Open access • Journal Article • DOI:10.1021/CT1002225

Auxiliary Density Matrix Methods for Hartree-Fock Exchange Calculations.

Manuel Guidon, Jürg Hutter, Joost VandeVondele

Institutions: University of Zurich

Published on: 01 Jul 2010 - Journal of Chemical Theory and Computation (American Chemical Society)

Topics: Fock matrix

Auxiliary density matrix methods for Hartree-Fock exchange calculations

Guidon, M; Hutter, J; VandeVondele, J

Auxiliary Density Matrix Methods for Hartree-Fock exchange calculations

Manuel Guidon, Jürg Hutter, and Joost VandeVondele*

Physical Chemistry Institute, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Abstract

The calculation of Hartree-Fock exchange (HFX) is computationally demanding for large systems described with high quality basis sets. In this work, we show that excellent performance and good accuracy can nevertheless be obtained if an auxiliary density matrix is employed for the HFX calculation. Several schemes to derive an auxiliary density matrix from a high quality density matrix are discussed. Key to the accuracy of the auxiliary density matrix methods (ADMM) is the use of a correction based on standard generalized gradient approximations for HFX. ADMM integrates seamlessly in existing HFX codes, and in particular can be employed in linear scaling implementations. Demonstrating the performance of the method, the effect of HFX on the structure of liquid water is investigated in detail using Born-Oppenheimer molecular dynamics simulations (300 ps) of a system of 64 molecules. Representative for large systems are calculations on a solvated protein (Rubredoxin), for which ADMM outperforms the corresponding standard HFX implementation by approximately a factor 20.

1 Introduction

The success of density functional theory (DFT) can be attributed to the fact that it can provide an accurate description of the electronic structure at a moderate computational cost. DFT has become a unique tool to describe systems containing hundreds to thousands of atoms. Not only is it possible to describe molecules in the gas phase, properties of condensed phase systems such as liquids and solids can also be computed. For these systems, using contemporary computer resources, it has become possible to go beyond a static description of matter, and finite temperature effects can be included directly through ab initio molecular dynamics (MD) simulations. Large scale, condensed phase and dynamical simulations have mostly adopted a relatively simple form for the exchange and correlation functional, namely the semi-local generalized gradient approximation (GGA). However, it becomes increasingly clear that an improved description of the electronic structure, and thus more accurate results, can only be obtained by functionals that go beyond the GGA form, and incorporate a non-local term such as Hartree-Fock exchange (HFX). The computational cost of these non-local terms is typically much larger than that of the local terms. Consequently, there is significant interest in finding efficient approaches to deal with these non-local forms.

The efficiency of a HFX calculation depends strongly on the algorithm employed. A straightforward implementation based on localized basis sets scales with the fourth power of the system size. However, integral screening[1] reduces the scaling with system size to quadratic, and for short range operators, such as screened[2] or truncated exchange,[3–6] to linear scaling. Non-metallic systems furthermore allow for a screening on the density matrix,[7] which leads to linear scaling also for long-range operators. Using these techniques, HFX can be evaluated also for condensed phase systems containing a few thousand atoms,[3] and can be used to perform ab initio molecular dynamics simulations.[8] Despite the favorable scaling with system size, HFX calculations scale very poorly with basis set quality. This is an important issue, since high quality results not only require accurate functionals, but also good basis sets. There are several reasons why the cost of HFX depends very strongly on the basis employed. Indeed, even in a linear scaling code, the cost increases

with the fourth power of the number of (primitive) basis functions per atom. Basis sets with a high l-quantum number (polarization functions) are therefore costly, as the number of basis functions per atom grows quadratically with l. Heavily contracted basis functions, such as the molecularly optimized basis sets proposed in Ref. 9 , are expensive since for each quartet of basis functions a very larger number of primitive integrals needs to be considered. Very flexible basis sets, or basis sets with diffuse primitives are costly for several reasons. First, diffuse primitives are non-zero in a larger part of space, and thus screening becomes less efficient. This is particularly important in condensed phase systems, where periodic boundary conditions provide a potentially unlimited number of interacting atomic sites. Second, uncontracted diffuse primitives influence the condition number of the overlap matrix (S) strongly, and a poor condition number in turn implies that a tighter screening threshold has to be employed[3] to obtain a stable self consistent calculation. Third, the sparsity of the matrix representation of the density matrix (P) also depends strongly on the condition number of S, making density matrix screening less efficient for poorly conditioned basis sets. Table 1 illustrates this problem by providing costs and maximal thresholds needed in order to get converged results for a water cluster containing 20 water molecules. Clearly, a technique which reduces the impact of the basis set on the computational cost is a significant progress. During the last decade, much effort has been invested into solving this problem and many different techniques have been proposed. Among them are methods that apply an approximate resolution of identity, for example RI[18] or Cholesky decomposition.[19] These schemes rely on the introduction of auxiliary basis functions in terms of which the four center integrals can be approximated by corresponding two- and three-center terms. In order to improve efficiency, Sodt. et al.[20] developed a local variant of RI, atomic resolution of identity (ARI). A slightly different post-Hartree-Fock approach in a dual basis was introduced in Ref. 21 where a reference calculation in a small basis set is perturbatively corrected to a large basis set. A different approximation for the two-electron integrals has been proposed by Friesner and co-workers[22] and has been termed "pseudo-spectral" method. Recently, Neese et al.[23] presented an algorithm called COSX that is a combination of semi-numerical methods and RI. Furthermore, there exist several schemes to achieve linear scal-

Table 1: Impact of the basis set quality for the wavefunction optimization of a 20 water cluster. The condition number $\kappa(S)$ of the overlap matrix determines the maximal possible screening threshold. The latter needs to be chosen more tightly, if $\kappa(S)$ gets large. This is reflected in the cost of a calculation, which is given once by the number of Cartesian four-center electron repulsion integrals (ERIs) that need to be evaluated and the time in seconds that is spent in building the Fock matrix in the first self consistent field (SCF) step. 3-21G*, 6-31G**, 6-311G++G** refer to basis sets by Pople and co-workers,[10–13] the polarization consistent (pc) basis sets have been developed by Jenssen,[14–16] the def2-QZVP basis by Ahlrichs and co-workers.[17] Timings are obtained on 128 cores of a CRAY-XT5.

basis	$\kappa(S)$	threshold	cost [ERIs]	cost [s]
3-21G*	4.9E+01	1.0E-04	2.3E+07	0.06
6-31G**	2.1E+02	1.0E-05	5.2E+08	0.35
6-311G++G**	1.2E+05	1.0E-07	1.1E+10	11.71
pc-0	5.2E+01	1.0E-04	1.7E+07	0.07
pc-1	4.5E+03	1.0E-05	4.4E+08	0.50
pc-2	5.7E+05	1.0E-07	2.0E+10	11.21
aug-pc-1	1.4E+06	1.0E-08	5.0E+10	53.23
aug-pc-2	3.9E+08	1.0E-09	1.5E+12	766.92
def2-QZVP	7.1E+04	1.0E-08	3.2E+11	127.16
aug-def2-QZVP	8.5E+05	1.0E-08	6.2E+11	331.61

ing in the context of plane wave basis sets such as the multiwavelet based ansatz of Harrison et al.[24] or FFT based algorithms as presented in Refs. 25–27.

In this work, we propose to employ an auxiliary density matrix to evaluate the expensive non-local part of the functional, while all other energy components are computed with the primary (original) density matrix. The auxiliary density matrix will be constructed in a way that allows for a rapid evaluation of the HFX energy, using any algorithm, including traditional or linear scaling approaches. In order to ensure that the quality of the calculation is influenced as little as possible by the quality of the auxiliary density matrix, a correction term is added to the exchange and correlation functional. Based on a GGA for exchange, this correction takes the difference between auxiliary and primary density matrix into account. All terms of the resulting density functional are straightforward to compute, but there is considerable freedom in how to obtain from a given primary density matrix a suitable auxiliary density matrix. In this paper, various procedures are

discussed and tested. Tests are presented in Sec. 3 and include gas phase thermochemistry, basis set superposition error, electronic structure including band gaps, large systems, and liquid water. The theory is introduced in the following section, but for mathematical derivations and technical details we refer to the appendices.

2 Theory

2.1 Basic Concepts

In Kohn-Sham DFT, the total energy of a system consisting of N_e electrons can be written in terms of the electron density

$$\rho(\mathbf{r}) = \sum_{i=1}^{N_\mathrm{e}} |\psi_i(\mathbf{r})|^2, \tag{1}$$

where ψ_i denote the single particle wavefunctions, which are assumed to be real valued. The total energy is then expressed in terms of a functional of the electron density as

$$E[\rho] = T_s[\rho] + J[\rho] + E_\mathrm{xc}[\rho] + \int v(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}, \tag{2}$$

with the standard abbreviations for kinetic, Hartree and exchange-correlation energy and the part due to the external potential. In hybrid DFT, the exchange-correlation functional is augmented by a certain fraction of Hartree-Fock exchange based on the wavefunctions $\{\psi_i\}$

$$E_\mathrm{xc}[\rho] = \alpha E_x^\mathrm{HFX}[\{\psi_i\}] + (1-\alpha)E_x^\mathrm{DFT}[\rho] + E_c^\mathrm{DFT}[\rho], \tag{3}$$

where α denotes the fraction of HFX and E_x and E_c are the density functionals for exchange and correlation, respectively. In the presence of an atomic centered basis set $\{\phi_\mu(\mathbf{r})\}$

$$\psi_i(\mathbf{r}) = \sum_\mu C^{\mu i}\phi_\mu(\mathbf{r}), \tag{4}$$

the Hartree-Fock exchange energy can be expressed in terms of a density matrix and two-electron integrals (ERIs)

$$E_x^{\mathrm{HFX}}[P] = -\frac{1}{2}\sum_{\lambda\sigma\mu\nu} P^{\mu\sigma}P^{\nu\lambda}\,(\mu\nu|\lambda\sigma)\,, \tag{5}$$

where the density matrix elements $P^{\mu\nu}$ are obtained from the molecular (MO) coefficients as

$$P^{\mu\nu} = \sum_i C^{\mu i}C^{\nu i} \Leftrightarrow P = CC^T \tag{6}$$

and the ERIs are defined as

$$(\mu\nu|\lambda\sigma) = \int\int \phi_\mu(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)g(|\mathbf{r}_2-\mathbf{r}_1|)\phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)d\mathbf{r}_1\mathbf{r}_2, \tag{7}$$

with the interaction potential $g(r)$ that is Coulombic $(1/r)$ in standard Hartree-Fock theory. The fourth order scaling of HFX with basis set size can be directly inferred from Eq. 5.

By introducing an auxiliary density matrix $\hat{P} \approx P$ that is either smaller in size or more rapidly decaying than the original one, the evaluation of HFX can be sped up significantly. The HFX energy can be written as

$$\begin{aligned} E_x^{\mathrm{HFX}}[P] &= E_x^{\mathrm{HFX}}[\hat{P}] + \left(E_x^{\mathrm{HFX}}[P] - E_x^{\mathrm{HFX}}[\hat{P}]\right) \\ &\approx E_x^{\mathrm{HFX}}[\hat{P}] + \left(E_x^{\mathrm{DFT}}[P] - E_x^{\mathrm{DFT}}[\hat{P}]\right). \end{aligned} \tag{8}$$

The assumption behind this approximation is that *the difference* in the exchange energy between primary and auxiliary density matrix is well captured by a GGA, even in those cases where GGA exchange and HFX might be qualitatively different. Eq. 8 amounts to computing the HFX energy with an auxiliary density matrix, while a GGA correction is introduced which takes the difference between auxiliary and primary density matrix into account. As shown in Sec. 3, applying this correction indeed improves upon uncorrected results. Clearly, our approach yields the original HFX energy as either the quality of the auxiliary density matrix or of the correcting functional

improves. In this work, we have based the GGA correction on PBE exchange,[28,29] and have not explored other parameterizations or other functionals. The introduction of Eq. 8 in hybrid density functionals is natural and straightforward, and usually, because only a fraction of exchange is needed, will introduce a smaller error. If hybrid functionals employ a non-Coulombic operator the exchange functional needs to be chosen consistently with the shape of the interaction potential ($g(r)$) in the ERI calculation. Currently, the GGA correction for the standard Coulomb potential, the short range (erfc) and the truncated Coulomb potential have been implemented and tested.

2.2 Auxiliary density matrices

The performance and accuracy of the ADMM scheme depends on how the auxiliary density matrix is constructed, and various approaches seem possible. In this section, we present methods that either rely on the use of an auxiliary basis set, or directly manipulate the sparsity of the density matrix.

