RESEARCH ARTICLE | FEBRUARY 27 2015

High-speed prediction of crystal structures for organic molecules FREE

Shigeaki Obata; Hitoshi Goto

AIP Conf. Proc. 1649, 130–134 (2015)
https://doi.org/10.1063/1.4913557

High-Speed Prediction of Crystal Structures for Organic Molecules

Shigeaki Obata[1, a)] and Hitoshi Goto[1, 2]

[1]*Educational Programs on Advanced Simulation Engineering, Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi 441-8580, Japan*
[2]*Computer Science and Engineering, Graduate School of Engineering, Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi 441-8580, Japan*

[a)] Corresponding author: obata@adsim.tut.ac.jp

Abstract. We developed a master-worker type parallel algorithm for allocating tasks of crystal structure optimizations to distributed compute nodes, in order to improve a performance of simulations for crystal structure predictions. The performance experiments were demonstrated on TUT-ADSIM supercomputer system (HITACHI HA8000-tc/HT210). The experimental results show that our parallel algorithm could achieve speed-ups of 214 and 179 times using 256 processor cores on crystal structure optimizations in predictions of crystal structures for 3-aza-bicyclo(3.3.1)nonane-2,4-dione and 2-diazo-3,5-cyclohexadiene-1-one, respectively. We expect that this parallel algorithm is always possible to reduce computational costs of any crystal structure predictions.

INTRODUCTION

Crystal structure prediction (CSP) is to theoretically find stable (low-energy) crystal structures of a given molecule from only a chemical formula of the molecule [1-5]. In the research fields of organic solid-state materials and drug discovery, CSP method has received great attention as a useful theoretical tool for designing functional molecules and for screening polymorphs [6]. Therefore, we have developed a method to realize a high-speed and high-accurate CSP for organic molecular crystal structures and implemented the method into a computational chemistry program, CONFLEX, that we have developed [7-9]. This method has been applied to a semiconducting molecule, rubrene, and has succeeded to predict its three polymorphs [10]. Developments of CSP methods generally tend to focus on reliability for the exploring algorithm of possible crystal structures and on reproducibility for experimental crystal structures and relative stability between polymorphs. In this paper, we pay considerable attention to an introduction of massive parallel computing technique as another important development.

Computational procedure in CSP generally consists of three steps: (i) generation of a great number of trial crystal structures with unique molecular packing arrangements, (ii) optimization of each trial, and (iii) evaluation of their optimized crystal structures. The most time-consuming part is the optimization step for all generated trial crystal structures (generally ten thousand structures at least), because each optimization includes an enormous number of calculations for intermolecular interactions among molecules in crystal ensemble. In the case of our CSP method, the rate of computational time of the optimization step for total computation time is 99.9 %. Therefore, we thought that a key point to improvement of whole CSP calculation time is development of high-speed processing technique for the crystal structure optimization step.

In this work, in order to improve the performance of our CSP method, we develop a master-worker type parallel algorithm for allocating tasks of optimizations for a great number of trial crystal structures to distributed compute nodes of supercomputer systems. We also demonstrate performance tests of the parallel algorithm on TUT-ADSIM super computer system (HITACHI HA8000-tc/HT210) at Toyohashi University of Technology.

The Irago Conference 2014
AIP Conf. Proc. 1649, 130-134 (2015); doi: 10.1063/1.4913557

PARALLE ALGORITHM FOR CRYSTAL STRUCTURE PREDICTION

As the mentioned above, our CSP method also performs the following three steps: (i) generation of trial crystal structures, (ii) optimization of each trial, and (iii) evaluation of their optimized crystal structures. Before constructing trial crystal structures, an isolated molecule is optimized in gas phase and various oriented molecules are generated by stepwise rotating the optimized isolated molecule around the x, y, and z axes. Each oriented molecular structure is regarded as asymmetric unit. The trial crystal structures are generated by means of symmetry operations for the asymmetric unit based on pre-specified space groups. Each trial crystal structure is subjected to the crystal structure optimization. Finally, the optimized crystal structures are evaluated by their crystal energies given in Eq. 1:

$$E_{\text{crystal}} = E_{\text{intra}} + \sum_{i}^{N}\sum_{S}^{M}\sum_{J}^{N} E_{\text{inter}}(i;S,J) \qquad (1).$$

