eQE — An Open-Source Density Functional Embedding Theory Code for the Condensed Phase

Alessandro Genova,[†] Davide Ceresoli,[‡,†] Alisa Krishtal,[†] Oliviero Andreussi,[¶]
Robert A. DiStasio, Jr.,[§] and Michele Pavanello[*,†]

†*Department of Chemistry, Rutgers University, Newark, NJ 07102, USA*

‡*CNR-ISTM: Institute of Molecular Sciences and Technologies, Milano, Italy*

¶*Theory and Simulations of Materials (THEOS) and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, Station 12, CH-1015 Lausanne, Switzerland*

§*Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853 USA*

E-mail: m.pavanello@rutgers.edu

Abstract

In this work, we present the main features and algorithmic details of a novel implementation of the Frozen Density Embedding (FDE) formulation of subsystem Density Functional Theory (DFT) that is specifically designed to enable *ab initio* molecular dynamics (AIMD) simulations of large-scale condensed-phase systems containing 1000s of atoms. This code (available at `http://eqe.rutgers.edu`) has been given the moniker of embedded Quantum ESPRESSO (eQE) as it is a generalization of the open-source Quantum ESPRESSO (QE) suite of programs. The strengths of eQE reside in a hierarchical parallelization scheme that allows for an efficient and fully self-consistent treatment of the electronic structure (*via* the addition of an additional DIIS extrapolation layer) while simultaneously exploiting the inherent symmetries and periodicities in the system (*via* sampling of subsystem-specific first Brillouin zones and utilization of subsystem-specific basis sets). While bulk liquids and molecular crystals are two classes of systems that exemplify the utility of the FDE approach (as these systems can be partitioned into weakly interacting subunits), we show that eQE has significantly extended this regime of applicability by outperforming standard semilocal Kohn–Sham DFT (KS-DFT) for large-scale heterogeneous catalysts with quite different layer-specific electronic structure and intrinsic periodicities. eQE features very favorable *strong* parallel scaling for a model system of bulk liquid water composed of 256 water molecules, which allows for a significant decrease in the overall time to solution when compared to KS-DFT. We show that eQE achieves speedups greater than one order of magnitude ($> 10\times$) when performing AIMD simulations of such large-scale condensed-phase systems as: (1) molecular liquids *via* bulk liquid water represented by 1024 independent water molecules (3072 atoms with a 25.3× speedup over KS-DFT), (2) polypeptide/biomolecule solvation *via* $(\textsc{gly})_6$ solvated in $(H_2O)_{395}$ (1230 atoms with a 38.6× speedup over KS-DFT), and (3) molecular crystals *via* a 3×3×3 periodic supercell of pentacene (1940 atoms with a 12.0× speedup over KS-DFT). These results represent a significant improvement over the current state-of-the-art and now enable subsystem DFT based AIMD simulations of realistically-sized condensed-phase

systems of interest throughout chemistry, physics, and materials science.

1 Introduction

The Kohn–Sham (KS) formulation of Density Functional Theory (DFT)[1] is currently the most widely employed electronic structure method in the fields of chemistry, physics, and materials science. This is largely due to the fact that KS-DFT employing semilocal exchange-correlation (xc) functionals produces models of remarkable accuracy and predictive capability[2] with a relatively low associated computational cost (that in general scales as $\mathcal{O}(N^3)$, with N representing the size of the electronic system). Although the computational complexity associated with KS-DFT is significantly lower than the quantum chemical hierarchy of wavefunction-based methodologies, substantial complications arise when performing KS-DFT calculations on realistically-sized systems (*e.g.*, containing 1000s of atoms), and such issues are only compounded when one seeks to perform KS-DFT based molecular dynamics.

Efforts to reduce the computational scaling of KS-DFT are currently underway in many research groups and in many different directions,[3–10] most of which are combining algorithmic advances with the efficient utilization of modern (super)-computer architectures to attain efficient linear-scaling or $\mathcal{O}(N)$ implementations of KS-DFT. In order to do so, such linear-scaling KS-DFT codes exploit nearsightedness in the electronic structure, *i.e.*, the computational savings are achieved by screening some of the one- and two-electron integrals (or equivalently, if such integrals are not employed in the formalism, an effective cutoff distance is applied to the off-diagonal elements of the one-body reduced density matrix), coupled with multipolar expansions of the interactions between far-field charge densities.

In this work, we present our contribution to these efforts, which focuses on the divide-and-conquer strategy provided by the Frozen Density Embedding (FDE) formulation of subsystem DFT.[11] In general, we find the FDE approach to be particularly flexible, as the nearsightedness in the electronic structure (which is often globally imposed on the entire system in linear-scaling approaches) can be naturally tailored on a subsystem basis. In essence, this intrinsic feature of the FDE approach can be simultaneously exploited to capture potential long-range coherences in the electronic structure of a given subsystem (*e.g.*, as found in metals and semiconductors) and reduce the associated computational cost when treating relatively localized fragments or subsystems. As such, subsystem DFT is particularly suited to tackle systems for which it is possible to take advantage of certain inherent symmetries. For example, consider the hydrodesulfurization model catalyst depicted in Figure 1, in which thiophene molecules interact with a molybdenum disulfide (MoS_2) monolayer deposited on an alumina support. As discussed below (see "Applicability and performance of eQE" section and Table 1 below), the FDE approach is able to treat each layer of this heterogeneous catalyst as a single subsystem. In this way, not only can the first Brillouin zones in the x and y directions be sampled on a subsystem basis, but also the nearsightedness of the electronic structure can be appropriately exploited along the z direction *only.*

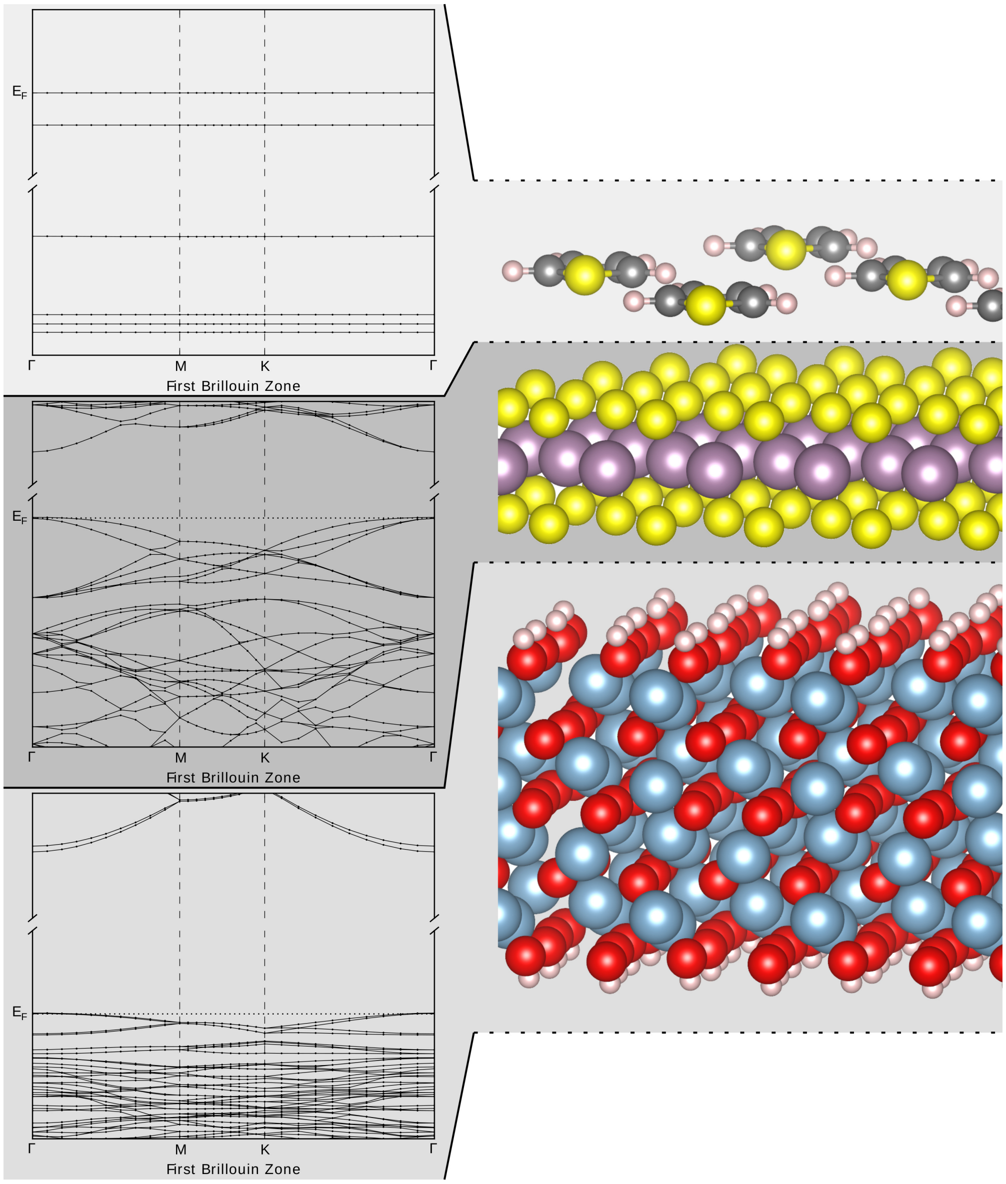
We note in passing here that the theoretical and algorithmic developments in the FDE and linear-scaling DFT communities can be quite complementary, a fact exemplified by the recent work of Andermatt *et al.*,[12] in which the combination of linear-scaling DFT with FDE enabled a first-principles based geometry optimization of the satellite tobacco mosaic virus in solution (a system which contains 10,000s of atoms). Quite simply put, this calculation would have remained intractable for the foreseeable future without the efficient utilization of the advances in both of these approaches.

