Domain decomposition for implicit solvation models

Eric Cancès, Yvon Maday, Benjamin Stamm

Domain decomposition for implicit solvation models

Eric Cancès,[1] Yvon Maday,[2] and Benjamin Stamm[3]

[1] *Université Paris-Est, CERMICS, Project-team Micmac, INRIA-Ecole des Ponts,*

6 & 8 avenue Blaise Pascal, 77455 Marne-la-Vallée Cedex 2,

France

[2] *UPMC Univ. Paris 06, UMR 7598 LJLL, Paris, F-75005 France;*

Institut Universitaire de France and

Division of Applied Mathematics, Brown University, Providence, RI,

USA

[3] *UPMC Univ. Paris 06, UMR 7598 LJLL, Paris, F-75005 France;*

CNRS, UMR 7598 LJLL, Paris, F-75005 France

(Dated: May 10, 2013)

This article is the first of a series of papers dealing with domain decomposition algorithms for implicit solvent models. We show that, in the framework of the COSMO model, with van der Waals molecular cavities and classical charge distributions, the electrostatic energy contribution to the solvation energy, usually computed by solving an integral equation on the whole surface of the molecular cavity, can be computed more efficiently by using an integral equation formulation of Schwarz's domain decomposition method for boundary value problems. In addition, the so-obtained potential energy surface is smooth, which is a critical property to perform geometry optimization and molecular dynamics simulations. The purpose of this first article is to detail the methodology, set up the theoretical foundations of the approach, and study the accuracies and convergence rates of the resulting algorithms. The full efficiency of the method and its applicability to large molecular systems of biological interest is demonstrated elsewhere.

I. INTRODUCTION

Most of the physical and chemical phenomena of interest in chemistry and biology take place in the liquid phase, and it is well known that solvent effects play a crucial role in these processes. There are basically two different approaches to account for solvent effects in the computation of the properties of a solvated molecule or ion. The first approach is to use an explicit solvation model, in which the simulated chemical system is composed of the solute molecule and of a large number of solvent molecules. The second approach is to use an implicit solvation model, in which the solute molecule under study, sometimes together with a small number of solvent molecules weakly bonded to the solute, is embedded in a cavity Ω surrounded by a continuous medium modelling the solvent (Fig. 1).
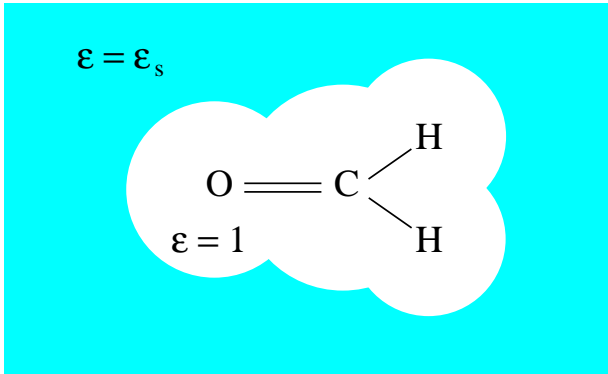

Figure 1: A formaldehyde molecule embedded in a polarizable continuum medium

Implicit solvent models are widely used in practice. For brevity, we will not elaborate here on their capabilities and limitations, nor on their applications in chemistry and biochemistry, and refer the reader to the monograph and the review articles[1–3], and references therein.

In the original implicit solvent model that goes back to Born, Kirkwood and Onsager, the continuous medium is a homogeneous isotropic dielectric medium of relative permittivity $\epsilon_s > 1$. The electrostatic energy of the charge distribution ρ (classical point charges, electric dipoles and multipoles in force-field models, classical nuclear charges and quantum electronic charge density in first-principle or semi-empirical models) carried by the solute is modified by the presence of the solvent, and an extra term, called the electrostatic contribution to the solvation energy, and denoted here by E^s, must be added to the electrostatic energy computed *in vacuo* (note that other terms accounting for short-range interaction must also

be taken into account[4]). The contribution E^{s} can be written as

$$E^{\mathrm{s}} = \frac{1}{2} \int_{\mathbb{R}^3} \rho(\mathbf{r})\, V^{\mathrm{r}}(\mathbf{r})\, d\mathbf{r},$$

where V^r is the reaction-field potential generated by the charge distribution ρ in presence of the dielectric continuum. If the charge distribution ρ is entirely (resp. almost entirely) supported in the cavity Ω, the reaction-field potential V^{r} can be represented (resp. approximated) by a single layer potential generated by a surface charge distribution σ supported by the boundary $\Gamma := \partial\Omega$ of Ω:

$$\forall \mathbf{r} \in \mathbb{R}^3, \quad V^{\mathrm{r}}(\mathbf{r}) = \int_{\Gamma} \frac{\sigma(\mathbf{s})}{|\mathbf{r} - \mathbf{s}|}\, d\mathbf{s}.$$

We thus obtain

$$E^{\mathrm{s}} = \frac{1}{2} \int_{\mathbb{R}^3} \rho(\mathbf{r}) \left(\int_{\Gamma} \frac{\sigma(\mathbf{s})}{|\mathbf{r} - \mathbf{s}|}\, d\mathbf{s} \right) d\mathbf{r} = \frac{1}{2} \int_{\Gamma} \sigma(\mathbf{s}) \Phi(\mathbf{s})\, d\mathbf{s},$$

where

$$\Phi(\mathbf{r}) = \int_{\mathbb{R}^3} \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}'$$

is the potential generated by ρ *in vacuo* (defined in the whole space $\mathbb{R}^3$). In practice, Φ is given by an explicit formula and the problem of computing E^{s} boils down to finding the surface charge density σ. This can be done by solving an integral equation set on the boundary Γ (see e.g. Ref.[1] (Section 1.2)). Introducing the integral operators $\mathcal{S}_{\Gamma}$, $\mathcal{D}_{\Gamma}$ and $\mathcal{D}_{\Gamma}^*$ defined by

$$\forall \mathbf{s} \in \Gamma, (\mathcal{S}_{\Gamma}\sigma)(\mathbf{s}) = \int_{\Gamma} \frac{\sigma(\mathbf{s}')}{|\mathbf{s} - \mathbf{s}'|}\, d\mathbf{s}',$$

$$(\mathcal{D}_{\Gamma}\sigma)(\mathbf{s}) = \int_{\Gamma} \frac{(\mathbf{s} - \mathbf{s}') \cdot \mathbf{n}_{\Gamma}(\mathbf{s}')}{|\mathbf{s} - \mathbf{s}'|^3}\, \sigma(\mathbf{s}')\, d\mathbf{s}',$$

$$(\mathcal{D}_{\Gamma}^*\sigma)(\mathbf{s}) = -\int_{\Gamma} \frac{(\mathbf{s} - \mathbf{s}') \cdot \mathbf{n}_{\Gamma}(\mathbf{s})}{|\mathbf{s} - \mathbf{s}'|^3}\, \sigma(\mathbf{s}')\, d\mathbf{s}',$$

where $\mathbf{n}_{\Gamma}(\mathbf{s})$ is the outward pointing normal vector at $\mathbf{s} \in \Gamma$, the surface charge density σ can be obtained by solving the integral equation

$$\forall \mathbf{s} \in \Gamma, \quad \left[\left(2\pi \frac{\epsilon_{\mathrm{s}} + 1}{\epsilon_{\mathrm{s}} - 1} - \mathcal{D}_{\Gamma}^* \right) \sigma \right](\mathbf{s}) = \frac{\partial \Phi}{\partial n}(\mathbf{s}), \tag{1}$$

where $\frac{\partial \Phi}{\partial n}(\mathbf{s}) := \nabla\Phi(\mathbf{s}) \cdot \mathbf{n}_{\Gamma}(\mathbf{s})$ is the normal derivative of Φ at $\mathbf{s} \in \Gamma$. In the case when ρ is entirely supported in Ω, the above equation is equivalent to the integral equation

$$\forall \mathbf{s} \in \Gamma, \quad (\mathcal{A}\sigma)(\mathbf{s}) = -\left((2\pi - \mathcal{D}_{\Gamma})\Phi\right)(\mathbf{s}) \quad \text{where} \quad \mathcal{A} = \left(2\pi \frac{\epsilon_{\mathrm{s}} + 1}{\epsilon_{\mathrm{s}} - 1} - \mathcal{D}_{\Gamma} \right) \mathcal{S}_{\Gamma}. \tag{2}$$

Although (2) is formally more complicated than (1), it is in some respects better suited to numerical calculation: the integral operator $\mathcal{A}$ is self-adjoint and positive definite on $L^2(\Gamma)$ (so that Eq. (2) has a simple variational formulation[5], whose discretization can be solved by the preconditioned conjugate gradient method), and the error made in first principle calculations by assuming that the charge distribution ρ is entirely supported in Ω (which is not the case for the electronic charge density) is smaller[6].

Another popular implicit solvation model is the COnductor-like Screening MOdel (COSMO)[7,8]. In this model, the electrostatic contribution to the solvation energy is given by

$$E_C^s = \frac{1}{2} f(\epsilon_s) \int_\Omega \rho(\mathbf{r}) W(\mathbf{r}) \, d\mathbf{r},$$

where $f(\epsilon_s) = \frac{\epsilon_s - 1}{\epsilon_s + k}$ is an empirical function of ϵ_s (k is a parameter taken equal to 0.5 in COSMO), and where W is the solution to the boundary value problem

$$\begin{cases} -\Delta W = 0 & \text{in } \Omega, \\ W = -\Phi & \text{on } \Gamma. \end{cases} \tag{3}$$

The usual method to compute E_C^s is to represent W by a single layer potential:

$$\forall \mathbf{r} \in \overline{\Omega} := \Omega \cup \Gamma, \quad W(\mathbf{r}) = \int_\Gamma \frac{\sigma_C(\mathbf{s})}{|\mathbf{r} - \mathbf{s}|} \, d\mathbf{s},$$

where the surface charge density σ is obtained by solving

$$\forall \mathbf{s} \in \Gamma, \quad (\mathcal{S}_\Gamma \sigma_C)(\mathbf{s}) = -\Phi(\mathbf{s}). \tag{4}$$

COSMO can be seen as an approximation of the original implicit solvent model, valid in the case when $\epsilon_s \gg 1$ (recall that $\epsilon_s \simeq 78$ for water). The two models indeed coincide in the limit $\epsilon_s = \infty$ of a conducting continuum medium. This point is discussed in detail in Ref.[6].

