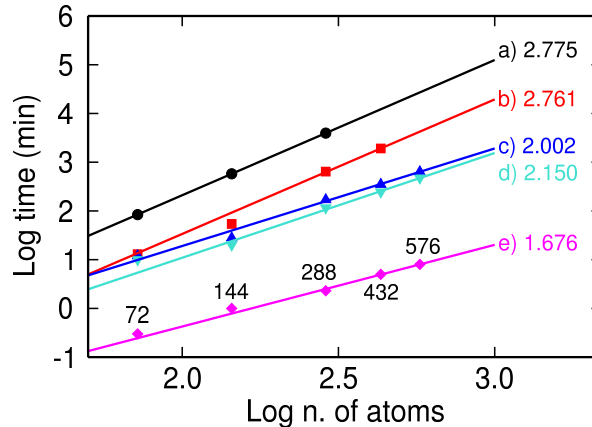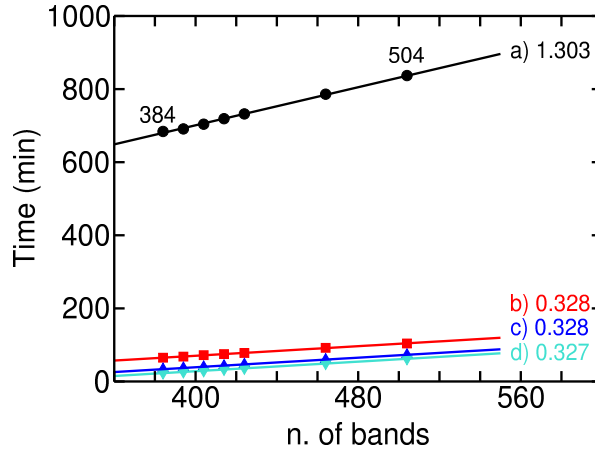

Figure 3. Wall time for a complete SCF (logarithmic scale) in Silica versus the logarithm of the number of atoms. a) Full, b) ACE, c) L-ACE(0.002), d) L-ACE(0.004), e) DFT. DFT calculations are with the PBE functional, all the others are with the PBE0 functional. The lines are the plot in logarithmic scale of the fitting function $y = \alpha x^\beta$ and β is reported for each method. The number of atoms is shown for clarity near the magenta (e) line.

It is then reasonable that the computational times of the two methods become more similar for increasing system size.

In Figure 4 the scaling with the number of processors is shown for the 144-atom system. In order to account for the unavoidable fluctuations in the computational load of the cluster, and for the different memory allocations when different number of nodes are chosen, 32 runs (on 1, 2, 3, 4, 5, 6, 8 and 10 nodes, using 15, 20, 25 and 30 cores per node)
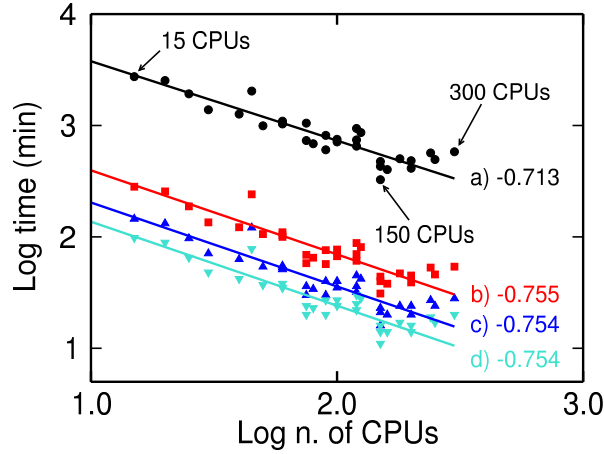
Figure 4. Wall time for a complete SCF (logarithmic scale) versus the logarithm of the number of CPUs used in the calculations, for the 144-atom Silica cell. a) Full, b) ACE, c) L-ACE(0.002), d) L-ACE(0.004). With pure DFT method SCF times were always lower than 3 min. The lines are the plot in logarithmic scale of the fitting function $y = \alpha x^\beta$ and β is reported for each method.