The FDE approach is achieved by partitioning the total supersystem electron density, $\rho(\mathbf{r})$, as follows:

$$\rho(\mathbf{r}) = \sum_{I}^{N_S} \rho_I(\mathbf{r}), \tag{1}$$

where N_S is the number of subsystems chosen. If the partitioning above can be achieved, information regarding the interactions (both static and dynamic) between the subsystems becomes available in a highly-distributed fashion (both in terms of computational FLOPs as well as data input/output), thereby opening the door to new prospectives for interpreting

Figure 1: A realistic model system for the catalytic hydrodesulfurization process in which thiophene interacts with a monolayer of molybdenum disulfide (MoS_2) deposited on a hydroxylated alumina support. The different nature of the electronic structure of each region induces a layer-specific intrinsic periodicity as exemplified by the band structures plotted on the left. Reproduced with permission from Ref. 13.



any physicochemical phenomenon.[14]

The density partitioning in Eq. (1) is achieved by FDE *via* the solution to the following subsystem-specific KS equations,

$$-\frac{1}{2}\nabla^2\phi_i^I(\mathbf{r}) + \left[v_{\text{KS}}^I[\rho_I](\mathbf{r}) + v_{\text{emb}}^I[\{\rho_J\}](\mathbf{r})\right]\phi_i^I(\mathbf{r}) = \varepsilon_i^I\phi_i^I(\mathbf{r}), \tag{2}$$

with the embedding potential given by

$$\begin{aligned} v_{\text{emb}}^I[\{\rho_J\}](\mathbf{r}) = \sum_{J\neq I}^{N_S}\left[\int\frac{\rho_J(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}' - \sum_{\alpha\in J}\frac{Z_\alpha}{|\mathbf{r}-\mathbf{R}_\alpha|}\right] \\ + \frac{\delta T_{\text{s}}^{\text{nadd}}[\{\rho_J\}]}{\delta\rho_I(\mathbf{r})} + \frac{\delta E_{\text{xc}}^{\text{nadd}}[\{\rho_J\}]}{\delta\rho_I(\mathbf{r})}. \end{aligned} \tag{3}$$

In this expression, the Coulomb (Hartree) interaction is accounted for exactly while the nonadditive kinetic ($T_{\text{s}}^{\text{nadd}}$) and exchange-correlation ($E_{\text{xc}}^{\text{nadd}}$) functionals, defined as $F^{\text{nadd}}\left[\rho_J\right] \equiv F[\rho] - \sum_J^{N_S} F[\rho_J]$, are evaluated in terms of approximate density functionals.[11,15–17] Finally, the total energy of the supersystem is evaluated as:

$$\begin{aligned} E_{\text{FDE}} = \sum_I\left(T_{\text{s}}[\rho_I] + E_{eN}^I[\rho_I] + E_H[\rho_I] + E_{\text{xc}}[\rho_I]\right) + \sum_I^{N_S}\sum_{K\neq I}^{N_S} E_{eN}^K[\rho_I] \\ + T_{\text{s}}^{\text{nadd}}[\{\rho_J\}] + E_H^{\text{nadd}}[\{\rho_J\}] + E_{\text{xc}}^{\text{nadd}}[\{\rho_J\}] + V_{NN}, \end{aligned} \tag{4}$$

where $E_{eN}^K[\rho_I]$ is the Coulomb interaction energy between the nuclei of fragment K and the electron density of fragment I and V_{NN} is the nuclear-nuclear repulsion term. While Eq. (4) is probably one of the most common and informative ways to write the total FDE energy expression, calculating the nonadditive Hartree energy, E_H^{nadd}, is certainly not efficient and is never done so in practice (as the full Hartree potential and energy are computed). More details on the calculation of the total FDE energy can be found in Appendix 2 of Ref. 18.

FDE and other similar embedding methods have found many advocates in recent years. On the FDE side, the Amsterdam Density Functional (ADF) code[19] holds perhaps the most celebrated implementation.[20–25] Another notable implementation[26,27] resides in the CP2K code.[28] CASTEP[29] also has an implementation of FDE[30] which was employed in simulations involving two subsystems, one of which was treated at the correlated wavefunction level.[31–33] In addition, Turbomole[34] has its own implementation by the Della Sala group.[35–39] We also mention here that other embedding methods, which can be categorized as exact density embedding, exact orbital embedding, or electrostatic embedding, are now found in ADF,[40] MOLPRO,[41–45] Q-Chem,[46,47] CP2K,[48] NWChem,[49] and GAMESS.[50]

In this work, we present a novel implementation of the FDE approach that aims at filling the following gap that has persisted over the years, namely, the absence of a code that (1) has a proven strong parallel efficiency that consistently outperforms semilocal KS-DFT, (2) has the ability to run AIMD, and (3) is applicable to periodic systems such as conductors and semiconductors. Since we have incorporated our implementation into the open-source Quantum ESPRESSO (QE) suite of programs,[51] we dub our code embedded QE or eQE.

In the sections that follow, we present the implementation details of the eQE algorithm and provide the reader with theoretical and algorithmic strategies employed in its design and construction. This is followed by several illustrative applications which showcase the performance of eQE in treating large-scale condensed-phase systems of interest which contain 1000s of atoms. We conclude the manuscript with an outline of several potential research directions which we believe will further increase the range of applicability of subsystem DFT within the eQE framework.

2 Details of the eQE implementation

In this section we describe the key details of the eQE implementation. Specifically, we discuss the following five important developments that make eQE unique and extremely efficient: (1) self-consistency of the electronic structure, (2) guiding the SCF to convergence *via* the introduction of an additional DIIS layer, (3) the employed hierarchical parallelization scheme,

(4) sampling of the first Brillouin zone (FBZ), and (5) subsystem-specific basis sets.

2.1 Self-consistency of the electronic structure

One key element of the eQE implementation is that we seek a fully self-consistent field (SCF) coherent with the choice of the underlying electronic structure theory. This implies that each subsystem density, ρ_I, needs to be self-consistent as well as converged with respect to variation of the other subsystem densities, $\{\rho_J\}$, for $J \neq I$. To achieve this goal, we have adopted a strategy similar to the one employed in the CP2K implementation of FDE, which prescribes that each subsystem gains knowledge of the other subsystem densities at every SCF cycle.[26] Thus, the KS-like problem in Eq. (2) is run simultaneously for each subsystem.

The SCF is considered converged once all subsystem density errors from one SCF cycle to the next, σ, are smaller than a user-defined threshold. In our experience, a threshold of $\sigma < 10^{-9}$ Ry is sufficient for running AIMD simulations that achieve a satisfactory energy conservation in the NVE ensemble.