In early calculations, the cavity was spherical and ρ a classical charge distribution composed of point charges and dipoles. In this simple case, σ and σ_C can be computed explicitly by expansion on spherical harmonics. In order to better represent the solvent excluded volume, successive refinements of the cavity definition have been introduced: ellipsoidal cavities, for which σ and σ_C can still be computed by expansion on a basis of special functions, then molecular-shaped cavities[9,10], for which the integral equations (1), (2) and (4) must be solved by numerical methods. The notion of molecular-shaped cavity is not universal. Several construction methods are actually in use. In this contribution, we focus on van

der Waals cavities. For a solute composed of M atoms, the van der Waals cavity is made of a union of M balls, each of them centered on a nucleus and of radius βR_{VdW}, R_{VdW} denoting the UFF van der Waals radius of the corresponding atom and β being a fixed scaling parameter (taken equal to $\beta = 1.1$ in this article). Several numerical methods have been proposed to solve the integral equations (1), (2) and (4). These equations can be discretized by means of $\mathbb{P}_0$ boundary element methods and the resulting linear systems can be solved by direct or iterative algorithms[11–16], using fast multipole methods to achieve linear scaling[17]. Other discretizations of Eq. (4) based on approximations of the surface charge density by a collection of Gaussian functions whose centers are carried by the boundary Γ, have been recently proposed[18–21].

Our purpose is to introduce a new numerical approach, relying on an integral equation formulation of Schwarz's domain decomposition method, to compute the electrostatic contribution to the solvation energy in implicit solvent models. In the present article, we explain how to implement this approach in practice in the framework of COSMO, for van der Waals molecular cavities and classical charge distributions. For simplicity, we assume that ρ consists of point charges only (as in the fluctuating charge model[22,23] for instance), but the extension to other polarizable or non-polarizable force fields involving electric dipoles and multipoles is straightforward. Schwarz's domain decomposition method for solving the Laplace equation (3) is described in Section II, and its formulation in terms of integral equations is given in Section III. Discretization and implementation issues are addressed in Section IV. It is shown in particular that it is possible to regularize the problem in such a way that the resulting potential energy surface is smooth. This is essential to carry out geometry optimization and molecular dynamics. The accuracy of the method and the convergence rates of the different algorithms are then studied numerically on a set of simple benchmark molecules in Section V. In a companion article[24] in collaboration with F. Lipparini and B. Mennucci, we demonstrate that the proposed domain decomposition method is very efficient in practice: the electrostatic contribution to the solvation energy and the corresponding terms in the atomic forces (computed by means of analytical derivatives) of a large molecule can be obtained much faster than with the usual methods based on the discretization of Eq. (4). Extensions of our method to other kinds of molecular cavities (solvent accessible surfaces and solvent excluded surfaces), to quantum charge distributions, and to the original implicit solvent model, are works in progress.

II. SCHWARZ'S DOMAIN DECOMPOSITION METHOD FOR COSMO

As already mentioned in the introduction, we consider for simplicity a neutral solute molecule composed of M atoms and carrying a classical charge distribution

$$\rho(\mathbf{r}) = \sum_{j=1}^{M} q_j \delta_{\mathbf{R}_j}(\mathbf{r}),$$

where $q_j \in \mathbb{R}$ is the charge of the jth atom, $\mathbf{R}_j \in \mathbb{R}^3$ its position, and $\delta_{\mathbf{R}_j}$ the Dirac function at point $\mathbf{R}_j$. The case when the point charges are not located at the centers of the spheres is addressed in Remark 1.The van der Waals molecular cavity is

$$\Omega = \bigcup_{j=1}^{M} \Omega^j,$$

where $\Omega^j \subset \mathbb{R}^3$ is the ball with center $\mathbf{R}_j$ and radius $r_j := 1.1 \times R_{\text{vdW}}^j$, R_{vdW}^j denoting the UFF van der Waals radius[25] of the jth atom. The electrostatic contribution to the COSMO solvation energy of the system is then given by

$$E_{\text{C}}^{\text{s}} = \frac{1}{2} f(\epsilon_{\text{s}}) \sum_{j=1}^{M} q_j W(\mathbf{R}_j), \tag{5}$$

where W is the unique solution to the Laplace boundary value problem (3), with

$$\Phi(\mathbf{r}) = \sum_{j=1}^{M} \frac{q_j}{|\mathbf{r} - \mathbf{R}_j|}. \tag{6}$$

For $j = 1, \ldots, M$, we denote by $\Gamma^j := \partial\Omega^j$ the boundary of Ω^j (that is the sphere with center $\mathbf{R}_j$ and radius r_j), by $\Gamma^{j,\text{e}}$ the external part of the sphere Γ^j in contact with the solvent, and by $\Gamma^{j,\text{i}}$ the part of the sphere Γ^j interior to the cavity (Fig. 2). For $\mathbf{s} \in \Gamma^j$, we also define the set $N(j, \mathbf{s}) = \{1 \leq k \leq M \,|\, k \neq j \text{ and } \mathbf{s} \in \Omega^k\}$ of the balls Ω^k intersecting Ω^j and containing the point $\mathbf{s}$, and denote by $|N(j, \mathbf{s})|$ the cardinality of the set $N(j, \mathbf{s})$.

In the spirit of Schwarz's domain decomposition method[26], we introduce the restriction $W^j := W|_{\Omega^j}$ of W to the ball Ω^j. Then W is a solution to (3) if and only if the functions W^j's are solutions to

$$\forall 1 \leq j \leq M, \quad \begin{cases} -\Delta W^j = 0 & \text{in } \Omega^j, \\ W^j = -\Phi & \text{on } \Gamma^{j,\text{e}}, \\ W^j = W_{\mathbb{N}}^j & \text{on } \Gamma^{j,\text{i}}, \end{cases} \tag{7}$$
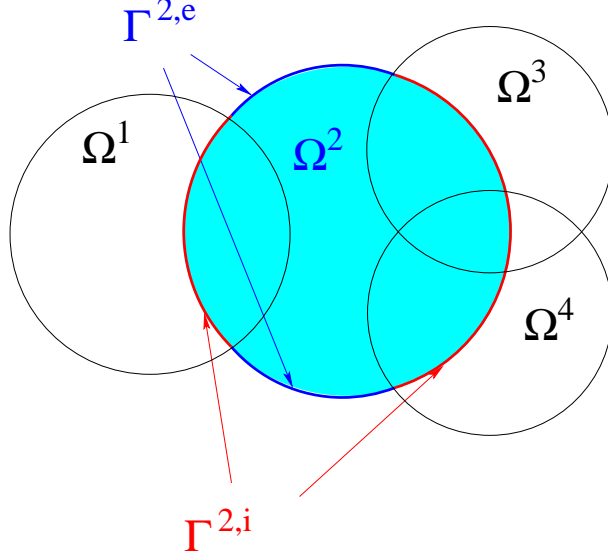
6

Figure 2: Domain decomposition with overlapping subdomains

where $W_{\mathtt{N}}^j$ is defined on $\Gamma^{j,\mathrm{i}}$ by

$$\forall \mathbf{s} \in \Gamma^{j,\mathrm{i}}, \quad W_{\mathtt{N}}^j(\mathbf{s}) = \frac{1}{|N(j,\mathbf{s})|} \sum_{k \in N(j,\mathbf{s})} W^k(\mathbf{s}). \tag{8}$$

This definition may seem akward, since all the terms in the above sum are equal: for all $\mathbf{s} \in \Gamma^{j,\mathrm{i}}$ and all $k \in N(j,\mathbf{s})$, $W^k(\mathbf{s}) = W(\mathbf{s})$. Actually, any partition of unity can be used to express (8); in particular, more regular partitions of unity than the piecewise constant functions proposed here, can be considered. The reason why we have rewritten the original Laplace problem (3) as the system of coupled Laplace problems (7) is that the latter can be solved efficiently by an iterative procedure, such as Jacobi or Gauss-Seidel algorithms:

$$\forall 1 \leq j \leq M, \quad \begin{cases} -\Delta W_n^j = 0 & \text{in } \Omega^j, \\ W_n^j = -\Phi & \text{on } \Gamma^{j,\mathrm{e}}, \\ W_n^j = W_{\star,n-1}^j & \text{on } \Gamma^{j,\mathrm{i}}, \end{cases} \tag{9}$$

where $n \geq 1$ is the iteration number, where W_0^j is the initial guess (the simplest choice is $W_0^j = 0$), and where $\star$ stands for either $\mathtt{J}$ (Jacobi) or $\mathtt{GS}$ (Gauss-Seidel) with

$$W_{\mathtt{J},n-1}^j(\mathbf{s}) = \frac{1}{|N(j,\mathbf{s})|} \sum_{k \in N(j,\mathbf{s})} W_{n-1}^k(\mathbf{s}) \qquad \forall \mathbf{s} \in \Gamma^{j,\mathrm{i}},$$

$$W_{\mathtt{GS},n-1}^j(\mathbf{s}) = \frac{1}{|N(j,\mathbf{s})|} \left[\sum_{\substack{k \in N(j,\mathbf{s}) \\ k < j}} W_n^k(\mathbf{s}) + \sum_{\substack{k \in N(j,\mathbf{s}) \\ k > j}} W_{n-1}^k(\mathbf{s}) \right] \qquad \forall \mathbf{s} \in \Gamma^{j,\mathrm{i}}.$$

Usually, the Gauss-Seidel approach yields faster convergence with respect to the number of iterations. The Jacobi approach is advantageous if a parallel implementation is pursued since communication is only necessary after each iteration, yielding M independent problems at each step.

The domain decomposition method described above is known in the mathematics literature as the Schwarz alternating method[26,27]. It is based on a decomposition of the domain Ω as a union of *overlapping* subdomains. Other popular domain decomposition methods[27], such as the Neumann-Dirichlet method, use non-overlapping subdomains, the information being exchanged between subdomains by means of boundary conditions. Note that domain decomposition algorithms are iterative methods: it is obviously impossible to obtain the exact solution to a boundary value problem set on a large domain, by solving only once the problems set on each subdomain. The convergence of the Schwarz method has been studied from a mathematical viewpoint by Mikhlin[28] and Lions[29].

