QUICKSTEP: Make the Atoms Dance

Matthias Krack and Michele Parrinello

Computational Science
Department of Chemistry and Applied Biosciences
ETH Zürich
USI-Campus, via Giuseppe Buffi 13
6900 Lugano, Switzerland
E-mail: krack@phys.chem.ethz.ch

1 Introduction

During the last decade density functional theory (DFT) has become a very efficient tool for electronic structure calculations. DFT methods were successfully applied to many problems in different fields ranging from material science to chemistry and bio-chemistry. Most of the applied methods use either plane waves or Gaussian-type functions for the expansion of the Kohn-Sham orbitals. Both types of basis functions have their own merits. Plane waves provide from the beginning an orthogonal basis set and are independent of the atomic positions which makes the force calculation very simple. Moreover, the calculation of the Hartree (Coulomb) potential can be efficiently performed by fast Fourier transformation (FFT). Unfortunately, there are also some disadvantages. The strong variations of the wave function close to the nuclei require a huge number of plane waves. Atomic pseudo potentials are usually employed to alleviate this problem, but for many applications the number of plane waves is still large. Furthermore, the whole space is equally filled with plane waves and therefore each point in space is described with the same accuracy, but this feature of the plane waves turns out to be rather inefficient for systems of low density like biological systems where the homogeneous description of empty and atom-filled regions results in a waste of basis functions and thus computer time. By contrast, Gaussian-type functions localized at the atomic centers are much more efficient in this respect, since they provide a more compact description of the atomic charge densities and basically there is no

need to employ atomic pseudo potentials. Nevertheless, Gaussian-type functions have also some disadvantages. The force calculation is more complicated, the Hartree term usually requires the computation of a large number of multi-center integrals, and possibly basis set superposition errors have to be considered. The Gaussian plane waves (GPW) method [1] tries to combine the merits of Gaussian-type functions and plane waves. In that way it becomes feasible to build the Kohn-Sham operator matrix with a computational cost scaling linearly for a growing system size.

A new implementation of the GPW method, called QUICKSTEP, was performed within the framework of the HPC-Chem project [2]. The goal of the project was to implement the GPW method in a fully modular and efficiently parallelized manner. QUICKSTEP is part of the open source project CP2K [3, 4] which ensures the continuation of the development even after the end of the HPC-Chem project. The next section will provide a short outline of the GPW method followed by a description of the pseudo potentials and the Gaussian basis sets employed by QUICKSTEP. Finally, the accuracy and the efficiency of the new parallelized QUICKSTEP implementation will be shown.

2 Gaussian and plane waves method

The energy functional for a molecular or crystalline system in the framework of the Gaussian plane waves (GPW) method [1] using the Kohn-Sham formulation of density functional theory (DFT) [5, 6] is defined as

$$
\begin{aligned}
E[n] =& E^{\mathrm{T}}[n] + E^{\mathrm{V}}[n] + E^{\mathrm{H}}[n] + E^{\mathrm{XC}}[n] + E^{\mathrm{II}} \qquad (1)\\
=& \sum_{\mu\nu} P_{\mu\nu}\ \langle\varphi_\mu(\boldsymbol{r})| -\frac{1}{2}\nabla^2|\varphi_\nu(\boldsymbol{r})\rangle \\
&+ \sum_{\mu\nu} P_{\mu\nu}\ \langle\varphi_\mu(\boldsymbol{r})|V^{\mathrm{PP}}_{\mathrm{loc}}(r)|\varphi_\nu(\boldsymbol{r})\rangle \\
&+ \sum_{\mu\nu} P_{\mu\nu}\ \langle\varphi_\mu(\boldsymbol{r})|V^{\mathrm{PP}}_{\mathrm{nl}}(\boldsymbol{r},\boldsymbol{r}')|\varphi_\nu(\boldsymbol{r}')\rangle \\
&+ 2\pi\,\Omega \sum_{\boldsymbol{G}} \frac{\tilde{n}^*(\boldsymbol{G})\,\tilde{n}(\boldsymbol{G})}{\boldsymbol{G}^2} \\
&+ \int \bar{n}(\boldsymbol{r})\,\varepsilon_{\mathrm{XC}}[\bar{n}]\,d\boldsymbol{r} \\
&+ \frac{1}{2}\sum_{I\neq J} \frac{Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|}
\end{aligned}
$$

where $E^{\mathrm{T}}[n]$ is the kinetic energy, $E^{\mathrm{V}}[n]$ is the electronic interaction with the ionic cores, $E^{\mathrm{H}}[n]$ is the electronic Hartree (Coulomb) energy and $E^{\mathrm{XC}}[n]$ is the exchange–correlation

energy. The interaction energies of the ionic cores with charges Z_I and positions $\boldsymbol{R}_I$ is denoted by E^{II}. The electronic interaction with the ionic cores is described by norm-conserving pseudo potentials with a potential split in a local part $V_{\mathrm{loc}}^{\mathrm{PP}}(r)$ and a fully non-local part $V_{\mathrm{nl}}^{\mathrm{PP}}(\boldsymbol{r}, \boldsymbol{r}')$ (see section 3).

The electronic density

$$n(\boldsymbol{r}) = \sum_{\mu\nu} P_{\mu\nu} \varphi_\mu(\boldsymbol{r}) \varphi_\nu(\boldsymbol{r}) \tag{2}$$

is expanded in a set of contracted Gaussian functions

$$\varphi_\mu(\boldsymbol{r}) = \sum_i d_{i\mu} g_i(\boldsymbol{r}) \tag{3}$$

where $P_{\mu\nu}$ is a density matrix element, $g_i(\boldsymbol{r})$ is a primitive Gaussian function, and $d_{i\mu}$ is the corresponding contraction coefficient. The density matrix $\boldsymbol{P}$ fulfills normalization and idempotency conditions

$$\mathrm{Tr}(\boldsymbol{PS}) = N \tag{4}$$

$$\boldsymbol{PS} = (\boldsymbol{PS})(\boldsymbol{PS}) \tag{5}$$

where $\boldsymbol{S}$ is the overlap matrix of the Gaussian basis functions

$$S_{\mu\nu} = \langle \varphi_\mu(\boldsymbol{r}) || \varphi_\nu(\boldsymbol{r}) \rangle \tag{6}$$

and N is the number of electrons.

In the original work by Lippert et al. [1] the same auxiliary basis approximation was used for the Hartree and exchange-correlation energy. It was useful to relax this constraint and use two independent approximations to the density, denoted $\tilde{n}(\boldsymbol{G})$ for the Hartree energy and $\bar{n}(\boldsymbol{r})$ for the exchange-correlation energy. Both approximate electronic charge densities are functions of the density matrix $\boldsymbol{P}$.

3 Pseudo potentials

The GPW method works like plane waves methods with atomic pseudo potentials, since an expansion of Gaussian functions with large exponents is numerically not efficient or even not feasible.

The current implementation of the GPW method uses only the pseudo potentials of Goedecker, Teter, and Hutter (GTH) [7, 8]. The separable dual-space GTH pseudo potentials consist of a local part

$$V_{\text{loc}}^{\text{PP}}(r) = -\frac{Z_{\text{ion}}}{r}\,\text{erf}\left(\alpha^{\text{PP}}r\right) + \sum_{i=1}^{4} C_i^{\text{PP}}\left(\sqrt{2}\alpha^{\text{PP}}r\right)^{2i-2}\exp\left[-\left(\alpha^{\text{PP}}r\right)^2\right] \qquad (7)$$

with

$$\alpha^{\text{PP}} = \frac{1}{\sqrt{2}r_{\text{loc}}^{\text{PP}}}$$

and a non-local part

$$V_{\text{nl}}^{\text{PP}}(\boldsymbol{r},\boldsymbol{r'}) = \sum_{lm}\sum_{ij}\langle\,\boldsymbol{r}\mid p_i^{lm}\,\rangle\,h_{ij}^{l}\,\langle\,p_j^{lm}\mid\boldsymbol{r'}\,\rangle \qquad (8)$$

with the Gaussian-type projectors

$$\langle\boldsymbol{r}\mid p_i^{lm}\rangle = N_i^l Y^{lm}(\hat{r})r^{l+2i-2}\exp\left[-\frac{1}{2}\left(\frac{r}{r_l}\right)^2\right]$$