The size of P obtained from a high quality primary basis set (PBS) $\{\phi_\mu(\mathbf{r})\}$ can be reduced by introducing an auxiliary basis set (ABS) $\{\hat{\phi}_\mu(\mathbf{r})\}$ for the description of the underlying wavefunction

$$\hat{\psi}_i(\mathbf{r}) = \sum_\mu \hat{C}^{\mu i}\hat{\phi}_\mu(\mathbf{r}), \tag{9}$$

i.e.

$$\hat{P}^{\mu\nu} = \sum_i \hat{C}^{\mu i}\hat{C}^{\nu i} \Leftrightarrow \hat{C}\hat{C}^T. \tag{10}$$

An optimal value for the MO coefficients can be obtained by requiring that the square difference for the occupied wavefunctions in ABS and PBS representation is minimized

$$\min_{\hat{C}} \sum_j \int \left(\psi_j(\mathbf{r}) - \hat{\psi}_j(\mathbf{r})\right)^2 d\mathbf{r}. \tag{11}$$

This yields the following expression for the auxiliary MO coefficients

$$\hat{C} = AC, \tag{12}$$

where A is defined as the projector between the two basis sets

$$A = \hat{S}^{-1} Q \tag{13}$$

with the overlap matrices

$$\hat{S}_{nn'} = \int \hat{\phi}_n(\mathbf{r}) \hat{\phi}_{n'}(\mathbf{r}) d\mathbf{r} \quad \text{and} \quad Q_{nm} = \int \hat{\phi}_n(\mathbf{r}) \phi_m(\mathbf{r}) d\mathbf{r}. \tag{14}$$

A slightly more complicated formula is obtained when the auxiliary wavefunctions are required to minimize Eq. 11 subject to the constraint that they remain orthonormal. This constraint can be enforced introducing Lagrangian multipliers (Λ_{kl}) in Eq. 11 as

$$\min_{\tilde{C}} \left[\sum_j \int \left(\psi_j(\mathbf{r}) - \tilde{\psi}_j(\mathbf{r}) \right)^2 d\mathbf{r} + \sum_{k,l} \Lambda_{kl} \left(\int \tilde{\psi}_k(\mathbf{r}) \tilde{\psi}_l(\mathbf{r}) d\mathbf{r} - \delta_{kl} \right) \right]. \tag{15}$$

The coefficients ($\tilde{C}$) that minimize this expression can be obtained as

$$\tilde{C} = \hat{C} \Lambda^{-1/2} \quad \text{with} \quad \Lambda = \hat{C}^T \hat{S} \hat{C}. \tag{16}$$

where $\hat{C}$ is defined by Eq. 12.

Of course, there is significant freedom in selecting the auxiliary basis set, and the choice need not to be homogenous in space. For example, for large systems with a chemically active region, such as enzymes, it is natural to retain the high quality primary basis where exchange matters most, while a lower quality auxiliary basis can be used for the bulk. Furthermore, note that the explicit shape of the basis functions (Gaussian functions, Slater functions, ...) is not important, and indeed need not to be the same in the auxiliary and primary basis sets. The method thus provides

an interesting approach for computing exchange contributions in programs that do not employ Gaussian basis functions and for which the calculation of exchange is relatively difficult.

To the two different sets of MO coefficients, $\hat{C}$ and $\tilde{C}$ correspond two different density matrices:

$$\tilde{P} = \tilde{C}\tilde{C}^T = \hat{C}\Lambda^{-1}\hat{C}^T, \tag{17}$$

and

$$\hat{P} = \hat{C}\hat{C}^T = APA^T \tag{18}$$

that can be used as an auxiliary density matrix. We will refer to the first as purified wavefunction fitting or ADMM1 and to the second as non-purified wavefunction fitting or ADMM2 (see Sec. 2.3 for explanation of nomenclature).

A strategy directly aimed at obtaining a sparse auxiliary density matrix relies on a blocking of the primary density matrix. This strategy is applicable if the system of interest can be divided into subsystems that have no important exchange interactions beyond what is captured with a GGA. In this case, the non-relevant blocks in the auxiliary density matrix can just be zeroed, and to some extend this method can be considered a subsystem based neglect of diatomic differential overlap (NDDO) If all inter-subsystem blocks are zeroed the approximate density matrix will be positive definite, however, we employ the slightly generalized form of the auxiliary density matrix as

$$\hat{P} = P \otimes B \tag{19}$$

where B is a blocking matrix with $B_{ij} \in \{1,0\}$ and $\otimes$ denotes the Hadamard product of two matrices. In this case, B can reflect the molecular topology, and allow for connections between subsystems. In the general case, the resulting $\hat{P}$ need not to be positive definite. This method of obtaining an auxiliary density matrix will be referred to as blocking or ADMM3 in the following.

2.3 Density matrix purification

As already mentioned, an approximate density matrix might not fulfill the properties of a pure density matrix:

$$P = P^T, \tag{20}$$

$$PSPS = PS, \tag{21}$$

$$\mathrm{tr}(PS) = N_e, \tag{22}$$

i.e. symmetry, idem-potency and particle conservation. For the three approximations mentioned in the previous section, all three conditions are only fulfilled by ADMM1, i.e. the purified wave-function fitting scheme. ADMM2 and the block diagonal version of ADMM3 fulfill a property of ensemble averaged (finite temperature) density matrices i.e. that the eigenvalues of $\hat{P}$ are bounded by 0 and 1, which is a relaxed version of the idempotency condition. In order to compute a GGA correction for exchange, it is essential that the approximate density matrix is at least positive semi-definite. Fortunately, there exist purification algorithms that can restore the idempotency of an approximate density matrix.

Well known is the McWeeny purification algorithm[30] which, in the presence of an overlap matrix, is defined as follows

$$\bar{P}_{n+1} = f(\bar{P}_n) = 3\bar{P}_n S\bar{P}_n - 2\bar{P}_n S\bar{P}_n S\bar{P}_n \tag{23}$$

for an initial guess $\bar{P}_0 = \hat{P}$. The pure density matrix is than given as

$$\tilde{P} = \lim_{n\to\infty} \bar{P}_n. \tag{24}$$

An interesting property of this algorithm is that it can be implemented in a linear scaling fashion.[31,32] In the current context, we prefer an extension of the McWeeny procedure based on a

Cauchy integral representation[33]

$$\tilde{P} = S^{-1}\left[\frac{1}{2\pi i}\oint \frac{\Theta(z-0.5)}{S^{-1}z-\hat{P}}dz\right]S^{-1}, \tag{25}$$

where $\Theta(z)$ denotes the Heaviside function. This scheme yields a pure density matrix for all input matrices, is non-iterative, but is not easily incorporated in a linear scaling procedure. Through Eq. 25 a purified $\tilde{P}$ can be interpreted as a matrix functional of a non-pure $\hat{P}$. This is an important property, which will be used to derive an expression for the Kohn-Sham matrix in the following section. Eq. 25 can be easily computed using basic linear algebra techniques as

$$\tilde{P} = S^{-1}RLR^{T}S^{-1}, \tag{26}$$

where R is the matrix of eigenvectors of the generalized eigenvalue problem

$$\hat{P}R = S^{-1}R\lambda \tag{27}$$

and L is the diagonal matrix $L_{ii} = \Theta(\lambda_i - 0.5)$ with the corresponding eigenvalues λ_i. At this point, and as shown in the appendix, we remark that purification by Eq. 25 of the density matrix obtained from non-purified wavefunction fitting (ADMM2) exactly yields the density matrix derived from the purified wavefunction fitting (ADMM1).

2.4 Kohn-Sham matrix and the SCF procedure

In a standard SCF procedure, an improved density matrix is obtained from a diagonalization of the Kohn-Sham matrix. The Kohn-Sham matrix itself is obtained as the derivative of the total energy with respect to the density matrix. In ADMM, the total energy can be considered to consist of two parts, one part depending explicitly on the primary density matrix ($E[P]$), and one part depending

explicitly on the auxiliary matrix ($\tilde{E}[\tilde{P}]$):

$$E_{\text{total}} = E[P] + \tilde{E}[\tilde{P}]. \tag{28}$$

The Kohn-Sham matrix associated with this expression

$$K_{\text{total}} = \frac{dE[P]}{dP} + \frac{d\tilde{E}[\tilde{P}]}{dP} = K + \frac{d\tilde{E}[\tilde{P}]}{dP}, \tag{29}$$

contains one non-trivial term

$$\frac{d\tilde{E}[\tilde{P}]}{dP} = \frac{d\tilde{E}}{dP} = \frac{d\tilde{E}}{d\tilde{P}}\frac{d\tilde{P}}{d\hat{P}}\frac{d\hat{P}}{dP} = \tilde{K}\frac{d\tilde{P}}{d\hat{P}}\frac{d\hat{P}}{dP}, \tag{30}$$

where $\tilde{K}$ is the Kohn-Sham matrix constructed from the purified density matrix. $\frac{d\hat{P}}{dP}$ is readily evaluated for wavefunction fitting and blocking, while $\frac{d\tilde{P}}{d\hat{P}}$ can be obtained through the Cauchy integral Eq. 25. We obtain in the case of purified wavefunction fitting (for details see App. B)

$$\frac{d\tilde{E}}{dP} = A^T R \left[\left(R^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} R \right) \otimes M \right] R^T A, \tag{31}$$

with R as defined above, and

$$M_{kj} = \begin{cases} \frac{\Theta(\lambda_k - 0.5) - \Theta(\lambda_j - 0.5)}{\lambda_k - \lambda_j}, & k \neq j \\ \delta(\lambda_k - 0.5), & k = j \end{cases}. \tag{32}$$

In the appendix, computationally more efficient expressions are presented for optimization schemes that only require the derivative of the energy with respect to the MO coefficients ($\frac{dE}{dC}$) or that exploit the special structure of $\hat{P}$.

At this point, it is important to point out that the eigenvalues of the Kohn-Sham matrix in ADMM might be very different from the eigenvalues of the Kohn-Sham matrix in the primary basis. This is not an indication of the inaccuracy of the scheme, nor is it a problem for the SCF

procedure, but is related to the fact that purification as part of the energy functional partially accounts for the orthonormality constraint of the wavefunction (see also App. G). In order to use the eigenvalues of the ADMM Kohn-Sham matrix directly as orbital energies, e.g. to calculate the band gaps of a system, an ADMM scheme without purification needs to be employed. For the non-purified wavefunction fitting (ADMM2), the corresponding Kohn-Sham matrix is given by

$$K_{\text{total}} = K[P] + A^T \hat{K} A, \tag{33}$$

where $\hat{K}$ is build from $\hat{P}$. This simple expression suggests an expression for use with purified wavefunction fitting (ADMM1), i.e. orbital energies can be obtained from eigenvalues of

$$K_{\text{total}} = K[P] + A^T \tilde{K} A, \tag{34}$$

where $\tilde{K}$ is constructed from $\tilde{P}$. We will show in Sec. 3.6 that this expression can be accurate.

3 Assessment and validation of the method

3.1 Computational details

All algorithms have been implemented in CP2K,[34] a freely available molecular simulation package. CP2K is well suited for these calculations as the density functional module Quickstep,[35] implements a linear scaling and fast scheme for calculations based on local functionals. Indeed, the Gaussian and plane waves (GPW) scheme[36] and its augmented (GAPW) variant[37] provide an efficient method to evaluate the Coulomb energy for pseudopotential and all-electron calculations, respectively. These approaches use Fourier transform based techniques, i.e. a plane wave auxiliary basis, and scale favorably with basis set size. Recently, an efficient, massively parallel and linear scaling implementation of Hartree-Fock exchange has been incorporated into the CP2K code.[3,8] Despite this efficiency, calculations including HFX and employing high quality basis sets, are at

least one order of magnitude more expensive than calculations based on GGAs. ADMM aims at resolving this issue. Currently, the GGA correction term required for ADMM has only been implemented for use with the GPW method, and consequently all calculations are based on Goedecker, Teter, Hutter (GTH) pseudopotentials.[38] Pseudopotentials[39] constructed for the PBE functional have been used throughout. This is an approximation that poorly describes core-valence exchange and that is known to introduce errors in excess of 0.1eV in the computation of band gaps,[40,41] but appears to give reasonable results for ground state properties (see e.g. 3.2). In this work, both ADMM and the standard HFX implementation employ the same pseudopotential approximation, so that a meaningful comparison can be made. The all-electron implementation of ADMM and the development of pseudopotentials for hybrid functionals are beyond the scope of the current work.

Table 2: Cost for using the FIT3 basis sets on a cluster of 20 water molecules. For comparison and details, see Table 1.

basis	$\kappa(S)$	threshold	cost [ERIs]	cost [s]
cFIT3	1.3E+02	1.0E-04	1.8E+07	0.08
FIT3	1.5E+02	1.0E-04	1.6E+07	0.10
cpFIT3	1.7E+02	1.0E-04	1.0E+08	0.13
pFIT3	2.3E+02	1.0E-04	9.0E+07	0.15
aug-cFIT3	5.5E+04	1.0E-07	1.1E+09	3.38
aug-FIT3	6.1E+04	1.0E-07	1.1E+09	4.14
aug-cpFIT3	5.9E+04	1.0E-07	3.2E+09	6.09
aug-pFIT3	6.4E+04	1.0E-07	3.0E+09	6.78

Calculations based on pseudopotentials use split valence Gaussian basis sets as discussed in Ref. 35, the fully contracted molecularly optimized (MOLOPT) basis sets discussed in Ref. 9, or a reference basis (GTH-def2-QZVP), which combines the pseudo-atomic orbitals of the MOLOPT basis, with uncontracted valence, and polarization exponents of the Ahlrichs quadruple-ζ (aug-)def2-QZVP[17] basis set. The latter basis can be considered close to the basis set limit. The choice of auxiliary basis for the ADMM method will in general be dictated by accuracy and performance requirements of a particular calculation. Indeed, the gain in performance for the hybrid part of the calculation might allow for better primary basis sets, or large systems can be simulated by more

aggressively using a smaller auxiliary basis. Here, we are interested in exploring the accuracy of relatively small auxiliary basis sets, of which a library of eight different basis sets per atom have been constructed. This basis employs three Gaussian exponents for the valence orbitals, optimized in atomic calculations. We will refer to this uncontracted basis, without polarization functions, as FIT3, while a contraction of this basis (to double zeta quality) is referred to as cFIT3. In order to improve accuracy, polarization functions from the standard 6-31G** basis sets have been added yielding pFIT3 and cpFIT3 basis sets. Finally, an augmented version has been constructed by adding a 'diffuse' function (typical exponents are 0.03 for hydrogen and 0.09 for oxygen), yielding aug-FIT3, aug-cFIT3, aug-pFIT3, aug-cpFIT3.