Here, E_{intra} is the sum of the intramolecular interaction energies on the molecule in the asymmetric unit, and $E_{\text{inter}}(i;S,J)$ is the interatomic interaction energy between atom i in the asymmetric unit and atom J in a symmetry-related unit S. N is the number of atoms in the asymmetry unit. M is the total number of symmetry-related units that we take into account: More explicitly speaking, the molecule in the S is included in the calculation when the closest interatomic distance between the molecules in the asymmetric unit and S is less than or equal to a cut-off radius R_{crystal}. In this work, R_{crystal} is set to 20 Å. During the calculation of $E_{\text{inter}}(i;S,J)$, the energy contribution is halved to avoid double counting. Halgren's Merck Molecular Force Filed 94 (MMFF94) [11] is employed as our crystal force field.

To improve the performance of our CSP method, the optimization tasks for all trial crystal structures are allocated to distributed compute nodes by means of a master-worker type parallel algorithm using the MPI library [12]. The master process (one MPI process) manages all tasks composed of optimization processes for trial crystal structures, and the worker processes (remaining MPI processes) perform only the given optimization processes. Steps of our parallel algorithm are described as the follows:

1. The master process creates the collection of tasks.
2. The master process sends one task to an idle worker process.
3. The worker process receives the task from the master process and performs the given task.
4. The worker process sends results to the master process after completing the task.
5. The master process receives the results from the worker process.
6. The computation repeatedly carries out 2-5 steps among worker processes until completing all tasks.

It is expected that the master-worker type parallel algorithm can automatically resolve a load imbalance among the compute nodes unless there are enormous gaps of computational cost between tasks.

EXAMPLE APPLICATIONS

As performance experiments of our parallel algorithm, we demonstrate crystal structure predictions of two organic molecules, 3-aza-bicyclo(3.3.1)nonane-2,4-dione **IV** [2,13] and 2-diazo-3,5-cyclohexadiene-1-one **XVI** [5] which were used in the CSP blind test organized by CCDC [2,5]. In the both practical demonstrations, we assume the observed space group symmetries, that is, $P2_1/c$ for **IV** and *Pbca* for **XVI**. The rotational step for varying molecular orientation in the asymmetric unit is set to 20 degrees. Only chair formation of 6-membered ring of **IV** is considered. On the other hand, there is no conformational flexibility for **XVI**. In the both CSP demonstrations of **IV** and **XVI**, total 11 664 trials have been generated and subjected to crystal structure optimization for relaxing the dimension of unit cell and molecular arrangements as assuming the rigid body of molecules.

The performance experiments have demonstrated on TUT-ADSIM supercomputer system that consists of 28 compute nodes connected by InfiniBand FDRx4. Each compute node has two intel Xeon E5-2680 processers and 64 GB memories. The theoretical peak performance and the maximal LINPACK performance of TUT-ADSIM

supercomputer system show 10.368 and 9.656 TFlops, respectively. The demonstrations have been performed in the rage of 2 to 256 processor cores.

RESULTS AND DISCUSSTION

Table 1 summaries computing times in CSP demonstrations of **IV** and **XVI**. Figure 1 shows the speed-ups of crystal structure optimizations step in each demonstration. As shown in Figure 1, our parallelization of crystal structure optimization has achieved an excellent speed-up in the both demonstrations and can reach almost ideal speed-up using up to 127 workers (Figure 1). In the use of 256 processor cores, the both demonstrations have been finished within only one hour (Table 1). The results suggest that our parallelization of crystal structure optimization will work well on much more massive supercomputing system and realize CSP of larger and complex organic molecules.