III.   REFORMULATION IN TERMS OF INTEGRAL EQUATIONS

The coupled Laplace problems (7) can be transformed into a set of coupled integral equations by representing each function W^j by a single layer potential. For this purpose, we introduce the integral operators $\mathcal{S}_j : H^{-1/2}(\Gamma^j) \to H^{1/2}(\Gamma^j)$ and $\widetilde{\mathcal{S}}_j : H^{-1/2}(\Gamma^j) \to H^1(\Omega^j)$ respectively defined for all $\sigma \in H^{-1/2}(\Gamma^j)$ by

$$\forall \mathbf{s} \in \Gamma^j, \quad (\mathcal{S}_j \sigma)(\mathbf{s}) := \int_{\Gamma^j} \frac{\sigma(\mathbf{s}')}{|\mathbf{s} - \mathbf{s}'|} \, d\mathbf{s}',$$

and

$$\forall \mathbf{r} \in \Omega^j, \quad (\widetilde{\mathcal{S}}_j \sigma)(\mathbf{r}) := \int_{\Gamma^j} \frac{\sigma(\mathbf{s}')}{|\mathbf{r} - \mathbf{s}'|} \, d\mathbf{s}'.$$

Recall that the notation $H^s(\Omega^j)$ (resp. $H^s(\Gamma^j)$) stands for the Sobolev space[30] of order s on Ω^j (resp. Γ^j). We refer to Ref.[1] (Section 1.2) for a pedagogical introduction to the Sobolev spaces used in the mathematical formulation of implicit solvation models. We will see in Section IV that the Sobolev spaces $H^{1/2}(\Gamma^j)$ are well-suited to define relevant stopping criteria for the iterative algorithms we are going to introduce. The function W^j, being harmonic in Ω^j and belonging to the Sobolev space $H^1(\Omega^j)$, can be represented by a single layer potential: there exists a unique function $\sigma^j \in H^{-1/2}(\Gamma^j)$ such that $W^j = \widetilde{\mathcal{S}}_j \sigma^j$. Note

that the function

$$\widetilde{W}^j(\mathbf{r}) = \int_{\Gamma^j} \frac{\sigma^j(\mathbf{s})}{|\mathbf{r} - \mathbf{s}|} \, d\mathbf{s}$$

is well-defined in the whole space $\mathbb{R}^3$, but coincides with W^j, hence with W, only in $\overline{\Omega^j} := \Omega^j \cup \Gamma^j$, the closure of Ω^j. In particular, we therefore have $\widetilde{W}^j(\mathbf{R}_j) = W(\mathbf{R}_j)$ but $\widetilde{W}^j(\mathbf{R}_k) \neq W(\mathbf{R}_k)$ (in general) for $k \neq j$. This is nevertheless sufficient to obtain a new, useful expression of the electrostatic contribution to the COSMO solvation energy:

$$E_C^s = \frac{1}{2} f(\epsilon_s) \sum_{j=1}^M q_j W(\mathbf{R}_j) = \frac{1}{2} f(\epsilon_s) \sum_{j=1}^M q_j (\widetilde{\mathcal{S}}_j \sigma^j)(\mathbf{R}_j)$$

$$= \frac{1}{2} f(\epsilon_s) \sum_{j=1}^M q_j \int_{\Gamma^j} \frac{\sigma^j(\mathbf{s})}{|\mathbf{R}_j - \mathbf{s}|} \, d\mathbf{s} = \frac{1}{2} f(\epsilon_s) \sum_{j=1}^M \frac{q_j}{r_j} \int_{\Gamma^j} \sigma^j(\mathbf{s}) \, d\mathbf{s}, \tag{10}$$

where we have used the fact that Γ^j is the sphere with center $\mathbf{R}_j$ and radius r_j.

The functions $\sigma^1, \cdots, \sigma^M$ can then be obtained by solving the set of coupled integral equations

$$\forall 1 \leq j \leq M, \quad (\mathcal{S}_j \sigma^j)(\mathbf{s}) = \begin{vmatrix} -\Phi(\mathbf{s}) & \text{if } \mathbf{s} \in \Gamma^{j,e}, \\ \dfrac{1}{|N(j,\mathbf{s})|} \displaystyle\sum_{k \in N(j,\mathbf{s})} (\widetilde{\mathcal{S}}_k \sigma^k)(\mathbf{s}) & \text{if } \mathbf{s} \in \Gamma^{j,i}, \end{vmatrix}$$

which also reads

$$\forall 1 \leq j \leq M, \quad \forall \mathbf{s} \in \Gamma^j, \quad (\mathcal{S}_j \sigma^j)(\mathbf{s}) - \frac{1}{|N(j,\mathbf{s})|} \sum_{k \in N(j,\mathbf{s})} (\widetilde{\mathcal{S}}_k \sigma^k)(\mathbf{s}) = -\mathbb{1}_{\Gamma^{j,e}}(\mathbf{s}) \Phi(\mathbf{s}), \tag{11}$$

where $\mathbb{1}_{\Gamma^{j,e}}$ is the characteristic function of the set $\Gamma^{j,e}$ ($\mathbb{1}_{\Gamma^{j,e}}(\mathbf{s}) = 1$ if $\mathbf{s} \in \Gamma^{j,e}$, and 0 otherwise), and with the convention that $\frac{1}{0} \sum_\emptyset = 0$.

Let us emphasize that the surface charge densities σ^j computed by the domain decomposition algorithm are not straightforwardly related to the surface charge density σ_C defined by Eq. (4). In particular, σ^j and σ_C are not equal on $\Gamma^{j,e}$, and σ^j is not equal to zero on $\Gamma^{j,i}$.

It is useful (mainly to facilitate the calculation of analytical derivatives[24]) to set the M coupled integral equations in (11) on the unit sphere $\mathbb{S}^2 := \{\mathbf{r} \in \mathbb{R}^3 \,|\, |\mathbf{r}| = 1\}$. For this purpose, we introduce the functions $\sigma_j \in H^{-1/2}(\mathbb{S}^2)$ defined on $\mathbb{S}^2$ by

$$\forall \mathbf{s} \in \mathbb{S}^2, \quad \sigma_j(\mathbf{s}) = r_j \, \sigma^j(\mathbf{R}_j + r_j \mathbf{s}).$$

We thus obtain

$$E_C^s = \frac{1}{2} f(\epsilon_s) \sum_{j=1}^M q_j \int_{\mathbb{S}^2} \sigma_j(\mathbf{s}) \, d\mathbf{s},$$

9

where $(\sigma_1, \cdots, \sigma_M) \in (H^{-1/2}(\mathbb{S}^2))^M$ is the unique solution to Eq. (11) transported on the reference sphere $\mathbb{S}^2$, that is:

$$\forall 1 \leq j \leq M, \quad \forall \mathbf{s} \in \mathbb{S}^2, \quad (\mathcal{S}\sigma_j)(\mathbf{s}) - \frac{1}{|\mathcal{N}^j(\mathbf{s})|} \sum_{k \in \mathcal{N}^j(\mathbf{s})} \left(\widetilde{\mathcal{S}}_{jk}\sigma_k\right)(\mathbf{s}) = -\chi_j^E(\mathbf{s})\widetilde{\Phi}_j(\mathbf{s}), \quad (12)$$

where

$$\forall \mathbf{s} \in \mathbb{S}^2, \quad \mathcal{N}^j(\mathbf{s}) := \left\{ k \neq j \mid t^{jk}(\mathbf{s}) := \frac{|\mathbf{R}_j + r_j\mathbf{s} - \mathbf{R}_k|}{r_k} < 1 \right\},$$

and where the linear integral operators $\mathcal{S}, \widetilde{\mathcal{S}}_{jk} : H^{-1/2}(\mathbb{S}^2) \to H^{1/2}(\mathbb{S}^2)$ and the functions $\chi_j^E, \widetilde{\Phi}_j : \mathbb{S}^2 \to \mathbb{R}$ are defined as follows:

- the operator $\mathcal{S}$ is the usual single layer operator

$$\forall \mathbf{s} \in \mathbb{S}^2, \quad (\mathcal{S}\sigma)(\mathbf{s}) = \int_{\mathbb{S}^2} \frac{\sigma(\mathbf{s}')}{|\mathbf{s} - \mathbf{s}'|} \, d\mathbf{s}';$$

- the operator $\widetilde{\mathcal{S}}_{jk}$ is defined by

$$\forall \mathbf{s} \in \mathbb{S}^2, \quad (\widetilde{\mathcal{S}}_{jk}\sigma)(\mathbf{s}) = \int_{\mathbb{S}^2} \frac{\sigma(\mathbf{s}')}{|t^{jk}(\mathbf{s})\,\mathbf{s}^{jk}(\mathbf{s}) - \mathbf{s}'|} \, d\mathbf{s}',$$

  where

$$t^{jk}(\mathbf{s}) = \frac{|\mathbf{R}_j + r_j\mathbf{s} - \mathbf{R}_k|}{r_k} \in \mathbb{R}_+ \quad \text{and} \quad \mathbf{s}^{jk}(\mathbf{s}) := \frac{\mathbf{R}_j + r_j\mathbf{s} - \mathbf{R}_k}{|\mathbf{R}_j + r_j\mathbf{s} - \mathbf{R}_k|} \in \mathbb{S}^2;$$

- the function χ_j^E is such that

$$\forall \mathbf{s} \in \mathbb{S}^2, \quad \chi_j^E(\mathbf{s}) = \begin{vmatrix} 1 & \text{if for all } k \neq j, \ t^{jk}(\mathbf{s}) \geq 1, \text{ that is if } \mathcal{N}^j(\mathbf{s}) = \emptyset, \\ 0 & \text{otherwise}; \end{vmatrix}$$

- lastly, $\widetilde{\Phi}_j(\mathbf{s}) = \sum_{k=1}^{M} \frac{q_k}{|\mathbf{R}_j + r_j\mathbf{s} - \mathbf{R}_k|}.$

Introducing the functions $\chi_{jk}^I(\mathbf{s})$ defined by

$$\chi_{jk}^I(\mathbf{s}) = \chi(t^{jk}(\mathbf{s})) \quad \text{where} \quad \chi(t) = \begin{vmatrix} 1 & \text{if } t < 1, \\ 0 & \text{otherwise}, \end{vmatrix} \quad (13)$$

(so that that $\chi_{jk}^I(\mathbf{s}) = \chi(t^{jk}(\mathbf{s}))$ if $k \in \mathcal{N}^j(\mathbf{s})$ and 0 otherwise), Eq. (12) also reads

$$\boxed{\forall 1 \leq j \leq M, \quad \forall \mathbf{s} \in \mathbb{S}^2, \quad (\mathcal{S}\sigma_j)(\mathbf{s}) - \sum_{k \in \mathcal{N}^j(\mathbf{s})} \omega^{jk}(\mathbf{s}) \left(\widetilde{\mathcal{S}}_{jk}\sigma_k\right)(\mathbf{s}) = -\left(1 - \sum_{k \in \mathcal{N}^j(\mathbf{s})} \omega^{jk}(\mathbf{s})\right)\widetilde{\Phi}_j(\mathbf{s}),}$$

$$(14)$$

where

$$\omega^{jk}(\mathbf{s}) = \frac{\chi_{jk}^I(\mathbf{s})}{\sum_{k' \in \mathcal{N}^j(\mathbf{s})} \chi_{jk'}^I(\mathbf{s})},$$

with the convention that $\omega^{jk}(\mathbf{s}) = 0$ if all the $\chi_{jk'}^I(\mathbf{s})$ are equal to zero. Indeed, if $k \notin \mathcal{N}^j(\mathbf{s})$, then $\omega^{jk}(\mathbf{s}) = 0$, while if $k \in \mathcal{N}^j(\mathbf{s})$, then $\omega^{jk}(\mathbf{s}) = \frac{1}{|\mathcal{N}^j(\mathbf{s})|}$.