3.2 GMTKN24 database

Table 3: Shown are WTMAD and $\text{WTMAD}_{\text{ref}}$ in kcal/mol for the GMTKN24 database and the PBE0 functional. Whereas WTMAD refers to the weighted mean absolute deviation with respect to experimental and theoretical benchmark results, $\text{WTMAD}_{\text{ref}}$ refers to deviations with respect to PBE0 reference results obtained using a standard HFX implementation and the high quality GTH-def2-QZVP basis. Standard (STD) HFX calculations with the GTH-def2-QZVP, FIT3, and pFIT3 basis sets are employed to establish the quality of these basis sets as a primary basis set (PBS). Wavefunction fitting results with purification (ADMM1) and without purification (ADMM2) are provided using four different auxiliary basis sets (ABS), while the GTH-def2-QZVP has been employed as a primary basis in all these cases. The results in parenthesis have been obtained using ADMM, but ignoring the GGA correction.

Method	PBS	ABS	WTMAD	$\text{WTMAD}_{\text{ref}}$
STD	GTH-def2-QZVP	-	5.0	0.0
	FIT3	-	15.3	10.8
	pFIT3	-	7.1	4.2
ADMM1	GTH-def2-QZVP	cFIT3	5.3	1.0
	GTH-def2-QZVP	FIT3	5.3 (6.1)	0.7 (1.8)
	GTH-def2-QZVP	cpFIT3	5.0	0.7
	GTH-def2-QZVP	pFIT3	4.9 (5.5)	0.5 (1.2)
ADMM2	GTH-def2-QZVP	cFIT3	5.3	1.1
	GTH-def2-QZVP	FIT3	5.3	0.8
	GTH-def2-QZVP	cpFIT3	4.9	0.7
	GTH-def2-QZVP	pFIT3	4.9	0.5

The GMTKN24 database is a compilation of 24 different chemically relevant benchmarks collected and established by L. Goerigk and S. Grimme.[42,43] It is based on 1049 atomic and molecular single point energies that are combined to yield 731 relative energies. These energies can be compared to available benchmark data, derived from either theory or experiment. In order to judge the quality of a computational method using a single number, the authors defined a weighted total mean absolute deviation (WTMAD) that combines all mean absolute deviations (MADs). This convenient measure is adopted here to judge the quality of the wavefunction fitting methods ADMM1 and ADMM2 for various basis sets. Results, summarized in Table 3, are based on the hybrid PBE0 functional[44–46] without empirical dispersion correction.[47] In a first step, reference results using a standard HFX implementation have been generated for the GTH-def2-QZVP basis. As in Ref. 42, an augmented basis set has been used for two of the subsets. ADMM results can be directly compared to these reference results, and deviations with respect to this data is referred to as $\mathrm{WTMAD}_{\mathrm{ref}}$. WTMAD without subscript is used to refer to the deviations with respect to the experimental and theoretical benchmark results. Secondly, to quantify the expected poor quality of the FIT3 family as a primary basis, these basis sets have been used with a standard HFX implementation. These calculations yield a $\mathrm{WTMAD}_{\mathrm{ref}}$ in the range 4–11 kcal/mol, and WTMADs in the range 7–15kcal/mol, far worse than the typical performance of local functionals, with a good basis set, on this database.[42] Thirdly, ADMM calculations have been performed using the FIT3 family as auxiliary basis sets. Whenever the primary basis is augmented, an augmented auxiliary basis has been used as well. The results obtained with ADMM are in very close agreement with the reference calculations. In particular, both ADMM1 and ADMM2 using the better auxiliary basis set (pFIT3 or cpFIT3) are basically indistinguishable in terms of error with respect to the benchmark data (WTMAD), and have an error of less than 1 kcal/mol compared to the reference run ($\mathrm{WTMAD}_{\mathrm{ref}}$). In the case of FIT3, ADMM results improve by 10 kcal/mol as compared to standard HFX calculations with the same basis. In Table 3 it is also shown that including the GGA correction term in ADMM more than halves the $\mathrm{WTMAD}_{\mathrm{ref}}$, thus emphasizing the benefit of the correction term. Finally, we observe that ADMM1 and ADMM2 perform equally well, suggesting

that in this case the purification is not essential. In the cases we have verified, $\hat{P}$ had eigenvalues close to 0 and 1, even for the small cFIT3 basis. This data shows that results of def2-QZVP quality can be obtained at a cost similar to 6-31G**. The relatively modest cost of computing the full database with ADMM has been exploited to benchmark the quality of the the PBE0-TC-LRC functional proposed in Ref. 3. This functional uses a truncated operator for the calculation of exchange, but, like HSE,[48,49] corrects for the long range part using a density functional. As shown in Ref. 3, the PBE0-TC-LRC is useful in the condensed phase, but can also reduce the computational cost for (large) molecules. In Table 4 the effect of varying the range of exchange has been studied systematically, using ADMM1 with the pFIT3 basis, for PBE0-TC-LRC functionals including 20% and 25% of non-local exchange.

Table 4: Shown are WTMADs in kcal/mol for the GTMKN24 database and the PBE0_TC_LRC functional for several different cutoff radii in the range of 0.5 to 6.0 Å. The column denoted with ∞ refers to the standard PBE0 hybrid functional. All calculations have been performed twice for different fractions of Hartree-Fock exchange $\alpha = 0.2$ and $\alpha = 0.25$.

α	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	3.00	4.50	6.00	∞
0.20	5.5	5.3	5.6	5.6	5.3	5.0	4.9	4.8	4.7	4.7	4.8	4.8	N/A
0.25	5.5	5.5	6.1	6.3	5.9	5.4	5.2	5.0	4.9	4.9	4.9	4.9	5.0

These WTMADs clearly show that the range of the truncated operator can be reduced to 2Å without affecting the quality of the results. The lowest WTMAD, slightly smaller than the WTMAD for PBE0, is found for 20% non-local exchange and a range of 2.5Å.

3.3 Basis Set Superposition Error

In this section, the impact of ADMM on the basis set superposition error (BSSE) for the water dimer is investigated. Indeed, the BSSE is a concern as soon as small, lower quality basis sets are employed. Here, it is shown that small auxiliary basis sets introduce only a moderate BSSE, especially if compared to the BSSE in standard HFX calculations with the same basis. In order to quantify the BSSE, the counterpoise correction[50] has been computed for a water dimer at

Table 5: Shown are counterpoise corrections in kcal/mol to the PBE0 binding energy of a water dimer. STD refers to traditional hybrid calculations, using the shown basis set as primary basis set. ADMM1 and ADMM2 refer to the wavefunction fitting methods, using the TZV2P-MOLOPT basis set as primary basis and the shown basis as auxiliary basis set.

Basis set	STD	ADMM1	ADMM2
cFIT3	-3.112	0.771	0.223
FIT3	-3.128	0.520	-0.006
cpFIT3	-3.468	0.882	0.248
pFIT3	-3.448	0.604	0.004
aug-cFIT3	-1.889	-0.193	-0.346
aug-FIT3	-1.744	-0.095	-0.253
aug-cpFIT3	-1.023	-0.246	-0.325
aug-pFIT3	-1.005	-0.162	-0.247
TZV2P-MOLOPT	-0.123	-0.123	-0.123

fixed equilibrium geometry, for various methods. These results are summarized in Table 5. As expected, using the non-augmented FIT3 basis sets as primary basis leads to errors of approximately 3 kcal/mol. This error is large when compared to a basis using diffuse primitives, such as the TZV2P-MOLOPT basis, which has a BSSE of only 0.1 kcal/mol. However, within the ADMM scheme, the error reduces to 0.8 kcal/mol, approximately a four-fold reduction. Using the augmented auxiliary basis sets reduces the error to approximately 0.3 kcal/mol, similar to, but not quite as good as, the quality of the primary basis set. Note that, since an auxiliary basis set method is not necessarily variational in the auxiliary basis, the counterpoise corrections can be of both signs. This can lead to an error cancellation, which is presumably the reason why ADMM2 performs surprisingly well with the lower quality auxiliary basis sets. It can thus be concluded that both wavefunction fitting methods do not suffer from the large BSSE associated with the inferior quality of the auxiliary basis even though the BSSE does not reduce to the extent of the primary basis in all cases.

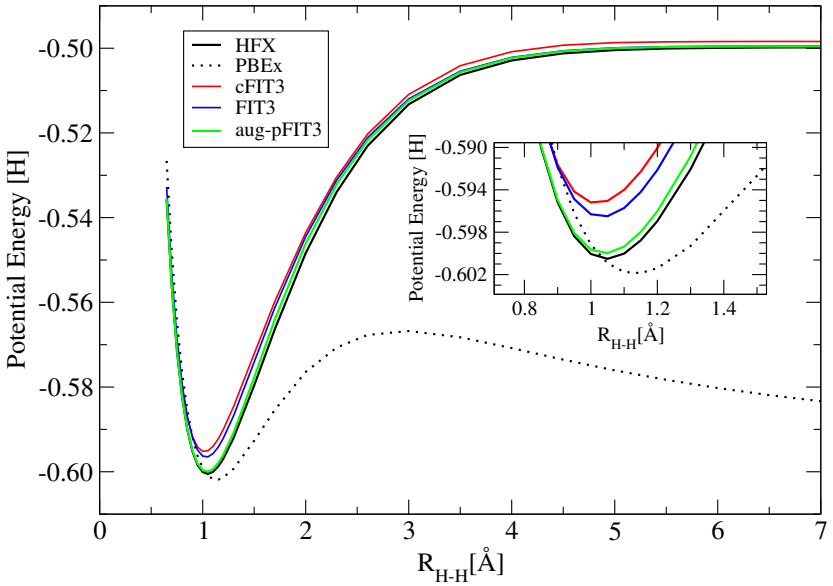

Figure 1: Shown are dissociation curves for H_2^+ obtained from different calculations. The black solid line depicts the reference Hartree–Fock run with the TZV2P-MOLOPT basis set. Red, blue and green lines represent results for ADMM1 for auxiliary basis sets of increasing quality, cFIT3, FIT3 and aug-pFIT3 respectively. The dotted black line shows the dissociation curve obtained from a pure GGA exchange calculation (PBEx). In the inset, a magnification of the potential energy around the minimum is presented.

3.4 H_2^+ dissociation curve

ADMM calculations that are GGA corrected might be biased from deficiencies of the underlying GGA functional. In order to investigate this effect, dissociation curves for H_2^+ at different levels of theory have been calculated. As is well known, GGA functionals, such as PBE exchange, describe the dissociation of this system incorrectly.[51] Figure 1 compares results obtained from a Hartree-Fock reference calculation, which is exact for this system, with results obtained from ADMM1. The primary basis was chosen to be the same as in the reference calculation (TZV2P-MOLOPT) while several different ABS have been applied. The results clearly show, that the wavefunction fitting is not biased by the GGA correction. Furthermore, as shown in the inset, better quality ABS consistently improve the description of the potential around the minimum. It can thus be concluded that the qualitatively important effects of HFX are properly retained and that the GGA correction

does not introduce artefacts of the underlying functionals.

3.5 The cationic hole in liquid water

In order to probe the effect of the dual basis set approach on the electronic structure directly, the spin density distribution of the cationic hole in bulk liquid water has been computed. The poor performance of local functionals for the radical cation water dimer was discussed in detail by Sodupe et al. in Ref. 52 and attributed to the self interaction error, which favors configurations with a delocalized spin density distribution. Hybrid functionals with a relatively large fraction of exchange, for example BH&HLYP,[53,54] perform significantly better. In Ref. 55, ionization of bulk liquid water has been probed, and the difficulty of DFT to properly describe the electronic structure has been discussed. In particular, it has been found that the electron hole, or similarly the spin density, is delocalized over the full simulation cell with local functionals, whereas it localizes on a single water molecule with Hartree-Fock exchange. Hybrid functionals with varying amounts of exchange yield intermediate degrees of localization. This is illustrated in Figure 2 for a bulk sample of liquid water (64 molecules), where the localization of the spin density is shown as a function of the amount of HFX in the PBE0 functional.

As quantitative measure, the maximum value of the Mulliken spin population is reported, ranging from approximately 0.1 in the local functional, to more than 1.0 in a functional containing 100% HFX. Contour plots of the spin density distribution emphasize this radical change in the electronic structure. Given this very strong dependence on the amount of Hartree-Fock exchange, this is a very stringent test for the auxiliary basis method presented in this work. Furthermore, this calculation has been performed with the relatively small FIT3 basis, i.e. without polarization functions. The results shown in Figure 2 are therefore very reassuring, since the spin distribution obtained with the auxiliary basis set approach essentially reproduces the reference density in all details for all fractions of exchange.