Figure 5. Wall time for a complete SCF versus the number of bands required in the calculations, for the 144-atom Silica cell (384 occupied states). a) Full, b) ACE, c) L-ACE(0.002), d) L-ACE(0.004). With pure DFT method SCF times were always lower than 1 min. The lines are the linear regression of the points and the slopes are reported for each method.

have been performed. All the WALL times are shown in Figure 4 as points. With all the methods the lowest computational times are achieved when using 150 cores (marked by an arrow for the black circles in the plot). The increase observed for more than 150 cores is due to limitations of the default FFT parallelization scheme used by QUANTUM ESPRESSO[8, 23], that becomes ineffective when the number of processors exceeds the number of FFT planes (144 in this case). Among the three different combinations of cores per nodes resulting in 150 CPUs, i.e. 10 nodes with 15 cores per node, 6 nodes with 25 cores per node and 5 nodes with 30 cores per nodes, the best performances are obtained using 10 nodes with 15 cores per node for all methods. Also in this case the

Atoms	Method	ΔE_{hyb}	Max	MAD	St. Dev.	$\Delta\varepsilon_{\mathrm{HL}}$	Time
72	Full	0.4059	0.01449	0.00398	0.00519	7.6355	84 (100%)
	ACE	0.4059	0.01450	0.00398	0.00519	7.6356	13 (100%)
	L-ACE(0.002)	0.4064	0.01450	0.00398	0.00520	7.6355	12 (91%)
	L-ACE(0.004)	0.4106	0.01451	0.00399	0.00520	7.6360	10 (71%)
	pure DFT	0	0	0	0	5.1161	0.3 (0%)
144	Full	0.5567	0.01772	0.00404	0.00520	7.2559	578 (100%)
	ACE	0.5567	0.01773	0.00404	0.00520	7.2560	54 (100%)
	L-ACE(0.002)	0.5592	0.01773	0.00404	0.00520	7.2562	28 (51%)
	L-ACE(0.004)	0.5663	0.01774	0.00404	0.00520	7.2576	20 (33%)
	pure DFT	0	0	0	0	4.7764	1 (0%)
288	Full	0.5588	0.01735	0.00400	0.00520	7.0843	3956 (100%)
	ACE	0.5588	0.01735	0.00400	0.00520	7.1154	641 (100%)
	L-ACE(0.002)	0.5588	0.01735	0.00400	0.00520	7.1155	170 (25%)
	L-ACE(0.004)	0.5694	0.01736	0.00401	0.00520	7.1156	116 (16%)
	pure DFT	0	0	0	0	4.6115	2.3 (0%)
432	ACE	0.5772	-0.02067	0.00405	0.00531	6.7921	1913 (100%)
	L-ACE(0.002)	0.5806	-0.02068	0.00405	0.00532	6.7963	350 (16%)
	L-ACE(0.004)	0.5878	-0.02069	0.00406	0.00532	6.7973	246 (11%)
	pure DFT	0	0	0	0	4.2792	5 (0%)
576	L-ACE(0.002)	0.5595	0.01812	0.00393	0.00507	7.1319	636 (12%)
	L-ACE(0.004)	0.5666	0.01815	0.00394	0.00507	7.1330	475 (8%)
	pure DFT	0	0	0	0	4.6377	8 (0%)

Table 1. Summary of results for Silica cells. Pure DFT calculations with PBE, hybrid calculations with PBE0. ΔE_{hyb} is the difference between PBE0 and PBE total energies per atom (kcal/mol), $\Delta\varepsilon_{\mathrm{HL}}$ is the HOMO-LUMO gap (eV), Max, MAD and St. Dev. are the maximum difference, the mean absolute difference, and the standard deviation, respectively, of the PBE0 force components (Ry/Bohr) with respect to the PBE ones, Time is the WALL time (min) for one SCF with 300 CPUs, in parenthesis the fraction of included integrals is reported.

points have been fitted with the function $y = \alpha x^{\beta}$, where y is the SCF time and x the total number of cores, in the range 15–150 cores, and slopes of about -0.7 have been found for all the methods, with slightly better performances for ACE and L-ACE with respect to the Full method.