as shown in Eq. 1 resulting in a fully analytical formulation which requires only the definition of a small parameter set for each element. Moreover, the GTH pseudo potentials are transferable and norm-conserving. Nevertheless, plane waves methods employ this pseudo potential type only for reference calculations or if no other reliable pseudo potentials are available, since this type requires relative high cut-off values, i.e. more plane waves. However, in the framework of the GPW method there are no such limitations, since all contributions are integrals over Gaussian functions which can be calculated analytically. Therefore the GTH pseudo potentials are particularly suited for the use with QUICKSTEP and that is why QUICKSTEP only supports GTH pseudo potentials, currently. The GTH pseudo potential parameters were optimized with respect to atomic all-electron wavefunctions obtained from fully relativistic density functional calculations using a numerical atom code. The optimized pseudo potentials include all scalar relativistic corrections via an averaged potential [8], because the consideration of relativistic effects is indispensable for applications involving heavier elements. A database with many GTH pseudo potential parameter sets optimized for different exchange-correlation potentials is already available [3]. It provides all parameter sets formatted for a direct usage with QUICKSTEP and it contains parameter sets for almost the whole periodic table based on the local density approximation (LDA). Moreover, there are also many optimized parameter sets for the exchange-correlation potentials based on the generalized gradient approximation (GGA) of Becke, Lee, Yang, and Parr (BLYP) [9, 10, 11], Becke and Perdew (BP) [9, 12], Hamprecht, Cohen, Tozer and Handy (HCTH/120, HCTH/407) [13] and Perdew, Burke and Ernzerhof (PBE) [14]. The following GTH pseudo potentials are currently available:

LDA:
H(1), He(2), Li(1), Li(3), Be(2), Be(4), B(3), C(4), N(5), O(6), F(7), Ne(8), Na(1), Na(9), Mg(10), Mg(2), Al(3), Si(4), P(5), S(6), Cl(7), Ar(8), K(1), K(9), Ca(10), Ca(2), Sc(11), Sc(3), Ti(12), Ti(4), V(13), V(5), Cr(14), Cr(6), Mn(15), Mn(7), Fe(16), Fe(8), Co(17), Co(9), Ni(10), Ni(18), Cu(1), Cu(11), Zn(12), Zn(2), Zn(20) Ga(13), Ga(3), Ge(4), As(5), Se(6), Br(7), Kr(8), Rb(1), Rb(9), Sr(10), Sr(2), Y(11), Y(3), Zr(12), Zr(4), Nb(13), Nb(5), Mo(14), Mo(6), Tc(15), Tc(7), Ru(16), Ru(8), Rh(17), Rh(9), Pd(10), Pd(18), Ag(1), Ag(11), Cd(12), Cd(2), In(13), In(3), Sn(4), Sb(5), Te(6), I(7), Xe(8), Cs(1), Cs(9), Ba(10), Ba(2), La(11), Ce(12), Pr(13), Nd(14), Pm(15), Sm(16), Eu(17), Gd(18), Tb(19), Dy(20), Ho(21), Er(22), Tm(23), Yb(24), Lu(25), Hf(12), Ta(13), Ta(5), W(14), W(6), Re(15), Re(7), Os(16), Os(8), Ir(17), Ir(9), Pt(10), Pt(18), Au(1), Au(11), Hg(12), Hg(2), Tl(13), Tl(3) Pb(4), Bi(5), Po(6), At(7), Rn(8)

BLYP:
H(1), He(2), Li(3), Be(4), B(3), C(4), N(5), O(6), F(7), Ne(8), Na(9), Mg(10), Al(3) Si(4), P(5), S(6), Cl(7), Ar(8), Ca(10), Ti(12), V(13), Cr(14), Mn(15), Fe(16), Co(17), Ni(18), Cu(11), Zn(12), Ge(4), Br(7), Zr(12), I(7), Ba(10), Ba(2), W(14)

BP:
H(1), He(2), Li(3), Be(4), B(3), C(4), N(5), O(6), F(7), Ne(8), Na(1), Na(9), Mg(10), Al(3), Si(4), P(5), S(6), Cl(7), Ar(8), Ca(10), Sc(11), Ti(12), V(13), Cr(14) Mn(15), Fe(16), Co(17), Ni(18), Cu(11), Zn(12), Zr(12), Cs(1), Cs(9)

HCTH/120:
H(1), O(6), Ar(8)

HCTH/407:
H(1), O(6)

PBE:
H(1), He(2), Li(3), Be(4), B(3), C(4), N(5), O(6), F(7), Ne(8), Na(9), Mg(10), Mg(2), Al(3), Si(4), P(5), S(6), Cl(7), Ar(8), Ca(10), Ti(12), Zr(12)

The numbers in brackets denote the number of valence electrons employed by the respective pseudo potential, i.e. the effective core charge. The pseudo potential data base is maintained within the CP2K project [3] and thus all the listed GTH pseudo potential data sets are available online.

4 Basis sets

The Kohn-Sham orbitals are expanded in Gaussian orbital functions in the framework of the GPW method as described in section 2. Therefore an appropriate set of Gaussian functions has to be defined as a basis set for each QUICKSTEP calculation. There is a plenty of Gaussian basis sets available in the literature. However, proper basis sets have to be optimized for the usage with the GTH pseudo potentials from the previous section. Therefore, the exponents of a set of primitive Gaussian functions were optimized for all first- and second-row elements with an atomic DFT code applying the appropriate GTH potential parameters for each element. The same set of exponents was employed for each angular momentum quantum number of the occupied valence states of the actual element which are only s and p orbitals for the elements from H to Ar. The optimization was performed for a growing numbers of primitive Gaussian functions in the set in order to obtain basis sets of increasing quality. The atomic DFT code allows for the calculation of first analytic derivatives of the total atomic energy with respect to the Gaussian orbital exponents. The second derivatives were calculated by an updated Hessian procedure (BFGS). Finally, the primitive Gaussian functions were contracted using the coefficients of the respective atomic wavefunctions. These basis sets were augmented by polarization functions which were taken from the all-electron basis sets cc-pVXZ (X = D, T, Q) of Dunning [15, 16]. In that way a new sequence of basis sets was created with an increasing number of primitive Gaussian functions and polarization functions for each first- and second-row element. The basis sets were labelled DZVP, TZVP, TZV2P, QZV2P, and QZV3P due to the applied splitting of the valence basis where DZ, TZ, and QZ denote double-, triple- , and quadruple-zeta, respectively, and the increasing number of polarization functions. The quality of the basis sets should improve systematically within this sequence. These basis sets can be further augmented by diffuse functions, if required, analogous to the aug-cc-pVXZ basis sets resulting in a sequence aug-DZVP, aug-TZVP, aug-TZV2P, aug-QZV2P, and aug-QZV3P analogous to the aug-cc-pVXZ basis sets. The inclusion of diffuse functions may improve the accuracy of certain molecular properties, however, they are prohibitive for condensed phase calculations, since they introduce linear dependencies into the basis set. The basis sets for H to Ar are collected in a basis set file which is included into the CP2K program package.

5 Wavefunction optimization

The total ground state energy (see Eq. 1) of a system for a given atomic configuration is minimized by an iterative self-consistent field (SCF) procedure. Three methods are currently available in QUICKSTEP to perform an SCF iteration procedure: a traditional diagonalization (TD) scheme, a pseudo diagonalization scheme [17], and an orbital transformation (OT) method [19]. For the sake of simplicity, we will restrict our description

of these methods in the following to closed-shell systems, however, the generalization to open-shell (spin-polarized) systems is straightforward and QUICKSTEP can deal with both types of systems using each of these methods.