2.2 Guiding the SCF to convergence: an additional DIIS layer

Straightforward implementation of the algorithm described above results in computations that are slower than standard KS-DFT. To achieve a fast and scalable FDE implementation, our efforts began with taking a closer look at the electron density mixing during the SCF procedure. Specifically, we have introduced a new and additional DIIS[52] layer that has substantially decreased the time to solution. Depending on the system, we can achieve a reduction in the number of SCF cycles by up to 50%. This new DIIS layer deals with density mixing at different rates for each of the different subsystems and is applied before the subsystem densities are summed through Eq. (1) to yield the supersystem density (needed for computing the new potentials for the next SCF cycle).

With $\rho_I^{out}(n_{\mathrm{SCF}})$ we indicate the electron density of subsystem I available in output after SCF cycle n_{SCF} has completed. $\rho_I^{out}(n_{\mathrm{SCF}})$ is mixed with subsystem electron densities of previous SCF cycles employing Broyden density mixing.[53] The density resulting from the Broyden mixing, $\rho_I^{\text{Broyden}}(n_{\text{SCF}})$, is then mixed to the previous SCF density to generate a new density, $\rho_I^{in}(n_{\text{SCF}}+1)$, that will be used to evaluate the KS-DFT potentials of the next SCF cycle, as follows

$$\rho_I^{in}(n_{\text{SCF}}+1) = \beta_I \rho_I^{\text{Broyden}}(n_{\text{SCF}}) + (1-\beta_I)\rho_I^{in}(n_{\text{SCF}}). \tag{5}$$

β_I is the Broyden mixing parameter which takes values between 0 and 1. It is typically dependent on the system's gap (i.e., large/small gaps afford large/small β_I).

In the new DIIS method, β_I is made SCF cycle dependent, namely, $\beta_I(n_{\text{SCF}})$, and it is evaluated in the following way for large values of σ_{FDE},

$$\beta_I(n_{\text{SCF}}) = \beta_I \cdot \max\left\{\frac{\sigma_{FDE}(n_{\text{SCF}}) - \sigma_I(n_{\text{SCF}})}{\sigma_{FDE}(n_{\text{SCF}})}, 0.2\right\}, \tag{6}$$

where σ_I is the density difference (or error) with the previous cycle for the subsystem density, and $\sigma_{FDE}(n_{\text{SCF}}) = \max_I\{\sigma_I(n_{\text{SCF}})\}$. While for low values of σ_{FDE} the original β_I is employed.

The effect of this new DIIS layer is such that if one subsystem features large density oscillations from one SCF cycle to the next, its new density will be mixed with the previous ones using a lower mixing parameter (as compared to the other subsystems).

For the layered system described in Figure 1 and Tables 1–2, the number of SCF cycles required for convergence was reduced by up to 50% by the new DIIS procedure depending on the specific FDE implementation (*vide infra*). For other systems we found a smaller but significant reduction (*e.g.*, 2 SCF steps are saved on average out of 10 for $(H_2O)_{64}$).

2.3 Hierarchical Parallelization Scheme

The simultaneous solution of Eq. (2) provides us with the opportunity to task-parallelize eQE at the subsystem level. The subsystem KS equations compute new subsystem densities (e.g.,

ρ_I) at every SCF cycle, which are then summed to yield the supersystem density (ρ). Given that the density is an array of low dimensionality, its transfer between subsystems involves relatively low data communication traffic.

The first MPI communicator that we define is the `inter_fragment` communicator. This communicator links one selected process for each subsystem (`ionode`, hereafter) and is mostly used for sharing information about the electron density. The `inter_fragment` communicator was inspired by the already existing image parallelization level in QE originally developed for nudged elastic band simulations.[51]

Each subsystem can take advantage of further parallelization. For this we have repurposed the existing `intra_image` communicator (originally developed for NEB calculations) to collect all processes that are assigned to a subsystem. These processes are further split into pools, in a way that is analogous to the standard QE parallelization architecture.

Figure 2 highlights the flexible and hierarchical parallelization strategy employed in eQE. Inset (a) of the figure depicts the parallelization strategy in the native QE code for nudged elastic band simulations. This preexisting strategy is problematic for subsystem DFT approaches, as each subsystem (denoted by the columns in the diagram) is forced to use the same number of processes. This of course is not ideal in those cases in which one subsystem is much larger than the others in terms of the number of valence electrons.

Inset (b) of Figure 2 depicts the parallelization strategy in eQE. The clear advantage over the strategy in inset (a) is the fact that the new MPI stack provides us with enough flexibility to assign a *subsystem-specific* number of MPI processes.

The new eQE MPI stack was designed to maximize speed while retaining as much flexibility as possible. For this reason, we also created the MPI communicator `large_comm`, which collects *up to* all available MPI processes to calculate potentials associated with the supersystem density, ρ. For some architectures (*e.g.*, those with relatively slow internode connectivity) the `large_comm` communicator can be disabled by running eQE with the `-nfp` flag, which stands for "no fancy parallelization". Experts can redefine `large_comm` depending on the specific computer architecture and only include the portion of available MPI processes that best exploits the available network architecture.

The number of MPI processes assigned to a given subsystem can be specified in the `fragments_procs.in` file, which needs to be copied into the execution folder before runtime. More specifically, line `n` of `fragment_procs.in` contains the desired number of MPI processes for subsystem `n`.

2.4 Sampling of the First Brillouin Zone (FBZ)

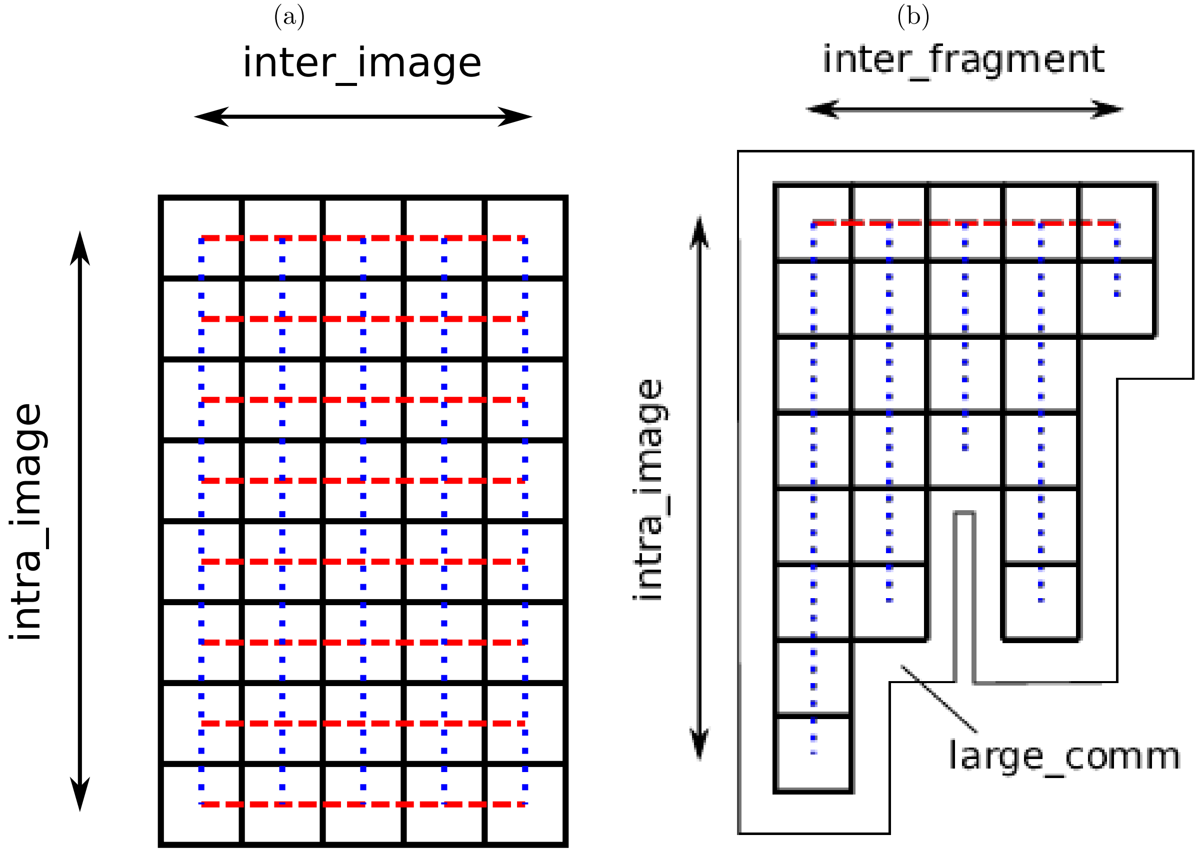
Figure 1 suggests that the FBZ of molecular systems and insulators can be sampled with fewer k-points than the FBZ of semiconductors and insulators. That is, molecular systems can be treated as non-periodic embedded systems.[27] In a recent work[13] we showed that it is possible to assign a set of k-points to each subsystem in order to sample the corresponding FBZ, and this number can be chosen according to the nature of the electronic structure of the subsystem. This finding allows us to represent each subsystem band with the smallest number of k-points needed to reach a target accuracy.