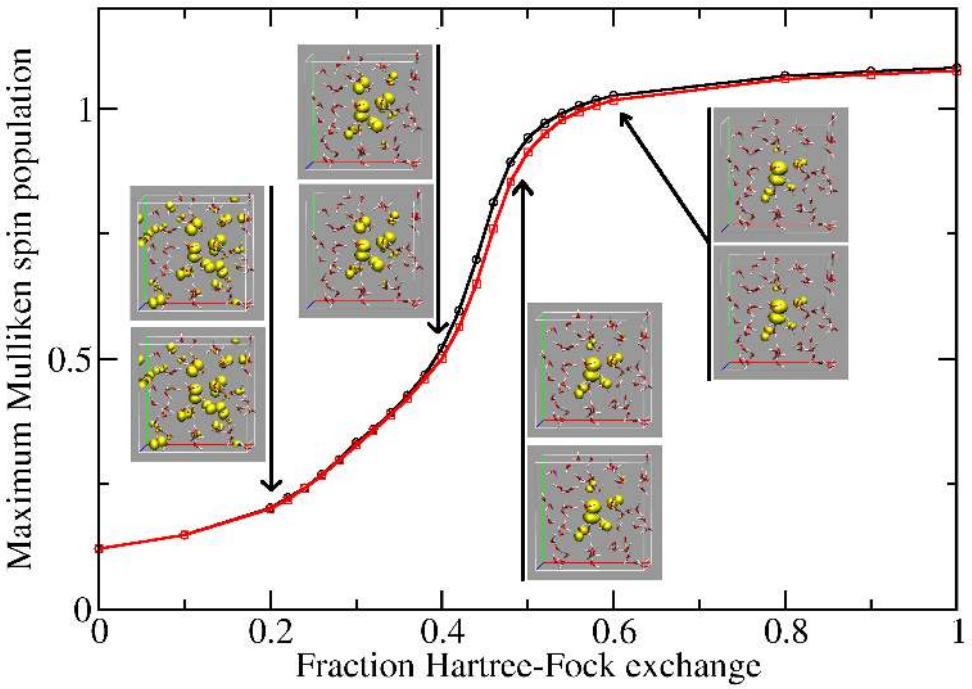

Figure 2: Shown is the localization of the spin density distribution after ionization of bulk liquid water as a function of the fraction of Hartree-Fock exchange employed in the density functional. The Mulliken spin populations of the oxygen atom on which the hole localizes is shown with a solid line, while the insets show a contour plot at 0.001 a.u. of the spin density for selected fractions (0.2, 0.4, 0.5, and 0.6) of exchange. Results obtained with the auxiliary FIT3 basis (black line, and upper panels of the inserts) are almost indistinguishable from the results obtained with the primary basis only, despite the pronounced sensitivity of this system towards the use of Hartree-Fock exchange.

3.6 Diamond band gap

Both wavefunction fitting methods (ADMM1 and ADMM2) have been benchmarked with respect to their capability of predicting the band gap in diamond. The basic cubic unit cell with lattice parameter $a = 3.576$ Å containing eight carbon atoms has been extended to a large super-cell in order to apply the Γ–point approximation. In a first step, PBE band gaps for super-cell sizes ranging from 1 x 1 x 1 to 6 x 6 x 6 repetitions of the basic unit cell have been determined with a high quality basis set. The band gap calculation was found to be converged for the 3 x 3 x 3 repetition, yielding a band gap of 4.17 eV in agreement with literature.[41] This super-cell has than been used to calculate the PBE0 reference band gap of this system applying the same high quality basis set. Since the condition number of the overlap matrix with the FIT3 basis is unfavorable in the case of bulk C

$(1.7 \cdot 10^5)$, an optimized FIT3 (optFIT3) basis has been constructed that served as ABS for the two wavefunction fitting methods. optFIT3 was obtained by minimization of the total energy of the PBE 2 x 2 x 2 super-cell with respect to the constraint of a well behaved overlap matrix (the final condition number is of order $\mathscr{O}(10^2)$). This allows for rather loose screening thresholds (10^{-6}) and thus significantly reduces the amount of work in the Fock matrix construction. Results are summarized in Table 6. Both wavefunction fitting methods are in good agreement the reference

Table 6: Shown are band gaps of diamond as obtained from different methods. All calculations have been performed using the 3 x 3 x 3 repetition of the basic unit cell in Γ–point approximation. For the hybrid PBE0 calculations, also the number of Cartesian integrals is shown. See text for details on the primary and auxiliary basis set ((PBS) and (ABS)). ADMM1 is purified wavefunction fitting and ADMM2 is non-purified wavefunction fitting.

method	number of integrals	gap [eV]
PBE (PBS)	-	4.17
PBE (ABS)	-	4.37
PBE0 (PBS)	40'787'850'778'591	6.07
PBE0 (ABS)	23'561'509'497	6.25
PBE0 ADMM1	24'816'897'009	6.03
PBE0 ADMM2	24'795'460'638	6.02

band gaps of the PBE0 run in the high quality basis. In order to illustrate the cost savings, the total number of Cartesian integrals that needs to be calculated has been added to the table. The ADMM calculations are by 3 orders of magnitude more efficient than the reference PBE0 run. Both methods give very similar results, suggesting that the approximate Kohn-Sham matrix (Eq. 34) is a valid approximation.

3.7 Performance and embedding for large systems

In order to illustrate the impact of ADMM for large systems, the electronic structure of Rubredoxin has been computed. Rubredoxin is a relatively small iron-sulfur protein that is an excellent benchmark system for electronic structure calculations, since it features an interesting active site. A realistic model including solvent and using periodic boundary conditions comprises of only 2825

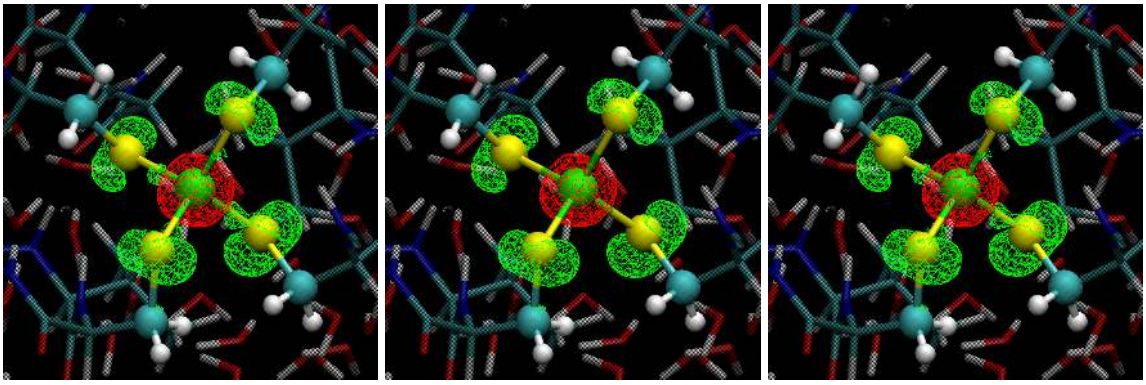

Figure 3: Shown are isosurfaces at +-0.001 a.u. of the difference between the spin density as computed with BLYP and B3LYP for the iron-sulfur protein Rubredoxin. Left panel: Traditional calculation using only a primary basis, middle panel: ADMM1 calculation using the cFIT3 auxiliary basis right panel: ADMM1 calculations using an embedding-like strategy, where the bulk of the system is described using the cFIT3 basis, but Fe and S use the primary basis as auxiliary basis. Both ADMM calculations clearly capture the effect of Hartree-Fock exchange, a reduced delocalization of the spin density, at a small fraction of the cost of the traditional approach. The embedding strategy faithfully reproduces all details, including the change in spin density along the Fe-S bonds.

atoms and fits in a unit cell with edges 31.1 x 28.1 x 30.5 $Å^3$. This system has been used extensively in our earlier work. In Ref. 56, ab initio simulations of the full system have been combined with statistical sampling to quantify the effect of mutations on the redox potential of the active site. In Ref. 9, the feasibility of computing the electronic structure with accurate, molecularly optimized, basis sets has been demonstrated. In Ref. 3, hybrid density functional calculations using an all-electron description and a polarized triple zeta valence basis set[57] have been performed. Molecularly optimized basis sets[9] have been employed as a primary basis for hybrid calculations (B3LYP[54,58,59]) of the same system, and the performance and accuracy of the ADMM scheme evaluated. Using a traditional HFX implementation with the MOLOPT basis set requires significant computational effort, despite the fact that the DZVP-MOLOPT-SR-GTH basis has been employed (22910 basis functions), which has fewer and less diffuse primitives than the basis sets originally presented in Ref.[9] Indeed, the reference calculation has been run using 48000 cores on a Cray XT5. The first SCF step required 45 minutes to compute $3.7 \cdot 10^{14}$ primitive Cartesian integrals after screening with a threshold of 10^{-6}. Successive SCF steps spent only 25 seconds in the

Hartree-Fock routines, since these calculations could be run in-core using integral compression[8] and 6.8Tb of RAM. Due to the contracted and diffuse nature of the basis sets, this calculation is significantly more expensive than the calculations performed in Ref. 3. The difference in spin density between the B3LYP and a BLYP calculation is shown in the left panel of Figure 3. ADMM1 calculations using the cFIT3 basis (12311 basis functions) require far fewer resources, and have been run on 1152 cores only. The Hartree-Fock routines used 75 seconds and 25 seconds in the first and successive SCF steps respectively, and in-core operation only required 5.2 Gb of RAM. The time spent in dense linear algebra for the wavefunction fitting (15s, Eq. 16) and corresponding derivative calculation (15s, Eq. 101) is similar to the time spent in the HFX, suggesting that this system might benefit from linear scaling techniques for this part of the calculation. For this system, ADMM thus improves the efficiency of the calculation by a factor 20 to 1000, depending on the measure. As shown in Figure 3, the obtained spin density reproduces the reference calculation very well, even though some small differences near the Fe-S bond can be observed. To improve the accuracy, we have employed the simple embedding strategy in which the auxiliary basis for the five central atoms (Fe and S) was set equal to the primary basis. These calculations can be performed without any significant increase in computational cost, and the right panel of Figure 3 shows that full quantitative agreement can be obtained in this way.

3.8 The effect of Hartree-Fock exchange on the structure of liquid water

In this section, ADMM is employed to study the effect of changing the fraction of Hartree-Fock exchange in the PBE0 functional on the structure of liquid water. Firstly, the accuracy of ADMM for describing bulk water is investigated. Secondly, we perform ab initio molecular dynamics simulations based on ADMM for various values of the fraction of exchange. The model system is a sample of 64 water molecules in a cubic box with edges 12.42 Å that has previously been equilibrated using PBE0.[8] The primary basis is in all cases a TZV2P basis (2560 basis functions in total). In order to investigate the accuracy of ADMM, the following procedure has been adopted. In a first step, a reference molecular dynamics trajectory of 2 ps starting from an equilibrated configuration

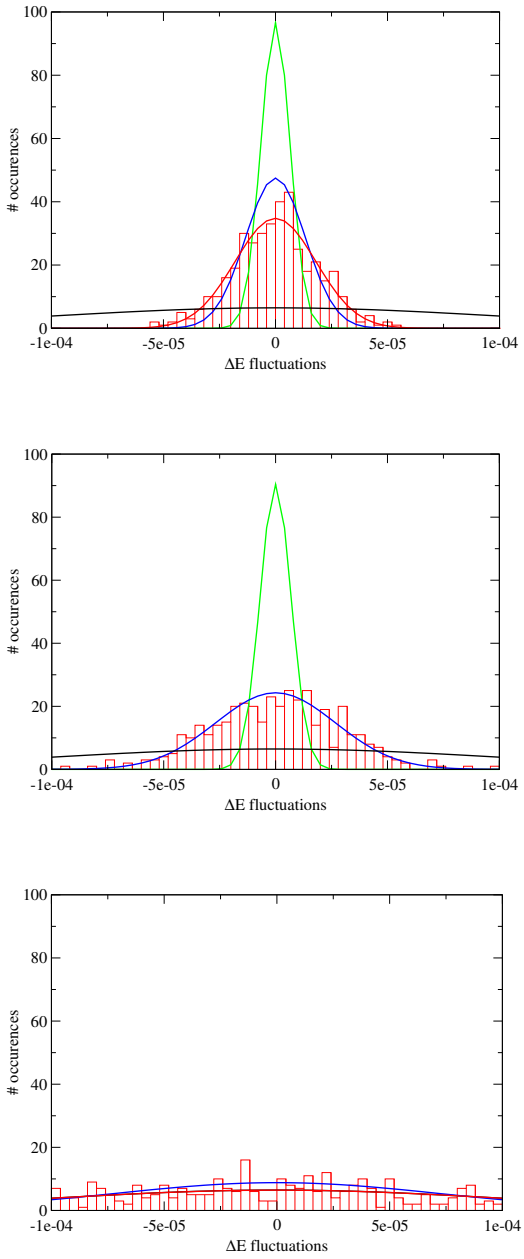

Figure 4: Shown are the centered distributions of energy differences between a standard PBE0 reference run and an ADMM method for bulk water. The top, middle, and bottom panel have been obtained with ADMM1, ADMM2, and ADMM3 respectively. ADMM1 and ADMM2 results have been computed using cFIT3 (red), FIT3 (blue), and aug-pFIT3 (green) auxiliary basis sets. ADMM3 employs blocking on a molecular level, with blocked (red) and full purification (blue). For clarity, Gaussian distributions are shown instead of binned data, except for one dataset per panel.

has been produced. In a second step, 400 equipartitioned configurations have been chosen and for all of these single point ADMM calculations have been performed. The error has been quantified by computing the distribution of the difference between the reference energy and the ADMM energy. The important quantity is the variance of this difference, i.e. the energy fluctuations between the two potential energy surfaces. ADMM1 and ADMM2 have been benchmarked for various basis sets. ADMM3, which starts from a blocked density matrix, has been employed with blocked purification (Eq. 70) or full purification (Eq. 69). Non-purified ADMM3 was found to be unstable. The subsystems have been defined as containing exactly one water molecule per block, i.e. the whole system consists of 64 diagonal sub-blocks. As shown in Figure 4, the fluctuations have approximately a Gaussian distribution. For ADMM1 and ADMM2, the associated variance gets consistently smaller when improving the quality of the auxiliary basis set. The variance for the purified wavefunction fitting (ADMM1) is slightly lower than the variance from non-purified wavefunction fitting (ADMM2). The variance of the energy fluctuations per water molecule is below 30 micro-Hartree for all auxiliary basis sets. This variance is significantly below the variance obtained applying the same procedure with a the pure density functional PBE, i.e. the difference between PBE0 and PBE is captured correctly with the ADMM1 and ADMM2 procedure. ADMM3 shows a relatively large variance, similar to direct use of the PBE functional, and its accuracy is not competitive.

