Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn

Wegelerstraße 6 • 53115 Bonn • Germany
phone +49 228 73-3427 • fax +49 228 73-7527
www.ins.uni-bonn.de

V. Weijo, M. Randrianarivony, H. Harbrecht, and L. Frediani

Wavelet formulation of the Polarizable Continuum Model

INS Preprint No. 0906

April 2009

WAVELET FORMULATION OF THE POLARIZABLE CONTINUUM MODEL

V. WEIJO, M. RANDRIANARIVONY, H. HARBRECHT, AND L. FREDIANI

Abstract. The first implementation of a wavelet discretization of the Integral Equation Formalism (IEF) for the Polarizable Continuum Model (PCM) is presented here. The method is based on the application of a general purpose wavelet solver on the cavity boundary to solve the integral equations of the IEF-PCM problem. Wavelet methods provide attractive properties for the solution of the electrostatic problem at the cavity boundary: the system matrix is highly sparse and iterative solution schemes can be applied efficiently; the accuracy of the solver can be increased systematically and arbitrarily; for a given system, discretization error accuracy is achieved at a computational expense that scales linearly with the number of unknowns. The scaling of the computational time with the system size N is formally quadratic but a $N^{1.5}$ scaling has been observed in practice. The current bottleneck is the evaluation of the potential integrals at the cavity boundary which scales linearly with the system size. To reduce this overhead, interpolation of the potential integrals on the cavity surface has been successfully employed.

1. Introduction

boundary between two different dielectrics is of broad interest in science and engineering. In quantum chemistry it is often used to obtain the electrostatic solvent reaction field on a molecular system. One such method is widely known as the Polarizable Continuum Model (PCM), which was first introduced by Miertus, Scrocco and Tomasi [1] in 1981. The goal was to be able to reproduce the electrostatic interaction between a molecular solute and a continuum solvent in a faithful way using information on a cavity surface only. What distinguished PCM from earlier models was the ability to deal with molecular-shaped cavities instead of spherical or ellipsoidal cavities [2, 3]. In order to tackle such a problem, it was necessary to abandon attempts to describe the electrostatic interaction analytically as most previous models did. An engineering approach was adopted by making use of the Boundary Element Method (BEM) [4]. Within the BEM approach, the molecular surface is discretized into portions called *tesserae* and the boundary condition of the electrostatic equations can then be rewritten as a conceptually simple matrix-vector problem where a vector of "unknowns", the Apparent Surface Charges (ASCs), is connected to known quantities, namely the electrostatic potential and the electrostatic field normal to

the cavity surface. The method has known a rapid growth and a widespread success and it is nowadays implemented in different "flavors" in several quantum chemistry packages [5, 6, 7, 8].

The two central components of the model are the cavity construction and the theoretical framework for the resolution of the electrostatic problem. The cavity is generally constructed as a set of interlocking spheres. The exposed portion of the spheres is then divided into tesserae [9, 10, 11]. A more elaborate but also more accurate description employs the Connolly surface in order to faithfully represent the Solvent Excluded Surface (SES) by "rolling" a sphere representing a solvent molecule over the cavity [12]. However, the realization of this simple idea is all but trivial. Among the challenges are the determination of the exposed surface [9, 13], a regular subdivision of the surface [14], the modeling of the SES [13, 12, 9], the differentiation of the surface with respect to geometrical modifications of the cavity [15, 16, 17], and the fulfillment of the molecular symmetry requirements [18, 19].

The resolution of the electrostatic problem is addressed by a convenient discretization of the underlying equations. We will here focus on the most recent and versatile version of the PCM, known as the Integral Equation Formalism for the Polarizable Continuum Model (IEF-PCM), where the problem is considered as a special case of a class of integral equations [20]. In this way it is actually possible to extend the applicability of the method to all cases where Green's function for the considered environment is known. This has been done for liquid crystals (anisotropic permittivity) [21], ionic solution (screened electrostatic potential) [20], sharp planar interfaces (image-charge approach) and diffuse planar interfaces (numerical integration) [22]. The common trait of all such models is the transformation of the integral equation into a matrix-vector equation through the discretization of the operator kernels and the functions supported on the cavity surface [20].

Despite all the evolutions from the original PCM to the state-of-the-art IEF-PCM, the method has always presented two main problems in all its implementations which are deeply rooted in its construction: one is the tessellation of the cavity and the other is the discretization of the operator. The main underlying idea behind the tessellation and consequent discretization of the operator is that this should ideally lead to a problem that systematically approximates the exact solution: the smaller the tesserae the more accurate the solution. On the other hand, the discretized operators include the inverse of the tessera areas leading to potential numerical instabilities. The discretized operator is also a potentially dense matrix thus not suited for fast numerical approaches. In other words, the refinement of the tessellation leads to a problem which becomes both numerically unstable and computationally expensive.

Moreover, the error committed by discretizing the equations is not known: even disregarding numerical instabilities one can only assume convergence to the exact solution but the actual error cannot be estimated. Several developments have been investigated in order to overcome the mentioned difficulties connected to the BEM resolution of the electrostatic problem at the cavity boundary. Scalmani *et al.* [23] have achieved almost linear scaling with the system size by employing a Fast Multiple Method technique to solve the linear system of equations iteratively. The transformation of the integral operators into a matrix form has also been deeply investigated. Purisima *et al.* [24] have shown that the evaluation of the diagonal elements plays a crucial role in the overall accuracy of the system. Recently Bardhan [25] has shown that by changing the integration order in the quadrature evaluation of the matrix elements, it is possible to obtain a substantial improvement of the accuracy.

In recent years, wavelet methods have been developed to solve a wide variety of interesting problems, including the integral equations that constitute the basis for the IEF-PCM. Wavelet methods lead to sparse operators thus allowing for fast, linearly scaling numerical solvers, accurate error control, and stability of the numerical solution. The possibility to employ wavelet-based methods for the resolution of the IEF-PCM problem was, however, limited by the lack of a suitable mesh generator for constructing nested meshes on molecular surfaces. This problem has been recently solved by Harbrecht and Randrianarivony [26, 27] who developed the necessary transformations to parametrize a molecular cavity made of interlocking spheres by four-sided patches. It has therefore been possible to combine the IEF-PCM with the wavelet-based solver developed by Harbrecht and coworkers [28, 29]

The remainder of this paper is organized as follows. In Sec. 2 the theoretical framework of the wavelet formulation of the IEF-PCM problem is presented. In Sec. 3 the results obtained with the present implementation are shown and discussed, and in Sec. 4 some conclusions are drawn with emphasis on the future development of the method.