5.1 Traditional diagonalization (TD)

The traditional diagonalization scheme uses an eigensolver from a standard parallel program library called ScaLAPACK to solve the general eigenvalue problem

$$\boldsymbol{K}\,\boldsymbol{c} = \boldsymbol{S}\,\boldsymbol{c}\,\epsilon \tag{9}$$

where $\boldsymbol{K}$ is the Kohn-Sham matrix and $\boldsymbol{S}$ is the overlap matrix of the system. The eigenvectors $\boldsymbol{c}$ represent the orbital coefficients, and the ϵ are the corresponding eigenvalues. Unfortunately, the overlap matrix $\boldsymbol{S}$ is not the unit matrix, since QUICKSTEP employs an atom-centered basis set of non-orthogonal Gaussian-type orbital functions. Thus we have to transform the eigenvalue problem to its special form

$$\boldsymbol{K}\,\mathbf{c} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\,\boldsymbol{c}\,\epsilon \tag{10}$$

$$(\boldsymbol{U}^{\mathrm{T}})^{-1}\,\boldsymbol{K}\,\boldsymbol{U}^{-1}\,\boldsymbol{c}' = \boldsymbol{c}'\,\epsilon \qquad \texttt{(pdsygst)} \tag{11}$$

$$\boldsymbol{K}'\,\boldsymbol{c}' = \boldsymbol{c}'\,\epsilon \qquad \texttt{(pdsyevx or pdsyevd)} \tag{12}$$

using a Cholesky decomposition of the overlap matrix

$$\boldsymbol{S} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} \qquad \texttt{(pdpotrf)} \tag{13}$$

as the default method for that purpose. Now, Eq. 12 can simply be solved by a diagonalization of $\mathbf{K}'$. The orbital coefficients $\boldsymbol{c}$ in the non-orthogonal basis are finally obtained by the back-transformation

$$\boldsymbol{c}' = \boldsymbol{U}\,\boldsymbol{c} \qquad \texttt{(pdtrsm)}. \tag{14}$$

The names in brackets denote the ScaLAPACK routines employed for the respective operation by QUICKSTEP.

Alternatively, a symmetric orthogonalization instead of a Cholesky decomposition can be applied by using

$$\boldsymbol{U} = \boldsymbol{S}^{1/2}. \tag{15}$$

However, the calculation of $\boldsymbol{S}^{1/2}$ involves a diagonalization of $\boldsymbol{S}$ which is computationally more expensive than a Cholesky decomposition. On the other hand, linear dependencies in the basis set introduced by small Gaussian function exponents can be detected when $\boldsymbol{S}$ is diagonalized. Eigenvalues of $\boldsymbol{S}$ smaller than 10^{-5} usually indicate significant linear dependencies in the basis set and a filtering of the corresponding eigenvectors might

help to ameliorate numerical difficulties during the SCF iteration procedure. Both orthogonalization schemes are implemented in QUICKSTEP. For small systems the choice of the orthogonalization has no crucial impact on the performance, since it has to be performed only once for each configuration during the initialization of the SCF run. By contrast, the eigenvectors and eigenvalues of the full Kohn-Sham matrix $\boldsymbol{K}'$ have to be calculated in each iteration step as indicated by Eq. 12 using a divide-and-conquer (`pdsyevd`) scheme or an expert driver (`pdsyevx`) which allows to request only the build of an eigenvector sub-set. The divide-and-conquer scheme is faster than the expert driver, if all eigenvectors have to be computed. However, for the construction of the new density matrix

$$\boldsymbol{P} = 2\,\boldsymbol{c}_{\mathrm{occ}}\boldsymbol{c}_{\mathrm{occ}}^{\mathrm{T}} \tag{16}$$

only the occupied orbitals are needed. In that case the expert driver is superior, since for standard basis sets only 10–20% of the orbitals are occupied and the orthonormalization of the requested eigenvectors is a time-consuming step, especially on parallel computers where it requires heavy communication between the processes.

The TD scheme is usually combined with methods to improve the convergence of the SCF iteration procedure. The most efficient SCF convergence acceleration is achieved by the direct inversion in the iterative sub-space (DIIS) [17, 20] exploiting the commutator relation

$$\boldsymbol{e} = \boldsymbol{K}\,\boldsymbol{P}\,\boldsymbol{S} - \boldsymbol{S}\,\boldsymbol{P}\,\boldsymbol{K} \tag{17}$$

between the Kohn-Sham and the density matrix where the error matrix $\boldsymbol{e}$ is zero for the converged density. The TD/DIIS scheme is an established method for electronic structure calculations. The DIIS method can be very efficient in the number of iterations required to reach convergence starting from a sufficiently pre-converged density which is significant, if the Kohn-Sham matrix construction is much more time consuming than the diagonalization. Nevertheless, the cost for the TD/DIIS scales as $O(M^3)$, where M is the size of the basis set. This implies that, even at fixed system size, increasing the number of basis functions results in a cubic growth of the computational cost. A further disadvantage of the DIIS is that the method might fail to converge or that a sufficiently pre-converged density cannot be obtained. This happens more frequently for electronically difficult systems. For instance spin-polarized systems or systems with a small energy gap between the highest occupied (HOMO) and the lowest unoccupied orbital (LUMO) like semiconductors or metals belong often to this kind of systems.

5.2 Pseudo diagonalization (PD)

Alternatively to the TD scheme, a pseudo diagonalization [17, 18] can be applied as soon as a sufficiently pre-converged wavefunction is obtained. The Kohn-Sham matrix $\boldsymbol{K}^{\mathrm{AO}}$ in the atomic orbital (AO) basis is transformed into the molecular orbital (MO) basis in each

SCF step

$$\boldsymbol{K}^{\mathrm{MO}} = \boldsymbol{c}^{\mathrm{T}} \boldsymbol{K}^{\mathrm{AO}} \boldsymbol{c} \qquad \text{(PDSYMM and PDGEMM)} \tag{18}$$

using the MO coefficients $\boldsymbol{c}$ from the preceding SCF step. The converged $\boldsymbol{K}^{\mathrm{MO}}$ matrix using TD is a diagonal matrix and the eigenvalues are its diagonal elements. Already after a few SCF iteration steps the $\boldsymbol{K}^{\mathrm{MO}}$ matrix becomes diagonal dominant. Moreover, the $\boldsymbol{K}^{\mathrm{MO}}$ matrix shows the following natural blocking

$$\left(\begin{array}{c|c} oo & ou \\ \hline uo & uu \end{array} \right) \tag{19}$$

due to the two MO sub-sets of $\boldsymbol{c}$ namely the occupied (o) and the unoccupied (u) MOs. The eigenvectors are used during the SCF iteration to calculate the new density matrix (see Eq. 16), whereas the eigenvalues are not needed. The total energy only depends on the occupied MOs and thus a block diagonalization which decouples the occupied and unoccupied MOs allows to converge the wavefunctions, i.e. only all elements of the block ou or uo have to become zero, since $\boldsymbol{K}^{\mathrm{MO}}$ is a symmetric matrix. Hence the transformation into the MO basis

$$\boldsymbol{K}^{\mathrm{MO}}_{ou} = \boldsymbol{c}_o^{\mathrm{T}} \boldsymbol{K}^{\mathrm{AO}} \boldsymbol{c}_u \qquad \text{(PDSYMM and PDGEMM)} \tag{20}$$

has only to be performed for that matrix block. Then the decoupling can be achieved iteratively by consecutive 2×2 Jacobi rotations

$$\theta = \frac{\epsilon_q - \epsilon_p}{2\, K^{\mathrm{MO}}_{pq}} \tag{21}$$

$$t = \frac{\mathrm{sgn}(\theta)}{|\theta| + \sqrt{1+\theta^2}} \tag{22}$$

$$c = \frac{1}{\sqrt{t^2+1}} \tag{23}$$

$$s = tc \tag{24}$$

$$\tilde{\boldsymbol{C}}_p = c\, \boldsymbol{C}_p - s\, \boldsymbol{C}_q \qquad \text{(DSCAL and DAXPY)} \tag{25}$$

$$\tilde{\boldsymbol{C}}_q = s\, \boldsymbol{C}_p + c\, \boldsymbol{C}_q \qquad \text{(DSCAL and DAXPY)} \tag{26}$$

where the angle of rotation θ is determined by the difference of the eigenvalues of the MOs p and q divided by the corresponding matrix element K^{MO}_{pq} in the ou or uo block. The

Jacobi rotations can be performed with the BLAS level 1 routines DSCAL and DAXPY. The oo block is significantly smaller than the uu block, since only 10–20% of the MOs are occupied using a standard basis set. Consequently, the ou or uo block also includes only 10–20% of the $\boldsymbol{K}^{\mathrm{MO}}$ matrix. Furthermore, an expensive re-orthonormalization of the MO eigenvectors $\boldsymbol{c}$ is not needed, since the Jacobi rotations preserve their orthonormality.