The k-point sampling in eQE is read from input *via* the `K_POINT` card. As each subsystem has its own input file, the individual `K_POINT` cards can be specified accordingly.

2.5 Subsystem-specific basis sets

Another aspect where the true potential of eQE is revealed resides in the definition of subsystem-specific basis sets. Due to the locality of the electronic structure of each subsystem, there is no need to use the entire plane wave set associated with the native simulation cell. Using subsystem based basis sets (also called monomer basis sets in the literature) is an important step for practical implementations of methods that exploit locality, such as subsystem DFT. In the literature, subsystem-based basis set implementations of Frozen Density Embedding are indicated by the acronym FDE(m), while FDE(s) indicates that the supersystem basis set is used for all subsystems.[20] While achieving this goal is (at least on

Figure 2: Comparison of QE and eQE MPI stacks. Each square represents a single MPI process. Blue dotted segments represent `intra_image` communicators and run along the columns (each representative of a subsystem). The red dashed segments represent communication across subsystems. Inset (a): The original QE MPI stack. All subsystems are restricted to have the same number of processes, and each process is part of a communicator with the corresponding process in the other subsystems (with corresponding topology defined in the `inter_image` MPI communicator). Inset (b): The new eQE MPI stack. The `inter_fragment` MPI communicator links only the first process belonging to each subsystem. As these are important processes, we call them `ionode`. All subsystems are given an arbitrary number of processes (as defined by the repurposed `intra_image` communicator) and the communication *between* subsystems is managed exclusively by the `ionode` set. Distributed quantities, such as densities and potentials, are transmitted across subsystems by first collecting the quantity on each `ionode` (through the `intra_image` communicator corresponding to a given subsystem), broadcasting it across the `ionode` set, and then redistributing it across the `intra_image` communicators. The role of the `large_comm` communicator is to compute potentials corresponding to the supersystem density (*i.e.*, the sum of the subsystem densities) using all available processes.

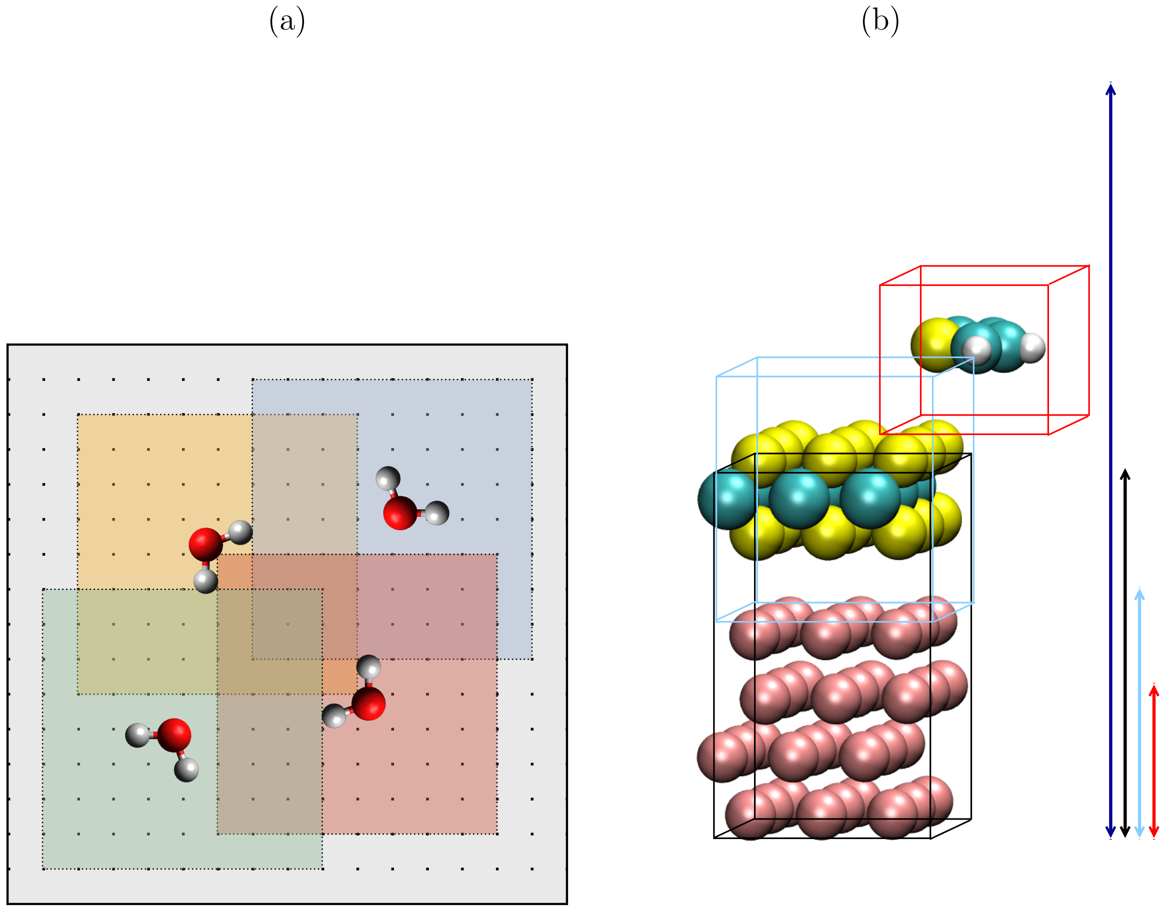


paper) relatively straightforward when atom-centered basis sets are employed, the fact that QE includes periodic boundary conditions and employs an originless plane-wave (PW) basis set, presents us with a challenge. If all of the N_S subsystems are represented on the same supersystem simulation cell and share the same kinetic energy cutoff in the PW expansion, the code would thus need to solve N_S coupled KS-like problems in the large (supersystem) basis set. This would significantly slow down the procedure even when compared to semilocal KS-DFT for the supersystem and defeat the purpose of using the subsystem approach. Hence, one way to employ subsystem-specific basis sets is to define smaller simulation cells whose PWs (the number of which is significantly reduced compared to the set associated with the native supersystem cell) are employed in the expansion of the subsystem orbitals or bands (see Figure 3). To achieve this goal, in eQE smaller and subsystem-centered cells are "carved out" of the native cell and used only to expand the subsystem molecular orbitals (*i.e.*, bands or waves). The cells are allowed to overlap so that the nonadditive potentials can be computed with no loss of accuracy.

In eQE, each subsystem lattice vector can be scaled independently from the others by assigning a value less than 1.0 to the keyword `fde_cell_split(n)`, with `n=1, 2, 3`. The subsystem cells are then generated by scaling the three lattice vectors of the supersystem cell by the values of `fde_cell_split(n)` in such a way that the grid points of each subsystem cell are imposed to match exactly with corresponding points in the supersystem cell. This allows us to efficiently (and exactly) transfer density and potentials in real space back and forth between the subsystem and supersystem cells rather than performing more expensive (and less accurate) interpolation schemes. As an illustrative example of how the subsystem cell approach is employed in eQE, consider the computation of the total electron density. This fundamental task in subsystem DFT is computed through Eq. (1) and a snippet of the corresponding algorithm in eQE is provided in the Supplementary Information (see Figure S1).

Iterative schemes to diagonalize the Hamiltonian for a few selected roots only (*i.e.*,

Figure 3: Graphical depiction of the scheme utilized for handling the PW expansion of subsystem molecular orbitals. Small, subsystem-centered cells are "carved out" of the supersystem cell and used to expand the subsystem molecular orbitals only. The subsystem cells are allowed to overlap so that the nonadditive potentials can be computed accurately. To ensure grid point matching in the subsystem and supersystem cells, there are some constraints in the choice of the scaling factors, `fde_cell_split(n)`. Inset (a) depicts liquid water, a system made of non-periodic subsystems, while inset (b) depicts the same system as in Figure 1 but with a gold substrate. For (b) the cells of the periodic subsystems (Au and MoS_2) can be reduced along the z-direction. The arrows in (b) depict the relative lengths of the various cells with the longest being associated with the supersystem cell.