With the aim of studying the effect of the fraction of exchange on the structure of the liquid, simulations employing the following functionals have been performed: PBE0 with various amounts of Hartree-Fock exchange ($\alpha \in \{0.12, 0.25, 0.37, 0.5, 0.62, 0.75, 1.00\}$), PBE, pure Hartree-Fock, PBE exchange (PBEx) and a revised parametrization[61] of PBEx (revPBEx). With these settings, trajectories longer than 30 ps have been obtained for all cases at a rate of 7 and 20 seconds per MD step (0.5 fs) for the pure and the hybrid functionals respectively on 64 cores of a Nehalem based cluster. Compared to a standard hybrid functional calculation in the PBS without ADMM and multiple time-step MD,[8] this is a speed-up of a factor 16 per MD step. All MD simulations have been done within the isokinetic ensemble[62] at a temperature of 330 K, using

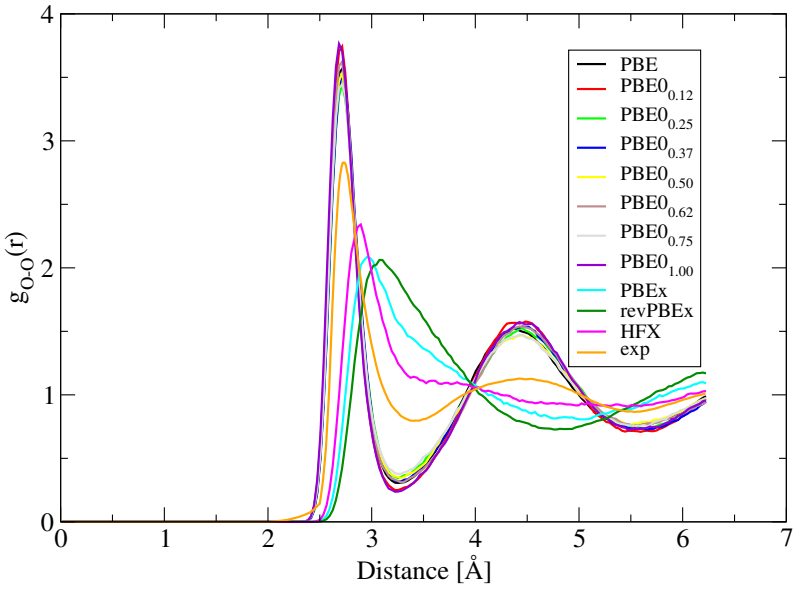

Figure 5: Oxygen-Oxygen pair correlation functions as obtained for bulk liquid water, based on a sample of 64 water molecules. Seven variants of PBE0, using various amounts of Hartree-Fock exchange, and PBE almost superimpose, but are overstructured as compared to the experimental result from Ref. 60. PBE exchange (PBEx) only, revised PBE exchange (revPBEx), and pure Hartree-Fock yield pair correlations that are similar, and understructured as compared to experiment. A detailed comparison of the maximum values of the pair correlation functions is shown in Figure 6.

ADMM1 and the FIT3 auxiliary basis. The structure has been analyzed using the oxygen-oxygen pair correlation function using the last 28 ps for each run, binning with a width of 0.03 Å. As shown in Figure 5 PBE and all variants of PBE0 yield very similar pair correlation functions. Compared to experiment,[60] the location of the peak is correct, but the liquid is overstructured. In order to quantify the structure, the maximum value of the pair correlation function is shown in Figure 6. For PBE and all variants of PBE0 the height of the first peak falls in the range 3.45-3.75. There is no systematic trend with respect to the fraction of exchange, and the differences between the peak heights must be attributed to the limited statistics that can be collected within 30ps for a structured liquid. Within these statistical uncertainties, these ADMM results agree with the PBE0 results obtained using traditional HFX and the same basis in Ref. 8, where a maximum height of

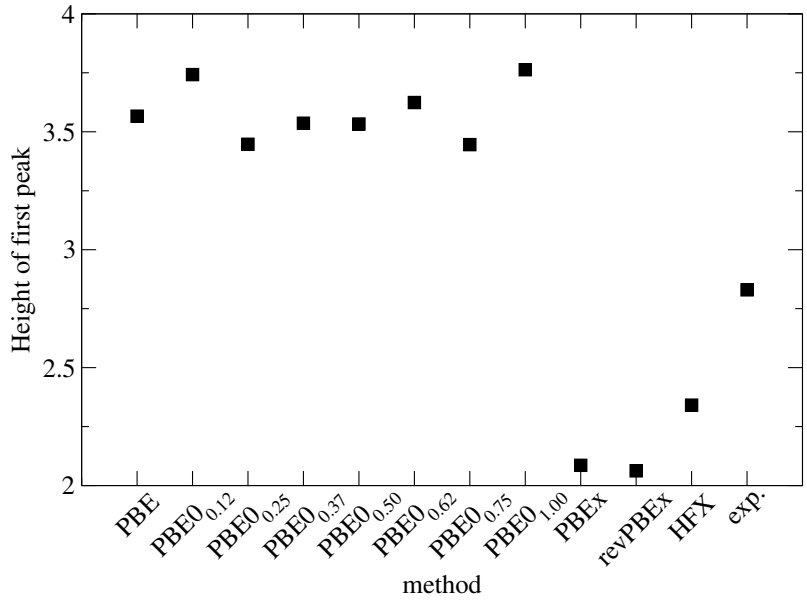

Figure 6: Shown is the height of the first peak for the oxygen-oxygen pair correlations shown in Figure 5. For the variants of PBE0, no trend in peak height with respect to the fraction of exchange can be observed.

3.4 was found for PBE and PBE0 ($\alpha = 0.25$) using 7.5ps of data. On the other hand, the liquid is significantly understructured for the pure Hartree-Fock, PBEx, and revPBEx runs. The maximum pair correlation height obtained from the Hartree-Fock simulation is in agreement with the results in Ref. 26, 2.34 and 2.35 respectively, where a plane waves basis set has been employed. The large difference between a Hartree-Fock simulation and a PBE0($\alpha = 1.00$) simulation can only be attributed to the correlation functional, since all other terms in the Kohn-Sham equations are the same. Consistent with this and the observations made above, we find that the PBEx and revPBEx simulations, which do not include a correlation functional and employ a density functional to model exchange, qualitatively reproduce the Hartree-Fock simulations. The deviation between the Hartree-Fock and the revPBEx pair correlation function is somewhat larger than the deviation between the Hartree-Fock and the PBEx results.

Finally, it is important to emphasize three limitations of our simulations. Firstly, due to the fact that the stress tensor is currently not implemented for hybrid functionals, these simulations have

been performed at constant volume and not at constant pressure. In recent work, see e.g. Ref. 63 and Ref. 64, it has been shown that constant volume simulations might differ significantly from constant pressure simulations for this system. Indeed, the density of the liquid, and several other macroscopic quantities,[65] most of them challenging to compute ab initio, might be more revealing about the quality of the underlying density functional than the pair correlation function. Secondly, whereas the structure of the liquid is for the PBE0 functional not depending strongly on the fraction of exchange, this dependence might be different for other hybrid functionals. Thirdly, the fact that the structure of the liquid is effectively unchanged as the fraction of exchange is varied does not imply that the properties as a solvent, i.e. the interaction of the liquid with solutes is unchanged. In the future, the efficiency of ADMM might contribute to addressing some of these important issues.

4 Summary

We presented auxiliary density matrix methods that aim at reducing the cost of simulations based on hybrid density functionals. By constructing an approximate density matrix, which allows for a fast calculation of exchange, and by correcting the error introduced using a density functional, significant speedups have been achieved while accuracy has been retained. Wavefunction fitting methods that employ a small auxiliary basis to reduce the size of the density matrix appear to be a simple yet successful way to obtain an approximate density matrix. The accuracy of this approach has been investigated using a variety of tests. Calculations on the GMTKN24 database suggest that the predictivity of calculations based on wavefunction fitting essentially equals that of the more expensive traditional approach. Test calculations specifically aimed at difficult systems, such as BSSE calculations for the water dimer and the dissociation profile of H_2^+, have demonstrated that neither the deficiencies of the small auxiliary basis nor of the correcting functional impact the quality of the results significantly. Two variants of wavefunction fitting, either with purification (ADMM1) or without purification (ADMM2) have been tested, and no significant differences in accuracy have been found so far. Whereas ADMM1 has the advantage of yielding a pure auxiliary

density matrix, ADMM2 is particularly simple to implement and is directly suitable for a linear scaling code. ADMM3, which relies on a blocking of the density matrix, has not been tested thoroughly yet, but might find its application in cases where clear subsystems, such as a solute in solution, can be easily defined. Exploiting the efficiency of the ADMM scheme, the effect of the range of exchange has been investigated for the PBE0-TC-LRC functional. The performance of this functional on the GMTKN24 database is optimal for 20% of exchange, and a range of 2.5Å. Furthermore, ADMM has been used to perform extensive simulations of bulk water, showing that for PBE0-like functionals the amount of Hartree-Fock exchange does not directly influence the structure of the liquid. In this case, the role of correlation is more significant. Finally, a calculation on a solvated protein has been used to demonstrate that speedups in excess of a factor of twenty can be observed in actual applications.

Acknowledgments

The authors acknowledge L. Goerigk and S. Grimme for making the GMTKN24 database publically available in a convenient format. Calculations were enabled by a 2008-2010 INCITE award on the CRAY XT5 using resources of the National Center for Computational Sciences at Oak Ridge National Laboratory (ORNL), which is supported by the Office of Science of the U.S. DOE under Contract No. DE-AC05-00OR22725, and by the Swiss National Supercomputer Centre (CSCS). This work has been funded by the Swiss University Conference through the High Performance and High Productivity Computing (HP2C) Programme.