2. Theory

2.1. **The IEF-PCM.** The IEF-PCM is a formalism to deal with the electrostatic problem of a charge distribution ρ placed inside a cavity C of a dielectric medium defined by macroscopic properties such as the dielectric permittivity or the ionic strength. The electrostatic potential generated by ρ will polarize the dielectric medium and the polarization can be described by an apparent surface charge density σ defined on the cavity boundary. The formal solution of the problem can be written in terms of the Poisson equation inside and outside the cavity boundary together

with the corresponding boundary conditions. For a uniform dielectric medium with relative permittivity ϵ it follows that

$$\begin{cases} \Delta V(r) = -4\pi\rho(r), & r \in C \\ \Delta V(r) = 0, & r \notin C \\ \lim_{r\to\Gamma^+} V(r) = \lim_{r\to\Gamma^-} V(r) \\ \epsilon \lim_{r\to\Gamma^+} \frac{\partial V(r)}{\partial n} = \lim_{r\to\Gamma^-} \frac{\partial V(r)}{\partial n}, \end{cases} \tag{2.1}$$

where the first two relations in (2.1) represent the Poisson equation inside and outside the cavity for the electrostatic potential $V(r)$ with a charge distribution $\rho(r)$, the third one is the continuity condition of the electrostatic potential on the cavity boundary Γ, and the last one is the jump condition of the electrostatic field normal to the cavity surface. We will use SI based atomic units throughout this work.

The problem can be solved by making use of the theory of integral equations: Cances and Mennucci [20] showed that the apparent surface charge $\sigma(s)$ can be found as the unique solution of the following equation:

$$\left[\left(\frac{1}{2} - D_e\right) S_i + S_e \left(\frac{1}{2} - D_i^*\right)\right] \sigma(s) = (D_e - \frac{1}{2})V(r) - S_e \frac{\partial V}{\partial n}. \tag{2.2}$$

In the above equation S_i, D_i and D_i^* correspond to the following Calderon operators:

$$(S_i \cdot u)(x) = \int_\Gamma G_i(x,y)u(y)\,\mathrm{d}y \tag{2.3}$$

$$(D_i \cdot u)(x) = \int_\Gamma \partial_y G_i(x,y)u(y)\,\mathrm{d}y \tag{2.4}$$

$$(D_i^* \cdot u)(x) = \int_\Gamma \partial_x G_i(x,y)u(y)\,\mathrm{d}y \tag{2.5}$$

where $G_i(x,y)$ is the Green's function $G_i(x,y) = 1/|x-y|$ corresponding to the differential operator Δ, whereas the corresponding operator with subscript e are obtained by making use of the Green's function for the operator describing the electrostatic problem outside the cavity: for example, for uniform, homogeneous dielectrics $G_e(x,y) = G_i(x,y)/\epsilon$.

In the traditional IEF-PCM approach the problem is solved numerically by discretizing Eq. (2.2) to a matrix form. Once the equation has been solved, the interaction energy between the ASCs and a charge density ρ' can be obtained:

$$E = \int_\Gamma V_{\rho'}(s)\sigma(s)\,\mathrm{d}s. \tag{2.6}$$

The integral above can be solved by quadrature and, in practice, the simple rectangle rule is employed:

$$E = \sum_i V_i \sigma_i a_i, \tag{2.7}$$

where a_i is the surface area associated with a point surface charge σ_i.

2.2. **The Quantum Mechanical Problem.** The inclusion of electrostatic solvation effect into quantum mechanical calculations leads to the formulation of the problem in terms of a nonlinear Hamiltonian H_{eff} which includes the potential due to the dielectric polarization:

$$H_{eff} = H_0 + V_\rho, \tag{2.8}$$

where H_0 is the Hamiltonian of an isolated molecule and V_ρ is the potential due to the polarization of the dielectric as described by Eq. (2.7).

A detailed exposition of the whole quantum mechanical problem is beyond the scope of the present work and the interested reader is referred to the literature on the subject [5, 6, 7]. We will limit the discussion to the aspects that are relevant in connection to the present work. In particular, we will consider the electronic problem of a molecular system in the clamped nuclei (Born–Oppenheimer) approximation. Moreover, we will use self-consistent field (SCF) method in connection with a mean-field Hartree–Fock (HF) theory. The current formalism is readily usable also in connection with the Kohn–Sham density functional theory.

The partition of the effective Hamiltonian in Eq. (2.8) induces a corresponding partition in the Fock matrix as follows:

$$F = F_0 + 1/2(J + Y) + X + U_{NN} \tag{2.9}$$

where F_0 is the Fock matrix for an isolated system. The four solvent polarization contribution are obtained by the usual partition of the molecular density into electronic and nuclear part, thus leading to a pure nuclear term U_{NN}, a pure electronic term X, and two cross terms Y and J.

By making use of the second quantization formalism [30], the solvent operators can be written in the following way [8]:

$$\hat{J} = \int_\Gamma \hat{V}^e(s)\sigma^N(s)\,\mathrm{d}s = \sum_{pq} \hat{E}_{pq} \int_\Gamma V^e_{pq}(s)\sigma^N(s)\,\mathrm{d}s = \sum_{pq} J_{pq}\hat{E}_{pq} \tag{2.10}$$

$$\hat{Y} = \int_\Gamma V^n(s)\hat{q}^e(s)\,\mathrm{d}s = \sum_{pq} \hat{E}_{pq} \int_\Gamma V^N(s) q^e_{pq}(s)\,\mathrm{d}s = \sum_{pq} Y_{pq}\hat{E}_{pq} \tag{2.11}$$

(2.12)
$$\hat{X}(0) = \int_\Gamma \hat{V}^e(s) \cdot \langle 0 \,|\hat{\sigma}^e(s)|\, 0 \rangle \, \mathrm{d}s = \sum_{pq} \hat{E}_{pq} \int_\Gamma V^e_{pq}(s) \, \langle 0 \,|\hat{\sigma}^e(s)|\, 0 \rangle \, \mathrm{d}s = \sum_{pq} X_{pq}(0) \hat{E}_{pq},$$

where the expectation value of $\hat{\sigma}^e$ is to be understood as the ASCs generated by the same ground-state electron density as in the operator $\hat{V}^e$.