The final equation (14) is an integral equation formulation, set on N copies of the unit sphere $\mathbb{S}^2$, of the boundary value problems (7)-(8). In the same line, the integral equation formulation of (9) for $\star = \mathbf{J}$ (Jacobi method) reads

$$\forall 1 \le j \le M, \quad \forall \mathbf{s} \in \mathbb{S}^2, \quad \left(\mathcal{S}\sigma_j^{\mathbf{J},n}\right)(\mathbf{s}) = \sum_{k \in \mathcal{N}^j(\mathbf{s})} \omega^{jk}(\mathbf{s}) \left(\widetilde{\mathcal{S}}_{jk}\sigma_k^{\mathbf{J},n-1}\right)(\mathbf{s}) - \left(1 - \sum_{k \in \mathcal{N}^j(\mathbf{s})} \omega^{jk}(\mathbf{s})\right) \widetilde{\Phi}_j(\mathbf{s}).$$

$$(15)$$

As the formulations (9), with $\star = \mathbf{J}$, and (15) are rigorously equivalent, the convergence of the iterative scheme (15) follows from the convergence results on the Schwarz alternating method for Laplace problems[29]. The same is true for the Gauss-Seidel method.

In the next section, we explain how to discretize Eq. (14) and solve the so-obtained discretized problem by Jacobi and Gauss-Seidel iterative methods.

Remark 1. *For a more general classical charge distribution of the form*

$$\rho(\mathbf{r}) = \sum_{j=1}^M \sum_{i=1}^{n_j} q_{j,i} \delta_{\mathbf{R}_{j,i}}(\mathbf{r})$$

with $\mathbf{R}_{j,i} \in \Omega^j$, the electrostatic contribution to the COSMO solvation energy is given by

$$E_C^s = \frac{1}{2} f(\epsilon_s) \sum_{j=1}^M \sum_{i=1}^{n_j} q_{j,i} \int_{\mathbb{S}^2} \frac{\sigma_j(\mathbf{s})}{|\mathbf{y}_{j,i} - \mathbf{s}|} d\mathbf{s}, \tag{16}$$

where $\mathbf{y}_{j,i} = r_j^{-1}(\mathbf{R}_{j,i} - \mathbf{R}_j)$, and where the surface charge densities σ_j are obtained by solving Eq. (14) with

$$\widetilde{\Phi}_j(\mathbf{s}) = \sum_{k=1}^M \sum_{i=1}^{n_k} \frac{q_{k,i}}{|\mathbf{R}_j + r_j\mathbf{s} - \mathbf{R}_{k,i}|}.$$

The integrals in Eq. (16) can be computed explicitly if the expansion of σ_j in the real spherical harmonics basis is known, which is the case in the numerical method we propose.

IV. DISCRETIZATION AND IMPLEMENTATION

It is natural to discretize the operators $\mathcal{S}$ and $\widetilde{\mathcal{S}}_{jk}$ in a truncated basis $(Y_l^m)_{0 \leq l \leq N,\, -l \leq m \leq l}$ of real spherical harmonics[31]. We use here the normalization convention:

$$\int_{\mathbb{S}^2} Y_l^m(\mathbf{s})\, Y_{l'}^{m'}(\mathbf{s})\, d\mathbf{s} = \int_0^\pi \int_{-\pi}^\pi Y_l^m(\theta, \phi)\, Y_{l'}^{m'}(\theta, \phi) \, \sin\theta \, d\theta \, d\phi = \delta_{ll'} \delta_{mm'},$$

where $\delta_{mm'}$ denotes the Kronecker symbol ($\delta_{mm'} = 1$ if $m = m'$, and 0 otherwise).

The functions χ_{jk}^I introduced in Eq. (13) are discontinuous. We therefore need to smooth them out in order to avoid discontinuities of the numerically computed energy $E_{\mathrm{C}}^{\mathrm{s}}$ with respect to nuclear coordinates, and obtain well-defined atomic forces for the discretized model. This can be done by introducing a smoothing χ_η of the function χ defined in Eq. (13):

$$\chi_\eta(t) = \left| \begin{array}{ll} 1 & \text{if } t \leq 1 - \eta, \\ p_\eta(t) & \text{if } 1 - \eta < t < 1, \\ 0 & \text{if } t \geq 1, \end{array} \right. \qquad \text{with} \qquad p_\eta(t) = \eta^{-4}(t-1)^2(t-1+2\eta)^2,$$

with $0 < \eta \ll 1$ a numerical parameter. The chosen regularization is asymmetric with respect to $t = 1$ in order to avoid technical issues when several balls are intersecting. The regularity of the function χ_η is sufficient to ensure that the second derivatives of the resulting potential energy surface are bounded. More regularity can be obtained by replacing p_η with a higher degree polynomial whose derivatives up to a given order vanish at $t = 0$ and $t = 1$.

We are now in position to detail the numerical procedure to compute numerically $E_{\mathrm{C}}^{\mathrm{s}}$. Each function σ_j is approximated by

$$\sigma_j^{\eta, N, N_g}(\mathbf{s}) = \sum_{l=0}^N \sum_{m=-l}^l [X_j]_l^m \, Y_l^m(\mathbf{s}), \tag{17}$$

where σ_j^{η, N, N_g} is the solution to the approximation of the regularized version of Eq. (14) (obtained by replacing χ with χ_η), in the space spanned by the discretization basis $(Y_l^m)_{0 \leq l \leq N,\, -l \leq m \leq l}$, and using the quadrature points $(\mathbf{s}_n)_{1 \leq n \leq N_g}$ to perform numerical quadratures on the sphere $\mathbb{S}^2$. Note that the approximation given in (17) is systematically improvable: it can be shown that σ_j^{η, N, N_g} converges to the exact solution σ_j to Eq. (14), in the Sobolev norm $H^{-1/2}(\Gamma^j)$, when N and N_g go to infinity while η goes to zero. The reason

L_g	15	17	19	21	23	25	27	29	31	35	41	47	53
N_g	86	110	146	170	194	230	266	302	350	434	590	770	974
L_g	59	65	71	77	83	89	95	101	107	113	119	125	131
N_g	1202	1454	1730	2030	2354	2702	3074	3470	3890	4334	4802	5294	5810

Table I: Order L_g and number of points N_g of Lebedev quadrature rules.

why we choose the $H^{-1/2}(\mathbb{S}^2)$ norm to measure the errors is that $\|\sigma\|_{H^{-1/2}(\mathbb{S}^2)}$ is in fact the Coulomb energy of the surface charge distribution σ.

For simplicity, we only consider here Lebedev quadrature (see Ref.[32] and references therein). Each Lebedev quadrature rule is completely characterized by enforcing the exact integration of spherical harmonics up to a given order named L_g (hence of all functions in the vector space Span $\{Y_l^m \,|\, 0 \leq l \leq L_g,\ -l \leq m \leq l\}$). This leads to a sequence of N_g associated quadrature points. Note that there does not exist a Lebedev formula for any value of L_g. The correspondence between the order L_g of the quadrature rule, and the number of integration points N_g, is given in Table I for the values of L_g used in our simulations.

Denoting by $\mathbf{X}_j \in \mathbb{R}^{(N+1)^2}$ the vector with entries $([X_j]_l^m)_{0 \leq l \leq N,\ -l \leq m \leq l}$, the calculation of the electrostatic contribution to the solvation energy can be done as follows.

Step 1: computation of the geometrical data.

For each $1 \leq j \leq M$, set $C_j = 1/r_j$. For each $1 \leq j \leq M$, $1 \leq n \leq N_g$, and $k \neq j$, compute

$$\mathbf{v}_n^{jk} = \mathbf{R}_j + r_j \mathbf{s}_n - \mathbf{R}_k, \quad v_n^{jk} = |\mathbf{v}_n^{jk}|, \quad t_n^{jk} = C_k v_n^{jk}, \quad z_n^{jk} = 1/v_n^{jk}, \quad \mathbf{s}_n^{jk} = z_n^{jk} \mathbf{v}_n^{jk},$$

where the points $(\mathbf{s}_n)_{1 \leq n \leq N_g}$ denote the quadrature points with associated weights $(w_n)_{1 \leq n \leq N_g}$. Use these computations to identify the sets

$$\mathcal{N}_n^j = \left\{ k \neq j \,|\, t_n^{jk} < 1 \right\} \quad \text{and} \quad \mathcal{N}^j = \left\{ k \neq j \,|\, \exists 1 \leq n \leq N_g \text{ s.t. } k \in \mathcal{N}_n^j \right\}.$$

For each $k \in \mathcal{N}_n^j$, set

$$\chi_n^{jk} = \chi_\eta(t_n^{jk}), \quad f_n^j = \sum_{k \in \mathcal{N}_n^j} \chi_n^{jk}, \quad d_n^j = \frac{\min(f_n^j, 1)}{f_n^j}, \quad W_n^{jk} = d_n^j \chi_n^{jk}, \quad U_n^j = 1 - \sum_{k \in \mathcal{N}_n^j} W_n^{jk}.$$

13

Step 2: calculation of the right-hand side by numerical quadratures.

For all $1 \leq j \leq M$ and $1 \leq n \leq N_g$, set $\Phi_n^j = q_j C_j + \sum_{k \neq j} q_k z_n^{jk}$. Then, for each $j = 1, \cdots, M$, and each l, m, set

$$[g_j]_l^m = -\sum_{n=1}^{N_g} w_n Y_l^m(\mathbf{s}_n) U_n^j \Phi_n^j.$$

Step 3: computation of the vector $(\mathbf{X}_1, \cdots, \mathbf{X}_M)$ by an iterative method.

The vector $(\mathbf{X}_1, \cdots, \mathbf{X}_M)$ is solution to the linear system

$$\underbrace{\begin{pmatrix} \mathbf{L}_{11} & \cdots & \mathbf{L}_{1M} \\ & \cdot & \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \mathbf{L}_{M1} & \cdots & \mathbf{L}_{MM} \end{pmatrix}}_{= \mathbf{L}} \begin{pmatrix} \mathbf{X}_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{X}_M \end{pmatrix} = \begin{pmatrix} \mathbf{g}_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{g}_M \end{pmatrix} \tag{18}$$

where

$$[\mathbf{L}_{jj}\mathbf{X}_j]_l^m = \frac{4\pi}{(2l+1)}[X_j]_l^m,$$

$$[\mathbf{L}_{jk}\mathbf{X}_k]_l^m = -\sum_{l'=0}^{N} \frac{4\pi}{(2l'+1)} \sum_{m'=-l'}^{l'} [c_{jk}]_{ll'}^{mm'}[X_k]_{l'}^{m'}, \qquad \forall k \in \mathcal{N}^j,$$

$$[\mathbf{L}_{jk}\mathbf{X}_k]_l^m = 0, \qquad \forall k \notin \mathcal{N}^j \cup \{j\},$$

where the coefficients $[c_{jk}]_{ll'}^{mm'}$ are given by

$$[c_{jk}]_{ll'}^{mm'} = \sum_{n=1}^{N_g} w_n Y_l^m(\mathbf{y}_n) W_n^{jk} (t_n^{jk})^{l'} Y_{l'}^{m'}(\mathbf{s}_n^{jk}) = \sum_{n\,|\,t_n^{jk}<1} w_n Y_l^m(\mathbf{s}_n) W_n^{jk} (t_n^{jk})^{l'} Y_{l'}^{m'}(\mathbf{s}_n^{jk}).$$

Note that, for a large molecule, the matrix $\mathbf{L}$ is block-sparse, since $\mathbf{L}_{jk} = 0$ if $\Omega^j \cap \Omega^k = \emptyset$. In practice, the linear system (18) is solved by an iterative procedure, such as block Jacobi's or block Gauss-Seidel's algorithms (more efficient algorithms can also be considered[24]). Let us emphasize that using the block Jacobi algorithm to solve (18) amounts to solving a discretized version of (15).