A Wavefunction fitting

The one-particle wavefunctions represented with the high quality primary basis set (PBS) $\{\phi_\mu\}$ can be written in terms of molecular coefficients $C^{\mu i}$

$$\psi_i(\mathbf{r}) = \sum_\mu C^{\mu i}\phi_\mu(\mathbf{r}). \tag{35}$$

These wavefunctions are assumed to be orthonormal, i.e.

$$\int \psi_i(\mathbf{r})\psi_j^*(\mathbf{r})d\mathbf{r} = \delta_{ij}. \tag{36}$$

For the wavefunction fitting, a lower quality auxiliary basis set (ABS) $\{\hat{\phi}_\mu\}$ is introduced which yields a second set of molecular coefficients $\hat{C}^{\mu i}$ and auxiliary one-particle wavefunctions in the following form

$$\hat{\psi}_i(\mathbf{r}) = \sum_\mu \hat{C}^{\mu i}\hat{\phi}_\mu(\mathbf{r}). \tag{37}$$

The molecular coefficients $\hat{C}^{\mu i}$ are a priori unknown but can be determined by requiring that the corresponding occupied wavefunctions resemble as well as possible the original ones by minimizing their square difference over all space

$$\sum_j \int (\psi_j(\mathbf{r}) - \hat{\psi}_j(\mathbf{r}))^2 d\mathbf{r}. \tag{38}$$

Optionally, the auxiliary wavefunctions can be restricted to obey the orthonormality constraint

$$\int \hat{\psi}_i(\mathbf{r})\hat{\psi}_j^*(\mathbf{r})d\mathbf{r} = \delta_{ij}. \tag{39}$$

These two possibilities give raise to two slightly different minimization problems:

$$\min_{\hat{C}} \left[\sum_j \int \left(\psi_j(\mathbf{r}) - \hat{\psi}_j(\mathbf{r})\right)^2 d\mathbf{r} \right], \tag{40}$$

and

$$\min_{\tilde{C}} \left[\sum_j \int \left(\psi_j(\mathbf{r}) - \tilde{\psi}_j(\mathbf{r}) \right)^2 d\mathbf{r} + \sum_{k,l} \Lambda_{kl} \left(\int \tilde{\psi}_k \tilde{\psi}_l - \delta_{kl} \right) \right], \tag{41}$$

where in the latter case, the Lagrangian multipliers Λ_{kl} enforce condition Eq. 39 and the notation $\tilde{C}$ has been introduce in order to distinguish the two different sets of molecular coefficients. The overlap matrices associated with the two basis set representations are given as

$$S_{mm'} = \int \phi_m(\mathbf{r}) \phi_{m'}(\mathbf{r}) d\mathbf{r} \quad \text{and} \quad \hat{S}_{nn'} = \tilde{S}_{nn'} = \int \hat{\phi}_n(\mathbf{r}) \hat{\phi}_{n'}(\mathbf{r}) d\mathbf{r}. \tag{42}$$

In order to retain a consistent notation, $\hat{S}$ and $\tilde{S}$ have been introduced, even though both matrices are identical. Furthermore, a mixed overlap matrix Q needs to be defined that takes the overlap of both sets of basis functions into account:

$$Q_{nm} = \int \hat{\phi}_n(\mathbf{r}) \phi_m(\mathbf{r}) d\mathbf{r}. \tag{43}$$

Within this notation, the Lagrange functions associated with the two minimization problems Eq. 40 and Eq. 41 can conveniently be expressed as

$$\hat{L} = \sum_j \left(\sum_{m,m'} C_{mj} C_{m'j} S_{mm'} + \sum_{n,n'} \hat{C}_{nj} \hat{C}_{n'j} \hat{S}_{nn'} - 2 \sum_{m,n} C_{mj} \hat{C}_{nj} Q_{nm} \right). \tag{44}$$

and

$$\begin{aligned} \tilde{L} \;=\; & \sum_j \left(\sum_{m,m'} C_{mj} C_{m'j} S_{mm'} + \sum_{n,n'} \tilde{C}_{nj} \tilde{C}_{n'j} \tilde{S}_{nn'} - 2 \sum_{m,n} C_{mj} \tilde{C}_{nj} Q_{nm} \right. \\ & + \left. \sum_{m,n} \sum_{k,l} \Lambda_{kl} \left(\tilde{C}_{nj} \tilde{C}_{mk} \tilde{S}_{nm} - \delta_{kl} \right) \right). \end{aligned} \tag{45}$$

Because of Eq. 36, or, equivalently $C^T S C = 1$ and due to Eq. 39 or $\tilde{C}^T \tilde{S} \tilde{C} = 1$ in the second case, this simplifies to

$$\hat{L} = \sum_j \left(\sum_{n,n'} \hat{C}_{nj} \hat{C}_{n'j} \hat{S}_{nn'} - 2 \sum_{m,n} C_{mj} \hat{C}_{nj} Q_{nm} \right) \tag{46}$$

and

$$\tilde{L} = -2\sum_{j}\sum_{m,n} C_{mj}\tilde{C}_{nj}Q_{nm} + \sum_{k,l}\Lambda_{kl}\left(\tilde{C}_{nk}\tilde{C}_{ml}\tilde{S}_{nm} - \delta_{kl}\right), \tag{47}$$

respectively. From that, the unknown auxiliary molecular coefficients can be determined by taking the partial derivatives and equating them to zero. This yields

$$\frac{\partial \hat{L}}{d\hat{C}_{pq}} = -2(QC)_{pq} + 2(\hat{S}\hat{C})_{pq} \doteq 0 \tag{48}$$

and

$$\frac{\partial \tilde{L}}{d\tilde{C}_{pq}} = -2(QC)_{pq} + 2(\tilde{S}\tilde{C}\Lambda)_{pq} \doteq 0. \tag{49}$$

Thus, the final results for the MO coefficients are given by

$$\hat{C} = \hat{S}^{-1}QC \quad \text{and} \quad \tilde{C} = \tilde{S}^{-1}QC\Lambda^{-1}, \tag{50}$$

with the matrix of the Lagrangian multipliers

$$\Lambda = \left[(QC)^T\tilde{S}^{-1}QC\right]^{1/2}. \tag{51}$$

Defining $A := \hat{S}^{-1}Q = \tilde{S}^{-1}Q$ to be the projector between the PBS and ABS directly yields Eq. 12 and Eq. 16 presented in Sec. 2.2.

B Purification

Most of the calculations that follow take advantage of the Cauchy integral theorem for matrix functions.[33] For an arbitrary matrix F, it states

$$f(F) = \frac{1}{2\pi i}\oint f(z)\frac{1}{zI - F}dz, \tag{52}$$

which, since

$$\frac{d}{dx}F^{-1} = -F^{-1}\frac{dF}{dx}F^{-1} \tag{53}$$

transforms into an explicit formula for the derivative of a matrix function

$$\frac{df(F)}{dx} = \frac{1}{2\pi i}\oint f(z)\frac{1}{F-zI}\frac{dF}{dx}\frac{1}{F-zI}dz. \tag{54}$$

Applying this formula, matrix function derivatives can be calculated through residues of its eigenvalues. Applying this to $f(x) = \Theta(z)$, where $\Theta(z)$ denotes the Heaviside function, the expression for the purified density matrix becomes

$$\tilde{P} = S^{-1/2}\left[\frac{1}{2\pi i}\oint \frac{\Theta(z-0.5)}{zI - S^{1/2}\hat{P}S^{1/2}}dz\right]S^{-1/2} \tag{55}$$

or, after some rearrangements

$$\tilde{P} = S^{-1}\left[\frac{1}{2\pi i}\oint \frac{\Theta(z-0.5)}{S^{-1}z-\hat{P}}dz\right]S^{-1}. \tag{56}$$

The evaluation of the contour integral can easily be performed via diagonalization. For that purpose, the following generalized eigenvalue problem needs to be solved:

$$\hat{P}R = S^{-1}R\lambda, \tag{57}$$

where R defines the matrix containing the generalized eigenvectors of $\hat{P}$. Indeed, inserting $RR^{-1} = 1$ and $(S^{-1}R)(S^{-1}R)^{-1}$ from left and right into Eq. 56 gives

$$\begin{aligned}
\tilde{P} &= S^{-1}\left[\frac{1}{2\pi i}\oint RR^{-1}\frac{\Theta(z-0.5)}{S^{-1}z-\hat{P}}(S^{-1}R)(S^{-1}R)^{-1}dz\right]S^{-1} \\
&= S^{-1}R\left[\frac{1}{2\pi i}\oint \frac{\Theta(z-0.5)}{zI-D}dz\right]R^{-1} \\
&= S^{-1}R\left[\frac{1}{2\pi i}\oint \frac{\Theta(z-0.5)}{zI-D}dz\right]R^{T}S^{-1}
\end{aligned} \tag{58}$$

where in the last step, the relation $R^T S^{-1} R = 1$ which is valid under the assumption that $\hat{P}$ is a symmetric matrix. The integral in brackets is evaluated using the Cauchy residue theorem and can be written in terms of a diagonal matrix L. Component-wise, this yields

$$L_{ii} = \frac{1}{2\pi i} \oint \frac{\Theta(z-0.5)}{z-\lambda_i} dz = \mathrm{Res}\left(\frac{\Theta(z-0.5)}{z-\lambda_i}, z=\lambda_i \right) = \Theta(\lambda_i - 0.5). \tag{59}$$

The final expression for $\tilde{P}$ is therefore

$$\tilde{P} = S^{-1} R L R^T S^{-1}. \tag{60}$$

In a similar fashion, the derivative of $\tilde{P}$ with respect to $\hat{P}$, that is needed in the expression for the Kohn-Sham matrix can be evaluated. After diagonalization of $\hat{P}$, this derivative reads

$$\frac{d\tilde{P}}{d\hat{P}} = S^{-1} R \left[\frac{1}{2\pi i} \oint \Theta(z-0.5) \left(\frac{1}{D - zI} \right) R^{-1} S \frac{d\hat{P}}{d\hat{P}} R \left(\frac{1}{D-zI} \right) dz \right] R^{-1}. \tag{61}$$

Again, the contour integral in brackets is computed via the Cauchy residue theorem. Since, in this case, the diagonal matrix D with the eigenvalues appears twice, the result is now a matrix M that also contains off-diagonal elements

$$\begin{aligned} M_{kj} &= \frac{1}{2\pi i} \oint \frac{\Theta(z-0.5)}{(\lambda_k - z)(\lambda_j - z)} dz = \frac{1}{2\pi i} \oint g(z) dz \\ &= \mathrm{Res}(g, \lambda_k) + \mathrm{Res}(g, \lambda_j) \\ &= \begin{cases} \frac{\Theta(\lambda_k - 0.5) - \Theta(\lambda_j - 0.5)}{\lambda_k - \lambda_j}, & k \neq j \\ \delta(\lambda_k - 0.5), & k = j \end{cases}. \end{aligned} \tag{62}$$

The derivative Eq. 61 thus becomes

$$\frac{d\tilde{P}_{cd}}{d\hat{P}_{ef}} = \left[S^{-1} R \left(M \otimes G_{ef} \right) R^{-1} \right]_{cd}, \tag{63}$$

with

$$G_{ef} = R^{-1} S \frac{d\hat{P}}{d\hat{P}_{ef}} R. \tag{64}$$

This result can now be applied to the purified wavefunction fitting in order to obtain an expression for the Kohn-Sham matrix. In this case, $\hat{P}$ is a function of P, i.e. it holds that $\hat{P} = APA^T$ and the Kohn-Sham matrix written in terms of a derivative of the energy with respect to P is given as

$$\frac{d\tilde{E}[\tilde{P}]}{dP_{ab}} = \frac{d\tilde{E}}{d\tilde{P}_{cd}} \frac{d\tilde{P}_{cd}}{d\hat{P}_{ef}} \frac{d\hat{P}_{ef}}{dP_{ab}} = \tilde{K}_{cd} \frac{d\tilde{P}_{cd}}{d\hat{P}_{ef}} \frac{d\hat{P}_{ef}}{dP_{ab}}, \tag{65}$$

where summation over same indices is assumed. The last derivative trivially amounts to

$$\frac{d\hat{P}_{ef}}{dP_{ab}} = \frac{d}{dP_{ab}} \left[APA^T \right]_{ef} = A_{ea} A_{fb}. \tag{66}$$

Under utilization of Eq. 63 the middle term simplifies to

$$\frac{d\tilde{E}[\tilde{P}]}{dP_{ab}} = \left[\left(A^T \tilde{S} R^{-1} \right) \left[\left(R^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} R^{-T} \right) \otimes M \right] R^T A \right]_{ab}. \tag{67}$$

Since $R^T \tilde{S}^{-1} R = 1$ this can be rewritten as

$$\frac{d\tilde{E}[\tilde{P}]}{dP} = A^T R \left[\left(R^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} R \right) \otimes M \right] R^T A. \tag{68}$$

If $\hat{P}$ is obtained from a blocking procedure, above expression needs to be filtered through B

$$\frac{d\tilde{E}[\tilde{P}]}{dP} = \left[A^T R \left[\left(R^T S^{-1} \tilde{K} S^{-1} R \right) \otimes M \right] R^T A \right] \otimes B \tag{69}$$

For the purification of the blocked density matrix, the McWeeny procedure based on the overlap matrix S can be replaced by a blocked McWeeny procedure where the overlap matrix is replaced

by its blocked counterpart $S^\dagger = S \otimes B$. Eq. 23 thus becomes

$$\bar{P}_{n+1} = f(\bar{P}_n) = 3\bar{P}_n S^\dagger \bar{P}_n - 2\bar{P}_n S^\dagger \bar{P}_n S^\dagger \bar{P}_n. \tag{70}$$

If the matrix B is chosen to be block diagonal, the eigenvalue problem Eq. 27 can thus be solved within the smaller diagonal subspaces only which significantly reduces the computational workload.