In order to illustrate how the IEF-PCM is practically employed in an SCF calculation, let us consider Eq. (2.12). The operator $\hat{X}(0)$ represents the interaction of the electrons with the ASC density they generate. This is explicitly shown in the second expression, where the first term represents the electronic potential operator and the second one is the expectation value of the electronic ASC. The last expression shows explicitly the separation between the singlet creation operators and the corresponding coefficients which are obtained by making use of IEF-PCM: the formal solution is obtained by integrating the potential produced by the overlap charge distribution $(\chi_r \chi_s)$ and the ASC produced by the total electronic density ρ on the cavity boundary. Traditionally, the solution has been obtained in the following way. For each tessera τ on the cavity boundary one representative point s_τ is considered; the potential is sampled on all such points both for the electronic density and for each overlap distribution $(\chi_r \chi_s)$; a discretized version of Eq. (2.2) is used to obtain the corresponding sampling of the apparent surface charge density; finally the integral in Eq. (2.12) is computed for each overlap distribution to obtain the operator.

2.3. **The Wavelet IEF-PCM.** Boundary integral equations are solved in general by the boundary element method. BEM is a well established tool in PCM [20, 22, 5]. However, traditional discretizations lead to densely populated and possibly ill-conditioned system matrices. Both features pose serious obstructions to the efficient numerical treatment of the IEF-PCM.

As first observed by Beylkin et al. [31], a Galerkin discretization with wavelet bases results in quasi-sparse matrices, i.e., most matrix entries are negligible and can be treated as zero. Discarding these non-relevant matrix entries is called *matrix compression*. It has been proven in Refs. [28, 32] that for a fixed surface the number of significant matrix coefficients scales only linearly in the number of ansatz functions. The compressed system matrix can be assembled within linear complexity if an appropriate *hp*-quadrature algorithm is used [29]. Therefore, involving wavelet preconditioning [33, 32], we arrive at an algorithm which solves a given boundary integral equation within discretization error accuracy, offered by the underlying Galerkin method, at a computational expense that is proven to stay proportional to the number of unknowns.

The wavelet construction starts with a hierarchy of nested trial spaces

$$\{0\} =: V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots, \qquad V_j = \operatorname{span}\{\varphi_{j,k} : k \in \triangle_j\}.$$

Instead of using only a single-scale j the idea of wavelet concepts is to keep track of the increment of information between two adjacent scales j and $j+1$. Since $V_{j-1} \subset V_j$ one decomposes $V_j = V_{j-1} \oplus W_j$ with some complementary space

$$W_j = \operatorname{span}\{\psi_{j,k} : k \in \nabla_j := \Delta_j \setminus \Delta_{j-1}\},$$

$W_j \cap V_{j-1} = \{0\}$, not necessarily orthogonal to V_{j-1}. Recursively one obtains the multiscale decomposition $V_j = \bigoplus_{\ell=0}^{j} W_\ell$ and thus the wavelet basis $\Psi_j = \{\psi_{\ell,k}\}_{k\in\nabla_\ell,\,\ell\le j}$.

The core properties of the wavelet basis are the vanishing moments, i.e.,

$$\int x^\alpha \psi_{\ell,k}(x) \underset{\cdot}{\mathrm{x}} = 0 \quad \text{for all} \quad |\alpha| < \widetilde{d}. \tag{2.13}$$

By using biorthogonal wavelet bases, the order $\widetilde{d}$ of vanishing moments can be chosen higher than the related approximation order. This is essential for deriving linear scaling and could not be realized by requiring orthonormality of the basis alone [28, 32].

General integral operators with kernel $G(\cdot,\cdot)$ satisfy estimates of the type

$$\left|\partial_x^\alpha \partial_y^\beta G(x,y)\right| \le c_{\alpha,\beta} |x-y|^{-2-2t-|\alpha|-|\beta|}, \tag{2.14}$$

where respectively $t = -1/2$ and $t = 0$ in case of single layer operators S_i/S_e and double layer operators D_i/D_e. By combining (2.13) with (2.14) we arrive at the essential estimate

$$\left| \int_\Gamma \int_\Gamma G(x,y) \psi_{\ell,k}(x) \psi_{\ell',k'}(y) \,\mathrm{d}x \,\mathrm{d}y \right| \le C \frac{2^{-(\ell+\ell')(1+\widetilde{d})}}{\operatorname{dist}(\operatorname{supp}\psi_{\ell,k}, \operatorname{supp}\psi_{\ell',k'})^{2(1+t+\widetilde{d})}}$$

which is the main foundation of compression estimates [28].

The construction of appropriate wavelets on surfaces, satisfying (2.13), is based on a parametric representation of the cavity surface. More precisely, let $\square := [0,1]^2$ denote the unit square. The surface Γ is partitioned into a finite number of four-sided patches

$$\Gamma = \bigcup_{i=1}^{M} \Gamma_i, \qquad \Gamma_i = \gamma_i(\square), \qquad i = 1, 2, \ldots, M,$$

where each $\gamma_i : \square \to \Gamma_i$ defines a diffeomorphism of $\square$ onto Γ_i. The intersection $\Gamma_i \cap \Gamma_{i'}$, $i \neq i'$, of the patches Γ_i and $\Gamma_{i'}$ is supposed to be either $\emptyset$, or a common edge or vertex. Additionally, the parametrization is claimed to be globally continuous.

The wavelets are constructed by the biorthogonal wavelet bases of Ref. [34] and adapted to the interval [35]. Tensor products yield corresponding dual pairs on $\square$.