Of course, the entries of the 6-order tensor $[c_{jk}]_{ll'}^{mm'}$ are not stored in memory, nor even computed. The terms $[\mathbf{L}_{jk}\mathbf{X}_k^n]_l^m$, where $\mathbf{X}_k^n$ is the approximation of $\mathbf{X}_k$ obtained at the nth

iteration, are computed on the fly using an optimal rearrangement of the various summations. This point is detailed in Ref.[24].

Step 4: computation of the solvation energy. Set

$$E_C^{\mathrm{s}}(\eta, N, N_g; N_{\mathrm{it}}) = \sqrt{\pi} f(\epsilon_{\mathrm{s}}) \sum_{j=1}^{M} q_j [X_j^{N_{\mathrm{it}}}]_0^0,$$

where $\mathbf{X}_j^{N_{\mathrm{it}}}$ is the approximation of $\mathbf{X}_j$ obtained at the N_{it}-th iteration of the Jacobi or Gauss-Seidel algorithm.

The notation $E_C^{\mathrm{s}}(\eta, N, N_g; N_{\mathrm{it}})$ stresses the fact that the so-obtained numerical approximation of the solvation energy E_C^{s} depends on four numerical parameters: the smoothing parameter η, the cut-off parameter N, the number N_g of quadrature points, and the number N_{it} of iterations of the Jacobi (or Gauss-Seidel) algorithm. These four numerical parameters do not have the same status. The first three of them are fixed in advance by the user. In practice, the maximum number of iterations is fixed by the user, but the algorithm usually (and hopefully) stops before this maximum number is reached, when a stopping criterion is satisfied. A possible stopping criterion is `stop when inc`$_n \leq$ `Tol`, where $0 < $ `Tol` $\ll 1$ is a fixed tolerance, and where the increment `inc`$_n$ between the iterations $n-1$ and n is defined as

$$\mathtt{inc}_n = \left(\frac{\sum_{j=1}^{M} \|\sigma_j^{\eta,N,N_g;n} - \sigma_j^{\eta,N,N_g;n-1}\|_{H^{-1/2}(\mathbb{S}^2)}^2}{\sum_{j=1}^{M} \|\sigma_j^{\eta,N,N_g;n-1}\|_{H^{-1/2}(\mathbb{S}^2)}^2} \right)^{1/2}$$

$$= \left(\frac{\sum_{j=1}^{M} \sum_{l=0}^{N} \sum_{m=-l}^{l} \frac{|[X_j^n]_l^m - [X_j^{n-1}]_l^m|^2}{l+1}}{\sum_{j=1}^{M} \sum_{l=0}^{N} \sum_{m=-l}^{l} \frac{|[X_j^{n-1}]_l^m|^2}{l+1}} \right)^{1/2}, \tag{19}$$

where

$$\sigma_j^{\eta,N,N_g;n}(\mathbf{s}) = \sum_{l=0}^{N} \sum_{m=-l}^{l} [X_j^n]_l^m Y_l^m(\mathbf{s})$$

is the approximation of σ_j obtained at the nth iteration.

V.   NUMERICAL RESULTS

We present computations performed with the domain decomposition method introduced in this paper for simple molecular cavities in order to study the convergence properties and the stability of the proposed method.

We emphasize that the present numerical tests aim at studying the influence on the computed solvation energy of the regularization (η), discretization (N) and numerical integration (N_g) parameters, as well as of the stopping criterion ($\text{inc}_n \leq \text{Tol}$), which controls the number of iterations N_{it} of the algorithm. Additional tests of the performance and scalability of our algorithm applied to large biological molecules are presented in a companion paper[24].

In all computations, unless otherwise stated, we use the following parameters:

o the linear system (18) is solved by Gauss-Seidel iterations;

o $\epsilon_s = 78.4$ (water). In any case, the value of $f(\epsilon_s)$ has no influence on the convergence properties of the proposed scheme as it is just a multiplicative factor of the solvation energy.

In all our computations, we report the energy and charges in atomic units. As mentioned earlier, the Van der Waals cavity is built using the UFF radii[25] (in particular, $r_H = 1.443\text{Å}$, $r_C = 1.925\text{Å}$, $r_N = 1.830\text{Å}$, $r_O = 1.75\text{Å}$ and $r_F = 1.682\text{Å}$) scaled by a factor 1.1. The QEq charges[33] are used for the solutes.

1. Influence of the regularization parameter η

Fig. 3 plots the variation of the solvation energy as a function of the regularization parameter η for sufficiently large discretization and numerical quadrature parameters ($N = 50$, $N_g = 5810$), and sufficiently tight stopping criterion ($\text{inc}_n \leq 10^{-10}$) for hydrogen fluoride, formaldehyde and benzene.

The relative error on the solvation energy is about 0.05% (hydrogen fluoride), 0.008% (formaldehyde), 2% (benzene) when η varies from 0 to 0.1. The value of η therefore has a limited impact on the value of the solvation energy.

A. Complete study of a diatomic molecule (hydrogen fluoride)

In this section, we study the influence of the above mentioned parameters on a simple diatomic molecule where we consider hydrogen fluoride with point charges of ± 0.33714 (in atomic units). As the error depends on the geometrical parameters (centers and radii of the balls), we will also investigate the robustness of the solvation energy with respect to those parameters.
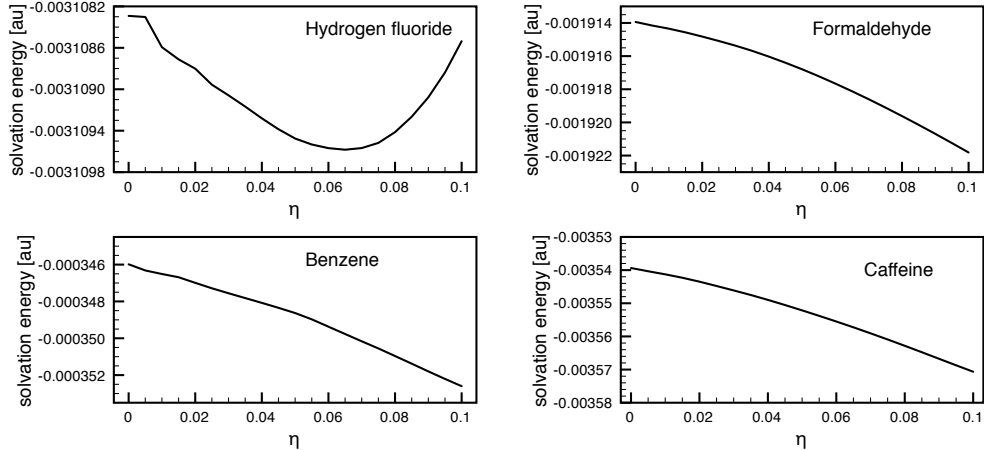
Figure 3: Influence of the regularization parameter η on the solvation energy for Hydrogen fluoride and Formaldehyde (with $N = 50$, $N_g = 5810$ and $\texttt{Tol} = 10^{-10}$).

1. Influence of the discretization parameter N

We investigate the error on the solvation energy $E_C^s(\eta, N, N_g; N_{it})$ and on the surface charge distributions $(\sigma_j^{\eta, N, N_g; N_{it}})_{1 \leq 1 \leq M}$ (measured in the $H^{-1/2}$-norm), as well as the number of iterations N_{it} to reach the stopping criterion, as functions of the discretization parameter N, for different values of the regularization parameter η (see Fig. 4). We still use $N_g = 5810$ Lebedev integration points and the stopping criterion $\texttt{inc}_n \leq 10^{-10}$ to ensure that the error is dominated by the effect of the degree N of the approximation. The errors are computed with respect to a reference calculation (for each η) obtained using an axisymmetric code with $N = 200$ and a very large number of Gauss-Legendre integration points.

We observe that the discretization errors due to the truncation of the spherical harmonics basis set decay,

- for $\eta = 0$, algebraically fast for both the solvation energy and the surface charge distribution (a least square fit yields rates of $N^{-1.38}$ for the energy and $N^{-0.93}$ for the surface charge distribution);
- for $\eta = 0.05, 0.1$, exponentially fast for both the solvation energy and the surface charge distribution.

We infer from the above analysis that, for η chosen equal to 0.05 or 0.1, the relative errors due to the regularization parameter η on the one hand, and the discretization parameter N on the other hand, are balanced for about $N = 10$.

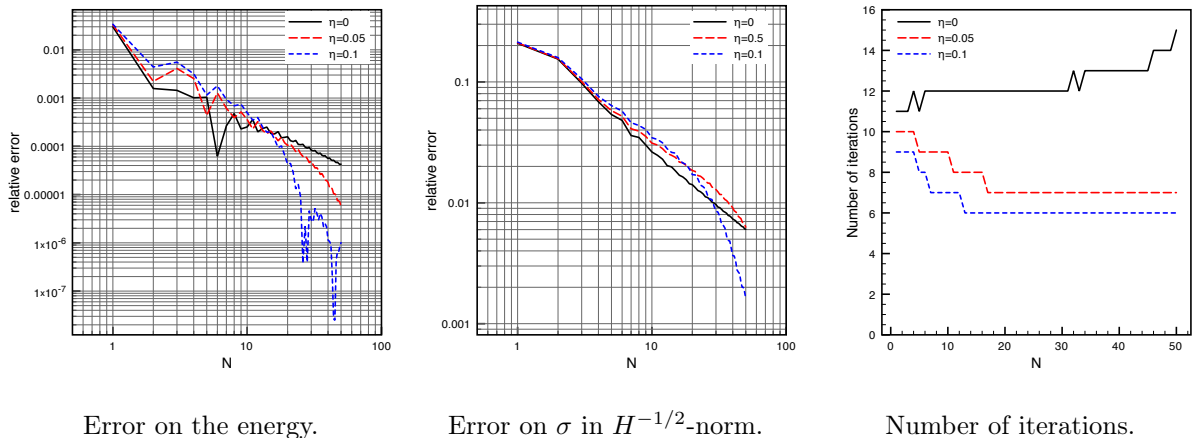| Error on the energy. | Error on σ in $H^{-1/2}$-norm. | Number of iterations. |

Figure 4: Numerical errors and number of iterations as functions of N for different values of $\eta = 0, 0.05, 0.1$ (with $N_g = 5810$ and `Tol` $= 10^{-10}$).