5.3 Orbital transformations (OT)

Finally, an orbital transformation method [19] is implemented in QUICKSTEP which performs a direct minimization of the wavefunctions. The OT method is guaranteed to converge and it scales, depending on the preconditioner, as $O(MN^2)$, where M is the total number of MOs or basis functions and N is the number of occupied MOs. A detailed description of the OT method is given in Ref. [19]. In the framework of the OT method the electronic energy $E(\boldsymbol{c})$ is minimized using the constraint

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{S}\,\boldsymbol{c} = \boldsymbol{I} \tag{27}$$

where $\boldsymbol{c}$, $\boldsymbol{S}$, and $\boldsymbol{I}$ are the matrix of the orbital coefficients, the overlap matrix, and the identity matrix, respectively. Given the constant start vectors $\boldsymbol{c}_0$ which fulfill the condition

$$\boldsymbol{c}_0^{\mathrm{T}}\boldsymbol{S}\,\boldsymbol{c}_0 = \boldsymbol{I} \tag{28}$$

a new set of vectors $\boldsymbol{c}(\boldsymbol{x})$ is obtained by

$$\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{c}_0\cos(\boldsymbol{U}) + \boldsymbol{x}\,\boldsymbol{U}^{-1}\sin(\boldsymbol{U}) \tag{29}$$

with

$$\boldsymbol{U} = (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\,\boldsymbol{x})^{1/2} \quad \text{and} \quad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\,\mathbf{c}_0 = \mathbf{0} \tag{30}$$

This implies

$$\boldsymbol{c}^{\mathrm{T}}(\boldsymbol{x})\,\boldsymbol{S}\,\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{I} \quad \forall\ \boldsymbol{x} \tag{31}$$

$\boldsymbol{x}$ can be used to optimize the energy $E(\boldsymbol{c}(\boldsymbol{x}))$ with standard methods like conjugate gradient in combination with line search, since the allowed $\boldsymbol{x}$ span a linear space. In that way, the OT method as a direct minimization method addresses both deficiencies of the TD or PD scheme, as the method is guaranteed to converge, and scales, depending on the preconditioner, as $O(MN^2)$. In more detail, the following scalings can be observed for the OT method:

- matrix product sparse-full like $\boldsymbol{S}\,\boldsymbol{X}$: $O(M^2N) \rightarrow O(MN)$
- matrix products full-full like $(\boldsymbol{K}\,\boldsymbol{C})^{\mathrm{T}}\boldsymbol{X}$: $O(MN^2)$
- diagonalization of the $N \times N$ matrix $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{S}\,\boldsymbol{X}$: $< O(N^3)$

The computational cost of the OT method is normally dominated by the computation of the $O(MN)$ terms $\boldsymbol{Hc}$ and $\boldsymbol{Sx}$, but is in principle $O(MN^2)$ with a sparse preconditioner, and $O(M^2N)$, if a non-sparse preconditioner is used. The relative efficiency of TD/DIIS and OT depends on many factors such as system size, basis set size, and network latency and bandwidth.

6 Accuracy

As a first accuracy test for QUICKSTEP, we employed the new basis sets described in section 4 for the geometry optimization of small molecules using the local density approximation (LDA). The CP2K geometry optimizer works with first analytic derivatives whereas the second derivatives are obtained via an updated Hessian method. In that way each molecule of the following test set of 39 small molecules:

> H_2, Li_2, LiH, BH_3, CH_4, C_2H_2, C_2H_4, C_2H_6, N_2, NH_3, HCN, H_2O, H_2O_2, CO, CO_2, CH_3OH, N_2O, F_2, HF, LiF, CH_3F, OF_2, AlH, SiH_4, SiO, P_2, PH_3, HCP, PN, S_3, H_2S, CS, CS_2, SO_2, COS, SF_6, HCl, CH_3Cl, LiCl

consisting of first- and second-row elements was optimized using Cartesian coordinates. Figure 1 compares the optimized bond distances obtained with QUICKSTEP using different basis sets with the NUMOL results of Dickson and Becke [21]. NUMOL is a purely numerical DFT code and thus considered to be free of basis set effects. The smallest basis set DZVP gives on average slightly too long bond distances, but already the TZVP basis set works fine for most of the molecules. Finally, the TZV2P, QZV2P, and QZV3P show an excellent agreement for all bond distances. Figure 2 shows the results for the optimized bond and dihedral angles. The agreement for the small DZVP and the TZVP basis set is already excellent. Only one data point is off which corresponds to the dihedral angle of H_2O_2. This angle is known to be very sensitive to the number of employed polarization functions. Thus one set of polarization functions is insufficient as shown by the results for the DZVP and TZVP basis set. However, for the TZV2P basis set the dihedral angle is already very close to the reference value and for the QZV3P basis set shows more or less a converged result. A comprehensive view of the numerical results of the geometry optimizations is provided by Table 1 which shows the maximum and the root mean square deviation of all bond distances and angle compared to the NUMOL results based on a statistics including 52 bond distances and 18 angles and dihedral angles. The errors become smaller for a growing basis set size as expected. The TZV2P basis set gives already an excellent overall agreement and for the QZV3P most distances coincide within the expected errors. Note, that a fully agreement with the NUMOL values is not possible, since NUMOL uses a slightly different LDA implementation and it employs a frozen core approximation for the elements beyond Beryllium that differs from the GTH pseudo potentials used by QUICKSTEP. These difference may cause a change of the bond distances of about 0.001 Å. This small error also shows that the effect of the pseudo potential is negligible compared to basis

Table 1: Maximum ($\Delta_{\max}$) and root mean square deviation (σ) of bond distances (Å), bond angles, and dihedral angles (°) compared to the NUMOL results for different basis sets.

basis set	distances [Å]		angles [°]	
	$\Delta_{\max}$	σ	$\Delta_{\max}$	σ
DZVP	0.048	0.018	6.4	1.6
TZVP	0.040	0.013	8.5	2.1
TZV2P	0.015	0.006	1.7	0.6
QZV2P	0.012	0.005	2.1	0.6
QZV3P	0.011	0.004	0.7	0.3

set effects concerning structural properties. Thus a basis set can be chosen tuned due to the accuracy requirements of the actual application, but finally the accuracy of QUICKSTEP is determined by the error of the employed exchange-correlation potential.

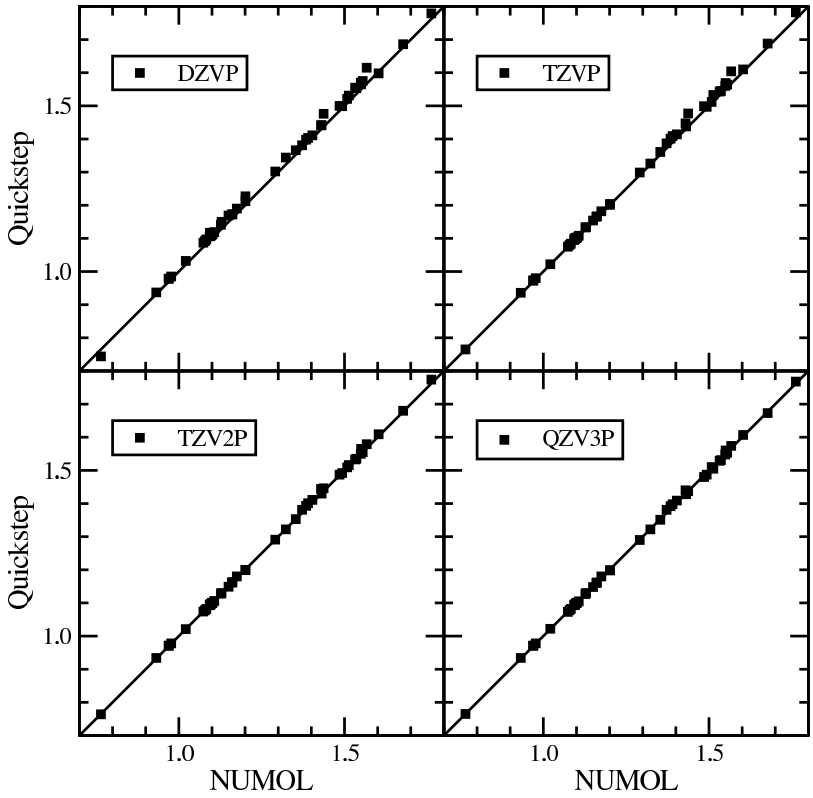

Figure 1: The optimized bond distances for 39 small molecules calculated with QUICKSTEP using different basis sets are compared to the NUMOL results of Dickson and Becke [21].