Using the parametric liftings γ_i one obtains the wavelets on the cavity surface [26, 36]. In the present implementation we employ piecewise constant wavelets with three vanishing moments.

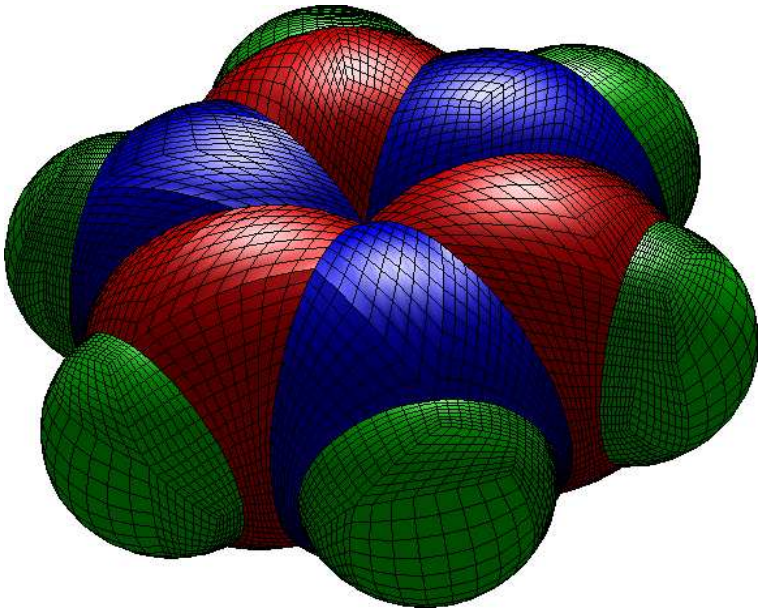


FIGURE 2.1. The parametrization by four-sided patches.

The surface representation is in contrast to the usual approximation of the surface by panels. It has the advantage that the geometry is exactly described. In Refs. [26, 27], an algorithm has been developed which decomposes a given surface defined by interlocking spheres – such as the van der Waals surface – into four-sided patches. The parametrization of the spherical patches is generated by using rational Bézier surfaces and homogeneous coordinates. The resulting parametrization of the Van der Waals surface for benzene is shown in Fig. 2.1.

3. Results

In order to investigate the performance and capabilities of the present implementation, we have considered several practical aspects of the wavelet PCM method described above.

We have employed benzene as a test molecule to analyze accuracy, CPU time and memory requirements of the method. The results were obtained at the HF level using two different Gaussian basis sets: a relatively small 6-31G basis set and a larger 6-311++G**, which includes polarization and diffuse functions in addition to the standard 6-311G split valence basis set.

The current scaling of the method with respect to the size of the system has been investigated by making use of a series linear-chain alkanes as our benchmark molecules. We have also considered the effect of interpolation of the potential. All calculations were carried out without any symmetry.

3.1. **Parameters of the wavelet solver.** Several parameters have been used to fine-tune the accuracy of the wavelet solver. We give here a brief account of them.

The cavity is divided in square patches as shown in Fig. 2.1. The *patch level* (PL) defines how many elements N_P per patch are present. Successive PLs are obtained by bisection of the current one thus obtaining that:

$$N_P = 4^P, \qquad P = 0, 1, 2, \ldots. \tag{3.15}$$

A larger number of patches leads to a larger number of degrees of freedom, and thus more accurate results (in the wavelet approach). The accuracy of the solution is also determined by the amount of compression in the system matrix. This is controlled by two parameters, a and d' (see Ref. [28] for details), which control the bandwidth and accuracy of the sparse representation of the system matrix, respectively.

In addition to those listed above, whose effects are analyzed in the present paper, there are parameters concerning both the *a posteriori* compression of the system matrix and the accuracy of the iterative solution of the linear system of equations. These parameters have been set in such way that the results are sufficiently accurate.

3.2. **Accuracy of the wavelet BEM in benzene.** We will begin by examining the effect of *a priori* compression to the accuracy at PL2 and PL3. The obtained results are presented in Tables 3.1 and 3.2.

The parameter a is varied in [1.0, 1.5, 2.0, 2.5] and the parameter d' in [1.125, 1.500, 2.000, 2.500]. We have reported both the nuclear polarization energy (the U_{NN} term in Eq. (2.9)) and the electronic polarization energy which corresponds to the expectation value of the bielectronic solvation operator defined in Eq. (2.12). The values reported are deviations of the energies from the $a = 2.5, d' = 2.5$ values, which can practically be considered as being sufficiently precise for a given patch level.

It is clear from the results in Table 3.1 and Table 3.2 that both increasing a and increasing d' leads to a higher accuracy, and especially increasing d' has a great effect on the precision. For example at PL2, employing $a = 1.0, d' = 2.500$ decreases the relative error of the nuclear polarization energy to 0.17 % from 0.92 % of the $a = 1.0, d' = 1.125$ pair. The best result is obtained by increasing both a and d'. At PL3, the situation is somewhat different. It can be noted that convergence is

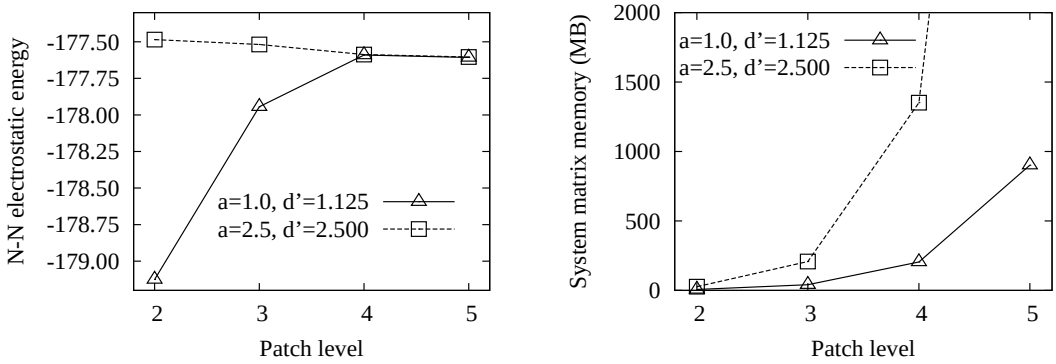
TABLE 3.1. **6-31G**: Effect of the amount of *a priori* compression in the system matrix on the accuracy of the nuclear and electronic polarization energies of benzene (mixed nuclear-electronic terms are not reported here). Absolute deviations with respect to the the choice of parameters $a = 2.5$ and $d' = 2.5$ are presented.