Lanczos-based approaches) are commonly employed in condensed-phase SCF algorithms and their complexity is proportional to the square of the simulation cell volume. As such, it is immediately clear that the subsystem-specific simulation cell (or subsystem-specific basis set) approach described above allows for a significant reduction in the computational cost and scaling of the diagonalization routine (see Section "Applicability and performance of eQE" for timings that compare the supersystem versus subsystem-specific basis set).

There are many types of condensed-phase systems for which such a subsystem-specific cell approach is useful. For example, simulations of molecular liquids, such as water, have been carried out (*vide infra* and Figure 3). Generally speaking, the subsystem-specific cell approach is useful in cases where the periodicity of a subsystem is less than or equal to two dimensions. The case of slabs (see for example the system depicted in Figure 1) is important as the simulation cell can be reduced in the vertical direction, effectively reducing the amount of vacuum that needs to be considered in the molecular orbital expansion and therefore speeding up the calculation considerably (see Figure 3 and Table 1).

2.6 Subsystem-specific spin

One of the aims of an embedding code is to be able to treat radical species in general environments that can be either open- or closed-shell. Thus, eQE has been designed to provide this flexibility and allows one to treat closed-shell subsystems embedded in open-shell environments and vice versa.[54] This is achieved using two keywords: `fde_nspin` and `nspin`. The general case has an open-shell supersystem density, achieved with `fde_nspin=2`, while the spin of the individual subsystems can be set as either closed- or open-shell by setting `nspin` to 1 or 2, respectively. Non-collinear spin options are not yet supported in eQE.

3 Applicability and performance of the eQE implementation

3.1 Periodic systems

We have tested the applicability of eQE employing standard GGA nonadditive kinetic energy functionals in the definition of the FDE energy functional and the embedding potential. We have considered an array of periodic systems and first assessed the performance of the FDE method in eQE as compared to standard (supersystem-based) semilocal KS-DFT in QE. In agreement with previous studies employing semilocal nonadditive functionals, we found that the discrepancies between FDE and KS-DFT increase as the overlap in the inter-subsystem densities increases.[18] For example, when two perylene diimide molecules are stacked on top of a Au(111) surface, the interaction is well characterized by FDE. Conversely, when molecular systems are placed on top of reactive metal surfaces (*e.g.*, water on Pt and methane on Pd) FDE reproduces the electronic structure of only those conformations that do not require substantial hybridization of the molecular orbitals with the bands on the respective metal.[18]

Arguably, employing local nonadditive energy functionals is not an ideal practice. The local density approximation leads to incorrect long-range behavior of the functionals (both the kinetic energy and the exchange–correlation). Regarding the nonadditive correlation functional, several groups have proposed solutions. Kevorkyants *et al.*[55] and Sinha *et al.*[56] have developed a fluctuation-dissipation theorem adiabatic connection formalism based on subsystem TDDFT that replaces the semilocal nonadditive correlation with a fully nonlocal (and in principle exact) RPA-like nonadditive correlation functional. This led to as accurate binding energies of weakly bonded molecular dyads as supersystem RPA [*e.g.*, mean unsigned errors of < 0.5 kcal/mol against CCSD(T) on the S22 set[57]]. Beyhan *et al.*[58] also proposed a Grimme-like long-range correction to the nonadditive correlation energy which led to much improved interaction energies. to account for dispersion interactions between subsystems, eQE can include London dispersion forces as implemented in the main QE code base. Although this has not yet been benchmarked for accuracy, it is a functionality available to the user.

In terms of timings, we summarize the computational speedups found from an eQE calculation of the complex catalyst system depicted in Figure 1 against standard supersystem-based KS-DFT in Table 1. All timings calculations were carried out on 80 CPUs distributed across nodes of dual 10-core E5-2680v2 Ivy Bridge 2.80 GHz processors. Comparison with KS-DFT shows that after eQE is optimized by the concurrent use of subsystem-specific k-point sampling and basis sets, the timing for one SCF cycle drops from 195s to 31s. As expected, a naïve implementation of FDE (*i.e.*, employing a fine k-point grid for all subsystems and not including subsystem-based basis sets) leads to poor timings even in comparison to KS-DFT.

In Table 2 we see that including the additional DIIS layer in eQE, *i.e.*, eQE-DIIS, for subsystem-specific mixing of the electron densities helps reduce the number of SCF cycles required to reach self-consistency in all cases but one (eQE^2, *i.e.* fragment-specific simulation cells and fine k-point grid for every subsystem). The additional eQE-DIIS layer performs best when the density mixing parameter is large (*e.g.*, $\beta_I = 0.7$).

Table 1: Computational timings and speedups of eQE compared to a KS-DFT reference for the C_4H_4S on MoS_2 on hydroxylated Al_2O_3 system depicted in Figure 1.[13] eQE^1: The k-point grids and simulation cells (basis sets) are subsystem-specific, achieving the best performance. eQE^2: Fragment-specific simulation cells (basis sets) but fine k-point grid for all subsystems. eQE^3: Supersystem (large) simulation cell (basis set) for all subsystems but subsystem-specific k-point grids. eQE^4: Supersystem (large) simulation cell (basis set) and fine k-point grid for all subsystems, achieving the slowest performance.

	eQE^1	eQE^2	eQE^3	eQE^4	KS-DFT
Speedup	6.2×	1.7×	1.8×	0.9×	1.0×
Time / SCF cycle (s)	31	113	108	220	195

Table 2: Number of SCF cycles required to reach convergence for the C_4H_4S on MoS_2 on hydroxylated Al_2O_3 system. We compare two values of the β mixing parameter as well as the use of the additional DIIS layer in eQE (eQE-DIIS). The eQE^x methods are described in the caption of Table 1.

	eQE^1	eQE^2	eQE^3	eQE^4
Mixing $\beta = 0.7$				
QE-DIIS	38	34	74	87
eQE-DIIS	**25**	41	**33**	**30**
Mixing $\beta = 0.2$				
QE-DIIS	31	**32**	44	60
eQE-DIIS	61	89	85	105

3.2 Molecular periodic systems (liquids and crystals)

To assess the ability of eQE to model molecular periodic systems, we have considered AIMD simulations of liquid water represented by a periodic box containing 64 and 256 water molecules in the NVT ensemble in a previous study.[54] For these systems, we have demonstrated that eQE reproduces the main features of the structure and dynamics of liquid water at room temperature, provided that an accurate nonadditive kinetic energy functional such as revAPBEK[37] is employed. In general, we obtained good agreement with experiment for the O–O and O–H radial distribution functions, as well as the O–O–O angular distribution function within the first solvation shell. In Ref. 54, we also showed that eQE correctly reproduces the cuts of the potential energy surface of the water dimer. In Ref. 54, we further determined the ability of eQE to reproduce forces from finite differences calculations. In this work, we further analyze the ability of eQE to conserve energy in a NVE dynamics. In the production run of the water 256 dynamics carried out in Ref. 54 only one velocity rescaling event took place. This provided us with the opportunity to estimate energy drift and fluctuations. Linear fitting of the energy between velocity rescaling events yields an energy drift of 1×10^{-4} kcal/mol/ps per water molecule, which corresponds to 5×10^{-3} K/dof/ps, and a standard deviation (fluctuations) of 8×10^{-2} K/dof. Drifts reported by Marx and Hutter on Table 2.6 of Ref. 28 for BO dynamics of water are 3.3×10^{-2} K/dof/ps with a standard deviation (fluctuations) of 1.13×10^{-1} K/dof. Thus, the energy drift of eQE is very low and should be considered satisfactory. The fluctuations are expected to be reduced in eQE compared to Ref. 28 due to the larger number of water molecules considered in the eQE simulation (256 molecules) compared to the benchmark data (32 molecules).

