POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

A Scalable Solver for the Linearized Poisson-Boltzmann Equation on Cartesian Grids with Hierarchical Local Refinement

Laurea Magistrale in Mathematical Engineering - Ingegneria Matematica

Author: Martina Politi

Advisor: Prof. Carlo de Falco

Co-advisors: Dr. Walter Rocchia, Dr. Sergio Decherchi

Academic year: 2020-2021

1. Introduction

This thesis deals with numerical methods and High Performance Computing (HPC) Techniques for the computation of the electrostatic potential at the surface of complex molecules (typically proteins) in aqueous solutions.
The mathematical modeling tool we choose to use for this purpose is the Poisson Boltzmann Equation (PBE), which, in mathematical terms, is a boundary value problem for a semi-linear elliptic operator with discontinuous coefficient and point sources.
The introduction of such model dates back almost exactly one century ago to the seminal work of Born [5], but its relevance for chemical and biological applications, *e.g.* for drug discovery, still holds [10], furthermore, when applied to large and complex molecules, such as, *e.g.* the SARS-CoV-2 spike protein [11], the solution of the PBE is a challenging benchmark for state–of–the–art HPC techniques. Among many popular open implementations of PBE solvers we note APBS [8] and Delphi [13, 14, 16, 17], the former implements both a Finite Element (FE) discretization method on adaptive, conforming, simplicial meshes [4, 9] (more accurate, less efficient) and a Finite Differences (FD) scheme on tensor product cartesian grids, while the latter strongly relies on the benefits of FD difference schemes on cartesian grids in terms of memory efficiency and scalability. Here we focus on the performance and scalability assessment of numerical methods for the PBE based on hierarchically refined cartesian Oct-tree grids [1, 2], which is a topic that received a growing interest in the research community in recent years [6, 11, 15].
One aspect in the modeling of protein electrostatics by means of the PBE that has been shown to play an important role both in terms of accuracy of the results and of computational efficiency, is that of the geometrical description of molecular surfaces in this study we consider and compare different approaches for the surface definition, ranging from those based on implicit level-set representations to those based on alpha-shapes [7, 18].

2. The Poisson–Boltzmann Equation

Figure 1 shows schematically the geometric representation of the solvated molecule that is the basis of the PBE modeling approach. In this representation Ω_m represents the interior of the molecule which is treated as a linear dielectric

continuum medium characterized by a relative permittivity ε_m, while the solvent occupies the region Ω_s which is also modeled as a continuum characterized by permittivity ε_s and by a volume density of charge $\rho^s(\varphi(x))$ and Γ denotes the molecular surface. The points x_i denote the centers of the atoms composing the molecule, where point charges q_i are located.

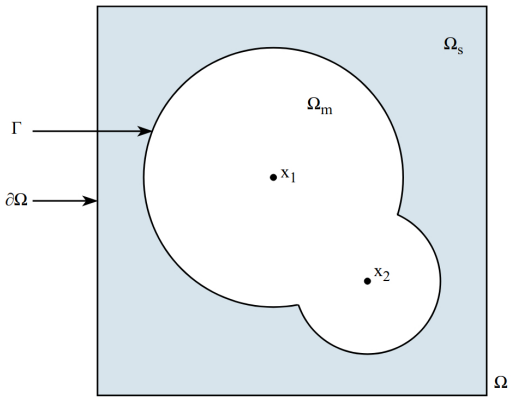


Figure 1: Schematic representation of the problem domain.