Next, we note on the slight decrease of speed-up performance in the case of the use of 255 workers (Figure 1). The practical speed-up rates for **IV** and **XVI** demonstrations are 214 and 179 times, respectively, using 255 workers. This degradation seems to be caused by large differences of computing times among crystal structure optimizations. For example, in the case of **IV** demonstration, although the 99 % (11, 597 trials) of optimizations have been completed within 4 minutes each, the remaining optimizations (1%, 67) have been taken the computing times to maximum time of 24 minutes. When a time-consuming task resides at latter of the sequence of tasks, the degradation of performance will occur even if the master-worker type parallel algorithm is used. In general, it is difficult to predict such time-consuming task. However, in the case of our CSP method, it may be possible to solve this problem by pre-evaluating all trial crystal structures and re-ordering such time-consuming tasks in the task pool of crystal structure optimization.

TABLE 1. Computing times in CSP demonstrations of **IV** and **XVI**.

Num. of workers	Computing Times[a] / hours	
	IV	**XVI**
1	196.2 (196.4)	174.7 (174.9)
3	65.4 (65.5)	58.4 (58.6)
7	28.1 (28.3)	25.1 (25.3)
15	13.2 (13.3)	11.7 (12.0)
31	6.4 (6.5)	5.8 (6.0)
63	3.2 (3.3)	2.9 (3.2)
127	1.7 (1.8)	1.6 (1.8)
255	0.9 (1.1)	1.0 (1.2)

[a]The values show total computing times of crystal structure optimizations in CSP demonstrations, and the values in parenthesis are total computing times of CSP demonstrations.

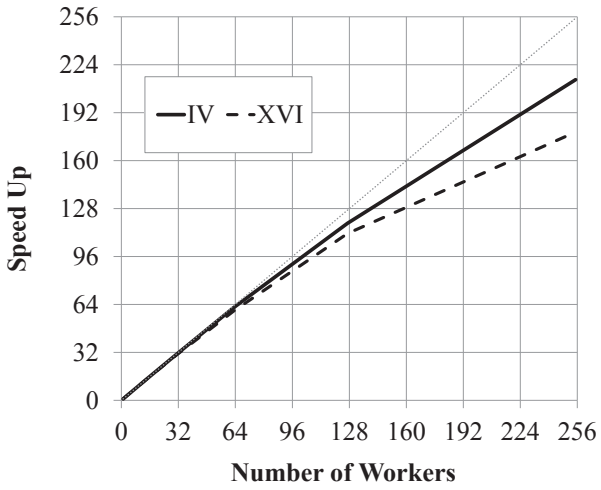

FIGURE 1. Speed-ups of crystal structure optimizations in CSP demonstrataions of **IV** (solid line) and **XVI** (dotted line).

Finally, we will briefly describe about the details of our CSP demonstration results. Our method has generated the equal numbers (11 664) of trial structures for **IV** and **XVI** molecules, and some trial structures have reached the identical optimized structure by the crystal structure optimization. As the results, we obtained 512 and 118 unique optimized structures of **IV** and **XVI**, respectively. The optimized crystal structures in good agreement with the experimentally observed ones could be found in the both demonstrations. Our CSP method could find the correct answers with respect to predictions of these two crystal structures of **IV** and **XVI**, Figure 2 shows superposition of the optimized and experimental crystal structures of **IV** and **XVI**. The root mean squared deviations of atomic positons excluding hydrogen atoms in the unit cell between the optimized and experimental crystal structures are 0.432 and 0.503 Å, respectively.