Patch level 2									
N-N energy (E = −177.48485)					E-E energy (E = −177.68344)				
	a					a			
d'	1.0	1.5	2.0	2.5	d'	1.0	1.5	2.0	2.5
1.125	-1.64085	-1.19504	-0.87917	-0.61201	1.125	-1.65850	-1.19981	-0.87716	-0.60896
1.500	-1.30931	-0.89788	-0.54175	-0.28702	1.500	-1.31787	-0.89632	-0.53879	-0.28570
2.000	-0.97140	-0.44873	-0.17970	-0.09319	2.000	-0.97100	-0.44593	-0.17871	-0.09280
2.500	-0.31715	-0.07002	-0.01507	0.00000	2.500	-0.31460	-0.06938	-0.01497	0.00000
Patch level 3									
N-N energy (E = −177.51844)					E-E energy (E = −177.70864)				
	a					a			
d'	1.0	1.5	2.0	2.5	d'	1.0	1.5	2.0	2.5
1.125	-0.42507	-0.23493	-0.20602	-0.19537	1.125	-0.42218	-0.23394	-0.20535	-0.19487
1.500	-0.02429	-0.00519	-0.00108	-0.00096	1.500	-0.02421	-0.00527	-0.00119	0.00085
2.000	-0.00761	-0.00252	0.00024	0.00174	2.000	-0.00766	-0.00254	0.00019	0.00169
2.500	-0.00612	-0.00297	-0.00046	0.00000	2.500	-0.00616	-0.00296	-0.00047	0.00000

TABLE 3.2. **6-311++G****: Effect of the amount of *a priori* compression in the system matrix on the accuracy of the nuclear and electronic polarization energies of benzene (mixed nuclear-electronic terms are not reported here). Absolute deviations with respect to the the choice of parameters $a = 2.5$ and $d' = 2.5$ are presented.

Patch level 2									
N-N energy (E = −177.48485)					E-E energy (E = −177.56699)				
	a					a			
d'	1.0	1.5	2.0	2.5	d'	1.0	1.5	2.0	2.5
1.125	-1.64085	-1.19504	-0.87917	-0.61201	1.125	-1.65744	-1.19895	-0.87649	-0.60837
1.500	-1.30931	-0.89788	-0.54175	-0.28702	1.500	-1.31685	-0.89560	-0.53819	-0.28537
2.000	-0.97140	-0.44873	-0.17970	-0.09319	2.000	-0.97020	-0.44537	-0.17849	-0.09271
2.500	-0.31715	-0.07002	-0.01507	0.00000	2.500	-0.31411	-0.06927	-0.01494	0.00000
Patch level 3									
N-N energy (E = −177.51844)					E-E energy (E = −177.59215)				
	a					a			
d'	1.0	1.5	2.0	2.5	d'	1.0	1.5	2.0	2.5
1.125	-0.42507	-0.23493	-0.20602	-0.19537	1.125	-0.42159	-0.23365	-0.20509	-0.19462
1.500	-0.02429	-0.00519	-0.00108	-0.00096	1.500	-0.02415	-0.00525	-0.00118	0.00084
2.000	-0.00761	-0.00252	0.00024	0.00174	2.000	-0.00762	-0.00253	0.00020	0.00168
2.500	-0.00612	-0.00297	-0.00046	0.00000	2.500	-0.00613	-0.00295	-0.00047	0.00000

FIGURE 3.2. (a): Nuclear polarization energy for different PLs using different *a priori* compression parameters. (b): Memory required (approximately) to hold all non-zero components of the system matrix at different patch levels.



(a) (b)

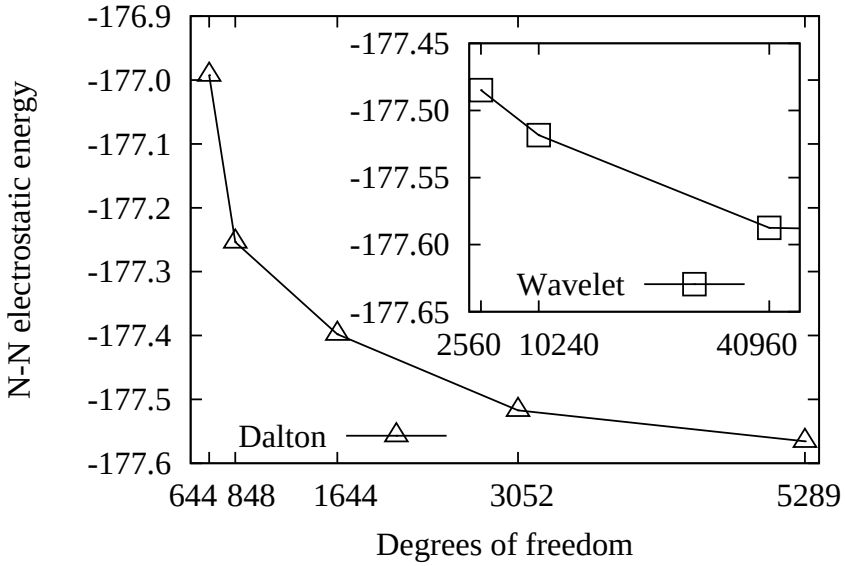
closer to begin with: a relative error of 0.24 % is obtained with the least accurate choice of parameters ($a = 1.0, d' = 1.125$). Nevertheless, increasing d' increases the accuracy but this time the effect is very large: the relative error drops by an order of magnitude with the two first stepwise increases of d', eventually leading to an error of 0.004 % at the $a = 1.0, d' = 2.000$ level. Further reduction of the error requires also increase of a. Similar results hold also for the electronic polarization energies showing that the method is not affected by dealing with a charge distribution, as opposite to a set of point charges.