The scalar electrostatic potential $\varphi(x)$ obeys

$$-\nabla \cdot (\varepsilon_0\ \varepsilon(x)\ \nabla\varphi(x)) = \ \rho^s(\varphi(x)) + \sum_i q_i\ \delta(x - x_i), \tag{1}$$

which must be complemented by suitable boundary conditions on $\partial\Omega$, the most common choice being that of homogeneous Dirichlet conditions. By assuming the charge density in the solvent to obey a Maxwell-Boltzmann equilibrium distribution ad by performing a suitable adimensionalization, (1) becomes

$$-\hat{\nabla} \cdot (\hat{\varepsilon}_0\ \varepsilon_r(\hat{x})\nabla\hat{\varphi}(\hat{x})) + \hat{\varepsilon}_0\ \varepsilon_r(\hat{x})\ \kappa^2(\hat{x}) \sinh(\hat{\varphi}(\hat{x})) = \hat{\rho}^f(\hat{x}), \tag{2}$$

where the symbol $\hat{\cdot}$ indicates scaled nondimensional quantities and will be dropped after this section to simplify the notation.
$\hat{\rho}^f$ in (2) represents the (scaled) fixed charge density, which is a sum of point sources, while $\hat{\varepsilon}_0\kappa^2$ is the (scaled) ionic strength.
When $\varphi(x)$ is small, one can linearize (2) so that it reduces to

$$-\hat{\nabla} \cdot \left(\hat{\varepsilon}_0\ \varepsilon_r\ \hat{\nabla}\hat{\varphi}\right) + \hat{\varepsilon}_0\ \varepsilon_r\ \kappa^2\ \hat{\varphi} = \hat{\rho}^f, \tag{3}$$

where the dependency on x has been dropped for sake of brevity.
Equation (3) is known as the *linearized PBE* and is the object of the study carried out in the present thesis.
At least two different values of ε_r are needed for a decent description of the system physic. Indeed, the solvent (in particular water) is more sensitive to the presence of an electric field, so that the corresponding ε_r will be higher than the solute one. The ratio between relative permittivity values in the solute and in the molecule is often of about two orders of magnitude, therefore the problem can be considered one with *high contrast*.
The most challenging issues to be dealt with in solving (3) are

1. the very large size and complex structure of the molecules being studied; which lead to very large scale algebraic systems of equations and demand for state-of-the-art HPC solver technology,
2. the complex geometry of the molecular surface, across which coefficient discontinuities occur;
3. the presence of point sources that lead to singularities in the problem coefficients and solutions.

We will briefly describe the approaches we adopted in order to tackle each of such issues in the following sections.

3. Treatment of Point Sources

The simplest approach to deal with the singularities introduced by the point sources, is to allow charges a *finite volume*, *i.e.* to modify the fixed charge density by replacing the Dirac δ distributions appearing in (1) with suitable shape functions with finite compact support

$$\rho^f(x) = q_i u_i(x), \quad \text{where} \int u_i(x)\ \mathrm{dx}^3 = 1 \tag{4}$$

so that the total (net) amount of charge is preserved.
When (3) is to be solved in a space of Finite Elements $V_h = \mathrm{Span}\{v_k\}$, usually the u_i are expressed in terms of the v_k so that

$$u_i(x) = \sum_k w_{ik} v_k(x). \tag{5}$$

In the thesis we discuss a simple choice for the weights w_{ik} that preserves total charge, while

minimizing dependence of numerical solutions on mesh rotation an translation.
It is worth noting that point sources can also be eliminated from the problem by using one of the so-called *regularized reformulations* of the PBE (see, *e.g.*, [12] for a review on this topic).
To provide an example of the reformulation technique, consider decomposing $\varphi(x)$ as $\varphi(x) = \varphi_s(x) + \varphi_m(x)$, where $\varphi_m(x)$ is the contribution to the electrostatic potential due only to point sources, while $\varphi_s(x)$ accounts for the correction due to the surface polarization charge related to the discontinuity in the permittivity.
By means of the latter decomposition, the linearized PBE becomes