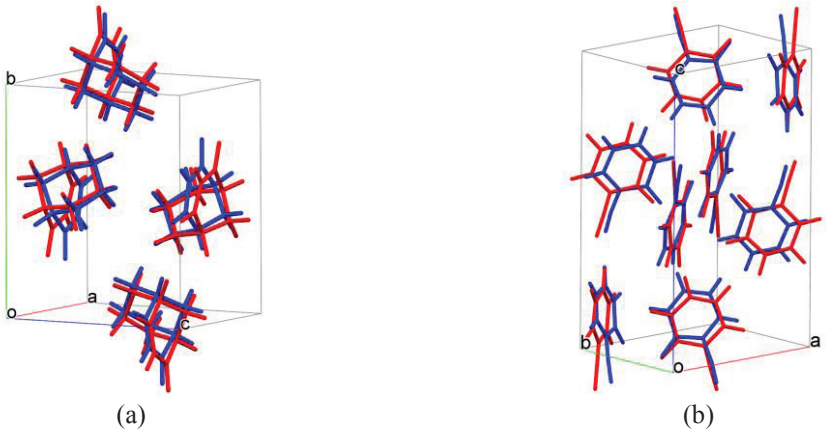

FIGURE 2. Superposition of the optimized (red) and experimental (blue) crystal structures. (a) **IV** and (b) **XVI**.

CONCLUSIONS

We developed the master-worker type parallel algorithm for allocating tasks of trial crystal structures optimizations to distributed compute nodes using the MPI library. The parallel algorithm could achieve the speed-ups of 214 and 179 times using 255 workers on the trial crystal structure optimizations in CSP demonstrations for **IV** and **XVI**, respectively. Furthermore, the CSP demonstrations could produce the optimieze crystal structures corresponding to experimental ones of **IV** and **XVI**. We expect that this new parallel algorithm is always possible to

reduce computational costs of any crystal structure predictions. Furthermore, the CSP method with the parallel algorithm can realize CSP of larger and complex organic molecules and increase the computational accuracy for the exploring possible crystal structures.

ACKNOWLEDGMENT

This work was supported by the president discretionary expenses of Toyohashi University of Technology. We also thank the promotion office for TUT programs on advanced simulation engineering (ADSIM) and the Information and Media Center (IMC) at Toyohashi University of Technology for the use of the TUT-ADSIM supercomputer system.

REFERENCES

1. J. P. M. Lommerse, W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, W. T. M. Mooij, S. L. Price, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer and D. E. Williams, Acta Cryst. B56, 697-714 (2000).
2. W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, J. P. M. Lommerse, W. T. M. Mooij, S. L. Price, H. Scheraga, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer and D. E. Williams, Acta Cryst. B58, 647-661 (2002).
3. G. M. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, B. P. van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, A. Torrisi, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt and P. Verwer, Acta Cryst. B61, 511-527 (2005).
4. G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. S. Tan, R. G. Della Valle, E. Venuti, J. Jose, S. R. Gadre, G. R. Desiraju, T. S. Thakur, B. P. van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, M. A. Neumann, F. J. J. Leusen, J. Kendrick, S. L. Price, A. J. Misquitta, P. G. Karamertzanis, G. W. A. Welch, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, J. van de Streek, A. K. Wolf and B. Schweizer, Acta Cryst. B65, 107-125 (2009).
5. D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. M. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle, G. R. Desiraju, B. P. van Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. J. J. Leusen, A. V. Maleev, A. J. Misquitta, S. Mohamed, R. J. Needs, M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, L. S. Price, S. L. Price, H. A. Scheraga, J. van de Streek, T. S. Thakur, S. Tiwari, E. Venuti and I. K. Zhitkov, Acta Cryst. B67, 535-551 (2011).
6. D. E. Braun, J. A. McMahon, L. H. Koztecki, S. L. Price, and S. M. Reutzel-Edens, Cryst. Growth Des., 14 (4), 2056–2072 (2014)
7. S. Obata and H. Goto, J. Comput. Aided Chem., 9, 8-16 (2008).
8. S. Obata and H. Goto, J. Comput. Chem., Jpn, 7, 151-164 (2008).
9. H. Goto, S. Obata, N. Nakayama, and K. Ohta, CONFLEX7; Conflex: Tokyo, Japan (2012).
10. S. Obata, T. Miura and Y. Shimoi, Jpn. J. Appl. Phys., 53, 01AD02 (2014).
11. T. A. Halgren, J. Comp. Chem., 17, 490-519 (1996).
12. Message Passing Interface Forum, http://www.mpi-forum.org/. Last access date: November 19, 2014.
13. R. A. Howie and J. M. S. Skakle, Acta Cryst. E57, o822-o824 (2001).