In general, we observe that very tight parameters (large values of a and d') for the compression of the system matrix are only necessary with lower quality cavities. This conclusion is also supported by the results illustrated in Fig. 3.2a where the convergence of the nuclear polarization energy at different PLs and two sets of *a priori* compression parameters is shown.

Tighter values of compression parameters affect significantly the electrostatic polarization energy when a coarse PL (PL2 or PL3) is used. When fine patches are employed instead, the choice of compression parameters is less important (PL4 and PL5) since the final result is practically independent of the choice of these parameters.

The memory requirement for the system matrix construction is shown in Figure 3.2b. The higher level of accuracy provided by the values $a = 2.5, d' = 2.500$ over $a = 1.0, d' = 1.125$ within a given PL requires a much larger memory allocation. Relative extra memory needed for higher accuracy is considerable at every patch level,

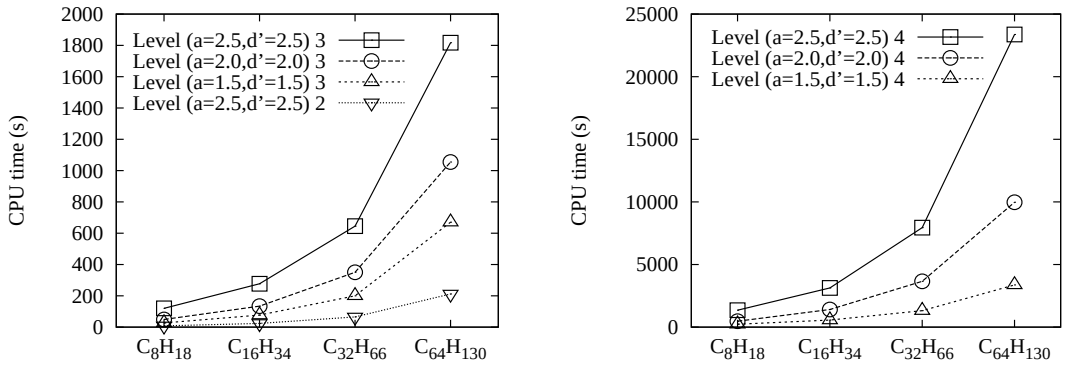
FIGURE 3.3. The N-N polarization energy versus the number of unknowns (equal to the number of tesserae) in traditional IEF-PCM and wavelet PCM implementations.



but at levels 2 and 3, the memory requirements are low for modern computers and the higher accuracy is worth the extra memory. This is not true anymore at PL4, as the difference in accuracy between compression parameter pairs is very small but at the same time a calculation with tighter compression parameters requires almost eight times as much memory to hold all the elements of the compressed system matrix than a lower quality calculation. System matrix construction took approximately 8 GB of memory at the patch level 5 with strict compression parameters, and therefore the large memory requirements clearly rule this combination out from the list of practical methods, at least for what concerns a non-parallel implementation. Moreover, the CPU time required to construct the system matrix and solve Eq. (2.2) is roughly proportional to the number of non-zero elements and hence to the memory requirements.

We have also compared the accuracy of the wavelet PCM approach to the traditional IEF-PCM implementation in the development version of the DALTON quantum chemistry program [37]. DALTON calculations were performed by making use of an initial tessera size for the cavity generation of 0.3, 0.2, 0.1, 0.05, and 0.025 Å^2 resulting in 644, 848, 1644, 3052, and 5289 tesserae, respectively. Corresponding wavelet PCM calculations have been performed with PL2–5 with 2560, 10240, 40960, and 163840 unknowns, respectively. The compression parameters were set to $a = 2.5, d' = 2.500$. Results are shown in Figure 3.3.

FIGURE 3.4. Overall CPU time spent in wavelet BEM to calculate the N-N polarization energy using different patch level and compression parameter combinations (see text for details).



Both methods clearly converge towards a certain common limit, which is close to −177.60. The traditional IEF-PCM implementation starts quite far away from the limit (tessera size 0.3) but converges close to it with the cavity of 5289 tesserae (tessera size 0.025). We can observe here that although the recommended tessera area is 0.3, this value is not adequate if high absolute accuracy is needed. The wavelet PCM method also converges towards the limit, but somewhat more slowly. At PL2, the absolute error, compared to the limit, is roughly 0.1, which is comparable to the traditional IEF-PCM with the same number of unknowns (interpolated from the neighboring values). Although the number of unknowns is increased fourfold at PL3, the accuracy is not significantly improved (compared to PL2). The result obtained by using PL4 is much closer to the limit as in the case of traditional IEF-PCM with smallest chosen tessera size (0.025Å^2), but the number of unknowns is eight times larger.

3.3. **Wavelet BEM in extended systems.** The CPU-time scaling of the wavelet PCM method with the size of the system is investigated by using the four linear alkanes C_8H_{18}, $C_{16}H_{34}$, $C_{32}H_{66}$, and $C_{64}H_{130}$.

The CPU time requirements of computations of the nuclear polarization energies have been investigated and the results are shown in Figure 3.4, where the CPU time for calculations at different PL/compression parameter combinations are shown. The scaling of the method does not depend considerably on the patch level as can be seen from the Figure 3.4. Both level 3 and level 4 curves have very similar slope, and the difference in CPU times between $C_{32}H_{66}$ and $C_{64}H_{130}$ is slightly less than a factor of three in all patch levels. In other words, the scaling of the calculation

TABLE 3.3. Errors in the N-N polarization energies with respect to the patch level 5 results in $C_{16}H_{34}$. The reference energy is -851.524 (Patch level 5, $a = 2.5, d' = 2.5$).

	Patch level			
Parameters	2	3	4	5
$a = 1.5, d' = 1.5$	-5.801	-0.161	-0.008	-0.003
$a = 2.0, d' = 2.0$	-2.927	0.025	-0.005	-0.001
$a = 2.5, d' = 2.5$	-0.611	0.040	0.007	0.000

is approximately proportional to $N^{1.5}$ where N is the number of carbon atoms. Such a scaling is roughly independent of the choice of PL and *a priori* compression. Independence of the CPU-time scaling on the patch level allows us to increase the accuracy of our calculations without changing the overall scaling of the method, although the more accurate calculations can take an order of magnitude more CPU time and, more importantly, memory to complete.