Finally since the exchange potential is applied to computed (occupied or empty) Kohn-Sham orbitals, the computational times grow with the number of such orbitals, as shown in Figure 5.

In Table 1 we report the results for selected physical quantities. Differences between PBE0 and PBE results are reported in order to have a unique reference, since PBE calculations are available for all systems. In the third column of Table 1 hybrid energy corrections (that is: differences between the PBE0 and PBE total energies per atom) are reported. The two exact methods, Full and ACE, give an hybrid energy correction of about 0.4 kcal/mol per atom for the 72 atoms system, of about 0.6 kcal/mol for all other systems. Results from L-ACE(0.002) and L-ACE(0.004) differ on the third and second significant digit, respectively, from exact results.

Columns 4 to 6 of Table 1 contain the Mean Absolute Differences (MAD), maximum (Max) differences and standard deviation of the differences between PBE0 and PBE forces, averaged over all the force components of all atoms. Having converged the SCF to 10^{-6} Ry we can trust the forces up to 10^{-4} Ry/Bohr, and five decimal digits are reported.

The maximum differences between the PBE0 force components computed with the Full and ACE method and the PBE ones in the system of 72 atoms is 0.01449 Ry/Bohr, while for the systems with 144 and 288 atoms they are 0.01772 and 0.01735 Ry/Bohr, respectively. The loss of accuracy introduced by L-ACE turns out to be very small: in all cases the maximum differences are almost exactly the same as for ACE and occur for the same force component.

A similar analysis can be performed for the MAD and the standard deviation. In all cases the differences between the ACE forces and L-ACE ones fall on the fifth decimal digit. This suggest that the truncation scheme used in L-ACE for the potential is reliable enough to provide good forces, that can be used for geometry optimizations and likely for molecular dynamics as well.

While energies and forces can be obtained from occupied orbitals only, other quantities such as e.g. the band gaps also require virtual orbitals. The band gaps for the Silica systems reported in Table 1 have been calculated by adding ten virtual orbitals to the number of occupied orbitals. Since the ACE operator is perfectly equivalent to the exact exchange operator only when it is applied to a function belonging to the space in which the projection is valid, N_P in Eq. 1 has been set as the number of occupied orbitals plus ten virtual orbitals. In the L-ACE scheme the occupied manifold is localized and the truncation scheme is applied, while the virtual orbitals are just those obtained from the KS equations, and all the exchange integrals involving at least one virtual orbital are evaluated. For this reason the speed-up of L-ACE with respect to ACE decreases when virtual orbitals are included. In all cases the errors in the band gaps (sixth column of Table 1) are of the order of 10^{-3} eV, which is accurate enough for any application.

4.2. Anthocyanine in water

As an example of organic molecules, the Cyanidin 3-O-glucoside ($C_{21}H_{20}O_{11}$) molecule (shortcut C3OG) has been considered. We extracted one snapshot from a Car-Parrinello simulation of the C3OG molecule surrounded by 171 water molecules, in an orthorhombic cell with parameters $a = 33.9933$ Bohr, $b = 36.7973$ Bohr and $c = 31.0310$ Bohr. The molecule is in the neutral state, the water molecules form various hydrogen bonds with the –OH moieties, and the whole system is shown in Figure 1b, with the aromatic and sugar rings of C3OG highlighted in orange and green, respectively.