Regarding the number of iterations to reach the stopping criterion, we notice that for $\eta = 0$, the number of iterations steadily increases with N, while for the cases $\eta = 0.05$ and $\eta = 0.1$, we observe a plateau.

2. *Influence of the stopping criterion*

An important question to analyze is what tolerance value should be chosen for the stopping criterion $\text{inc}_n \leq \text{Tol}$. A too small value of the tolerance might yield too many iterations which will not improve the accuracy, since the latter will be dominated by the error due to the chosen values of N and η. A too large value of the tolerance might limit the accuracy of the energy. The problem is even more stringent for the computation of forces that involve derivatives of the energy[24].

In Fig. 5, we present some numerical results where we analyze the influence of the stopping criterion on the solvation energy and the relative error (with respect to the solution obtained with `Tol` $= 10^{-10}$), for different values of η. Since the stopping criterion is based on the increment of the unknown functions $(\sigma_j^{\eta,N,N_g;N_{\text{it}}})_{1 \leq 1 \leq M}$, its impact on the precision of the energy is not straightforward. The Gauss-Seidel algorithm yields more stable results (with respect to the energy) than the Jacobi algorithm. It is worth noticing that for a given value of `Tol`, the two algorithms do not yield the same accuracy on the energy. We can see on this example that a loose convergence criterion (`Tol` $= 10^{-2}$ for Gauss-Seidel, `Tol` $= 10^{-3}$
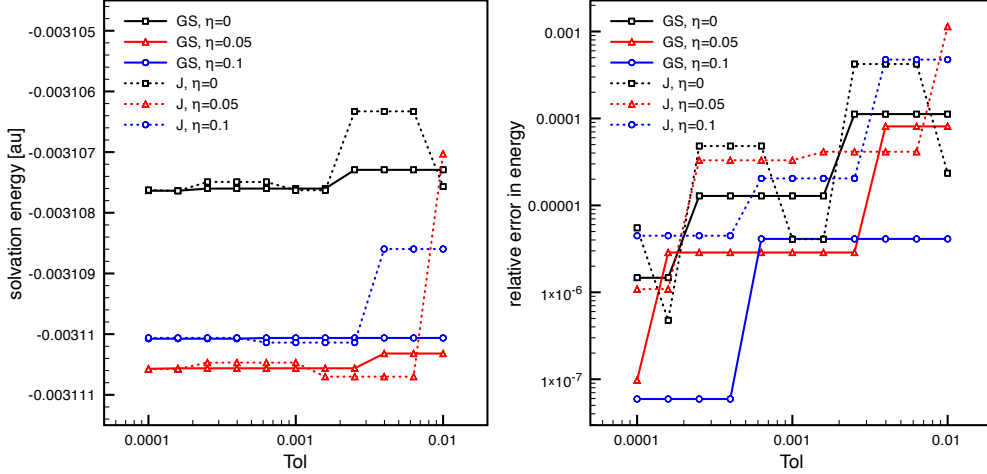
Figure 5: Influence of the stopping criterion $\text{inc}_n \leq \text{Tol}$ on the solvation energy and the relative error for different values of η (with $N = 20$ and $N_g = 5810$).

for Jacobi) is enough to obtain a very good approximation of the solvation energy. As expected, tighter convergence criteria are necessary to correctly approximate the atomic forces computed by analytical derivatives. This point is addressed in Ref.[24].

3. Influence of the number of integration points

We now investigate the influence on the solvation energy of the number N_g of Lebedev integration points, or in other words, of the order L_g of the numerical quadrature. Choosing an appropriate number of integration points is crucial since low regularity functions need to be integrated.

Fig. 6 illustrates the variations of the solvation energy in absolute and relative quantities, as a function of the order L_g of the numerical quadrature scheme. The numerical tests are obtained using the Gauss-Seidel algorithm with $\text{inc}_n \leq 10^{-2}$, $\eta = 0$ or 0.1, and $N = 10$, 20, or 30. The reference solution is computed using an axisymmetric code with a very large number of Gauss-Legendre integration points.

We observe that the number of integration points necessary to reach a given precision in the relative error does not depend on N (in the range $10 \leq N \leq 30$). For $\eta = 0$, it depends in fact on the regularity of the functions σ_j solutions to Eq. (14). For $\eta > 0$, it depends on the regularity of the functions σ_j^η solutions to the regularized version of Eq. (14) obtained
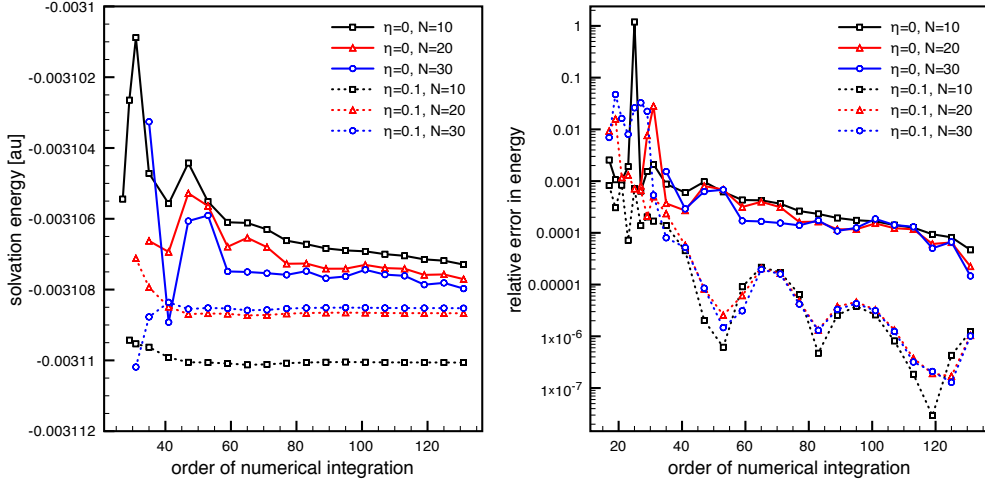
Figure 6: Influence of the integration order L_g on the solvation energy and on its relative error for different values of η and N (with $\texttt{Tol} = 10^{-2}$).

by replacing χ with χ_η in the definition (13) of the functions χ_{jk}^I. As the functions σ_j^η are more regular than the functions σ_j, using $\eta = 0.1$ allows us to choose a smaller value of the integration order L_g in the asymptotic regime of large values of N. For the chosen values of $\eta = 0.1$ and $N = 10$ any choice of L_g above or equal to 21 (corresponding to $N_g = 170$) leads to a relative error lower than 0.1% on the energy. Note that if N is large and L_g small the algorithm might fail to converge.

Using Lebedev quadrature leads to rotational symmetry breaking, just as using cubic integration grids leads to translational symmetry breaking (the so-called eggbox effect in planewave DFT calculations). Fig. 7 illustrates, for different values of η and N, the maximum energy difference obtained when rotating the hydrogen fluoride molecule, originally oriented along the z-axis, around the x-axis. Of course, the energy remains constant for $L_g = \infty$, that is, if integrals are computed exactly. For a finite value of L_g, and in the case when $\eta = 0$, the energy slightly jumps when integration points (which do not rotate with the molecule) cross the molecular surface. On the other hand, the energy remains continuous with respect to the rotation angle α when $\eta > 0$, and its variations go to zero when L_g tends to infinity, the decay rate increasing with η. A loose convergence criterion ($\texttt{Tol} = 10^{-2}$) has been used to get the results in Fig. 7, but almost identical curves are obtained if a smaller value of $\texttt{Tol}$ is employed.
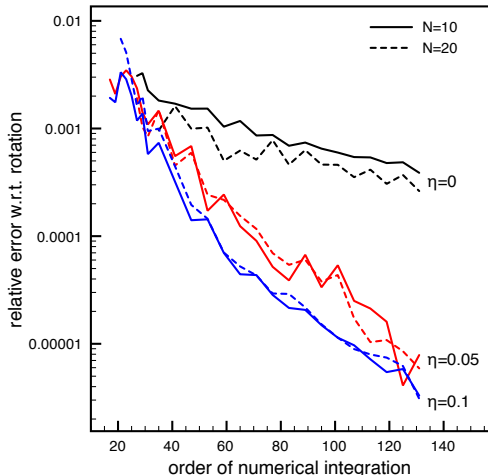
Figure 7: Maximal relative variation of the solvation energy of hydrogen fluoride for different values of η and N (with `Tol` $= 10^{-2}$).

4. Robustness with respect to geometrical parameters

The robustness of the domain decomposition scheme with respect to geometrical parameters, namely the centers of the balls and their radii, is extremely important in the perspective of computing parametric derivatives (such as atomic forces). On the basis of the numerical results obtained in the previous sections, we take $N = 10$, $\eta = 0.1$, `Tol` $= 10^{-2}$, and let L_g vary. In each of the upcoming tests, we compute a reference solution with parameters $N = 10$, $\eta = 0.1$, `Tol` $= 10^{-2}$, $L_g = 131$ and study the solvation energy and its relative error with respect to the reference computation.

Fig. 8 presents the influence of the distance in $[0.887, 0.980]$Å between the centers of two atoms of hydrogen fluoride. First, we note that the higher the order of numerical integration, the smaller the amplitude of the relative error, which is only around 0.13% even for $L_g = 17$. In additional tests we have observed that the scheme converges slower in the absence of the regularization ($\eta = 0$).

In both cases, we observe a smooth dependency of the solvation energy on the geometrical parameters for any value of the numerical integration order.