In practice, however, it is not necessary to perform expensive calculations at PL4. The accuracy in extended systems is probed by presenting values of $C_{16}H_{34}$ calculations in Table 3.3. These numbers provide a representative picture of the relative accuracy of the method, and similar values have been obtained for the other molecules (see Table 3.4 for details). The obtained results are in line with what has been observed for benzene in Table 3.1. PL2 results are off by -5.8 to -0.6 au as compared to the most accurate results, and all PL4 and PL5 results are very close to each other, differences being less than 0.01 au. At the PL3, tightening of the compression thresholds from $a = 1.5, d' = 1.5$ to $a = 2.0, d' = 2.0$ reduces the error from -0.161 to 0.025 au, with a relative error being 3×10^{-5}. Further increase in the parameters does not increase the accuracy anymore, and in fact, the accuracy is slightly decreased.

The system matrix construction at the PL3 takes about 15 min to complete and the solution of the linear equation takes approximately one minute in case of the $C_{64}H_{130}$ molecule at that level. Compared to the construction of the Fock matrix at each SCF iteration, the computational effort to obtain the ASC (and hence the solvent reaction field) does not affect the overall computation significantly.

Based on the previous observations, we can conclude that PL3 with $a = d' = 2.0$ provides a good combination of accuracy and speed. However, we have not considered the CPU time used to calculate the electrostatic potentials at the cavity surface yet. Assuming that the evaluation of the potential at a single point takes approximately one second to complete, the total CPU time used in the potential evaluation over the

TABLE 3.4. Results of interpolating the potential from the patch level 2 to finer levels. "Full" results are calculated without interpolation at the patch level specified on left, and "Interp." are the results interpolated to the patch level on left. Compression parameters were set to $a = 2.0, d' = 2.0$. Differences to "Full" level 5 results in N-N interaction energies are reported.

	C_8H_{18}		$C_{16}H_{34}$		$C_{32}H_{66}$		$C_{64}H_{130}$[a]	
Level	Full	Interp.	Full	Interp.	Full	Interp.	Full	Interp.
2	-0.92	-0.93	-2.93	-2.96	-4.66	-4.70	-7.39	-7.36
3	0.02	0.04	0.03	0.07	0.05	0.25	0.12	0.83
4	0.00	0.02	0.01	0.04	-0.01	0.19	0.00	0.71
5	0.00	0.02	0.00	0.03	0.00	0.19	—	—

[a] Level 5 calculations could not be performed because of the memory requirements. Present results are with respect to level 4 values.

cavity surface is 20 h in the system in question. Although the potential evaluation can be done parallel practically without overhead, it will still be a significant bottleneck.

3.4. **Extended systems using interpolation of the potential from coarse to fine.** To alleviate the problem of calculating hundreds of thousands of potential energy integrals, it may be possible to interpolate the potential at the surface. As noted before, patches of different levels cover exactly the same surface but with different tilings. This provides a natural way to perform the interpolation: the potential is evaluated at a coarse PL and then interpolated to a finer one where the actual computations are performed.

The results of interpolation in polyalkanes are shown in Table 3.4, where "Full" and "Interp." columns refer to results without and with interpolation, respectively. Interpolation is performed from the PL2, where the potential is evaluated at the nodal points of the mesh. This is in contrast to the "Full" case, in which the potential is evaluated on the patches using the center-point rule. These differences in the potential evaluation are the reason for the small discrepancy, approximately 0.03 au, between the interpolated and non-interpolated results at the PL2.

The greatest gain is achieved by interpolating from level 2 to level 3, where the error decreases by an order of magnitude. After that, interpolated results converge to their PL5 values very similarly to the non-interpolated results. In addition to the small error due to the potential evaluation differences, there seems to be also a certain inherent error in the interpolation, which can be seen from the converged results at

PL5. The error ranges from 0.02 au (C_8H_{18}) to 0.71 au ($C_{64}H_{130}$) and stays relatively constant when compared to the respective non-interpolated results at the levels 3–5. The residual error can easily be interpreted as the error committed by estimating the potential through interpolation instead of calculating its true value. It is clearly shown that beyond two levels the interpolation error dominates the overall error.

The interpolation scheme is therefore highly recommendable as a tool to decrease the number of potential evaluations. The same interpolation could in principle be applied at the Fock matrix evaluation where the potential of each overlap charge distribution needs to be computed at the cavity surface. However, the current implementation in DALTON calculates the integrals over basis functions in batches, which makes applying the interpolation scheme to the calculation of Fock matrix contribution inconvenient at present. However, the validity of the interpolation approach has been demonstrated.

4. Summary

We have presented the first implementation of a wavelet-based method to solve the IEF-PCM equations to obtain the electrostatic solvent effect on molecular solute.

The method combines the IEF-PCM formalism developed originally by Cances and Mennucci [20] with the general-purpose wavelet-based solver for boundary problems developed by Dahmen, Harbrecht and Schneider [28]. In order to apply the wavelet-solver on the cavity surface to solve the IEF-PCM equations, it is necessary to make use of a cavity made of patches that have a well defined topology. This has been provided by the work of Randrianarivony and Brunnett [27] both for surfaces defined by interlocking spheres as well as the Connolly surface obtained by "rolling" a solvent molecule over the VdW cavity.

This method is able to overcome potential problems of the traditional discretization approach such as non-sparsity of the operator, ill-conditioning, and absence of accurate error control.

In the present work we have focused on the accuracy and computational cost of the method by presenting some benchmark examples on benzene and a set of linear saturated alkanes C_nH_{2n+2} ($n = 8, 16, 32, 64$) molecules.

We observed that the wavelet PCM method provides an accurate and stable solution of the equations. This is in line with the formal results that have been proved theoretically for this approach.