On such geometry we run SCF calculations using the B3LYP[35] functional and cutoffs of 80 Ry, 320 Ry, 160 Ry for the orbital, density and exchange grids, respectively. For pure DFT calculations the PBE functional has been used. Calculations have been performed on the C3OG molecule both in vacuum and surrounded by the water

molecules. The calculations in vacuum were performed on the same geometry as in water just removing the solvent molecules. This allowed to compute the solvation energies as the differences between the total energy of the molecule in water and the total energies of the molecule and the waters alone, and the HOMO-LUMO band gaps in vacuum and in water. Such data are reported in Table 2, together with the computational times and with the hybrid energies corrections, analogously to the Silica systems (see previous paragraph).

Method	ΔE_{hyb}	ΔE_S	$\Delta\varepsilon_{\mathrm{HL}}^{\mathrm{gas}}$	$\Delta\varepsilon_{\mathrm{HL}}^{\mathrm{water}}$	Time
ACE	-24.9794	-161.3846	2.0521	2.6091	249 (100%)
L-ACE (0.001)	-24.9790	-161.3457	2.0521	2.6089	92 (36%)
L-ACE (0.002)	-24.9780	-161.2540	2.0521	2.6085	71 (26%)
L-ACE (0.003)	-24.9767	-161.1331	2.0521	2.6080	62 (21%)
L-ACE (0.004)	-24.9750	-161.0085	2.0521	2.6075	56 (18%)
L-ACE (0.005)	-24.9730	-160.8220	2.0521	2.6071	52 (16%)
pure DFT	0	-169.4025	1.1085	1.5666	5 (0%)

Table 2. Summary of results for C3OG molecule in water (565 atoms). Pure DFT calculations with PBE, hybrid calculations with B3LYP. ΔE_{hyb} is the difference between B3LYP and PBE total energies per atom (kcal/mol), ΔE_S is the solvation energy (kcal/mol), $\Delta\varepsilon_{\mathrm{HL}}^{\mathrm{gas}}$ and $\Delta\varepsilon_{\mathrm{HL}}^{\mathrm{water}}$ are the HOMO-LUMO gaps (eV) in the gas phase and in water respectively, Time is the WALL time (min) for one SCF with 160 CPUs, in parenthesis the fraction of included integrals is reported.

The SCF with the ACE method took 249 min using 160 CPUs on 8 nodes. With L-ACE(0.001), only 36% of the integrals are retained in the calculations, and the WALL time for the SCF is reduced by almost the same factor (37%) with respect to the ACE method. The error in the hybrid energy correction is of the order of 10^{-4} kcal/mol per atom; the errors in the solvation energy is -0.05 kcal/mol; the error in the HOMO-LUMO gap in water is just 0.0002 eV. Increasing the threshold of L-ACE the computational times are further reduced: with L-ACE(0.005) only 16% of the exchange integrals are included in the SCF, and the WALL time is 80% smaller than for ACE. Errors in the energies and gaps also increase, but still remain small: the error on the solvation energy is just 0.5626 kcal/mol, amounting to a relative error of about 0.3%, while the error on the HOMO-LUMO gap in water is just 0.002 eV.

4.3. TiO$_2$ nanoparticles

As a final example, a TiO$_2$ (anatase) nanoparticle of 465 atoms has been used. This has the classical bipyramidal shape exposing the (101) and (001) surfaces, with diameter of 1.5 and 3 nm, in the short and the long dimensions, respectively. The geometry has been taken from Ref.[47] and is shown in Figure 1c, together with the box – a tetragonal cell with $a = 50$ Bohr and $c = 70$ Bohr – in bold black. The minimum distance between two atoms belonging to different replicas is 18.36 Bohr and is found between replicas along the c (longest) direction.

The SCF calculations for such systems have been done using the HSE[3] functional and the same cutoffs as in the previous Section. For pure DFT calculations the PBE functional has been used.