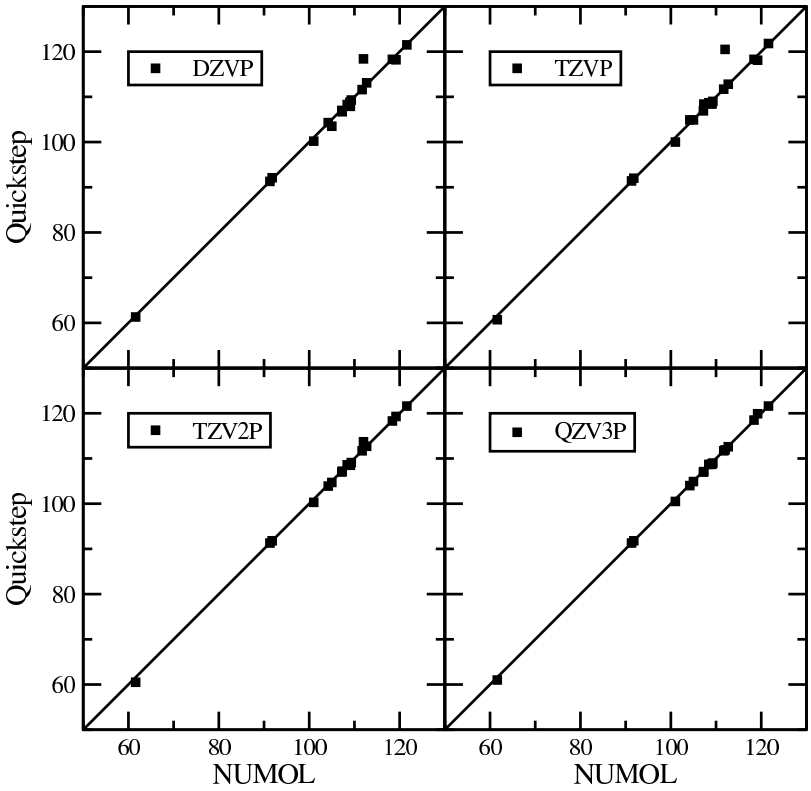

Figure 2: The optimized bond angles and dihedral angles for 39 small molecules calculated with QUICKSTEP using different basis sets are compared to the NUMOL results of Dickson and Becke [21].

7 Benchmarks

After proving the accuracy of QUICKSTEP in the previous section, it will be shown in this section that QUICKSTEP can achieve that accuracy with high computational efficiency. For that purpose, we firstly selected liquid water at ambient conditions as a benchmark system to show both the serial performance of QUICKSTEP and its scalability on a parallel computer. Moreover, we will report the performance results of geometry optimizations for some molecular and a crystalline system.

7.1 Liquid water

Liquid water is often used as a benchmark system, since it can easily be scaled by simply doubling the number of water molecules in the unit cell which is equivalent to a doubling of the unit cell at the same time. For instance, liquid water is employed as a standard benchmark system for the CPMD code [22] to check its performance and scalability on various parallel computers. Furthermore, water is an important ingredient of many bio-chemical applications involving water as the natural solvent and molecular dynamics (MD) simulations are performed to study the properties and the behavior of such systems. Therefore,

Table 2: Detailed characteristics of the employed benchmark systems for liquid water at ambient conditions (300 K, 1 bar). The edge length of the cubic simulation cell, the number of atoms, electrons, Gaussian-type orbitals (M), occupied orbitals (N), and plane waves, i.e. grid points, is listed.

system	cell [Å]	atoms	electrons	M	N	grid points ($\times 10^6$)
32 H_2O	9.9	96	256	1280	128	1.3
64 H_2O	12.4	192	512	2560	256	2.0
128 H_2O	15.6	384	1024	5120	512	4.1
256 H_2O	19.7	768	2048	10240	1024	9.3
512 H_2O	24.9	1536	4096	20480	2048	16.0
1024 H_2O	31.3	3072	8192	40960	4096	32.8

MD runs for pure liquid water at ambient conditions (300 K, 1 bar) were conducted for benchmarking using realistic input parameters as they would also be chosen for production runs. A GTH pseudo potential and a TZV2P basis set for hydrogen and oxygen were employed in all benchmark runs including 40 contracted spherical Gaussian-type orbital functions per water molecule. The high accuracy of the TZV2P basis set was already shown in section 6. Table 2 lists the detailed characteristics of the employed benchmark systems ranging from 32 water molecules in a cubic unit cell of edge length 9.9 Å up to 1024 water molecules in a cubic unit cell of 31.3 Å edge length. These unit cell sizes required up to $32.8 \cdot 10^6$ plane waves, i.e. grid points, as an auxiliary basis set given a density cut-off of 280 Ry for the expansion of the electronic density. This density cut-off was used for all the benchmark calculations of liquid water. Equally, the orbital basis set is linearly growing from 1280 to 40960 Gaussian-type orbital functions. However, the involved matrices like the overlap or Kohn-Sham matrix are growing quadratically for this entity. Thus the Kohn-Sham matrix calculation for 1024 H_2O requires to deal with matrices of the size 40960×40960 and it is therefore indispensable to take advantage of the localized character of the atomic interactions as efficient as possible. Table 3 shows the occupation of the overlap matrix for each benchmark system using a TZV2P basis set and a numerical threshold value of 10^{-12} a.u. for the overlap integral between two primitive Gaussian functions. For the systems with 32 and 64 H_2O each water molecule interacts with each other in the unit cell. Starting from roughly 200 water molecules, the interaction sphere of a water molecule is completely confined in the unit cell, i.e. for larger systems more and more water molecules inside the unit cell do not interact any longer with each other. This can be retrieved from the overlap matrix occupations starting with 256 H_2O, since the occupation is halved for each doubling of the simulation cell. Thus beginning with 256 H_2O in the unit cell the number of interactions grows linearly and similarly the sparsity of the matrices increases continuously. QUICKSTEP takes efficiently advantage of the matrix sparsity, however, this becomes only effective for more than 200 water molecules in the simulation cell. It is further important to recognize that the number of occupied orbitals

Table 3: Occupation of the overlap matrix applying a numerical threshold value of 10^{-12} for the overlap contribution of two primitive Gaussian orbital functions.

system	occupation
32 H_2O	100.0 %
64 H_2O	99.6 %
128 H_2O	85.1 %
256 H_2O	51.3 %
512 H_2O	25.8 %
1024 H_2O	12.9 %

N is significantly smaller than the total number of orbitals M (see Table 2). In this benchmark using the TZV2P basis set only 10 % of the orbitals are occupied. Thus any operation only dealing with the occupied orbitals (MN) is favorable compared to (M^2) for the full matrix. This is a crucial performance issue when comparing the eigensolvers implemented in QUICKSTEP. Figure 3 shows the timings for the benchmark systems of Table 2 using the IBM Regatta p690+ system at the Research Centre Jülich, called Jump. The Jump system consists of 39 compute nodes. Each node provides 32 Power4+ (1.7 GHz) processors. The processors are interconnected by an IBM High Performance Switch[1] (HPS). The results are given using a double logarithmic scale to show the scaling of the TD and the PD scheme. Each MD step included a full wavefunction optimization followed by a calculation of the forces on each atom. The total energy of the system was converged to 10^{-7} a.u. and the deviation of the electron count for the converged density was less than 10^{-5}. Ten MD steps were performed for each benchmark system (except 1024 H_2O) using a time step of 0.5 fs. The CPU timings of the last 5 MD steps were averaged. Figure 3 displays the obtained CPU timings per MD step for various CPU numbers and system sizes using the TD and the PD scheme. The missing data points are due to the limited memory per CPU which did not allow to run larger systems using only a small number of CPUs. The small systems with 32 and 64 H_2O can efficiently be run on a small number of CPUs. 64 H_2O need roughly one CPU minute per MD step, i.e. 2 CPU minutes per fs simulation time, when using 16 CPUs. The larger systems with 128 and 256 H_2O run efficiently on 32 and 64 CPUs, respectively. However, 14 minutes per MD step for 256 H_2O does not allow to obtain appropriate MD trajectories in reasonable time. It was not possible to run 512 H_2O, even if using 256 CPUs, since the TD scheme which is based on ScaLAPACK/BLACS requires to deal with a distribution of several full matrices during the SCF procedure exceeding the available memory.