C Wavefunction fitting with and without purification

As mentioned in Sec. 2.3 applying the purification scheme Eq. 23 to the density matrix $\hat{P}$ obtained from wavefunction fitting without the orthonormality constraint yields exactly the density matrix $\tilde{P}$ obtained through the fitting procedure including the constraint. This can easily be seen by plugging $\hat{P}$ into the McWeeny purification algorithm. The first two iterations amount to

$$\begin{aligned}
\bar{P}_1 &= 3\hat{P}\hat{S}\hat{P} - 2\hat{P}\hat{S}\hat{P}\hat{S}\hat{P} \\
&= 3\hat{C}\hat{C}^T\hat{S}\hat{C}\hat{C}^T - 2\hat{C}\hat{C}^T\hat{S}\hat{C}\hat{C}^T\hat{S}\hat{C}\hat{C}^T \\
&= 3\hat{C}\Lambda\hat{C}^T - 2\hat{C}\Lambda^2\hat{C}^T \\
&= \hat{C}\left(3\Lambda - 2\Lambda^2\right)\hat{C}^T =: \hat{C}g_1\hat{C}^T
\end{aligned} \tag{71}$$

and

$$\begin{aligned}
\bar{P}_2 &= 3\hat{P}_1\hat{S}\hat{P}_1 - 2\hat{P}_1\hat{S}\hat{P}_1\hat{S}\hat{P}_1 \\
&= 3\hat{C}\left(3\Lambda - 2\Lambda^2\right)\hat{C}^T\hat{S}\hat{C}\left(3\Lambda - 2\Lambda^2\right)\hat{C}^T \\
&- 2\hat{C}\left(3\Lambda - 2\Lambda^2\right)\hat{C}^T\hat{S}\hat{C}\left(3\Lambda - 2\Lambda^2\right)\hat{C}^T\hat{S}\hat{C}\left(3\Lambda - 2\Lambda^2\right)\hat{C}^T \\
&= 3\hat{C}g_1^2\Lambda\hat{C}^T - 2\hat{C}g_1^3\Lambda^2\hat{C}^T.
\end{aligned} \tag{72}$$

Recursively, that yields

$$\bar{P}_{n+1} = \hat{C} g_{n+1} \hat{C}^T, \tag{73}$$

with

$$g_{n+1} = 3g_n^2\Lambda - 2g_n^3\Lambda^2. \tag{74}$$

If the McWeeny procedure converges in the limit for $n \to \infty$, $g_n(\Lambda)$ is required to become a fixed point. Obviously this condition is fulfilled in the case of $g_n \to \Lambda^{-1}$, i.e.

$$\bar{P} = \hat{C}\Lambda^{-1}\hat{C}^T = \tilde{P}. \tag{75}$$

which matches exactly the corresponding equation for the purified wavefunction fitting.

D Diagonalization in the occupied subspace for wavefunction fitting

As shown in App. B, in order to obtain an expression for the Kohn-Sham matrix, a general eigenvalue problem of size N_{ABS}

$$\hat{P}R = \tilde{S}^{-1}R\lambda \tag{76}$$

needs to be solved. This might become the bottleneck in a ADMM calculation because N_{ABS} is not necessarily a small quantity. However, if $\hat{P}$ can be expressed in terms of molecular coefficients, as it is the case e.g. in wavefunction fitting, it is sufficient to diagonalize the occupied subspace only, which is typically much smaller than N_{ABS}. This can be achieved by introducing the following substitution $R \to R_\Lambda$:

$$R = \tilde{S}\hat{C}\Lambda^{-1/2}R_\Lambda = \tilde{S}\tilde{C}R_\Lambda. \tag{77}$$

Eq. 76 transforms thus into

$$\begin{aligned}
\hat{P}R &= \tilde{S}^{-1}R\lambda \\
\hat{C}\hat{C}^T R &= \tilde{S}^{-1}R\lambda \\
\hat{C}\hat{C}^T\tilde{S}\tilde{C}R_\Lambda &= \tilde{S}^{-1}\tilde{S}\tilde{C}R_\Lambda\lambda \\
\tilde{C}\Lambda^{1/2}\Lambda^{1/2}\tilde{C}^T\tilde{S}\tilde{C}R_\Lambda &= \tilde{C}R_\Lambda\lambda \\
(\tilde{S}\tilde{C})^T\tilde{C}\Lambda\tilde{C}^T\tilde{S}\tilde{C}R_\Lambda &= (\tilde{S}\tilde{C})^T\tilde{C}R_\Lambda\lambda \\
\Lambda R_\Lambda &= R_\Lambda\lambda, \qquad (78)
\end{aligned}$$

where in the last step, the fact that $\tilde{C}\tilde{S}\tilde{C}^T = 1$ has been used. This eigenvalue problem is of the size $N_{mo} \times N_{mo}$ with N_{mo} the number of occupied orbitals in the system, and therefore significantly smaller in size than the general one. Unfortunately, its solution will only provide the eigenvectors of the occupied subspace

$$R_o = \tilde{S}\hat{C}\Lambda^{-1/2}R_\Lambda, \qquad (79)$$

and the eigenvectors R_n for the null-space are unknown. However, since $RR^T = \tilde{S}$ it follows that for the decomposition into occupied and unoccupied subspaces that

$$R_n R_n^T + R_o R_o^T = \tilde{S} \qquad (80)$$

which motivates the notation $R = (R_n \quad R_o)$. Furthermore the matrix M has a very characteristic structure,

$$M = \begin{pmatrix} M_n & M_{no} \\ M_{on} & M_o \end{pmatrix} \qquad (81)$$

with

$$M_o = M_n = \begin{pmatrix} 0 & 0 & \dots \\ 0 & 0 & \dots \\ \vdots & \vdots & \end{pmatrix}, \quad M_{no} = M_{on}^T = \begin{pmatrix} M_{no} & M_{no} & M_{no} \dots \end{pmatrix}, \tag{82}$$

which directly follows from its definition in Eq. 62. Using this decomposition, Eq. 68 can be rewritten in terms of occupied and unoccupied parts:

$$\begin{aligned} \frac{d\tilde{E}[\tilde{P}]}{dP} &= A^T \begin{pmatrix} R_n & R_o \end{pmatrix} \left[\begin{pmatrix} R_n^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} R_n & R_n^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} R_o \\ R_o^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} R_n & R_o^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} R_o \end{pmatrix} \otimes M \right] \begin{pmatrix} R_n & R_o \end{pmatrix}^T A \\ &= A^T \begin{pmatrix} R_n & R_o \end{pmatrix} \begin{pmatrix} 0 & R_n^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} \tilde{R}_o \\ (R_n^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} \tilde{R}_o)^T & 0 \end{pmatrix} \begin{pmatrix} R_n & R_o \end{pmatrix}^T A, \end{aligned} \tag{83}$$

with $\tilde{R}_o$ denoting the eigenvectors of the occupied subspace with the columns scaled by the vectors M_{no}. This can further be simplified, yielding an expression that only depends on the eigenvectors R_o of the occupied subsystem:

$$\begin{aligned} \frac{d\tilde{E}[\tilde{P}]}{dP} &= A^T \begin{pmatrix} R_o \left(R_n^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} \tilde{R}_o \right)^T & R_n R_n^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} \tilde{R}_o \end{pmatrix} \begin{pmatrix} R_n & R_o \end{pmatrix}^T A \\ &= A^T \left(R_o \tilde{R}_o^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} R_n R_n^T + R_n R_n^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} \tilde{R}_o R_o^T \right) A \\ &= A^T \Big(R_o \tilde{R}_o^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} (\tilde{S} - R_o R_o^T) \\ &\quad + (\tilde{S} - R_o^T R_o) \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} \tilde{R}_o R_o^T \Big) A. \end{aligned} \tag{84}$$

An explicit expression for the products $R_o \tilde{R}_o^T$ is still required. Again, due to the special structure of M, the matrix $\tilde{R}_o$ resulting from a column scaling of R_o with M_{no} can conveniently be obtained from the eigenvalues of Λ

$$\tilde{R}_o = R_o D_\Lambda^{-1} \tag{85}$$

where D_Λ contains the eigenvalues of the matrix Λ and fulfills $R_\Lambda D_\Lambda R_\Lambda^T = \Lambda$ and similarly $R_\Lambda D_\Lambda^{-1} R_\Lambda^T = \Lambda^{-1}$, yielding

$$\begin{aligned} R_o \tilde{R}_o^T &= \tilde{S}\hat{C}\Lambda^{-1/2} R_\Lambda (\tilde{S}\hat{C}\Lambda^{-1/2} R_\Lambda D_\Lambda^{-1})^T \\ &= \tilde{S}\hat{C}\Lambda^{-1/2} R_\Lambda D_\Lambda^{-1} R_\Lambda^T \Lambda^{-1/2} \hat{C}^T \tilde{S} \\ &= \tilde{S}\hat{C}\Lambda^{-1/2} \Lambda^{-1} \Lambda^{-1/2} \hat{C}^T \tilde{S} \\ &= \tilde{S}\hat{C}\Lambda^{-2} \hat{C}^T \tilde{S}. \end{aligned} \tag{86}$$

Inserting this result into Eq. 84 yields

$$\begin{aligned} \frac{d\tilde{E}[\tilde{P}]}{dP} &= A^T \left(R_o \tilde{R}_o^T \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} (\tilde{S} - R_o R_o^T) \right. \\ &\quad \left. + (\tilde{S} - R_o^T R_o) \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} \tilde{R}_o R_o^T \right) A \\ &= A^T \left(\tilde{S}\hat{C}\Lambda^{-2} \hat{C}^T \tilde{S} \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} (\tilde{S} - R_o R_o^T) \right. \\ &\quad \left. + (\tilde{S} - R_o^T R_o) \tilde{S}^{-1} \tilde{K} \tilde{S}^{-1} \tilde{S}\hat{C}\Lambda^{-2} \hat{C}^T \tilde{S} \right) A. \end{aligned} \tag{87}$$

For the last step, this equation needs to be back-transformed applying the substitution from Eq. 77. Since Λ is symmetric, its eigenvectors are orthonormal, i.e. $R_\Lambda R_\Lambda^T = 1$ and one finds

$$\begin{aligned} R_o R_o^T &= \tilde{S}\hat{C}\Lambda^{-1/2} R_\Lambda R_\Lambda^T \Lambda^{-1/2} \hat{C}^T \tilde{S} \\ &= \tilde{S}\hat{C}\Lambda^{-1} \hat{C}^T \tilde{S} = \tilde{S}\tilde{P}\tilde{S}. \end{aligned} \tag{88}$$

Thus, the final expression for the Kohn-Sham matrix is given as

$$\begin{aligned} \frac{d\tilde{E}[\tilde{P}]}{dP} &= A^T \left(\tilde{S}\hat{C}\Lambda^{-2} \hat{C}^T \tilde{K} (1 - \tilde{P}\tilde{S}) \right. \\ &\quad \left. + (1 - \tilde{S}\tilde{P}) \tilde{K}\hat{C}\Lambda^{-2} \hat{C}^T \tilde{S} \right) A \end{aligned} \tag{89}$$

which indeed depends only on the inverse square of Λ which has the size $N_{\text{mo}} \times N_{\text{mo}}$ which can be evaluated through efficient Cholesky decomposition and does not require a diagonalization.

E MO derivatives

For wavefunction optimization algorithms that do not rely on the existence of a Kohn-Sham matrix but rather utilize the MO derivatives

$$U = \frac{dE}{dC}, \tag{90}$$

such as the orbital transformation (OT) method[66] in Quickstep,[35] the explicit construction of a Kohn-Sham matrix can be omitted. This is certainly the case for the purified wavefunction fitting because in that case the corresponding auxiliary density matrix can be obtained from molecular coefficients. Instead of calculating the derivative of the energy with respect to the density matrix, it is thus sufficient to compute the MO derivatives

$$U_{\text{total}} = \frac{dE[P]}{dC} + \frac{d\tilde{E}[\tilde{P}]}{dC}, \tag{91}$$

where only the second term is of interest here. Notice, that the auxiliary density matrix $\tilde{P}$ can be expressed either in terms of purified molecular coefficients

$$\tilde{P} = \tilde{C}\tilde{C}^T \quad \text{with} \quad \tilde{C} = \hat{C}\Lambda^{-1} = AC\Lambda^{-1} \tag{92}$$

or in terms of non-purified molecular coefficients

$$\tilde{P} = \hat{C}\Lambda^{-1}\hat{C}^T. \tag{93}$$

As a consequence, there exist two different approaches for calculating the desired MO derivative. The first method involves the auxiliary Kohn-Sham matrix

$$\frac{d\tilde{E}[\tilde{P}]}{dC} = \frac{d\tilde{E}}{dC} = \frac{d\tilde{E}}{d\tilde{P}}\frac{d\tilde{P}}{d\hat{C}}\frac{d\hat{C}}{dC} = \tilde{K}\frac{d\tilde{P}}{d\hat{C}}\frac{d\hat{C}}{dC}, \tag{94}$$

and the second method directly takes derivative of the energy with respect to the purified MO coefficients into account

$$\frac{d\tilde{E}[\tilde{P}]}{dC} = \frac{d\tilde{E}}{dC} = \frac{d\tilde{E}}{d\tilde{C}}\frac{d\tilde{C}}{dC} = \tilde{U}\frac{d\tilde{C}}{dC}, \tag{95}$$

with

$$\tilde{U} = \frac{d\tilde{E}[\tilde{P}]}{d\tilde{C}}. \tag{96}$$

The first case is algebraically straight forward, leading to

$$\frac{dE[\tilde{P}]}{dC} = 2\left(A^T\tilde{H}\hat{C}\Lambda^{-1}\right) - 2\left(A^T\tilde{S}\hat{C}\Lambda^{-1}\hat{C}^T\tilde{H}\hat{C}\Lambda^{-1}\right). \tag{97}$$

The second case is slightly more involved and requires the usage of the Cauchy integral formalism. The required derivative $\frac{d\tilde{C}}{d\tilde{C}}$ involves terms such as

$$\frac{d\Lambda^{-1/2}}{dC} \tag{98}$$

which can conveniently by expressed by

$$\frac{d\Lambda^{-1/2}}{dC} = \frac{1}{2\pi i}\oint f(z)\frac{1}{\Lambda - zI}\frac{d\Lambda}{dC}\frac{1}{\Lambda - zI}dz, \tag{99}$$

with $f(z) = z^{-1/2}$. For the evaluation of the contour integral, the matrix Λ needs to be diagonalized. Using the same notation as in App. D, i.e. $R_\Lambda D_\Lambda R_\Lambda^T = \Lambda$ and D_Λ defined through the eigenvalues

μ_i of Λ, the resulting matrix reads

$$\begin{aligned} N_{kj} &= \frac{1}{2\pi i}\oint \frac{z^{-1/2}}{(\mu_k - z)(\mu_j - z)} dz = \frac{1}{2\pi i}\oint g(z)dz \\ &= \mathrm{Res}(g,\mu_k) + \mathrm{Res}(g,\mu_j) \\ &= \begin{cases} \frac{\mu_k^{-1/2} - \mu_j^{-1/2}}{\mu_k - \mu_j}, & k \neq j \\ -\frac{1}{2}\mu_k^{-3/2}, & k = j \end{cases} . \end{aligned} \tag{100}$$

The final result for the MO derivatives is thus given by

$$\frac{d\tilde{E}[\tilde{P}]}{dC} = A^T \tilde{U} \Lambda^{-1/2} + Q^T A C (Y + Y^T), \tag{101}$$

with

$$Y = R_\Lambda \left(\left[R_\Lambda^T C^T A^T \tilde{U} R_\Lambda \right] \otimes N \right) R_\Lambda^T. \tag{102}$$

The first method, Eq. 97, has the advantage that only the inverse of Λ is needed whereas the second method, Eq. 101, requires the diagonalization of Λ.

