

Fast surface generation for biomolecular computations with the Voronoi Interface Method

Miles Detrixhe

*Department of Mechanical Engineering
University of California Santa Barbara, Santa Barbara, CA, 93106*

Mohammad Mirzadeh

MIT

Arthur Guittet

Frédéric Gibou

*Department of Mechanical Engineering
Department of Computer Science
Department of Mathematics
University of California Santa Barbara, Santa Barbara, CA, 93106*

Abstract

We improve upon the state-of-the-art level set surface generation technique for biomolecules by employing a fixed radius nearest neighbor search algorithm. By choosing an appropriate data structure we show that the time complexity of this algorithm is independent of the number of atoms in a particular biomolecule.

Keywords: Parallel computing, Poisson-Boltzmann equation, octree data structure, level set

1. Introduction
2. Poisson-Boltzmann equations
3. Biomolecular surface description
4. fast surface generation

The method of surface generation used in [4] was a brute force implementation of a modified version of Nearest Neighbor Search.

In the classic nearest neighbor search, given a query point, one aims to find the nearest element of a set of points. In computing the value of the level set function at a point, one aims to find the maximum value of the following function for all atoms in the molecule.

$$\phi(x, y, z) = \max_i \left\{ r_i - \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} \right\}$$

Because the reinitialization procedure is used to propagate the solution of ϕ away from the interface, it is sufficient to have an accurate method of computing ϕ near the interface.

This reduces it to the fixed-radius nearest neighbor problem. This is an old problem (even for molecular applications [3]) and there exist efficient ways of solving it e.g. [2, 1].

Email addresses: `mdetrixhe@engineering.ucsb.edu` (Miles Detrixhe), `(Mohammad Mirzadeh)`, `(Arthur Guittet)`, `fgibou@engineering.ucsb.edu` (Frédéric Gibou)

We use an octree data structure with each cell containing those atoms within a fixed radius of the center of the cell. The radius is related to the maximum atom radius plus the probe radius.

For instance, consider the 3J6D protein from the PDB database. It has 131664 individual atoms. If we store those atoms in an octree such that each leaf cell contains all the atoms within a fixed radius from the cell center, then we get the figure below. This is for an atom tree with max level 7. The maximum number of atoms to be queried is 622. If no atoms are within the fixed radius of a leaf cell, then we populate that cell with the nearest atom.

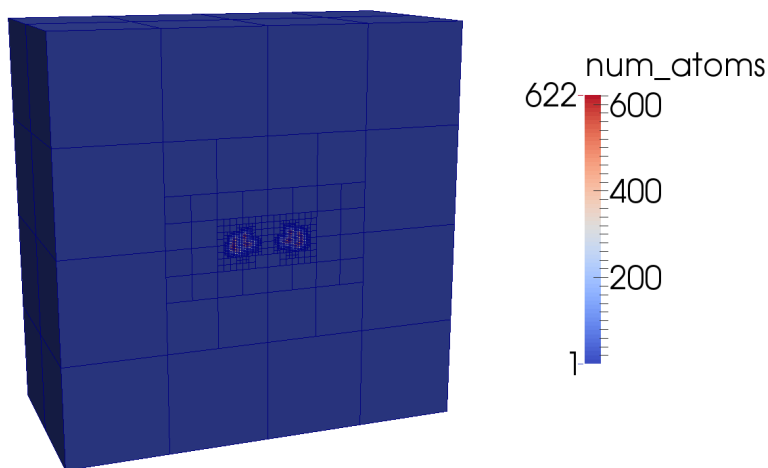


Figure 1: Figure showing the atom tree data structure. It is a quadtree with max level 7.

5. solving PB on octree - Voronoi Interface method

6. Numerical examples

6.1. Fast surface generation

MSMS [5]

6.2. Parallel scaling

There is a bug causing crashes when I run this on large numbers of processors. I haven't been able to deduce it yet.

6.3. Large example

This is a preliminary example because of the crashing issue.

Preliminary results 3J6D molecule has 131664 atoms.

