

Computational Approaches for the High-Throughput Analysis of Porous Materials for Energy-related Applications

M Haranczyk,¹ R L Martin,¹ Prabhat,¹ J A Sethian^{1,2}, E W Bethel¹

¹Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

²Department of Mathematics, University of California, Berkeley, CA 94720, USA

Email: mharanczyk@lbl.gov

Abstract. Crystalline porous materials are an important class of materials that can be used for oil refinement, hydrogen or methane storage, as well as carbon dioxide capture. Selecting optimal materials for these important applications requires analysis and screening of millions of potential candidates. Recently, we proposed an automatic approach based on the fast marching method for analyzing void space inside materials, a critical step preceding expensive molecular simulations. This article summarizes our progress in the development of parallel multicore CPU and GPU versions of our tool that enables high-throughput material screening on modern high-performance computing platforms.

1. Introduction

Crystalline porous materials such as zeolites have found wide use in industry since the late 1950s. They are used as chemical catalysts, in particular as cracking catalysts in oil refinement, as membranes for separations, and as water softeners [1]. Their value in these applications is estimated at \$350 billion per year [3]. There is an increasing interest in utilizing these and similar materials as membranes or adsorbents for carbon dioxide capture applications as well as for hydrogen or natural gas storage [4,5]. The need for optimal materials, which would be inexpensive, safe, and efficient, has stimulated researchers to develop databases containing millions of predicted material structures [6]. Development of such databases holds great promise for discovery of new and better materials. However, it is now being realized that in order to make such discoveries possible, new computational and cheminformatics techniques have to be developed to characterize, categorize, and screen such large databases. In particular, traditional approaches that involve visual analysis of each of the investigated structures have to be replaced with fast, automatic approaches based on mathematical algorithms and computation.

An important aspect of analysis of porous materials and their void space (Fig. 1) is the detection of inaccessible pockets, which can be occupied by guest molecules in computer

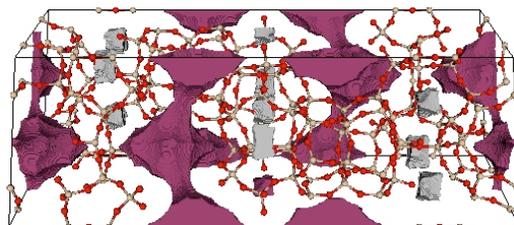


Figure 1. Shown is a periodic unit cell of a DDR zeolite crystal. Purple and gray isosurfaces highlight the surface of pores corresponding to channels and inaccessible pockets, respectively.

calculations, even though such pockets are inaccessible in adsorption experiments. It is important to account for, and often artificially block or exclude, these pockets in the calculation of guest-accessible volumes and surface areas or the prediction of guest-related properties, using molecular simulation techniques. Our group has provided a number of approaches that facilitate such analysis of void space. These include tools for visual analysis of abstract structure representations called chemical hieroglyphs [7] as well as automated computational geometry (e.g., Voronoi decomposition-based) approaches [8]. However, both approaches as well as others presented in the literature attempt to study paths of a spherical probe representing the molecule inside a convex hull constructed from atoms of a material's framework.

Recently, we developed advanced algorithms to determine accessibility of the void space for probes that mimic complex and flexible shapes of molecules [9]. We model complex objects built from solid blocks connected by flexible links (molecular worms), which are able to change orientation and/or shape during the traversal of a chemical structure, allowing them to reach areas not accessible to either a single large spherical probe or rigid real-shape probes. Our approach is based on a partial differential equations (PDEs)-based front propagation technique and can be fully automated, thereby enabling execution of molecular simulations-based calculation of materials' properties in an unsupervised, high-throughput manner [10]. The new capabilities offered by our approach come at the price of increased computational cost associated with the discretization and analysis of the configuration space of a guest molecule inside the material. This article summarizes our current efforts in mitigating the cost by porting the code to parallel platforms and integrating the tools into the material discovery pipeline.

2. Approach, Algorithms, and Implementations

We cast this problem of determination of accessibility of a molecule as a Hamilton-Jacobi-type Eikonal equation in configuration space describing the guest molecule inside a material:

$$|\nabla U| = C(x).$$

Here, U is the minimal total cost, and $C(x)$ is a cost function defined at each point x in the domain, corresponding to its ability to be occupied. Abstractly, this cost function is defined at the beginning of the problem, and the solution $U(x)$ to the above problem represents the total cost, which is the smallest obtainable integral of $C(x)$, considered over all possible trajectories throughout the computational domain from a start point to a finish point. The latter feature can then be used to construct practical techniques enabling, for example, the determination of shortest paths [9] and prediction of the accessibility of sections of the void space [10].