Fig. 9 shows the dependency of the energy and its relative error with respect to the reference computation with $L_g = 131$ on the multiplicative factor $\beta \in [1, 1.2]$ of the UFF Van der Waals radii of hydrogen fluoride. We can observe again that the variation in the
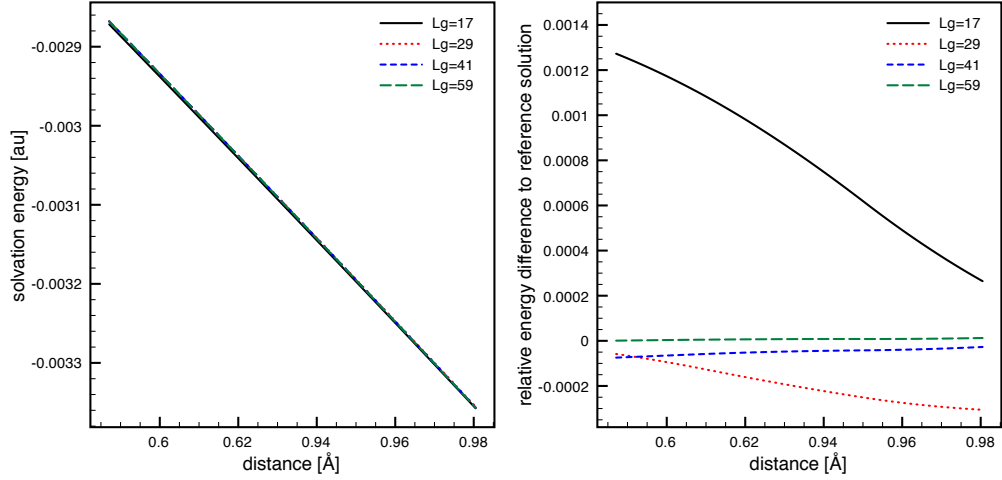
21

Figure 8: Solvation energy and its relative error when the distance between the nuclei of hydrogen fluoride varies in the range $[0.887, 0.980]$ Å, for different values of L_g (with $\mathtt{Tol} = 10^{-2}$, $N = 10$ and $\eta = 0.1$).
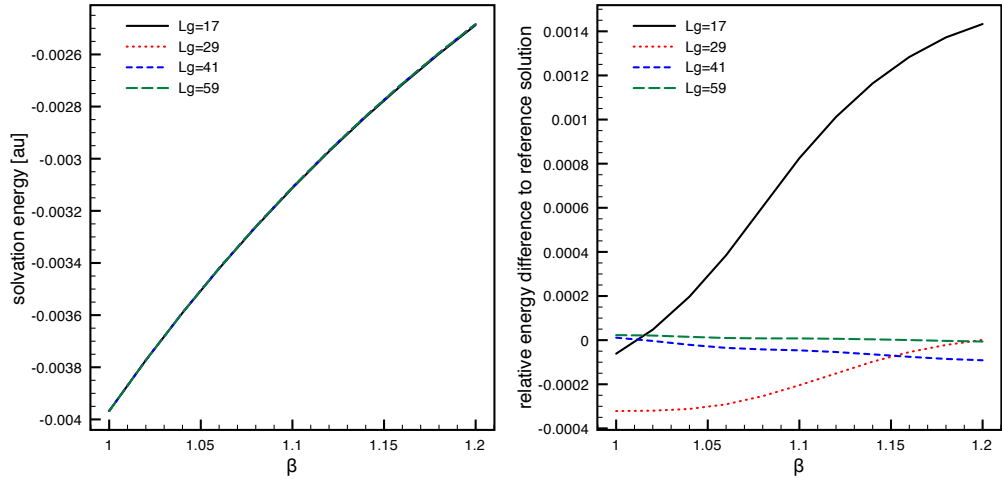

Figure 9: Solvation energy and its relative error of hydrogen fluoride depending on the multiplicative factor $\beta \in [1, 1.2]$ of the UFF Van der Waals radii, for different values of L_g (with $\mathtt{Tol} = 10^{-2}$, $N = 10$ and $\eta = 0.1$).

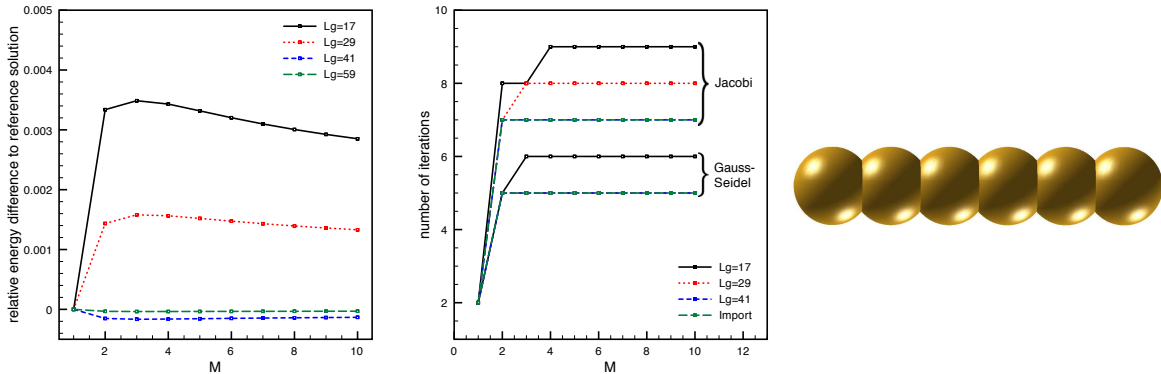relative error can be controlled by the order of the integration scheme.

Figure 10: Relative error on the solvation energy and number of iterations for a chain of M unit spheres separated by a distance $d = 1.5$, for different values of L_g (with `Tol` $= 10^{-2}$, $N = 10$ and $\eta = 0.1$). A particular configuration with $M = 6$ spheres is illustrated on the right.

B.   Multiple spheres configurations

In this section, we study the convergence properties of the proposed scheme depending on the structure and topology of the molecule. In order to carefully analyze the behavior of the numerical method in extreme cases, we allow ourselves to consider benchmark cavities not directly connected to real molecules.

Consider a chain of M aligned unit spheres, each containing a unit point charge (in atomic units) at its center, such that the distance between the centers of two adjacent spheres is equal to $d = 1.5$. Fig. 10 illustrates the dependency on M of the relative error on the solvation energy (with respect to reference calculations performed with $L_g = 131$) and the number of iterations to reach convergence. We observe that, in the range $1 \leq M \leq 10$, the number of iterations is independent of M. Supplementary tests for $M = 10, 20, \ldots, 200$ show that the number of iterations is the same for all values of M (up to $M = 200$).

We now study the dependency of the number of iterations with respect to the number of simultaneously intersecting balls. For this purpose, we consider M balls of radius 1.25 whose centers are uniformly distributed on a unit circle (Fig. 11). We observe that the number of iterations seems to stagnate (we have observed in additional tests that this is not the case if $\eta = 0$).

Finally, we consider a Y-shaped geometry as illustrated in Fig. 12 and study the variation of the solvation energy and number of iterations with the angle α. Note that when α varies
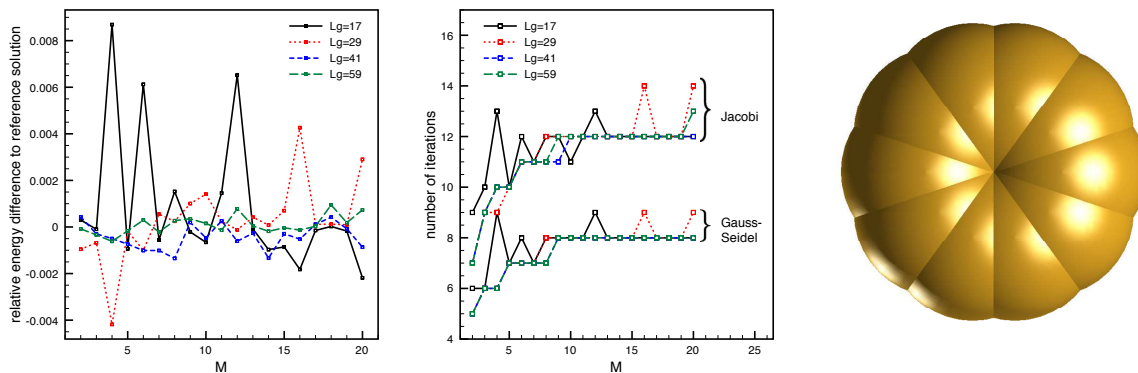
23

Figure 11: Relative error on the solvation energy and number of iterations for a cavity consisting of M intersecting balls of radius 1.2 whose centers are uniformly distributed on the unit circle, for different values of L_g (with $\texttt{Tol} = 10^{-2}$, $N = 10$ and $\eta = 0.1$). A particular configuration with $M = 10$ balls is illustrated on the right.

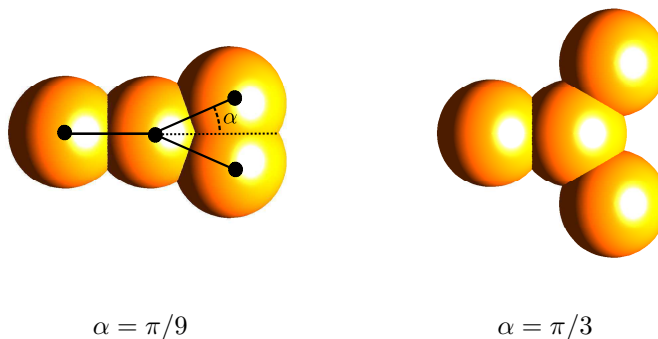$\alpha = \pi/9$ $\qquad\qquad\qquad\qquad$ $\alpha = \pi/3$

Figure 12: Configurations of the Y-shaped domain for $\alpha = \pi/9$ and $\alpha = \pi/3$.

from $\alpha = \pi/9$ to $\alpha = \pi/3$, a topological change occurs, in the sense that the triple intersection transforms to two double intersections. We do not observe any numerical artefacts in the energy due to this topological change in our numerical results (see Fig. 13).

C.   Application to small molecules

We finally present some test cases obtained with the van der Waals cavities of the formaldehyde, benzene and caffeine molecules, in their equilibrium geometries.

In Table II, we display the nuclear component of the solvation energy and the number of iterations to reach convergence, for different values of N and η. The tolerance is chosen
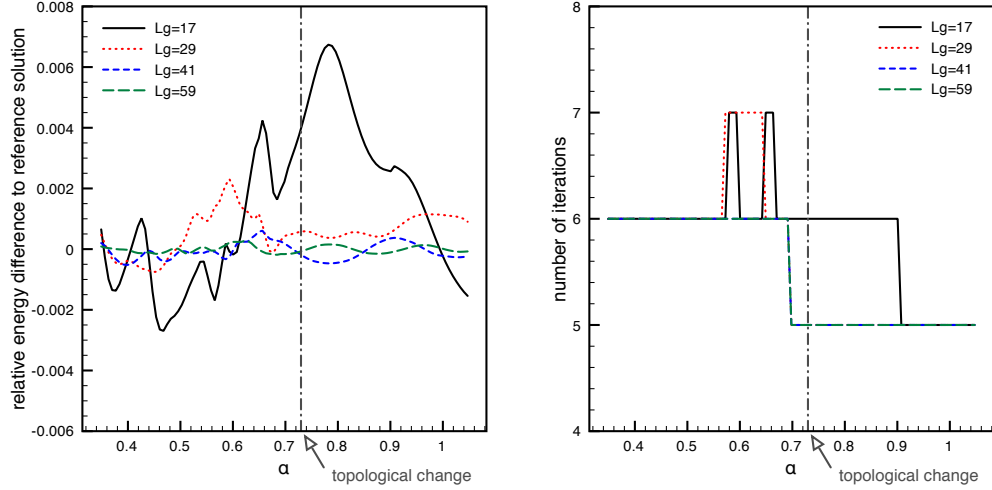
Figure 13: Relative error on the solvation energy and number of iterations for the Y-configuration as a function of the angle α, for different values of L_g (with $\texttt{Tol} = 10^{-2}$, $N = 10$ and $\eta = 0.1$).