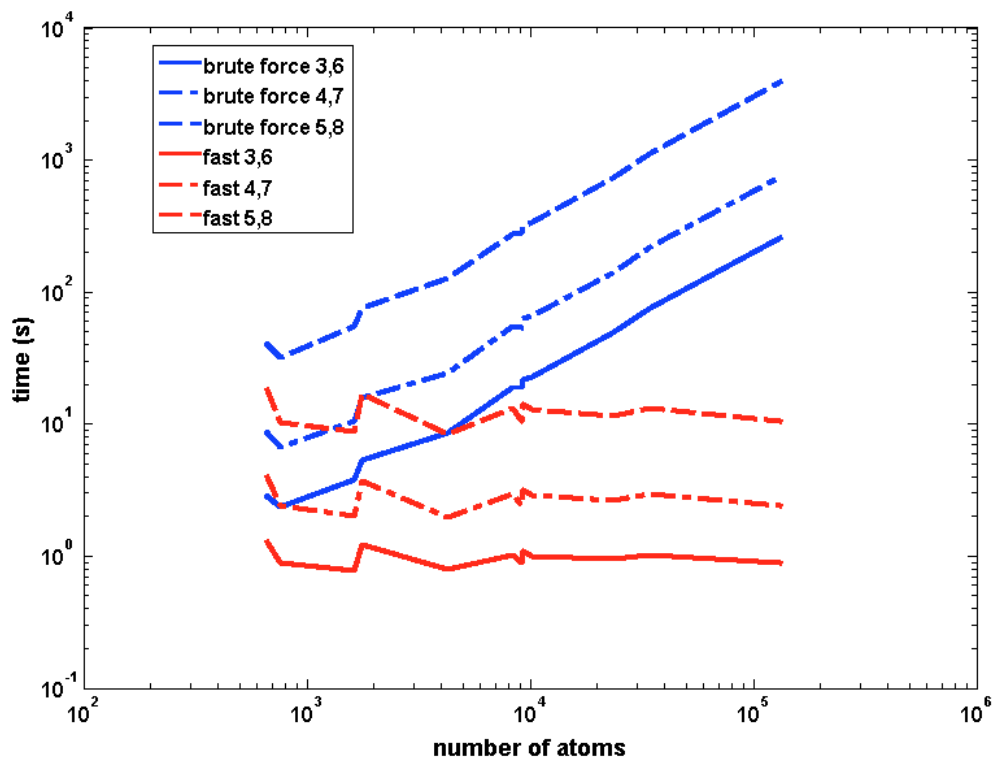


Figure 2: Surface generated using 10 different molecules from the PDB database. Exact, brute-force method used as a benchmark.

7. Conclusion

- [1] Jon L. Bentley, Donald F. Stanat, and E. Hollins Williams. The complexity of finding fixed-radius near neighbors. *Information Processing Letters*, 6(6):209–212, 1977.
- [2] Jon Louis Bentley. A Survey of techniques for fixed radius near neighbor searching. Technical report, 1975.
- [3] Cyrus Levinthal. Molecular model-building by computer. *Scientific American*, 214(6):42–52, 1966.
- [4] Mohammad Mirzadeh, Maxime Theillard, Asdis Helgadóttir, David Boy, and Frédéric Gibou. An adaptive, finite difference solver for the nonlinear Poisson-Boltzmann equation with applications to biomolecular computations. *Communications in Computational Physics*, 13(1):150–173, 2012.
- [5] Michel F Sanner, Arthur J Olson, Jean-Claude Spehner, and Michel D Sanner. REDUCED SURFACE: an Efficient Way to Compute Molecular Surfaces. *Biopolymers*, 38(383):305–320, 1996.