Method	ΔE_{hyb}	$\Delta \varepsilon_{\mathrm{HL}}$	Time
ACE	14.8840	4.3692	1877 (100%)
L-ACE (0.001)	14.8844	4.3728	838 (38%)
L-ACE (0.002)	14.8859	4.3751	754 (31%)
L-ACE (0.003)	14.8879	4.3772	691 (27%)
L-ACE (0.004)	14.8900	4.3783	643 (25%)
L-ACE (0.005)	14.8929	4.3792	651 (23%)
pure DFT	0	2.9434	112 (0%)

Table 3. Summary of results for the TiO_2 cluster (465 atoms). Pure DFT calculations with PBE, hybrid calculations with HSE. ΔE_{hyb} is the difference between HSE and PBE total energies per atom (kcal/mol), $\Delta \varepsilon_{\mathrm{HL}}$ is the HOMO-LUMO gap (eV), Time is the WALL time (min) for one SCF with 288 CPUs, in parenthesis the fraction of included integrals is reported.

The system has 2448 electrons, and 1300 KS states have been used in order to compute the HOMO-LUMO gap. In Table 3 the gaps and the hybrid energy corrections are displayed together with the WALL times for the SCF. We observe that the SCF with the L-ACE(0.005) method is three times faster than the ACE method, approaching the performances of the pure DFT. The error in the HOMO-LUMO gap is just 0.01 eV.

5. Conclusions

We have introduced a new method, L-ACE, for the fast and reliable calculation of the electronic structure with hybrid functionals and a PW basis set, effectively combining the ACE method and a localized representation of molecular orbitals. For the latter, the SCDM approach has been used. We have shown how to integrate L-ACE into the double-loop self-consistent algorithm of QUANTUM ESPRESSO and how to get rid of the unpredictable energy fluctuations present in other algorithms using truncation.

Compared with other similar strategies based on orbital localization, [13, 16, 17, 18] the present methodology has the advantage of combining the reduction of the number of evaluations of the exchange operator, granted by ACE, with the truncation of a large fraction of terms in each evaluation of the exchange operator, allowed by localization. Moreover, the algebraic, non-iterative character of SCDM localization makes the algorithm robust.

In conclusion, L-ACE allows to perform hybrid-functional calculations on systems that were previously unfeasible or exceedingly costly. Further work is under way to extend L-ACE to calculations using k-points (ACE without localization is already available for this case) and to compute each term in the truncated exchange potential via a real-space solution of the Poisson equation, getting rid of FFTs altogether.

[1] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, 1988.

[2] C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, 1988.

[3] J. Heyd, G. E. Scuseria, and M. Ernzerhof. Hybrid functionals based on a screened coulomb potential. *J. Chem. Phys.*, 118:8207–8215, 2003.

[4] Sanjay Chawla and Gregory A. Voth. Exact exchange in ab initio molecular dynamics: An efficient plane-wave based algorithm. *J. Chem. Phys.*, 108:4697–4700, 1998.

[5] Nicholas M. Boffi, Manish Jain, and Amir Natan. Efficient computation of the hartree-fock exchange in real-space with projection operators. *J. Chem. Theory Comput.*, 12:3614–3622, 2016.

[6] A. Sorouri, W. M. C. Foulkes, and N. D. M. Hine. Accurate and efficient method for the treatment of exchange in a plane-wave basis. *J. Chem. Phys.*, 124:064105, 2006.

[7] F. Gygi and A. Baldereschi. Self-consistent hartree-fock and screened-exchange in solids: Application to silicon. *Phys. Rev. B*, 34:4405–4408, 1986.

[8] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L. Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquerello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M. Wentzcovitch. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21:395502, 2009.

[9] Manuel Guidon, Jürg Hutter, and Joost VandeVondele. Auxiliary density matrix methods for hartree-fock exchange calculations. *J. Chem. Theory Comput.*, 6:2348–2364, 2010.

[10] M. Marsili and P. Umari. Method for the fast evaluation of fock exchange for nonlocalized wave functions. *Phys. Rev. B*, 87:205110, 2013.

[11] Nicola Varini, Davide Ceresoli, Layla Martin-Samos, Ivan Girotto, and Carlo Cavazzoni. Enhancement of dft-calculations at petascale: Nuclear magnetic resonance, hybrid density functional theory and car-parrinello calculations. *Comput. Phys. Commun.*, 184(8):1827–1833, 2013.