The resulting practical approaches involve three major steps:

- (i) Calculating $C(x)$ cost at each grid point x , representing considered configurations. The time requirement scales linearly with N , where N is the total number of mesh points in the computational domain. Here, we consider $(6 + k)$ -dimensional configuration space representing a guest molecule inside the material. The coordinates define the position of a molecule in the real three-dimensional space (3 coordinates), its orientation (3 coordinates), and its conformation (k coordinate(s)). The configuration space is then discretized by using specified step sizes. Then the cost function $C(x)$, where x is a grid

point, is defined as follows: $C = 1$ for each point x that can be occupied, $C = \infty$ otherwise. Here, an “occupiable” point means a position in which the probe is not colliding with any atom of the material’s framework.

- (ii) Solving the Eikonal equation by using a variant of the fast marching method (FMM) [11], a Dijkstra-like method to solve the boundary value problems of the form of the Eikonal equation. Starting with an initial position for the front, the method systematically marches the front outwards one grid point at a time, exploring all continuous pathways in the configuration space. FMM provides the solution in $O(N\log(N))$ time. Depending on the definition of initial conditions, step (ii) will provide information on shortest paths and/or accessible sections of the void space.
- (iii) Performing analysis of the resulting $U(x)$ grid, which usually requires investigation of only a small subset of N .

We profiled typical “production” runs of our serial baseline implementation [12] and discovered that step (i) is the most time consuming, typically requiring an order of magnitude more time than steps (ii) or (iii). Given this runtime profile, our strategy was to improve the most time-consuming task, namely, step (i), by enabling parallel cost calculation at each independent configuration.

We implemented optimized versions of the cost-calculation step and tested its performance on a high-end workstation with Opteron Magny-Cours processors and an NVIDIA Quadro FX 5800 card. We also tested performance of the code on an Intel workstation with an NVIDIA Fermi C2050 card. We chose these particular hardware platforms because they are representative of single nodes in our larger HPC facilities. Our code has been developed in C++ and used pthreads for the multi-threaded implementation. We used CUDA for porting our code to the GPUs and implemented several optimizations such as utilizing constant memory and overlapping computation with memory transfers. Further details of the implementation can be found in [12].

3. Results and Performance

We are now routinely using our tool to study accessibility of guest molecules in zeolites. Figure 2 presents an interesting case of WEN zeolite, which we identified as a structure that requires our FMM-based approach in order to model diffusion pathways of a CO_2 molecule. A high-resolution analysis of WEN is necessary to be confident that the nonspherical guest molecule cannot traverse the pore network. However, running a lower-resolution simulation is sufficient to determine that WEN is a “problem case,” and the computational benefit of this approach enables the high-throughput screening of databases to identify the minority of cases that require a follow-up, higher-resolution analysis.

We generated performance data for the optimized serial, multicore, and GPU implementations and present our results in Figure 3. We compare the execution time of the cost grid step for each case, as this is the focus of our optimizations. This performance test involves a medium-sized problem—the zeolite WEN, at a 0.1Å grid resolution and a 30-degree rotation angle for

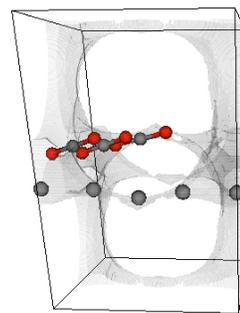


Figure 2. Prediction of accessibility of CO_2 molecule in WEN zeolite (for clarity, only the surface of the pore network is displayed) by two models: Spherical probe model can traverse the channel (from left to right) while a more complex model based on our FMM approach in 6D gives a qualitatively different result – CO_2 is stuck in the channel.

a CO₂ guest molecule. This yields a four-dimensional grid consisting of 65.86 million points, at each of which we require the periodic distance to the nearest atom in the WEN structure. Figure 3 provides the absolute time required to perform the cost grid calculation and presents these timings as speedups over the optimized serial implementation. These results show that the cost grid step scales almost linearly with the number of threads in the multicore approach, but also that significant further speedups are obtainable by using GPUs; in particular, the Fermi implementation yields a 9.74x speedup over the 24-core approach, which itself is a 23.54x speedup over serial threaded implementation.