A direct comparison of the two panels of Figure 3 shows that the PD scheme scales slightly

[1]This benchmark was run on the Jump system before the major software update (PTF7) in July 2004 which improved the MPI communication performance significantly.

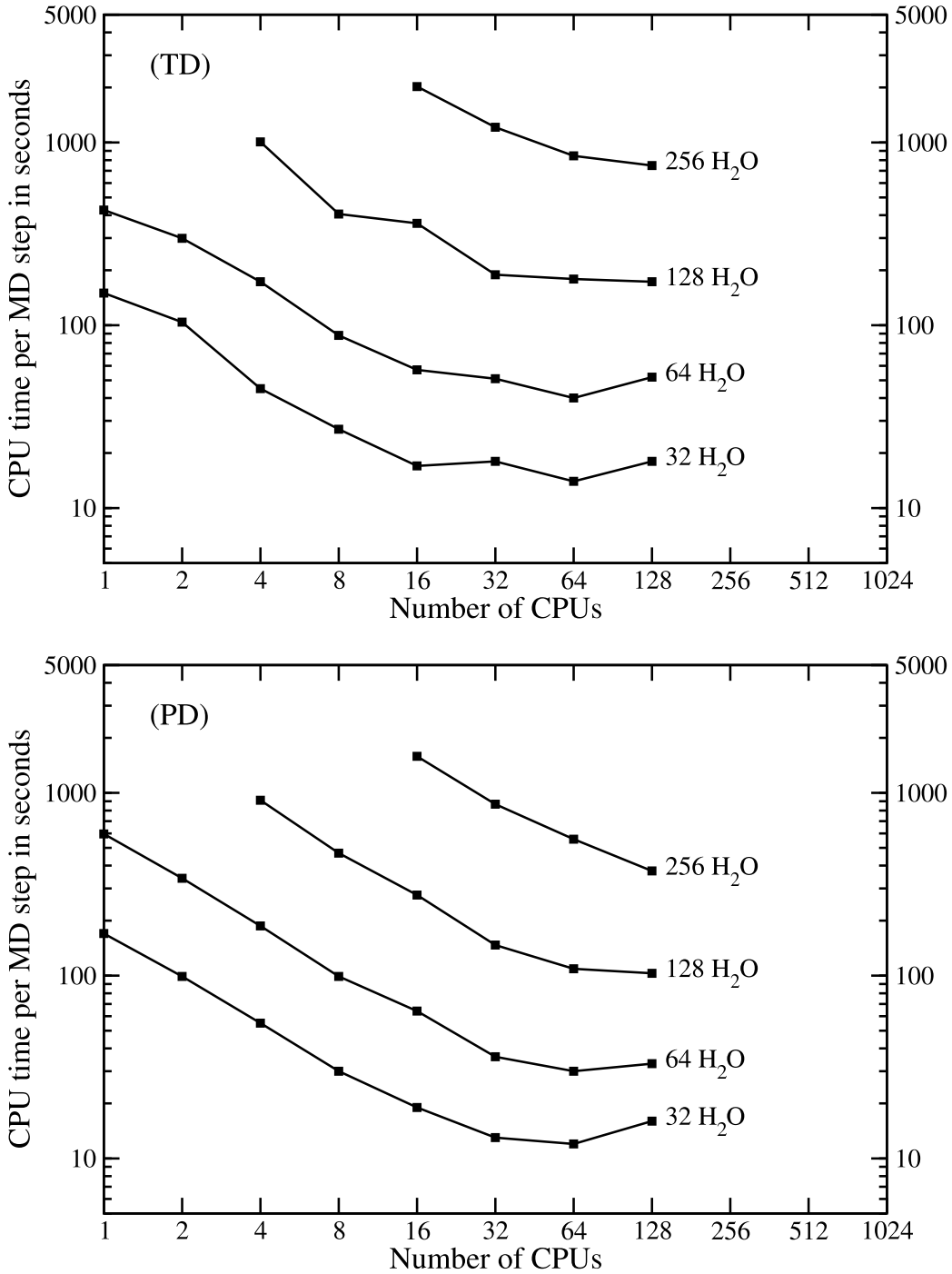

Figure 3: Scaling of the CPU time per MD step using the traditional diagonalization (TD) scheme and the pseudo diagonalization (PD) scheme for the benchmarks systems of Table 2. The calculations were performed on an IBM Regatta p690+ system with 32 Power4+ (1.7 GHz) per node interconnected by an IBM High Performance Switch (HPS).

better than the TD scheme. The small systems with 32 and 64 H_2O scale up to 32 CPUs and the largest system with 256 H_2O scales even up to 128 CPUs using the PD scheme. However, the absolute CPU times per MD step for the TD and the PD scheme are very close, even if the PD scheme requires less communication than the TD scheme. The PD scheme shows only for 256 H_2O on average significant shorter CPU times per MD step compared to the TD scheme. As described in section 5, the PD scheme can only be applied to sufficiently pre-converged wavefunctions. The TD scheme is employed until this convergence is achieved and thus no speed-up with respect to the TD scheme is obtained for the first SCF iteration steps. Furthermore, the initialization of the PD scheme requires at least once a diagonalization of the Kohn-Sham matrix including the calculation of *all* eigenvectors. This step turns out to be rather expensive. It is known that the orthonormalization of a large eigenvector set is computationally expensive step that involves a lot of communication. In fact, this SCF step may consume two or three times more CPU time than a normal TD SCF step and turns out to be a bottleneck for the larger systems. However, once the PD scheme is set up, the following iteration steps are less expensive than a TD step. Moreover, the PD steps are becoming cheaper and cheaper, since the number of matrix elements which have to be processed by the Jacobi rotations decrease continuously. However, a typical MD step only involves approximately 8 SCF iteration steps and at least two or three of these are normal TD steps followed by an expensive TD step providing the full eigenvector set. Thus there are only four or five SCF steps left for the faster PD scheme and finally nothing is gained compared to the pure TD scheme for most of the test systems.

By contrast, the OT method shows a much better performance as shown in Figure 4. The OT method needs less memory than the TD and the PD scheme, because it does not deal with full matrices during the SCF iteration and therefore it allows to run larger benchmark systems with up to 1024 water molecules in the unit cell. Also the scaling behavior of the OT method is much better. The small systems with 32 and 64 H_2O scale nicely up to 32 CPUs. A scaling beyond 32 CPUs cannot be expected, since the data blocks per CPU become too small to keep an SP4+ processor efficiently busy and the calculation will be completely dominated by the communication between the processes. At least one or two H_2O molecules per CPU are needed, formally. Also the larger benchmark systems show a better scaling with OT as indicated by the slope. The 512 H_2O system shows a continuous scaling up to 128 CPUs including 4 compute nodes of the Jump system. This shows that the scaling behavior of QUICKSTEP is also preserved when the processors of more than one compute node are employed.