We have also demonstrated that it is possible to carry out AIMD simulations using eQE on systems with constrained spin on selected subsystems (diabatic dynamics). As an example, we studied the $OH^{\bullet}$ radical system solvated in $(H_2O)_{63}$. The spin was constrained to be localized on the radical throughout 20ps of dynamics.[54] Although more extensive testing is required to fully assess the eQE approach in radical systems, our internal tests and this $OH^{\bullet}$ dynamics study point to the ability of eQE to carry out diabatic dynamics with localized spins and charges on a single subsystem. In Figure 4, we compare the spin-density obtained from a semilocal KS-DFT calculation with one obtained from eQE, in which the spin-density was constrained to remain localized on the $OH^{\bullet}$ subsystem.

In terms of time to solution, we present newly obtained computational timings and parallel scaling in Table 3 and Figure 5, respectively, for a system composed of 256 water molecules. The main conclusion here is that eQE performs about four times faster than semilocal KS-DFT for this system when 256 MPI processes are employed and the parallelization scheme is optimized (all calculations are carried out on 256 CPUs distributed across nodes of dual 10-core E5-2680v2 Ivy Bridge 2.80 GHz processors). Figure 5 further shows that eQE is better able to utilize a larger number of employed processes than QE (*i.e.*, a more favorable strong scaling performance), widening the gap with KS-DFT in the large process limit.

The additional DIIS layer, which was useful in bringing down the number of SCF cycles to solution in the periodic system of Figure 1, did not provide an equivalently appreciable speedups for bulk water. For $(H_2O)_{64}$, we have assessed the DIIS on 10 AIMD steps and found that DIIS on average improves (2 SCF steps per MD step, or 20% of the average number of SCF cycles, are saved with DIIS). Instead, for water 256 at the geometry we considered, we found no improvement (11 SCF steps were recorded with or without the new

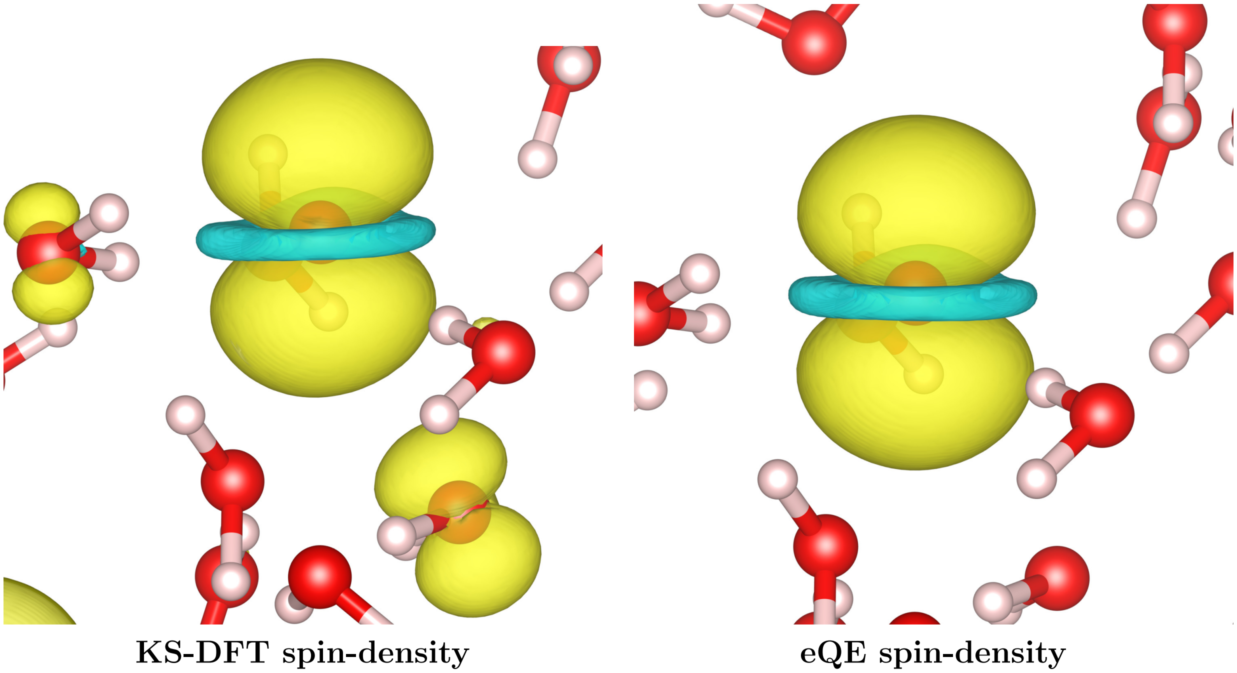


Figure 4: AIMD snapshot of OH$^\bullet$ in $(H_2O)_{63}$ carried out with QE and eQE. The spin-density isosurface plots (cutoff of 10^{-3}) obtained from eQE and standard supersystem-based KS-DFT in QE are compared. Note that eQE produces a (correctly) localized spin-density on the OH$^\bullet$ radical. PBE was used for both QE and eQE, and additionally, the LC94[59] nonadditive kinetic energy functional was employed in eQE.

DIIS layer).

Table 3: Computational timings for an eQE calculation on $(H_2O)_{256}$ as compared to supersystem-based semilocal KS-DFT in QE. All 256 subsystems were computed at the Γ-point. eQE^1: Fully optimized eQE employing subsystem specific basis sets, and calculating the GGA xc and kinetic potentials of the total density only once on the supersystem cell utilizing the `large_comm` communicator (*i.e.*, "fancy parallelization"). eQE^2: GGA xc and kinetic potentials of the supersystem density are computed independently by each subsystem on the small, subsystem-specific cells with the `intra_image` communicator. eQE^3: In addition to eQE^2, the "no fancy parallelization" flag, `-nfp`, is invoked (*i.e.*, no use of `large_comm`). eQE^4: In addition to eQE^3 the supersystem PW basis is employed to expand the KS orbitals of each subsystem, predictably further slowing down the calculation.

	eQE^1	eQE^2	eQE^3	eQE^4	KS-DFT
Speedup	4.1×	2.9×	1.9×	0.1×	1.0×
Time / SCF cycle (s)	115	161	242	4500+	469

In this work, we have also investigated the efficiency of eQE compared to standard KS-DFT in QE for three large-scale condensed-phase systems of interest, each containing over 1000 atoms: (1) MOLECULAR LIQUIDS *via* bulk liquid water represented by 1024 independent water molecules (containing 3072 atoms, see Figure 6(a)), (2) POLYPEPTIDE/BIOMOLECULE SOLVATION *via* $(\mathrm{GLY})_6$ solvated in $(H_2O)_{395}$ (containing 1230 atoms, see Figure 6(b)), and (3) MOLECULAR CRYSTALS *via* a 3×3×3 periodic supercell of pentacene (containing 1940 atoms, see Figure 6(c)).

In Table 4, we report the computational timings for these three large-scale condensed-phase systems. The most important aspect of this data is that **eQE surpasses KS-DFT by more than one-order-of-magnitude for all of these realistic condensed-phase systems**. In addition, we note that different systems feature different levels of speedup. This is because the subsystems have different sizes in these three cases (*e.g.*, pentacene is a large molecule compared to water). Generally speaking, eQE will be less efficient if any of the subsystems approach the size of the supersystem simulation cell.

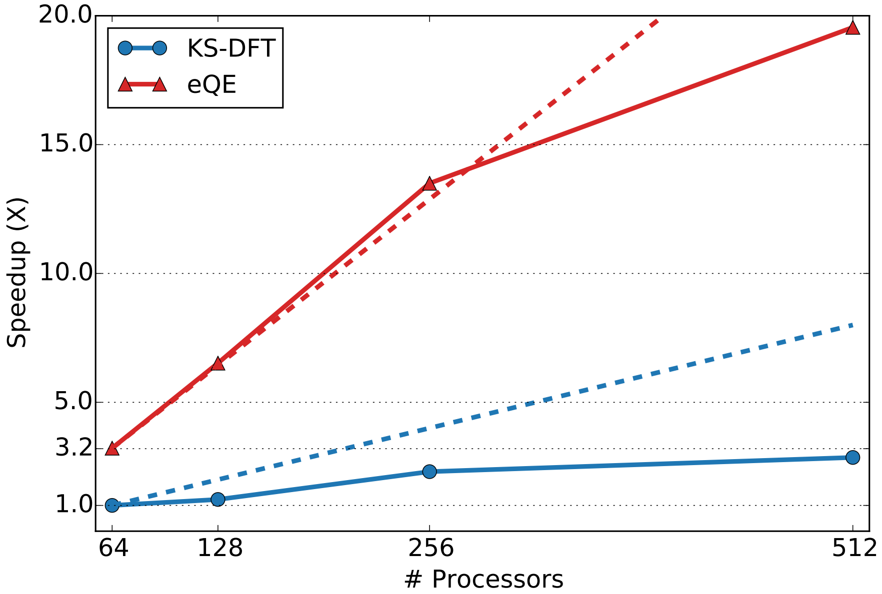


Figure 5: Strong parallel scaling of eQE for a system composed of 256 water molecules as a function of the number of MPI processes employed. The dashed lines represent the ideal speedups for eQE (red) and supersystem KS-DFT (blue). The 1.0× reference corresponds to the supersystem KS-DFT baseline on 64 MPI processes.