F Analytical ionic forces

The derivative of the energy with respect to the atomic positions $\mathbf{R}$

$$\frac{dE}{d\mathbf{R}} = \frac{dE[P]}{d\mathbf{R}} + \frac{d\tilde{E}[\tilde{P}]}{d\mathbf{R}} \tag{103}$$

can be calculated via the MO derivatives given in Eq. 101. That is, for the interesting term,

$$\frac{d\tilde{E}[\tilde{P}]}{d\mathbf{R}} = \frac{d\tilde{E}[\tilde{P}]}{d\tilde{C}} \frac{d\tilde{C}}{d\mathbf{R}} = \tilde{U} \frac{d\tilde{C}}{d\mathbf{R}}. \tag{104}$$

Component-wise, this yields

$$\frac{d\tilde{E}[\tilde{P}]}{d\mathbf{R}} = \tilde{U}_{ab}\left[\frac{dA}{d\mathbf{R}}C\Lambda^{-1/2}\right]_{ab} + \tilde{U}_{ab}\left[AC\frac{d\Lambda^{-1/2}}{d\mathbf{R}}\right]_{ab}, \tag{105}$$

with implicit summation over repeated indices. In order to evaluate the second term, it is possible to apply the same mathematical formalism as in Eq. 99 and the final result reads

$$\begin{aligned}\frac{d\tilde{E}[\tilde{P}]}{d\mathbf{R}} &= -\tilde{U}_{ab}\left[\tilde{S}^{-1}\frac{d\tilde{S}}{d\mathbf{R}}\tilde{S}^{-1}QC\Lambda^{-1/2}\right]_{ab} \\ &\quad +\tilde{U}_{ab}\left[\tilde{S}^{-1}\frac{dQ}{d\mathbf{R}}C\Lambda^{-1/2}\right]_{ab} \\ &\quad +Y_{ab}\left[C^T\frac{dQ^T}{d\mathbf{R}}AC\right]_{ab} \\ &\quad -Y_{ab}\left[C^TQ^T\tilde{S}^{-1}\frac{d\tilde{S}}{d\mathbf{R}}\tilde{S}^{-1}QC\right]_{ab} \\ &\quad +Y_{ab}\left[C^TQ^T\tilde{S}^{-1}\frac{dQ}{d\mathbf{R}}C\right]_{ab},\end{aligned} \tag{106}$$

with Y_{ab} as in Eq. 102. This can further be simplified in terms of traces and becomes

$$\frac{d\tilde{E}}{d\mathbf{R}} = -\mathrm{tr}\left(W_{\tilde{S}}^T\frac{d\tilde{S}}{d\mathbf{R}}\right) + \mathrm{tr}\left(W_Q^T\frac{dQ}{d\mathbf{R}}\right), \tag{107}$$

with the weighted density matrices

$$W_{\tilde{S}} = \tilde{S}^{-1}\tilde{U}\Lambda^{-T/2}C^TA^T + ACYC^TA^T \tag{108}$$

and

$$W_Q = \tilde{S}^{-1}\tilde{U}\Lambda^{-T/2}C^T + ACY^TC^T + ACYC^T. \tag{109}$$

G Eigenvalues

For purified wavefunction fitting, the Kohn-Sham matrix obtained through the McWeeny procedure or the Cauchy integral is not suitable for the calculation of orbital energies. This problem can be illustrated, by evaluating Eq. 89 for identical primary and auxiliary basis sets, i.e. $A = 1$, $\Lambda = 1$, $\hat{C} = C$, $\tilde{S} = S$ and $\tilde{P} = CC^T$. In that case, the eigenvalues are given as

$$\begin{aligned} C^T \frac{d\tilde{E}[\tilde{P}]}{dP} C &= C^T \left[A^T (\tilde{S}\hat{C}\Lambda^{-2}\hat{C}^T\tilde{K}(1-\tilde{P}\tilde{S}) + (1-\tilde{S}\tilde{P})\tilde{K}\hat{C}\Lambda^{-2}\hat{C}^T\tilde{S})A \right] C \\ &= C^T \left[SCC^T\tilde{K} - SCC^T\tilde{K}PS + \tilde{K}CC^TS - SP\tilde{K}CC^TS \right] C \\ &= C^T\tilde{K}C - C^T\tilde{K}PS + C^T\tilde{K}C - SP\tilde{K}C \\ &= C^T\tilde{K}C - C^T\tilde{K}C + C^T\tilde{K}C - C^T\tilde{K}C, \end{aligned} \tag{110}$$

which is identically zero. Thus, Eq. 89 is not a suitable candidate for the calculation of the orbital energies and a different approach needs to be taken into account. The obvious choice is to derive a similar expression as in Eq. 33 for the non-purified wavefunction fitting, i.e. omitting the purification procedure. This assumption leads to the approximated Kohn-Sham matrix given in Eq. 34.

References

(1) Strout, D. L.; Scuseria, G. E. *J. Chem. Phys.* **1995**, *102*, 8448–8452.

(2) Izmaylov, A. F.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **2006**, *125*, 104103.

(3) Guidon, M.; Hutter, J.; VandeVondele, J. *J. Chem. Theory Comput.* **2009**, *5*, 3010–3021.

(4) Spencer, J.; Alavi, A. *Phys. Rev. B* **2008**, *77*, 193110.

(5) Williamson, A. J.; Rajagopal, G.; Needs, R. J.; Fraser, L. M.; Foulkes, W. M. C.; Wang, Y.; Chou, M.-Y. *Phys. Rev. B* **1997**, *55*, R4851–R4854.

(6) Kent, P. R. C.; Hood, R. Q.; Williamson, A. J.; Needs, R. J.; Foulkes, W. M. C.; Rajagopal, G. *Phys. Rev. B* **1999**, *59*, 1917–1929.

(7) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 1663–1669.

(8) Guidon, M.; Schiffmann, F.; Hutter, J.; VandeVondele, J. *J. Chem. Phys.* **2008**, *128*, 214104.

(9) VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2007**, *127*, 114105.

(10) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939–947.

(11) Pietro, W. J.; Francl, M. M.; Hehre, W. J.; Defrees, D. J.; Pople, J. A.; Binkley, J. S. *J. Am. Chem. Soc.* **1982**, *104*, 5039–5048.

(12) Harihara, P. C.; Pople, J. A. *Theoret. Chimica Acta* **1973**, *28*, 213–222.

(13) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–654.

(14) Jensen, F. *J. Chem. Phys.* **2001**, *115*, 9113.

(15) Jensen, F. *J. Chem. Phys.* **2002**, *116*, 7372.

(16) Jensen, F. *J. Phys. Chem. A* **2007**, *111*, 11198–11204.

(17) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(18) Weigend, F. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

(19) Boman, L.; Koch, H.; de Merás, A. S. *J. Chem. Phys.* **2008**, *129*, 134107.

(20) Sodt, A.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 104106.

(21) Liang, W.; Head-Gordon, M. *J. Phys. Chem. A* **2004**, *108*, 3206–3210.

(22) Friesner, R. A. *Chem. Phys. Lett.* **1985**, *116*, 39 – 43.

(23) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. *Chem. Phys.* **2009**, *356*, 98–109.

(24) Yanai, T.; Fann, G. I.; Gan, Z.; Harrison, R. J.; Beylkin, G. *J. Chem. Phys.* **2004**, *121*, 6680–6688.

(25) Gygi, F.; Baldereschi, A. *Phys. Rev. B* **1986**, *34*, 4405–4408.

(26) Todorova, T.; Seitsonen, A. P.; Hutter, J.; Kuo, I.-F. W.; Mundy, C. J. *J. Phys. Chem. B* **2006**, *110*, 3685–3691.

(27) Sorouri, A.; Foulkes, W. M. C.; Hine, N. D. M. *J. Chem. Phys.* **2006**, *124*, 064105.

(28) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(29) Ernzerhof, M.; Perdew, J. P. *J. Chem. Phys.* **1998**, *109*, 3313–3320.

(30) McWeeny, R. *Rev. Mod. Phys.* **1960**, *32*, 335–369.

(31) Palser, A. H. R.; Manolopoulos, D. E. *Phys. Rev. B* **1998**, *58*, 12704–12711.

(32) Niklasson, A. M. N.; Tymczak, C. J.; Challacombe, M. *J. Chem. Phys.* **2003**, *118*, 8611–8620.

(33) Rinehart, R. F. *Proc. Amer. Math. Soc* **1956**, *7*, 2–5.

(34) The CP2K developers group, `http://cp2k.berlios.de/`, accessed June 3, 2010.

(35) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103.

(36) Lippert, G.; Hutter, J.; Parrinello, M. *Theor. Chem. Acc.* **1999**, *103*, 124.

(37) Krack, M.; Parrinello, M. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2105–2112.

(38) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703–1710.

(39) Krack, M. *Theor. Chem. Acc.* **2005**, *114*, 145–152.

(40) Gómez-Abal, R.; Li, X.; Scheffler, M.; Ambrosch-Draxl, C. *Phys. Rev. Lett.* **2008**, *101*, 106404.

(41) Duchemin, I.; Gygi, F. *Comput. Phys. Commun.* **2010**, *181*, 855 – 860.

(42) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2010**, *6*, 107–126.

(43) Goerigk, L.; Grimme, S. `http://toc.uni-muenster.de/GMTKN/GMTKNmain.html`, accessed June 3, 2010.

(44) Perdew, J. P.; Ernzerhof, M.; Burke, K. *J. Chem. Phys.* **1996**, *105*, 9982–9985.

(45) Perdew, J. P.; Ernzerhof, M.; Burke, K. *Int. J. Quantum Chem.* **1997**, *64*, 285–295.

(46) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029–5036.

(47) Grimme, S. *J. Comp. Chem.* **2006**, *27*, 1787–1799.

(48) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. *J. Chem. Phys.* **2003**, *118*, 8207–8215.

(49) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. *J. Chem. Phys.* **2006**, *124*, 219906, Erratum.

(50) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–&.

(51) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I. *J. Phys. Chem. A* **2005**, *109*, 11006–11014.

(52) Sodupe, M.; Bertran, J.; Rodriguez-Santiago, L.; Baerends, E. *J. Phys. Chem. A* **1999**, *103*, 166–170.

(53) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(54) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(55) Maršálek, O.; Elles, C. G.; Pieniazek, P. A.; VandeVondele, J.; Bradforth, S. E.; Jungwirth, P. in preparation.

(56) Sulpizi, M.; Raugei, S.; VandeVondele, J.; Carloni, P.; Sprik, M. *J. Phys. Chem. B* **2007**, *111*, 3669.

(57) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.

(58) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5852.

(59) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(60) Hura, G.; Sorenson, J. M.; Glaeser, R. M.; Head-Gordon, T. *J. Chem. Phys.* **2000**, *113*, 9140–9148.

(61) Zhang, Y.; Yang, W. *Phys. Rev. Lett.* **1998**, *80*, 890.

(62) McGrath, M. J.; Siepmann, J. I.; Kuo, I. F. W.; Mundy, C. J.; VandeVondele, J.; Hutter, J.; Mohamed, F.; Krack, M. *ChemPhysChem* **2005**, *6*, 1894–1901.

(63) Schmidt, J.; VandeVondele, J.; Kuo, I. F. W.; Sebastiani, D.; Siepmann, J. I.; Hutter, J.; Mundy, C. J. *J. Phys. Chem. B* **2009**, *113*, 11959–11964.

(64) McGrath, M. J.; Siepmann, J. I.; Kuo, I. F. W.; Mundy, C. J.; VandeVondele, J.; Hutter, J.; Mohamed, F.; Krack, M. *J. Phys. Chem. A* **2006**, *110*, 640–646.

(65) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragones, J. L. *Faraday Discuss.* **2009**, *141*, 251–276.

(66) VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2002**, *118*, 4365–4369.