[12] T. Barnes, T. Kurth, P. Carrier, N. Wichmann, D. Prendergast, P. R. C. Kent, and J. Deslippe. Improved treatment of exact exchange in quantum espresso. *Comput. Phys. Commun.*, 241:52–58, 2017.

[13] Xifan Wu, Annabella Selloni, and Roberto Car. Order-n implementation of exact exchange in extended insulating systems. *Phys. Rev. B*, 79:085102, 2009.

[14] François Gygi and Ivan Duchemin. Efficient computation of hartree-fock exchange using recursive subspace bisection. *J. Chem. Theory Comput.*, 9:582–587, 2013.

[15] Jeff Mountjoy, Michelle Todd, and Nicholas J. Mosey. Exact exchange with non-orthogonal generalized wannier functions. *J. Chem. Phys.*, 146:104108, 2017.

[16] C. J. García-Cervera, Jianfeng Lu, Yulin Xuan, and Weinan E. Linear-scaling subspace-iteration algorithm with optimally localized nonorthogonal wave functions for kohn-sham density functional theory. *Phys. Rev. B*, 79:115110, 2009.

[17] Carlos J. García-Cervera, Jianfeng Lu, and Weinan E. A sub-linear scaling algorithm for computing the electronic structure of materials. *Commun. Math. Sci.*, 5:999–1026, 2007.

[18] Stephan Mohr, Laura E. Ratcliff, Luigi Genovese, Damien Caliste, Paul Boulanger, Stefan Goedecker, and Thierry Deutsch. Accurate and efficient linear scaling dft calculations with universal applicability. *Phys. Chem. Chem. Phys.*, 17:31360, 2015.

[19] Robert A. DiStasio, Biswajit Santra, Zhaofeng Li, Xifan Wu, and Roberto Car. The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water. *J. Chem. Phys.*, 141:084502, 2014.

[20] Nicola Marzari and D. Vanderbilt. Maximally localized generalized wannier functions for composite energy bands. *Phys. Rev. B*, 56:12847–12865, 1997.

[21] Nicola Marzari, A. A. Mostolfi, J. R. Yates, I. Souza, and D. Vanderbilt. Maximally localized wannier functions: Theory and applications. *Rev. Mod. Phys.*, 84:1419–1475, 2012.

[22] Lin Lin. Adaptively compressed exchange operator. *J. Chem. Theory Comput.*, 12:2242–2249, 2016.

[23] Paolo Giannozzi, Oliviero Andreussi, Thomas Brumme, Oana Bunau, Marco Buongiorno Nardelli, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Matteo Cococcioni, Nicola Colonna, Ivan Carnimeo, Andrea Dal Corso, Stefano de Gironcoli, Pietro Delugas, Robert A. DiStasio Jr., Andrea Ferretti, Andrea Floris, Guido Fratesi, Giorgia Fugallo, Ralph Gebauer, Uwe Gerstmann, Feliciano Giustino, Tommaso Gorni, J. Jia, M. Kawamura, H.-K. Ko, Anton Kokalj, Emine Küçükbenli, Michele Lazzeri, Margherita Marsili, Nicola Marzari, Francesco Mauri, N. L. Nguyen, H.-V. Nguyen, Alberto Otero-de-la Roza, Lorenzo Paulatto, Samuel Poncé, Sario Rocca, Riccardo Sabatini, Biswajit Santra, Martin Schlipf, Ari Seitsonen, Alexander Smogunov, Iurii Timrov, Timo Thonhauser, Paolo Umari, Nathalie Vast, Xifang Wu, and Stefano Baroni. Advanced capabilities for materials modelling with quantum espresso. *J. Phys.: Condens. Matter*, 29:465901, 2017.

[24] Ivan Duchemin and Franois Gygi. A scalable and accurate algorithm for the computation of hartree–fock exchange. *Comput. Phys. Comm.*, 181:855, 2010.