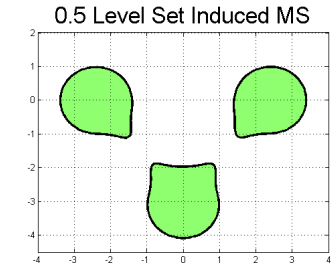
$$\begin{cases} -\nabla\cdot(\varepsilon_s\nabla\varphi_s) = \\ = \dfrac{1}{4\pi\varepsilon_0\varepsilon_m}\nabla\varepsilon_s\cdot\displaystyle\sum_i \nabla\Big(\frac{q_i}{|x-x_i|}\Big) + \\ + \dfrac{\rho_s}{\varepsilon_0}\Big(\varphi_s + \dfrac{1}{4\pi\varepsilon_0\varepsilon_m}\displaystyle\sum_i \frac{q_i}{|x-x_i|}\Big) & \text{in } \Omega_s \\ -\Delta\varphi_s = 0 & \text{in } \Omega_m \\ [[-\varepsilon\nabla\varphi_s\cdot n]]_\Gamma = [[\varepsilon]]_\Gamma\nabla\varphi_m\cdot n & \text{on } \Gamma \\ \varphi_s(x) = \dfrac{1}{4\pi\varepsilon_0\varepsilon_m}\displaystyle\sum_i \frac{q_i}{|x-x_i|} & \text{on } \partial\Omega \end{cases}$$

The obtained system has no point sources, indeed their effect is now transferred to the third and fourth equations, where the interface condition and the boundary conditions are imposed. At the interface the equality between the jumps $[[\cdot]]_\Gamma$ of the given quantities across Γ is enforced. Notice that such jums can be interpreted as a distribution of surface charges on Γ.
The main advantage of this regularization technique is to avoid the need of using small mesh sizes in Ω_m.

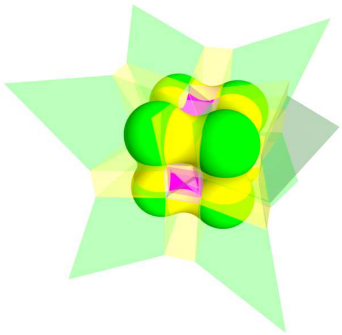
4. Geometric Representation of the Molecular Surface

The molecular surface is the 2D manifold across which the relative permittivity ε has its discontinuity.
A set of different approaches have been proposed in the literature to construct the geometry of the surface of a molecule, given the atomic centers and radii.
In this thesis, we consider and compare three different surface definitions, schematically shown in Figure 2, *i.e.* :

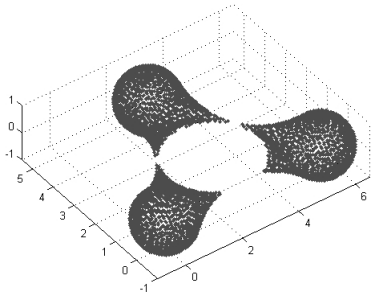
1. *Blobby surface*;
2. *Skin surface*;
3. *Solvent Excluded Surface* (SES), most commonly known as Connolly surface.



(a) 0.5 level-set, Blobby surface



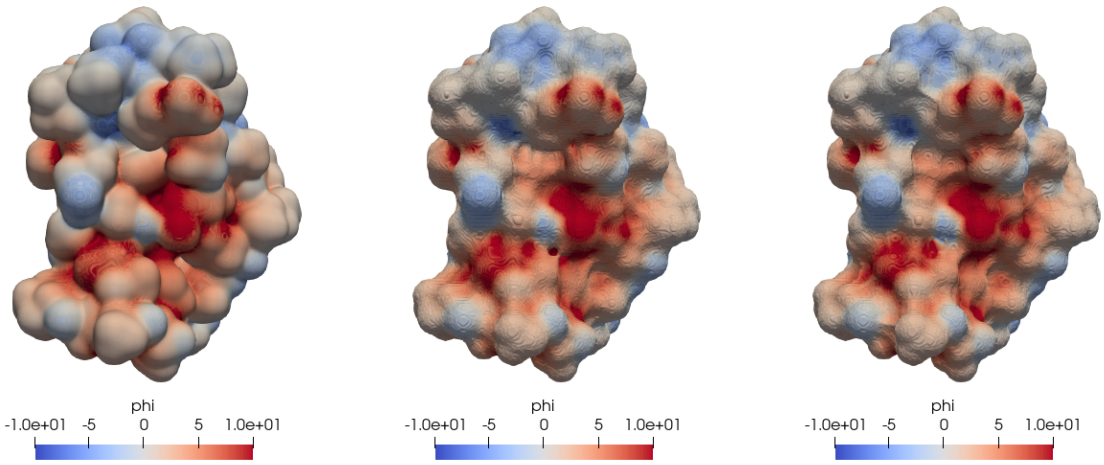
(b) Skin surface



(c) Connolly surface

Figure 2: Different molecular surfaces [7].