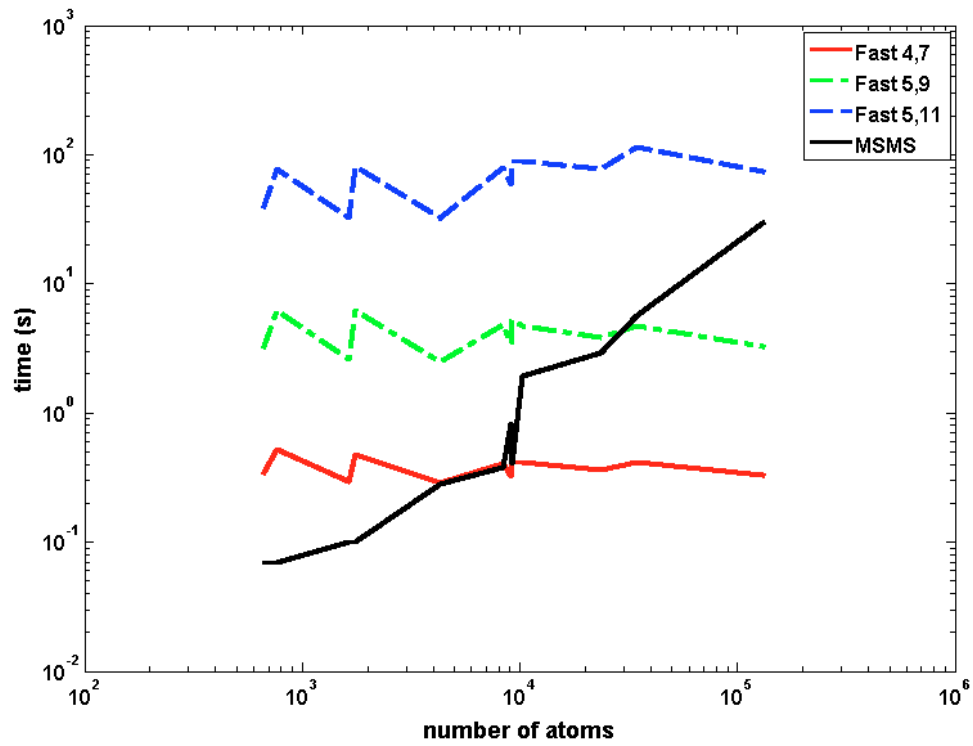


Figure 3: Time to generate surface on various grid levels. MSMS timings used as a benchmark. This shows that for large atoms, this method is faster than MSMS even on relatively fine octrees (max level 11).