equal to $\texttt{Tol} = 10^{-2}$, and the order of the Lebedev quadrature according to the empirical rule (for any N)

$$\eta = 0 \; \Rightarrow \; L_g = 119, \qquad \eta = 0.05 \; \Rightarrow \; L_g = 113, \qquad \eta = 0.1 \; \Rightarrow \; L_g = 89. \qquad (20)$$

We have chosen large values for L_g in order to make sure that the Gauss-Seidel algorithm converges for values of $0 \leq \eta \leq 0.1$ and $5 \leq N \leq 30$ (as mentioned in Section V A 3, the algorithm might fail to converge for large values of N and small values of L_g). Fig. 14 shows the corresponding solutions for $N = 20$ and $\eta = 0.1$.
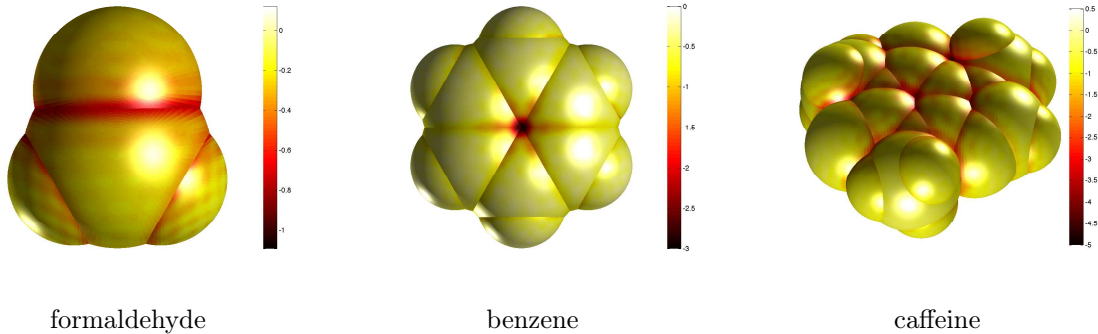

formaldehyde

benzene

caffeine

Figure 14: Graphical representations of the functions σ^j for $N = 20$ with $\eta = 0.1$ for formaldehyde, benzene and caffeine.

25

VI. CONCLUSION

In this paper we have presented the basics of a new paradigm to compute the electrostatic component of the solvation energy in implicit solvent models. The paradigm is developed here on the COSMO model, with classical charge distributions and van der Waals molecular cavities. It is inspired from the Schwarz domain decomposition method, which is widely used in fluid and solid mechanics to solve problems arising from the discretization of partial differential equations with finite element, finite volume, or spectral methods. In our approach, the Schwarz method is combined, for the first time in this context, with the integral equation method. For van der Waals molecular cavities, this approach is extremely powerful as the resulting integral equations (Eq. (14)) are set on M copies on the unit sphere, and can be efficiently discretized using spherical harmonics basis sets. In particular, it is meshless, and does not require to solve an integral equation on the whole molecular cavity. We believe that this approach, which i) is mathematically sound, ii) provides smooth potential energy surfaces, and iii) performs very well on large molecules (see Ref.[24], where linear scaling on biological molecules with more than 10^4 atoms is reported), is a method of choice for the numerical simulation of implicit solvation models.

In this first paper, we have detailed the theoretical foundations of the method, and its practical implementation. We have thoroughly tested the respective effects of the various numerical parameters entering in the definition of the method, namely (i) the regularization parameter η, (ii) the spherical harmonics cut-off N, (iii) the number of Lebedev quadrature points N_g, and (iv) the tolerance `Tol` used in the stopping criterion for the iterations of the Schwarz algorithm. We also checked the robustness of the method with respect to the geometrical parameters (atomic positions and sphere radii).

This numerical study shows that good choices for the parameters η and N are $\eta = 0.1$ and $N = 10$. The regularization parameter η turns out to have a major impact on the performance of the discretization scheme: the value $\eta = 0.1$ provides a good approximation of the solvation energy, and allows us to obtain smooth potential energy surfaces with a relatively low number of Lebedev quadrature points, and with a controlled number of iterations, basically independent of the geometry of the molecule. To some extend, the parameter η can be considered as a parameter of the model rather than a numerical parameter, of the same nature as the radius of the rolling sphere used to define solvent accessible and solvent excluded

surfaces in implicit solvation models[10]. We do not comment further here on the choices of the parameters N_g and `Tol`, and refer to Ref.[24], where the effects of these parameters on the accuracy of the atomic forces (and not only of the energy) are investigated.

VII. ACKNOWLEDGEMENTS

REFERENCES

[1] B. Mennucci and R. Cammi (eds.), *Continuum Solvation Models in Chemical Physics*, Wiley, New York, 2007.

[2] J. Tomasi, B. Mennucci, and R. Cammi, *Chem. Rev.* 105 (2005) 2999.

[3] B. Mennucci, *J. Phys. Chem. Lett.* 10 (2010) 1666.

[4] A.V. Marenich, C.J. Cramer and D.G. Truhlar, *J. Phys. Chem. B* 113 (2009) 6378.

[5] F. Lipparini, G. Scalmani, B. Mennucci, E. Cancès, M. Caricato and M.J. Frisch, *J. Chem. Phys.* 133 (2010) 014106.

[6] E. Cancès and B. Mennucci, J. Chem. Phys. 115 (2001) 6130.

[7] A. Klamt and G. Schüürmann, *J. Chem. Soc., Perkin Transactions* 2 (1993) 799.

[8] V. Barone and M. Cossi, J. Phys. Chem. A 102 (1998) 1995.

[9] J.L. Pascual-Ahuir, E. Silla and I. Tunon, *J. Comput. Chem* 15 (1994) 1127.

[10] J. Tomasi and M. Persico, *Chem. Rev.* 94 (1994) 2027.

[11] S. Miertuš, E. Scrocco and J. Tomasi, *Chem. Phys.* 55 (1981) 117.

[12] R. Cammi and J. Tomasi, *J. Comput. Chem.* 12 (1995) 1449.

[13] E. Cancès, B. Mennucci and J. Tomasi, *J. Chem. Phys.* 107 (1997) 3032.

[14] B. Mennucci, E. Cancès and J. Tomasi, *J. Phys. Chem. B* 49 (1997) 10506.

[15] M. Cossi, G. Scalmani, N. Rega and V. Barone, *J. Chem. Phys.* 117 (2002) 43.

[16] M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Comput. Chem.* 24 (2003) 669.

[17] G. Scalmani, V. Barone, K.N. Kudin, C.S. Pomelli, G.E. Scuseria, and M.J. Frisch, *Theor. Chem. Acc.* 111 (2004) 90.

[18] D.M. York and M. Karplus, *J. Phys. Chem. A* 103 (1999) 11060.

[19] G. Scalmani and M. Frisch, *J. Chem. Phys.* 132 (2010) 114110.

[20] A.W. Lange and J.M. Herbert, *J. Chem. Phys.* 133 (2010) 244111.

[21] A.W. Lange and J.M. Herbert, *J. Phys. Chem. Lett.* 1 (2010) 556.

[22] S.W. Rick and B.J. Berne, *J. Am. Chem. Soc.* 118 (1996) 672.

[23] F. Lipparini and V. Barone, *Chem. Theory Comput.* 11 (2011) 3711.

[24] F. Lipparini, B. Stamm, E. Cancès, Y. Maday and B. Mennucci, submitted.

[25] A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard and W.M. Skiff, *J. Am. Chem. Soc.* 114 (25) (1992) 10024-10035.

[26] H.A. Schwarz, Vierteljahrsschrift der Naturforschenden Gesellschaft in ZÃŒrich 15 (1870) 272.

[27] A. Quarteroni and A. Valli, *Domain decomposition methods for partial differential equations*, Oxford Science Publications, 1999.

[28] S.G. Mikhlin, Dokl. Akad. Nauk SSSR 77 (1951) 569 (in Russian).

[29] P.-L. Lions, in: R. Glowinski, G.H. Golub, G.A. Meurant and J. Périaut, eds., *First international symposium on domain decomposition methods for partial differential equations*, SIAM, Philadelphia, 1988.

[30] R.A. Adams, *Sobolev spaces*, Academic Press, Boston, 1975.

[31] M.A. Wieczorek and F.J. Simons, Geophys. J. Int. 162 (2005) 655.

[32] V.I. Lebedev and D. N. Laikov, Dokl. Math. 59 (1999), 477.

[33] A.K. Rappe and W.A. Goddard, *J. Phys. Chem.* 95 (8) (1991) 3358–3363

Formaldehyde:

$\eta = 0$		
N	Energy	# Iterat.
5	-0.00190623	4
10	-0.00191091	4
15	-0.00191277	4
20	-0.00191337	4
25	-0.0019136	4
30	-0.00191379	4

$\eta = 0.05$		
N	Energy	# Iterat.
5	-0.00191451	4
10	-0.00191651	4
15	-0.00191709	4
20	-0.00191711	4
25	-0.00191702	4
30	-0.00191697	4

$\eta = 0.1$		
N	Energy	# Iterat.
5	-0.00192264	4
10	-0.0019225	4
15	-0.00192226	4
20	-0.00192205	4
25	-0.00192192	4
30	-0.00192188	4

Benzene:

$\eta = 0$		
N	Energy	# Iterat.
5	-0.000332549	7
10	-0.00034019	7
15	-0.000342704	7
20	-0.000343872	7
25	-0.000344457	7
30	-0.000344804	7

$\eta = 0.05$		
N	Energy	# Iterat.
5	-0.000340895	6
10	-0.000346098	6
15	-0.000347498	6
20	-0.000347901	6
25	-0.000348055	6
30	-0.000348111	6

$\eta = 0.1$		
N	Energy	# Iterat.
5	-0.00034783	6
10	-0.000351349	6
15	-0.000352029	6
20	-0.000352108	6
25	-0.000352107	6
30	-0.000352104	6

Caffeine:

$\eta = 0$		
N	Energy	# Iterat.
5	-0.00349124	7
10	-0.00351916	7
15	-0.00352696	7
20	-0.00353052	7
25	-0.00353237	7
30	-0.00353352	7

$\eta = 0.05$		
N	Energy	# Iterat.
5	-0.0035257	7
10	-0.00354396	7
15	-0.0035479	7
20	-0.00354907	7
25	-0.00354956	7
30	-0.00354981	7

$\eta = 0.1$		
N	Energy	# Iterat.
5	-0.00355698	7
10	-0.00356799	7
15	-0.00356935	7
20	-0.00356928	7
25	-0.0035692	7
30	-0.00356916	7

Table II: Energy and number of iterations for formaldehyde, benzene and caffeine depending on N for different values of $\eta = 0, 0.05, 0.1$.