The first one is based on an implicit level-set representation while the latter two rely on alpha-shapes. Alpha-shapes are a parametrized generalization of the convex hull concept.
Figure 3 shows the electric potential on a crambin protein simulated using the three different approaches to describe the surface.
Efficiently constructing the surface and evaluating its interior and exterior region has a significant impact on the overall simulation complexity, for this reason much effort during the development of the thesis has been put in constructing an effective interface with the NanoShaper library [7] to deal with this task.



(a) Blobby with B = -2.5. (b) Skin with s = 0.45. (c) SES with probe radius 1.4Å.

Figure 3: Surface for a crambin molecule using different surface representations.

5. Oct-tree grids

The large size and complex structure of the examined molecules leads to very large scale algebraic systems of equations, which necessitates the usage HPC Techniques.
Most existing PBE solvers adopt two different kind of grids, either tensor product cartesian grids, or adaptive, simplicial, conforming meshes. Both these options have some advantages and some limitations, in particular the first is characterized by high memory efficiency, but it is not so accurate because it uses simple numerical methods, the second, since the resulting mesh is composed of a very large number of degrees of freedom, is more accurate, but less efficient. The idea for our work was to choose a sort of compromise between the two, solving the linearized PBE with Finite Element method on Oct-tree, which are hierarchically refined, non conforming, cartesian grids.
In this way we obtained accurate results, given by the adaptive refinement of the mesh, coupled with efficiency, given by the lower number of degrees of freedom in cartesian meshes.
The structure of an Oct-tree grid can be easily represented as a tree, as the name suggests, where each octant (*i.e.* each element of the mesh) is a either a leaf or has 8 children, as the Figure 4 shows.

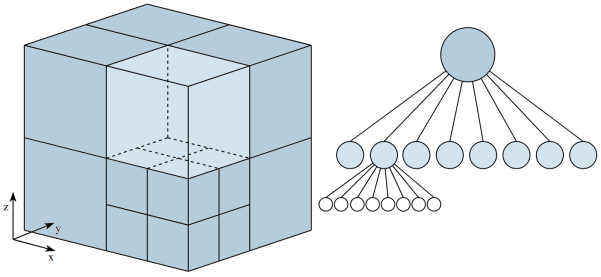


Figure 4: Oct-tree mesh with the corresponding tree.

A refining-coarsening iterative cycle is performed, starting from a uniform grid, at the end of which the resulting Oct-tree mesh has benefits in terms of efficient parallel partitioning and balancing. In particular, parallel partitioning means redistribute the octants present inside the mesh according to a given number of elements per processor or according to some prescribed weights associated to those elements. This operation is performed in an efficient way because the resulting element numbering limits communications among processors. For what concerns balancing, it ensures at most 2:1 edge size relations between neighboring octants. This is done in order to properly deal with the *hanging nodes*, points belonging to an edge or a face which is refined for two or four elements, but not for the neighboring one. They have to be considered in a different way because don't belong to the set of the degrees of freedom, but are identified through their parents indices.