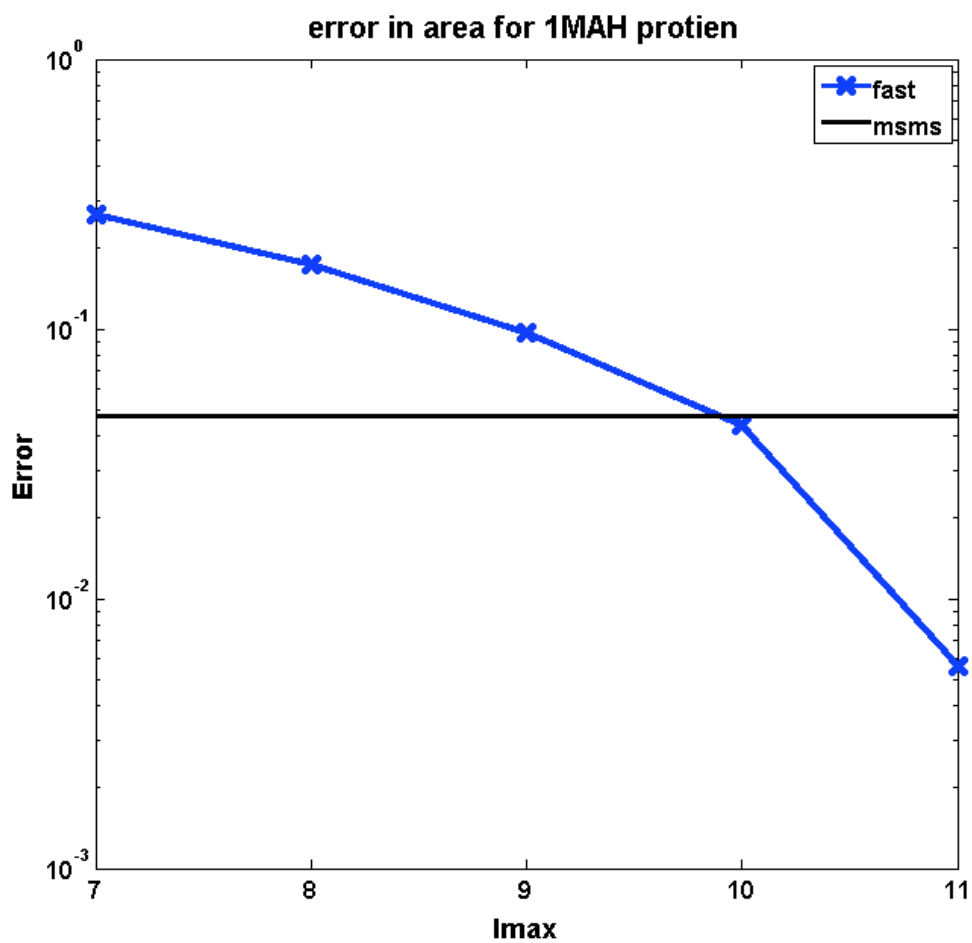


Figure 4: Error in total computed surface area for the 1MAH protien level set method compared to MSMS. This shows that the surface area accuracy is competitive with MSMS for moderately refined grids (max level 10).

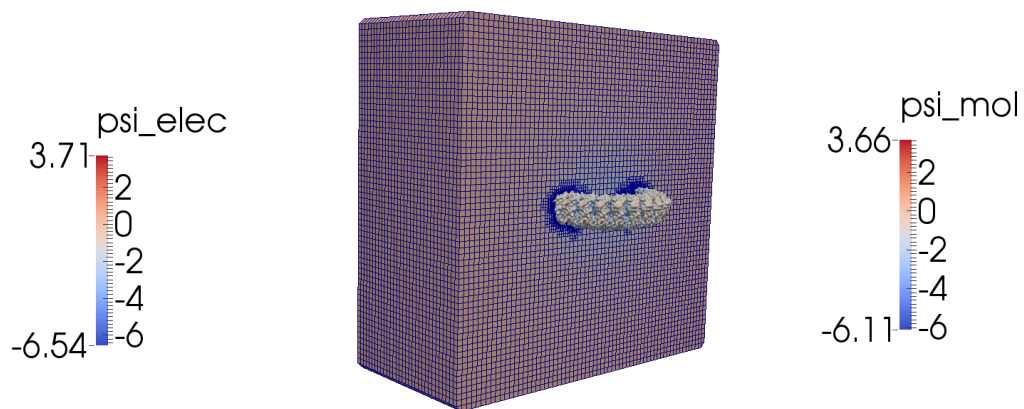


Figure 5: A cutaway of the computational domain (colored by bulk potential) and the computed SES surface of the 3J6D protien colored by the electric potential on its surface.

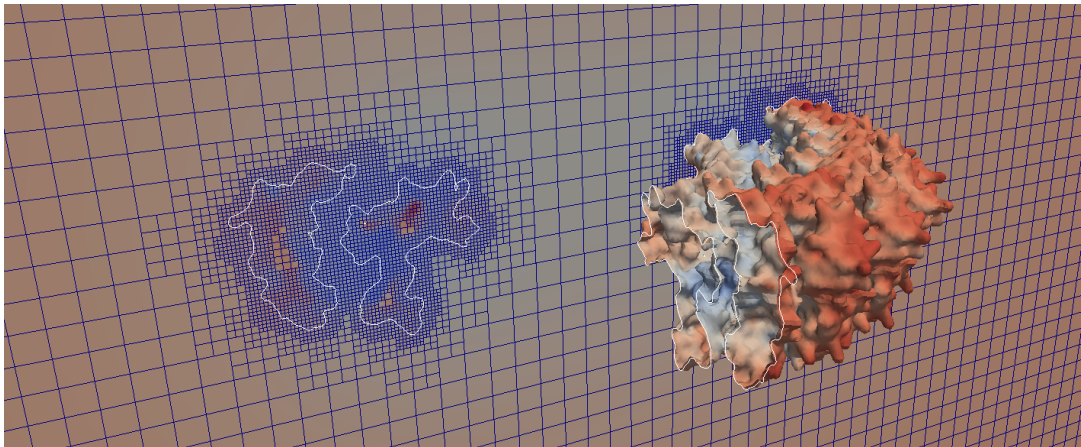


Figure 6: Zoomed in version