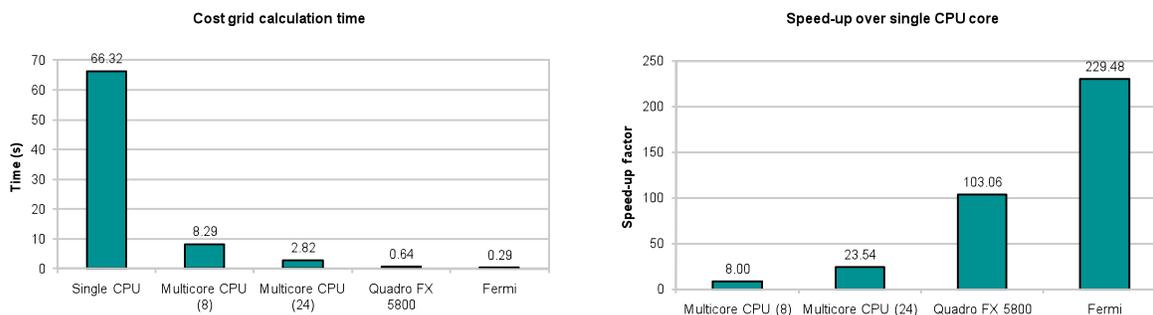


Figure 3. Performance results for the cost grid calculation for different platforms. Absolute computation time is plotted on the left, and the corresponding speedups are shown on the right.

4. Summary

Our implementation of the distance grid calculation step on multicore and GPU architectures has demonstrated the suitability of this algorithm for acceleration through exploitation of its parallel nature. The distance grid can be used in the FMM step to segment and characterize the distinct features present in a structure, and this information can be used in subsequent steps to generate other important data such as shortest paths through void space, channel and pocket volumes, and surface areas. In some applications, such as detection of inaccessible pockets, the FMM step can be exchanged for a simpler and faster flood fill algorithm.

Our collaborators at the Energy Frontier Research Center on Gas Separations now routinely use this parallel tool to analyze very large sets of materials on the order of hundreds of thousands of structures.

Acknowledgments

We thank Berend Smit, Jihan Kim, David D. Donofrio and Samuel W. Williams for valuable suggestions and NERSC for access to GPU computing facilities. This work was supported by the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy under Contract No. DE-AC02-05CH11231 through the SciDAC-e project “Accelerating Discovery of New Materials for Energy-related Gas Separations through PDE-based Mathematical and Geometrical Algorithms and Advanced Visualization Tools.” This work was supported in part (to J.A.S) by the Applied Mathematical Science subprogram of the Office of Energy Research, U.S. Department of Energy, under Contract Number DE-AC03-76SF00098, and by the Computational Mathematics Program of the National Science Foundation.

References

- [1] Auerbach S M, Carrado K A, Dutta P, Marcel K, eds. 2004. Handbook of Zeolite Science and Technology, Dekker, New York.
- [2] Smit B, Maesen T L M. 2008. *Chemical Reviews*, Molecular Simulations of Zeolites: Adsorption, Diffusion, and Shape Selectivity **108** pp 4125-4184.
- [3] Phan A, Doonan C J, Uribe-Romo F J, Knobler C B, O'Keeffe, Yaghi O M. 2010. Synthesis, structure, and carbon dioxide capture properties of zeolitic imidazolate frameworks, *Acc. Chem. Res.*, **43** pp 58-67.
- [4] Banerjee R, Phan A, Wang B, Knobler C, Furukawa H, O'Keeffe M, Yaghi O.M. 2008. High-throughput synthesis of zeolitic imidazolate frameworks and application to CO₂ capture, *Science* **319** pp 939-94.
- [5] Sumida K, Hill M R, Horike S, Dailly A, Long J R. 2009. Synthesis and Hydrogen Storage Properties of Be-12(OH)(12)(1,3,5-benzenetribenzoate) *J. Am. Chem. Soc.* (4) **131** pp 15120.
- [6] Deem M W, Pophale R, Cheeseman P A, Earl D J. 2009. Computational Discovery of New Zeolite-Like Materials, *J. Phys. Chem. C* **113** pp 21353-21360.
- [7] Theisen K, Smit B, Haranczyk M. 2010. Chemical Hieroglyphs: Abstract Depiction of Complex Void Space Topology of Nanoporous Materials, *J. Chem. Inf. Model.* **50** pp 461-469.
- [8] Willems T F, Rycroft C H, Kazi M, Meza J C, Haranczyk M. Algorithms and tools for High-Throughput Geometry-based Analysis of Crystalline Porous Materials, *Microporous and Mesoporous Materials*, submitted.
- [9] Haranczyk M, Sethian J A. 2009. Navigating Molecular Worms inside Chemical Labyrinths, *Proc. Natl. Acad. Sci. USA (PNAS)* **106** pp 21472-21477.
- [10] Haranczyk M, Sethian J A. 2010. Automatic Structure Analysis in High-Throughput Characterization of Porous Materials, *J. Chem. Theory Comput.* **6** pp 3472-3480.
- [11] Sethian J A. 1999. Level Set Methods and Fast Marching Methods, 2nd ed., Cambridge University Press, New York, pp 86-99.
- [12] Martin R L, Prabhat, Donofrio D D, Sethian J A, Haranczyk M. 2011. Accelerating Analysis of Void Space in Porous Materials on Multicore and GPU Platforms, *International Journal of High Performance Computing Applications*, under review.