The construction of the operator matrix and the iterative solution of the equations are efficient and do not pose a significant overhead to the overall calculation. Instead,

the bottleneck has been identified in the evaluation of the potential integrals over the cavity surface: the wavelet PCM method is in this respect more demanding because the number of evaluation points might be significantly larger than in the traditional IEF-PCM approach. In order to overcome this bottleneck, we have introduced an interpolation scheme which provides a considerable reduction in the overhead due to potential evaluations. At the same time the loss of accuracy has been found to be negligible.

5. Acknowledgments

The financial support from the Mohn foundation is greatly acknowledged. The computational resources for this project have been provided by the NOTUR high performance computing program.

References

[1] S. Miertus, E. Scrocco, and J. Tomasi, *Chem. Phys.* **55**, 117 (1981).
[2] J. L. Rivail and D. Rinaldi, *Chem. Phys.* **18**, 233 (1976).
[3] J. G. Kirkwood, *J. Comput. Phys.* **2**, 351 (1934).
[4] D. E. Beskos, editor, *Boundary Element Methods in Mechanics*, North Holland, Amsterdam, 1987.
[5] J. Tomasi, B. Mennucci, and R. Cammi, *Chem. Rev.* **105**, 2999 (2005).
[6] J. Tomasi and M. Persico, *Chem. Rev.* **94**, 2027 (1994).
[7] R. Cammi and J. Tomasi, *J. Computational Chem.* **16**, 1449 (1995).
[8] R. Cammi, L. Frediani, B. Mennucci, J. Tomasi, K. Ruud, and K. V. Mikkelsen, *J. Chem. Phys.* **117**, 13 (2002).
[9] J. L. Pascualahuir, E. Silla, and I. Tunon, *J. Computational Chem.* **15**, 1127 (1994).
[10] E. Silla, I. Tunon, and J. L. Pascualahuir, *J. Computational Chem.* **12**, 1077 (1991).
[11] J. L. Pascualahuir and E. Silla, *J. Computational Chem.* **11**, 1047 (1990).
[12] M. L. Connolly, *J. Appl. Crystallography* **16**, 548 (1983).
[13] C. S. Pomelli and J. Tomasi, *J. Computational Chem.* **19**, 1758 (1998).
[14] C. S. Pomelli and J. Tomasi, *Theoretical Chem. Accounts* **99**, 34 (1998).
[15] B. Mennucci, R. Cammi, and J. Tomasi, *J. Chem. Phys.* **110**, 6858 (1999).
[16] E. Cances and B. Mennucci, *J. Chem. Phys.* **109**, 249 (1998).
[17] M. Cossi, B. Mennucci, and R. Cammi, *J. Computational Chem.* **17**, 57 (1996).
[18] L. Frediani, R. Cammi, C. S. Pomelli, J. Tomasi, and K. Ruud, *J. Computational Chem.* **25**, 375 (2004).
[19] C. S. Pomelli, J. Tomasi, and R. Cammi, *J. Computational Chem.* **22**, 1262 (2001).
[20] E. Cances and B. Mennucci, *J. Math. Chem.* **23**, 309 (1998).

[21] E. Cances, B. Mennucci, and J. Tomasi, *J. Chem. Phys.* **107**, 3032 (1997).
[22] L. Frediani, R. Cammi, S. Corni, and J. Tomasi, *J. Chem. Phys.* **120**, 3893 (2004).
[23] G. Scalmani, V. Barone, K. N. Kudin, C. S. Pomelli, G. E. Scuseria, and M. J. Frisch, *Theoretical Chem. Accounts* **111**, 90 (2004).
[24] E. O. Purisima and S. H. Nilar, *J. Computational Chem.* **16**, 681 (1995).
[25] J. P. Bardhan, *J. Chem. Phys.* **130**, 094102 (2009).
[26] H. Harbrecht and M. Randrianarivony, Wavelet BEM on molecular surfaces — parametrization and implementation, *INS Preprint 0902*, Institute for Numerical Simulation, University of Bonn, Germany, 2009, *submitted*.
[27] M. Randrianarivony and G. Brunnett, Molecular Surface Decomposition using Graphical Modeling, in Bildverarbeitung für die Medizin: Algorithmen — Systeme — Anwendungen, Proceedings des Workshops vom 6. bis 8. April 2008 in Berlin, pp. 197–201, Berlin Heidelberg, 2008, Springer-Verlag.
[28] W. Dahmen, H. Harbrecht, and R. Schneider, *Siam J. On Numerical Analysis* **43**, 2251 (2006).
[29] H. Harbrecht and R. Schneider, *Siam J. On Scientific Computing* **27**, 1347 (2006).
[30] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular Electronic-Structure Theory*, Wiley, 2005.
[31] G. Beylkin, R. Coifman, and V. Rokhlin, *Comm. Pure App. Math.* **44**, 141 (1991).
[32] R. Schneider, Multiskalen- und Wavelet-Matrixkompression: Analysisbasierte Methoden zur Lösung großer vollbesetzter Gleichungssysteme, 1998, Teubner, Stuttgart.
[33] W. Dahmen and A. Kunoth, *Numerische Mathematik* **63**, 315 (1992).
[34] A. Cohen, I. Daubechies, and J. C. Feauveau, *Comm. On Pure Appl. Mathematics* **45**, 485 (1992).
[35] W. Dahmen, A. Kunoth, and K. Urban, *Appl. Computational Harmonic Analysis* **6**, 132 (1999).
[36] H. Harbrecht and R. Schneider, *Mathematische Nachrichten* **269**, 167 (2004).
[37] DALTON, a molecular electronic structure program, Release 2.0 (2005), see http://www.kjemi.uio.no/software/dalton/dalton.html.

Ville Weijo, Luca Frediani: Department of Chemistry, University of Tromsø, 9037 Tromsø, Norway.

E-mail address: {luca.frediani,ville.weijo}@uit.no

Maharavo Randrianarivony, Helmut Harbrecht: Institut für Numerische Simulation, Universität Bonn, Wegelerstr. 6, 53115 Bonn, Germany.

E-mail address: {randrian,harbrecht}@ins.uni-bonn.de