7.2 Molecular and crystalline systems

As a final performance test for QUICKSTEP, geometry optimizations for a couple of molecular and crystalline systems were performed. The detailed characteristics of the employed molecular and crystalline benchmark systems is listed in Table 4. The DZVP basis set described in section 4 was used for all elements including hydrogen, even if the p-type

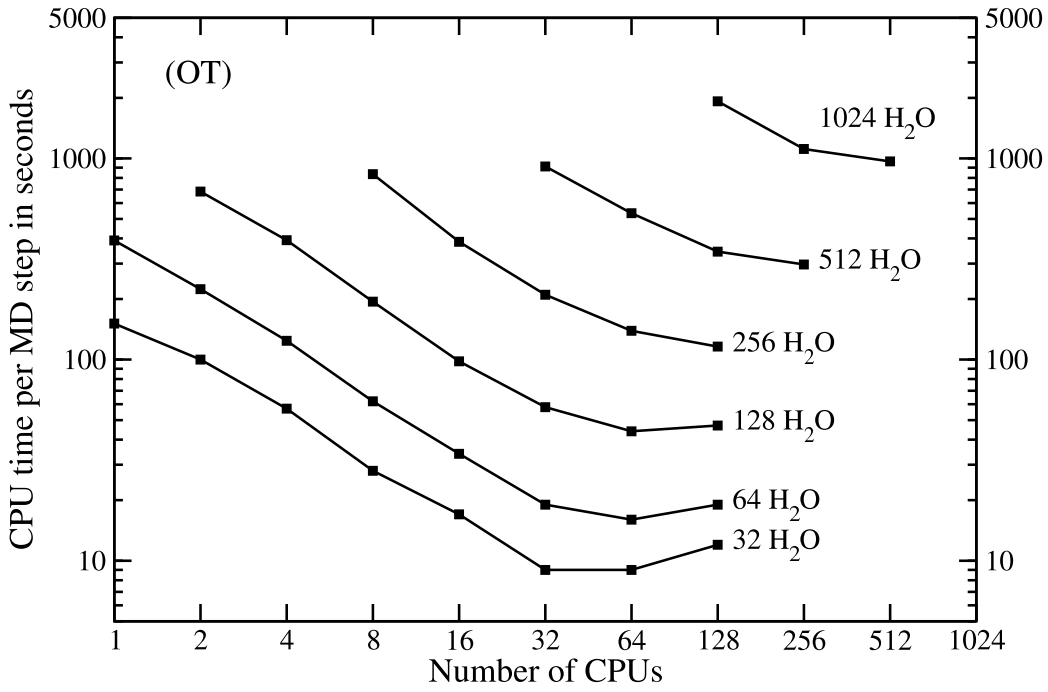

Figure 4: Scaling of the CPU time per MD step using the orbital transformation (OT) scheme for the benchmarks systems of Table 2. The calculations were performed on an IBM Regatta p690+ system with 32 Power4+ (1.7 GHz) per node interconnected by an IBM High Performance Switch (HPS).

polarization functions for hydrogen are not needed in most cases. It turned out that the quality of the DZVP basis set is sufficient for most of the applications. Optionally, a refinement of the structure can be obtained with the TZV2P basis set based on the structure pre-optimized with the DZVP basis set. It was shown in section 6 that the TZV2P basis set provides structures of high accuracy within the actual density functional approximation. The density cut-off for the plane waves expansion of the electronic density was chosen sufficiently large, i.e. in correspondence with the largest Gaussian function exponent of the

Table 4: Detailed characteristics of the employed molecular and crystalline benchmark systems. The number of atoms, electrons, Gaussian-type orbitals (M), and occupied orbitals (N) is listed. The employed exchange-correlation functional is given in brackets.

system	atoms	electrons	M	N	Cut-off [Ry]
C_{60} fullerene (LDA)	60	240	780	120	240
C_{180} fullerene (LDA)	180	720	2340	360	240
Grubbs catalysator (BP)	120	284	774	142	280
Taxol (BLYP)	113	328	908	164	280
[2]Catenan (BLYP)	164	460	1524	230	280
RNA duplex (BLYP)	368	1192	3444	596	320

Table 5: CPU time per geometry optimization step for the molecular and crystalline benchmark systems as described in Table 4. The calculations were performed on an IBM Regatta p690+ system with 32 Power4+ (1.7 GHz) per node interconnected via an IBM High Performance Switch (HPS).

system	4 CPUs	8 CPUs	16 CPUs
C_{60} fullerene	30	11	8
C_{180} fullerene	115	69	36
Grubbs catalysator	178	108	63
Taxol	208	118	74
[2]Catenan	246	138	92
RNA duplex	432	186	128

orbital basis set based on the accuracy of the computed electron count. The OT method was employed in all optimization runs.

C_{60} is the well-known hollow, soccer ball shaped molecule called buckminsterfullerene or simply bucky ball. The C_{180} fullerene is a bigger variety of the C_{60} which is also a hollow ball structure. Figure 5 shows the ruthenium based olefin metathesis catalysts also called after its inventor Grubbs catalysator. Taxol (see Figure 6) is a compound which is used as an anti-cancer drug. The [2]Catenan [23] is an electronically reconfigurable molecular switch which consists of two interlocked rings: (1) a tetracationic cyclophane that incorporates two bipyridinium units and (2) a crown ether containing a tetrathiafulvalene unit and a 1,5-dioxynaphthalene ring system located on opposite sides of the crown ether (see Figure 7). Finally, Figure 8 shows the unit cell of a fully hydrated RNA duplex crystal structure [24]. For all the molecular systems a sufficiently large unit cells were chosen to eliminate the interaction with the images. GTH pseudo potentials were employed for all structure optimization runs. For ruthenium and sodium the semi-core pseudo potential versions were used including 16 and 9 valence electrons, respectively.

The CPU times per geometry optimization step are listed in Table 5 using 4, 8, and 16 CPUs of one compute node of the Jump system at the Research Centre Jülich. Each geometry optimization step includes like an MD step a full wavefunction optimization followed by a calculation of the forces on all atoms. The timings depend not only on the size of the orbital basis set, but also on the selected exchange-correlation functional and the density cut-off. For instance, the C_{180} fullerene has a large orbital basis, however, the pseudo potential of carbon is softer than the pseudo potential of oxygen and thus it requires only the relatively small density cut-off of 240 Rydberg. Moreover, the gradient of the electronic density has not to be calculated in the framework of the local density approximation (LDA), whereas this is needed for the exchange-correlation functionals BLYP [9, 10, 11] and BP [9, 12] (see section 3) based on the generalized gradient approximation (GGA). A couple of geometry optimization steps can be performed for all the presented systems within the limits of an

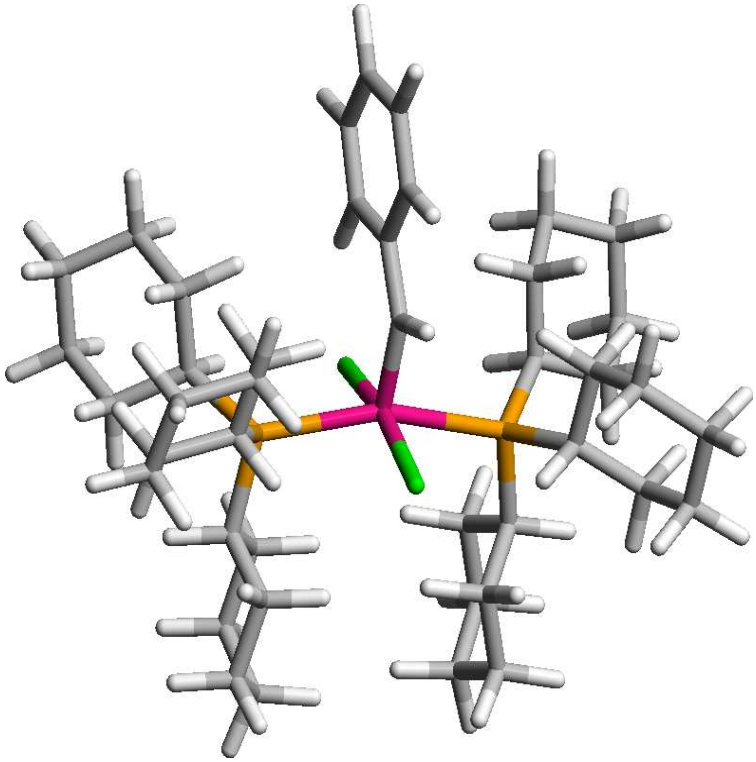

Figure 5: Grubbs catalysator ($RuC_{43}H_{72}P_2Cl_2$)