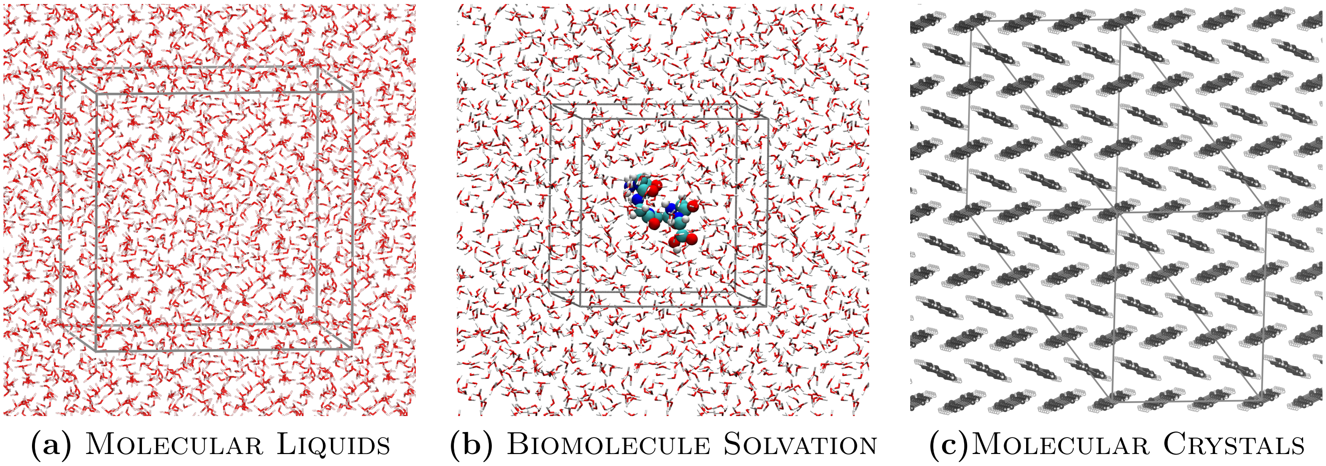

(a) Molecular Liquids **(b)** Biomolecule Solvation **(c)**Molecular Crystals

Figure 6: Graphical depictions (to scale) of three large-scale condensed-phase systems employed to investigate the efficiency of eQE. (a): Bulk liquid water represented by 1024 independent water molecules. (b): $(\textsc{gly})_6$ solvated in $(H_2O)_{395}$ (c): A 3×3×3 periodic supercell of pentacene.

Table 4: Computational timings for eQE calculations on the three large-scale condensed-phase systems in Figure 6. Timings are compared against semilocal KS-DFT of the supersystem and refer to the total CPU time needed to perform 10 steps of AIMD (except for the liquid water system). For $(H_2O)_{1024}$, the speedup was evaluated for the first 10 SCF cycles only (as we were unable to converge the SCF with KS-DFT in this system despite a tremendous amount of effort).

System	# of Atoms	Time eQE^1 (s)	Time KS-DFT (s)	Speedup
$(H_2O)_{1024}$	3072	115	2917	25.3×
$(\textsc{gly})_6$ in $(H_2O)_{395}$	1230	1462	56405	38.6×
Pentacene $3 \times 3 \times 3$	1940	7156	85718	12.0×

4 Conclusions and future directions

In this work, we have introduced a new density embedding code called eQE, which is based on the open-source QE software package. This code is the result of four years of collaborative work, and can now be utilized by the wider chemistry and condensed matter physics communities to perform subsystem DFT based calculations and AIMD simulations on large-scale condensed-phase systems of interest (available at `http://eqe.rutgers.edu`).

During this time, we have tested eQE for a wide variety of different condensed-phase systems of varying sizes and compositions and shown that eQE can be efficiently utilized to treat any system composed of weakly interacting subsystems. Furthermore, eQE can do so as accurately as standard supersystem-based KS-DFT as long as the subsystem electron densities do not strongly overlap.

One of the aims of this work was to introduce code developers to eQE. We have therefore provided them with a description of the salient ideas behind the eQE code, from parallelization strategies to the details of the generation of subsystem-specific basis sets. In addition, we have provided the potential eQE user with compelling computational timings and benchmarks, which undeniably show the superior performance of the hierarchical parallelization framework of eQE when compared to mainstream parallel KS-DFT.

There are many exciting future directions for eQE. On a technical note, we would like eQE to be as general as possible, and thus it should be able to employ any available pseudopotential type. Unfortunately, the current implementation is only general for ultrasoft pseudopotentials (US-PP). Norm-conserving pseudopotentials (NC-PP) can be used, with the stipulation that they include the core density in the pseudopotential file. As a result, we have seldom used NC-PP in the applications presented above. PAW-PP are also not supported yet.

Two limitations of eQE need to be addressed. One is related to the computation of exact exchange, and the other is related to the Hartree potential. As of now, eQE does not support the use of hybrid functionals. However, efficient real-space Poisson solvers are now available[60,61] and can be included in eQE to efficiently compute the Hartree-Fock exchange potential. In addition, the Hartree potential is computed exactly from the total density (*i.e.*, with the `v_h` routine of QE). However, methods based on the Fast Multipole Method (FMM)[62] can be implemented to further decrease computational complexity and improve timings, specifically for simulations of molecular liquids and crystals.

Other, more methodological avenues of improvement are related to the nonadditive exchange–correlation and kinetic energy functionals. For the nonadditive correlation, several avenues have already been explored based on Grimme-like corrections[58] or on RPA-like formalisms[55,63] which will be soon included in eQE. For the kinetic energy, the problem is much more complex and of no straightforward solution. The current eQE release only features LDA and GGA nonadditive kinetic energy functionals. However, several nonlocal kinetic energy functionals have been proposed and are considered to be much more accurate than currently available GGA functionals.[64–68] In the next release of eQE, we will include the possibility of using nonlocal nonadditive kinetic energy functionals for improved accuracy.

Another potential application of eQE is in the description of excited states. Recently, we have made progress in this direction[14,69,70] and expect to release the real-time time-dependent DFT version of eQE in the not too distant future.

In conclusion, the simulations that we have carried out in this work illustrate eQE's improvement over the current state-of-the-art, thereby paving the way to subsystem DFT based AIMD simulations of realistically-sized condensed-phase systems containing 1000s of atoms. As such, we hope that the eQE community of users and developers will continue to grow, enabling us to treat future systems of importance throughout chemistry, physics, and materials science.

Acknowledgments

All authors thank the Quantum ESPRESSO foundation for hosting the eQE project on `qe-forge.org/gf/fde` and for training our group members in the QE codebase. R.D. is grateful for financial support through a startup grant from Cornell University. M.P. gratefully acknowledges the donors of the American Chemical Society Petroleum Research Fund (grant number 54063-DNI6) for partial support of this research and Rutgers University-Newark for startup funds and the continued support of Hackathon workshops. This material is based upon work by supported by the National Science Foundation under Grant No. CHE-1553993 (M.P.). This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-06CH11357.

References

(1) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, 1133–1138.

(2) Burke, K. *J. Chem. Phys.* **2012**, *136*, 150901.

(3) Hohenstein, E. G.; Luehr, N.; Ufimtsev, I. S.; Martínez, T. J. *J. Chem. Phys.* **2015**, *142*, 224103.

(4) Bowler, D. R.; Miyazaki, T. *Rep. Prog. Phys.* **2012**, *75*, 036503.

(5) Mohr, S.; Ratcliff, L. E.; Genovese, L.; Caliste, D.; Boulanger, P.; Goedecker, S.; Deutsch, T. *Phys. Chem. Chem. Phys.* **2015**, *17*, 31360–31370.