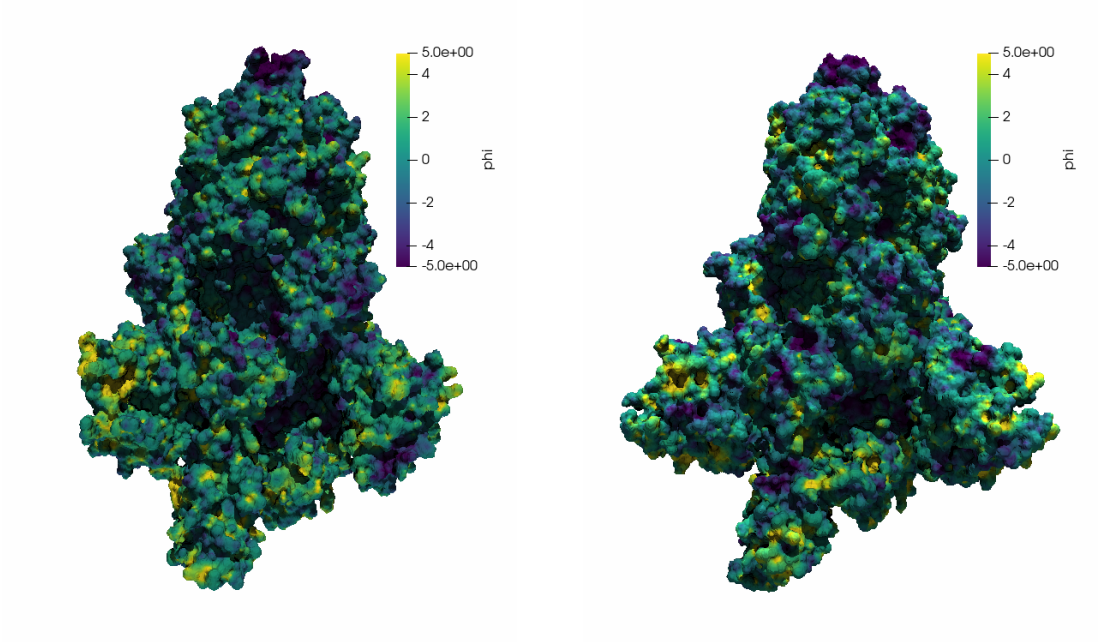


Figure 5: SARS-CoV-2, spike protein, ectodomain structure open state (left), Omicron variant of concern (right). One possible application of PBE simulations is that of identifying electrostatic potential patterns that can be useful for automatic drug discovery driven by Artificial Intelligence.

6. Results and Conclusions

Even though it has a very long history, the use of molecular electrostatics simulations in chemistry and biology is still an important tool in research today, especially as, by making use of cutting-edge HPC techniques, it allows to simulate ever larger structures.
During this thesis we have developed a parallel solver for the linearized PBE on adaptive meshes and have evaluated its parallel scalability (see Figure 6) and its viability for simulating very large structures of relevance, *e.g.*, for biomedical research (see Figure 5).

7. Acknowledgements

We kindly Acknowledge IIT ConceptLAB for interesting discussion and technical support.

References

[1] `http://p4est.github.io`.

[2] `http://optimad.github.io/bitpit/modules/index.html#PABLO`.

[3] `https://www.politesi.polimi.it/handle/10589/144819`.

[4] N. Baker, M. Holst, and F. Wang. Adaptive multilevel finite element solution of the poisson-boltzmann equation ii. refinement at solvent-accessible surfaces in biomolecular systems. *J. Comput. Chem.*, 21:1343–1352.

[5] M. Born. Volumen und hydratationswärme der ionen. *Z. Phys.*, 1(45), 1920.

[6] A. H. Boschitsch and M. O. Fenley. A fast and robust poisson–boltzmann solver based on adaptive cartesian grids. *Journal of Chemical Theory and Computation*, 7(5):524–1540, 05 2011.

[7] Sergio Decherchi, José Colmenares, Chiara Eva Catalano, Michela Spagnuolo, Emil Alexov, and Walter Rocchia. Between algorithm and model: different molecular surface definitions for the poisson-boltzmann based electrostatic characterization of biomolecules in solution. *Commun Comput Phys.*, 13:61–89, January 2013.

[8] E. Jurrus et al. Improvements to the APBS biomolecular solvation software suite. *Protein Science*, 27:112–128, 2018.