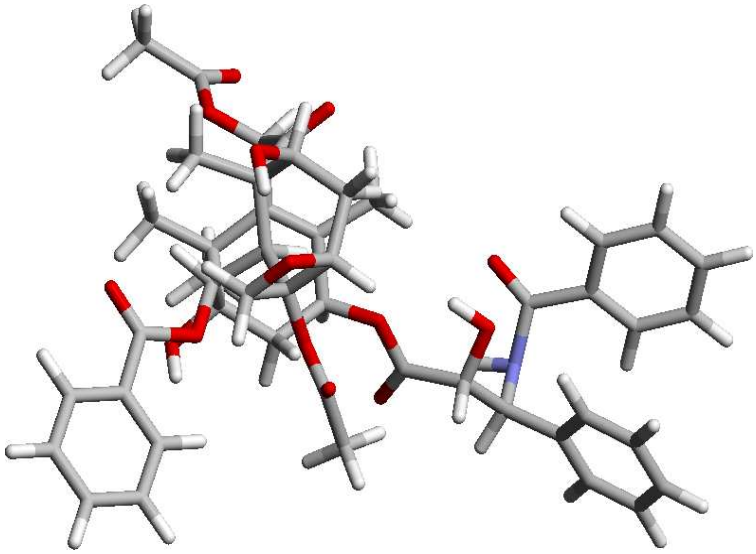

Figure 6: Taxol ($C_{47}H_{51}O_{14}N$)

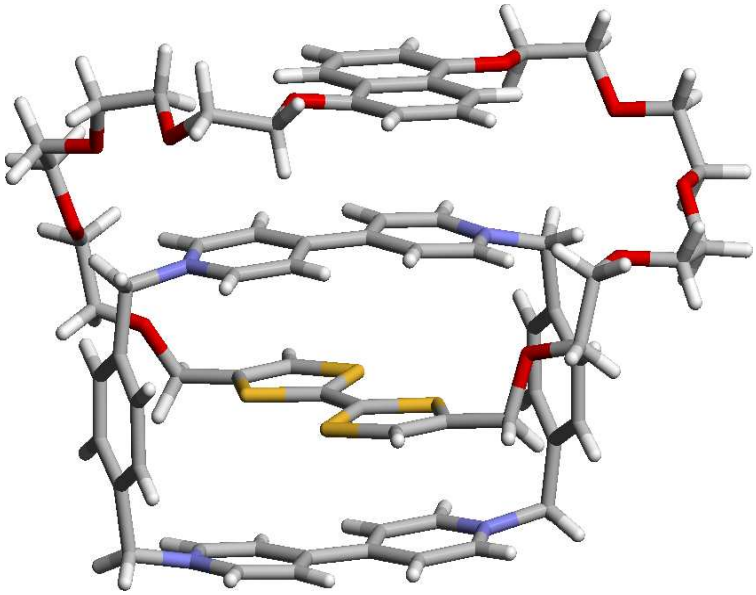

Figure 7: [2]Catenan ($C_{70}H_{76}O_{10}N_4S_4$)

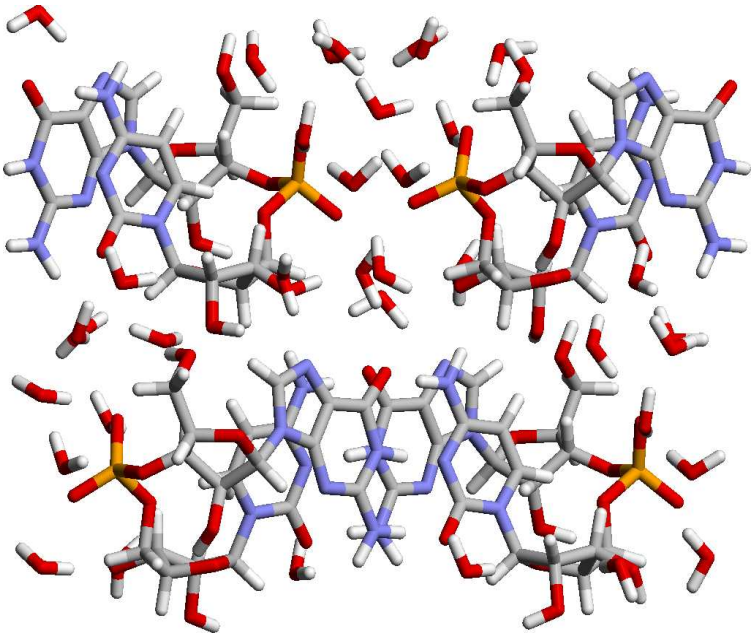

Figure 8: Unit cell of the fully hydrated RNA duplex ($C_{76}H_{168}N_{32}O_{84}Na_4P_4$) crystal structure

interactive job on the Jump system which provides up to 16 CPUs for 30 minutes. In that way, QUICKSTEP allows to optimize efficiently the structure of small and medium-sized molecular or crystalline systems.

8 Summary and outlook

It was shown that QUICKSTEP allows for fast and accurate density functional calculations of molecules and condensed phase systems. It provides the basic functionality needed to perform structure optimizations and to run Born-Oppenheimer molecular dynamics simulations. The nice scaling behavior of QUICKSTEP was proved using the new IBM parallel computer system Jump at the Forschungszentrum Jülich. The efficient parallelization of QUICKSTEP allows to obtain results in shorter time or to investigate larger systems.

Nevertheless, there are many possibilities to improve further the efficiency and functionality of QUICKSTEP. The extension of the GPW method to the Gaussian augmented plane waves (GAPW) method [25] will significantly speedup the calculations. Moreover, the GAPW method will also allow to perform all-electron density functional calculations [26].

Acknowledgment

The current state of the QUICKSTEP project is the result of the contributions from the whole CP2K developers team and in particular the QUICKSTEP developers, namely J. VandeVondele (University of Cambridge), T. Chassaing and J. Hutter (University of Zürich), and F. Mohamed and M. Krack (ETH Zürich).

Bibliography

[1] G. Lippert, J. Hutter, and M. Parrinello, *Mol. Phys.* **92**, 477 (1997).

[2] HPC-Chem (High Performance Computing in der Chemie), BMBF-Verbundprojekt, `http://www.fz-juelich.de/zam/hpc-chem`.

[3] The CP2K developers group, `http://cp2k.berlios.de`, 2004.

[4] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, *Comput. Phys. Comm.*, submitted (2004).

[5] P. Hohenberg and W. Kohn, *Phys. Rev. B* **136**, B864 (1964).

[6] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).

[7] S. Goedecker, M. Teter, and J. Hutter, *Phys. Rev. B* **54**, 1703 (1996).

[8] C. Hartwigsen, S. Goedecker, and J. Hutter, *Phys. Rev. B* **58**, 3641 (1998).

[9] A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).

[10] C. T. Lee, W. T. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).

[11] B. Miehlich, A. Savin, H. Stoll, and H. Preuss, *Chem. Phys. Lett.* **157**, 200 (1989).

[12] J. P. Perdew, *Phys. Rev. B* **33**, 8822 (1986).

[13] F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy, *J. Chem. Phys.* **109**, 6264 (1998).

[14] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).

[15] T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).

[16] D. E. Woon and T. H. Dunning, *J. Chem. Phys.* **98**, 1358 (1993).

[17] P. Pulay, *J. Comput. Chem.* **3**, 556 (1982).

[18] M. J. Fengler, in *Beiträge zum Wissenschaftlichen Rechnen – Ergebnisse des Gaststudentenprogramms 2002 des John von Neumann-Instituts für Computing*, Technical Report IB-2002-12, edited by R. Esser, 119 (2002).
`http://www.fz-juelich.de/zam/gaststudenten/ib-2002-12.pdf`

[19] J. VandeVondele and J. Hutter, *J. Chem. Phys.* **118**, 4365 (2003).

[20] P. Pulay, *Chem. Phys. Lett.* **73**, 393 (1980).

[21] R. M. Dickson and A. D. Becke, *J. Chem. Phys.* **99**, 3898 (1993).

[22] CPMD, Version 3.7, copyright IBM Corp. 1990–2003, copyright MPI für Festkörperforschung Stuttgart 1997-2001; `http://www.cpmd.org`.

[23] C. P. Collier, G. Mattersteig, E. W. Wong, Y. Luo, K. Beverly, J. Sampaio, F. M. Raymo, J. F. Stoddart, and J. R. Heath, *Science* **289**, 1172 (2000).

[24] J. Hutter, P. Carloni, and M. Parrinello, *J. Am. Chem. Soc.* **118**, 8710 (1996).

[25] G. Lippert, J. Hutter, and M. Parrinello, *Theor. Chem. Acc.* **103**, 124 (1999).

[26] M. Krack and M. Parrinello, *Phys. Chem. Chem. Phys.* **2**, 2105 (2000).