[25] Per-Olov Löwdin. Studies in perturbation theory. x. lower bounds to energy eigenvalues in perturbation-theory ground state. *Phys. Rev.*, 139:A357–372, 1965.

[26] Per-Olov Löwdin. Some properties of inner projections. *Int. J. Quant. Chem.*, 4:231–237, 1971.

[27] Francesco Aquilante, Linus Boman, Jonas Boström, Henrik Koch, Roland Lindh, Alfredo Sánchez de Merás, and Bondo Pedersen Thomas. Chapter 13: Cholesky decomposition techniques in electronic structure theory, in: Linear-scaling techniques in computational chemistry and physics. In *Challenges and Advances in Computational Chemistry and Physics*, volume 13, pages 1–44. Springer, Dordrecht, 2011.

[28] H. F. Beebe Nelson and Jan Linderberg. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. *Int. J. Quant. Chem.*, XII:683–705, 1977.

[29] Linus Boman, Henrik Koch, and Alfredo Sánchez de Merás. Method specific cholesky decomposition: Coulomb and exchange energies. *J. Chem. Phys.*, 129:134107, 2008.

[30] Florian Weigend. A fully direct ri-hf algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.*, 4:4285–4291, 2002.

[31] Damle Anil, Lin Lin, and Ying Lexing. Compressed representation of kohn-sham orbitals via selected columns of the density matrix. *J. Chem. Theory Comput.*, 11:1463–1469, 2015.

[32] Damle Anil, Lin Lin, and Ying Lexing. Computing localized representations of the kohn–sham subspace via randomization and refinement. *SIAM J. Sci. Comput.*, 39:B1178–B1198, 2017.

[33] Damle Anil, Lin Lin, and Ying Lexing. Scdm-k: Localized orbitals for solids via selected columns of the density matrix. *J. Comput. Phys.*, 334:1–15, 2017.

[34] A. D. Becke. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.*, 98:1372–1377, 1993.

[35] A. D. Becke. Density-functional thermochemistry .3. The role of exact exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.

[36] C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.*, 110:6158–6169, 1999.

[37] Lin Lin and Michael Lindsey. Convergence of adaptive compression methods for hartree-fock-like equations. arXiv:1703.05441, 2018.

[38] Pier Luigi Silvestrelli and Michele Parrinello. Water molecule dipole in the gas and in the liquid phase. *Phys. Rev. Lett.*, 82:3308–331, 1999.

[39] W. von Niessen. Density localization of atomic and molecular orbitals. i. *J. Chem. Phys.*, 56:4290–4297, 1972.

[40] H. V. Henderson and S. R. Searle. On deriving the inverse of a sum of matrices. *SIAM Review*, 23:53–60, 1981.

[41] K. S. Miller. On the inverse of the sum of matrices. *Mathematical Magazine*, 54:67–72, 1981.

[42] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, 1996.

[43] B. W. H. van Beest, G. J. Kramer, and R. A. van Santen. Force fields for silicas and aluminophosphates based on ab initio calculations. *Phys. Rev. Lett.*, 64:1955, 1990.

[44] N. S. Shcheblanov, B. Mantisi, P. Umari, and A. Tanguy. Detailed analysis of plastic shear in the raman spectra of sio_2 glass. *J. Non-Cryst. Solids*, 428:6, 2015.

[45] Loris Ercole, Aris Marcolongo, and Stefano Baroni. Accurate thermal conductivities from optimally short molecular dynamics simulations. *Scientific Reports*, 7:15835, 2017.

[46] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemitry*. Dover Publications Inc., 31 East 2nd Street, Mineola, N.Y. 11501, 1996.

[47] G. Mattioli, A. Amore Bonapasta, D. Bovi, and P. Giannozzi. Photocatalytic and photovoltaic properties of tio_2 nanoparticles investigated by ab initio simulations. *J. Phys. Chem. C*, 118:29928, 2014.