[9] M. Holst, N. Baker, and F. Wang. Adaptive multilevel finite element solution of the

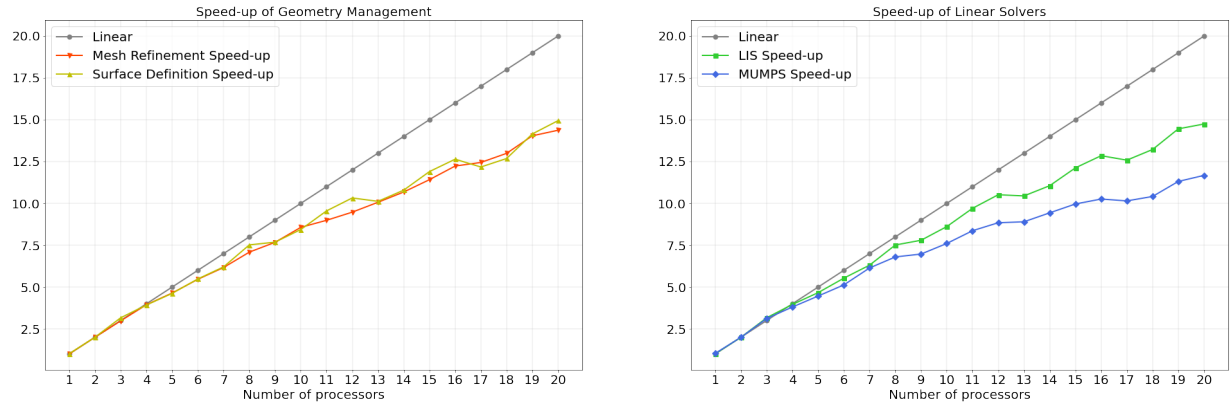


Figure 6: Results of parallel scalability tests that have been performed on the computing server GIGAT of the MOX lab of Politecnico di Milano (6 nodes, 12 Intel Xeon E5-2640v4 @ 2.40GHz, 120 cores, 64GB RAM per node, O.S. CentOS 7 interconnected by a dedicated Gigabit Ethernet).

poisson–boltzmann equation i. algorithms and examples. *J. Comput. Chem.*, 21:1319–1342.

[10] Barry Honig and Anthony Nicholls. Classical electrostatics in biology and chemistry. *SCIENCE*, 268, MAY 1995.

[11] Anna Kucherova, Selma Strango, Shahar Sukenik, and Maxime Theillard. Computational modeling of protein conformational changes - application to the opening sars-cov-2 spike. *Journal of Computational Physics*, 444:110591, 2021.

[12] Arum Lee, Weihua Geng, and Shan Zhao. Regularization methods for the poisson-boltzmann equation: Comparison and accuracy recovery. *Journal of Computational Physics*, 426:109958, 2021.

[13] C. Li, Z. Jia, A. Chakravorty, S. Pahari, Y. Peng, S. Basu, M. Koirala, S.K. Panday, M. Petukh, L. Li, and E. Alexov. Delphi suite: New developments and review of functionalities. *J Comput Chem.*, 40(28):2502–2508, Oct 2019.

[14] L. Li, C. Li, S. Sarkar, J. Zhang, S. Witham, Z. Zhang, L. Wang, N. Smith, M. Petukh, and E. Alexov. Delphi: a comprehensive suite for delphi software and associated resources. *BMC Biophys*, 4(1), May 2012.

[15] Mohammad Mirzadeh, Maxime Theillard, and Frédéric Gibou. A second-order discretization of the nonlinear poisson–boltzmann equation over irregular geometries using non-graded adaptive cartesian grids. *Journal of Computational Physics*, 230(5):2125–2140, 2011.

[16] A. Nicholls and B. Honig. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the poisson-boltzmann equation. *J. Comput. Chem.*, 12:435–445., 1991.

[17] E. Alexov W. Rocchia and B. Honig. Extending the applicability of the nonlinear poisson-boltzmann equation: Multiple dielectric constants and multivalent ions. *J. Phys. Chem. B*, 105(28):6507–6514.

[18] J. Anthony Wilson, Andreas Bender, Taner Kaya, and Paul A. Clemons. Alpha shapes applied to molecular shape characterization exhibit novel properties compared to established shape descriptors. *J. Chem. Inf. Model.*, 49(10):2231–2241, 2009.