(6) Sena, A. M. P.; Miyazaki, T.; Bowler, D. R. *J. Chem. Theory Comput.* **2011**, *7*, 884–889.

(7) Schlipf, M.; Gygi, F. *Comp. Phys. Comm.* **2015**, *196*, 36–44.

(8) Aidas, K. et al. *WIREs: Comput. Mol. Sci.* **2013**, *4*, 269–284.

(9) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. *J. Chem. Phys.* **2005**, *122*, 084119.

(10) VandeVondele, J.; Borštnik, U.; Hutter, J. *J. Chem. Theory Comput.* **2012**, *8*, 3565–3573.

(11) Wesolowski, T. A.; Warshel, A. *J. Chem. Phys.* **1993**, *97*, 8050.

(12) Andermatt, S.; Cha, J.; Schiffmann, F.; VandeVondele, J. *J. Chem. Theory Comput.* **2016**, *12*, 3214–3227.

(13) Genova, A.; Pavanello, M. *J. Phys.: Condens. Matter* **2015**, *27*, 495501.

(14) Krishtal, A.; Ceresoli, D.; Pavanello, M. *J. Chem. Phys.* **2015**, *142*, 154116.

(15) Jacob, C. R.; Neugebauer, J. *WIREs: Comput. Mol. Sci.* **2014**, *4*, 325–362.

(16) Wesolowski, T. A. In *Computational Chemistry: Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific: Singapore, 2006; Vol. 10; pp 1–82.

(17) Wesolowski, T. A.; Shedge, S.; Zhou, X. *Chem. Rev.* **2015**, *115*, 5891–5928.

(18) Genova, A.; Ceresoli, D.; Pavanello, M. *J. Chem. Phys.* **2014**, *141*, 174101.

(19) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; van Gisbergen, S. J. A.; Fonseca Guerra, C.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931–967.

(20) Jacob, C. R.; Neugebauer, J.; Visscher, L. *J. Comput. Chem.* **2008**, *29*, 1011–1018.

(21) Neugebauer, J. *J. Chem. Phys.* **2009**, *131*, 084104.

(22) Neugebauer, J. *J. Chem. Phys.* **2007**, *126*, 134116.

(23) Jacob, C. R.; Jensen, L.; Neugebauer, J.; Visscher, L. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2349–2359.

(24) Jacob, C. R.; Visscher, L. *J. Chem. Phys.* **2006**, *125*, 194104.

(25) Pavanello, M.; Van Voorhis, T.; Visscher, L.; Neugebauer, J. *J. Chem. Phys.* **2013**, *138*, 054101.

(26) Iannuzzi, M.; Kirchner, B.; Hutter, J. *Chem. Phys. Lett.* **2006**, *421*, 16–20.

(27) Luber, S. *J. Chem. Phys.* **2014**, *141*, 234110.

(28) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics*; Cambridge University Press, 2009; http://www.cambridge.org/gb/knowledge/isbn/item2327682/.

(29) Segall, M. D.; Lindan, P. J. D.; Probert, M. J.; Pickard, C. J.; Hasnip, P. J.; Clark, S. J.; Payne, M. C. *J. Phys.: Condens. Matter* **2002**, *14*, 2717–2744.

(30) Govind, N.; Wang, Y. A.; Carter, E. A. *J. Chem. Phys.* **1999**, *110*, 7677–7688.

(31) Huang, P.; Carter, E. A. *J. Chem. Phys.* **2006**, *125*, 084102.

(32) Klüner, T.; Govind, N.; Wang, Y. A.; Carter, E. A. *J. Chem. Phys.* **2002**, *116*, 42–54.

(33) Klüner, T.; Govind, N.; Wang, Y. A.; Carter, E. A. *Phys. Rev. Lett.* **2001**, *86*, 5954–5957.

(34) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(35) Fabiano, E.; Laricchia, S.; Della Sala, F. *J. Chem. Phys.* **2014**, *140*, 114101.

(36) Laricchia, S.; Constantin, L. A.; Fabiano, E.; Della Sala, F. *J. Chem. Theory Comput.* **2014**, *10*, 164–179.

(37) Laricchia, S.; Fabiano, E.; Constantin, L. A.; Della Sala, F. *J. Chem. Theory Comput.* **2011**, *7*, 2439–2451.

(38) Laricchia, S.; Fabiano, E.; Della Sala, F. *J. Chem. Phys.* **2012**, *137*, 014102.

(39) Laricchia, S.; Fabiano, E.; Della Sala, F. *J. Chem. Phys.* **2010**, *133*, 164111.

(40) Chulhai, D. V.; Jensen, L. *J. Chem. Theory Comput.* **2015**, *11*, 3080–3088.

(41) Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, III, T. F. *J. Chem. Theory Comput.* **2012**, *8*, 2564–2568.

(42) Barnes, T. A.; Goodpaster, J. D.; Manby, F. R.; Miller, III, T. F. *J. Chem. Phys.* **2013**, *139*, 024103.

(43) Goodpaster, J. D.; Ananth, N.; ; Manby, F. R.; Miller, III, T. F. *J. Chem. Phys.* **2010**, *133*, 084103.

(44) Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, III, T. F. *J. Chem. Phys.* **2012**, *137*, 224113.

(45) Goodpaster, J. D.; Barnes, T. A.; Miller, III, T. F. *J. Chem. Phys.* **2011**, *134*, 164108.

(46) Jacobson, L. D.; Herbert, J. M. *J. Chem. Phys.* **2011**, *134*, 094118.

(47) Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890–1900.

(48) Guidon, M.; Hutter, J.; VandeVondele, J. *J. Chem. Theory Comput.* **2010**, *6*, 2348–2364.

(49) Nafziger, J.; Wu, Q.; Wasserman, A. *J. Chem. Phys.* **2011**, *135*, 234101.

(50) Gordon, M. S.; Schmidt, M. W. *Theory and Applications of Computational Chemistry*; Elsevier BV, 2005; pp 1167–1189.

(51) Giannozzi, P. et al. *J. Phys.: Cond. Mat.* **2009**, *21*, 395502.

(52) Pulay, P. *Chem. Phys. Lett.* **1980**, *73*, 393–398.

(53) Broyden, C. G. *Mathematics of Computation* **1965**, *19*, 577–577.

(54) Genova, A.; Ceresoli, D.; Pavanello, M. *J. Chem. Phys.* **2016**, *144*, 234105.

(55) Kevorkyants, R.; Eshuis, H.; Pavanello, M. *J. Chem. Phys.* **2014**, *141*, 044127.

(56) Grafova, L.; Pitonak, M.; Rezac, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 2365–2376.

(57) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(58) Beyhan, S. M.; Götz, A. W.; Visscher, L. *J. Chem. Phys.* **2013**, *138*, 094113.

(59) Lembarki, A.; Chermette, H. *Phys. Rev. A* **1994**, *50*, 5328.

(60) Wu, X.; Selloni, A.; Car, R. *Phys. Rev. B* **2009**, *79*.

(61) DiStasio, R. A.; Santra, B.; Li, Z.; Wu, X.; Car, R. *J. Chem. Phys.* **2014**, *141*, 084502.

(62) White, C. A.; Johnson, B. G.; Gill, P. M.; Head-Gordon, M. *Chem. Phys. Lett.* **1996**, *253*, 268–278.

(63) Sinha, D.; Pavanello, M. *J. Chem. Phys.* **2015**, *143*, 084120.

(64) Wang, Y. A.; Govind, N.; Carter, E. A. *Phys. Rev. B* **1999**, *60*, 16350–16358.

(65) Huang, C.; Carter, E. A. *Phys. Rev. B* **2010**, *81*, 045206.

(66) Xia, J.; Carter, E. A. *Phys. Rev. B* **2012**, *86*, 235109.

(67) García-Gonzáles, P.; Alvarellos, J. E.; Chacón, E. *Phys. Rev. A* **1996**, *54*, 1897–1905.

(68) Chai, J.-D.; Weeks, J. D. *Phys. Rev. B* **2007**, *75*, 205122.

(69) Krishtal, A.; Pavanello, M. *J. Chem. Phys.* **2016**, *144*, 124118.

(70) Krishtal, A.; Sinha, D.; Genova, A.; Pavanello, M. *J. Phys.: Condens. Matter* **2015**, *27*, 183202.