

Unbiased Global Optimization of Lennard-Jones Clusters for $N \leq 201$ Using the Conformational Space Annealing Method

Julian Lee,^{1,*} In-Ho Lee,^{1,2} and Jooyoung Lee^{1,†}

¹*School of Computational Sciences, Korea Institute for Advanced Study, Seoul 130-012, Korea*

²*Korea Research Institute of Standards and Science, Daejeon 305-600, Korea*

(Received 27 November 2002; published 21 August 2003)

We apply the conformational space annealing method to the Lennard-Jones clusters and find all known lowest energy configurations up to 201 atoms, *without using extra information of the problem* such as the structures of the known global energy minima. In addition, the robustness of the algorithm with respect to the randomness of initial conditions of the problem is demonstrated by ten successful independent runs up to 183 atoms. Our results indicate that this method is a general and yet efficient global optimization algorithm applicable to many systems.

DOI: 10.1103/PhysRevLett.91.080201

PACS numbers: 02.60.Pn, 36.40.-c, 61.46.+w, 82.20.Kh

Finding the global minimum (GM) of a given function, called the global optimization, is an important problem in various fields of science and engineering. One of the simplest algorithms for global optimization is the simulated annealing (SA) method [1], which has been applied to many systems. Although SA is very versatile in that it can be applied to many problems, the drawback is that its efficiency is usually much lower than problem specific algorithms. This is especially problematic for nondeterministic polynomial-time hard (NP-hard) problems such as protein folding or molecular cluster optimizations. For this reason, it is important to find an algorithm which is as general as SA, and yet competitive with problem specific ones.

Recently, a powerful global optimization method called conformational space annealing (CSA) was proposed [2], and applied extensively and exclusively to the protein folding problem [3–9]. The benchmark tests [2–4] have demonstrated that it can not only find the known GM conformations with less computations than existing algorithms, but it also provides new GMs in some cases [3,4].

The CSA unifies the essential ingredients of three global optimization methods, Monte Carlo with minimization (MCM) [10], genetic algorithm (GA) [11], and SA. First, as in MCM, we consider only the phase space of local minima; i.e., all configurations are energy minimized by a local minimizer. Second, as in GA, we consider many configurations (called *bank* in CSA) collectively, and we perturb a subset of bank configurations (*seeds*) using other bank configurations. This procedure is similar to mating in GA. However, in contrast to the typical mating procedure in GA, we often replace *small* portions of a seed with the corresponding parts of bank configurations in order to search the neighborhood of the seed configuration, as is elaborated later. Finally, as in SA, we introduce a parameter D_{cut} , which plays the role of the temperature in SA. The diversity of sampling is directly controlled in CSA by introducing a distance

measure between two configurations and comparing it with D_{cut} , whereas in SA there are no such systematic controls. The value of D_{cut} is slowly reduced just as in SA, hence the name conformational space annealing. Maintaining the diversity of the population using a distance measure was also tried in the context of GA [12], although no annealing was performed.

It should be noted that although the CSA method has been applied, so far, only to the polypeptide systems, the structure of the algorithm is not specific to these systems. In fact, to apply the CSA to an optimization problem, only two things are necessary: a method for perturbing a seed configuration, and a distance measure between two configurations. This suggests that the CSA is a candidate for a versatile and yet powerful global optimization method. In this Letter, we demonstrate it by applying the CSA to Lennard-Jones (LJ) clusters.

The LJ cluster is a system consisting of identical atoms interacting by a pairwise LJ potential:

$$E(\{\mathbf{r}\}) = 4\epsilon \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right], \quad (1)$$

where N is the number of atoms and r_{ij} is the distance between atoms i and j . We use the reduced unit where $\epsilon = \sigma = 1$. It is not only interesting as a model for heavy inert gases but also serves as a popular benchmark system for optimization algorithms. In fact, despite the simple form of the interaction, finding the GM configuration has been a challenging problem even for small N [12–37].

Many of the powerful global optimization algorithms applied to this system are specific to the LJ cluster. In particular, they use the information on the structure of the known GMs of LJ clusters, favoring closely packed ones. For example, many GMs for $N \leq 147$ were discovered for the first time in Ref. [13] by using icosahedrally derived lattices. A similar strategy [34] was utilized for $N \leq 309$. Although these approaches are powerful, they fail when the GM configuration is of an unexpected

structure. Also, since they are designed specifically for the LJ cluster problem, it is not clear if they can be applied to other systems.

Recently there were a number of successful applications of unbiased search methods to LJ clusters. One of the most powerful methods in this category is the basin-hopping method [28,31]. Almost all GMs known at present [38] were reproduced for $N \leq 110$, and those for $N = 69, 78, 107$ were updated. This is an impressive result indeed. However, there are certain magic numbers ($N = 76, 77, 103, 104$) where the GMs could not be found by directly optimizing the system consisting of N atoms. They could be found only by adding and subtracting an atom from the GM configurations of cluster size $N \pm 1$. In Ref. [35], a variant of the basin-hopping method was applied for $N \leq 110$, which performed better than the original version, but the known lowest energy minima for $N = 75-77$ could be found only 4, 2, 8 times out of 1000 independent runs. Similarly, the unbiased search in Ref. [32] could not reproduce the GMs for $N = 75-77, 98$. In Ref. [12], all GMs for $N \leq 150$ were reproduced with a unbiased search method. The same method was also applied for larger cluster sizes, and some of the known lowest energy minima for $N \leq 309$ were found [37,38]. A global optimization method which combines the idea of basin hopping and the genetic algorithm was applied for the cluster size $N \leq 309$ but failed to reproduce the known minima [38] for $N = 185, 187$.

In this work, we apply the CSA to the LJ cluster problem. We find all known GM configurations for $N \leq 201$ [38]. In particular, for each LJ cluster for $N \leq 183$, we generate ten independent random configurations and succeed in finding the GMs for all cases *without an exception*. This is an exhaustive test unprecedented in the literature and shows that our algorithm is quite robust with respect to the change of initial conditions.

To elaborate on the details of the CSA method, we first randomly generate a certain number of initial configurations (50 in this work) whose energy is subsequently minimized. We call the set of these configurations the *first bank*. We make a copy of the first bank and call it the *bank*. The configurations in the bank are updated in later stages, whereas those in the first bank are kept unchanged. Also, the number of configurations in the bank is kept unchanged when the bank is updated. The initial value of D_{cut} is set as $D_{\text{ave}}/2$, where D_{ave} is the average distance between the configurations in the first bank. New configurations are generated by choosing a certain number (20 in this work) of *seed* configurations from the bank and by replacing parts of the seeds by the corresponding parts of configurations randomly chosen from either the first bank or the bank. Random perturbations are also performed. In this work 20 and 10 configurations are generated for each seed using the partial replacements and random perturbations, respectively. Then the energies of these configurations are subsequently minimized (trial configurations).

A newly obtained local minimum configuration α is compared with those in the bank to decide how the bank should be updated. One first finds the configuration A in the bank which is closest to α with the distance $D(\alpha, A)$. If $D(\alpha, A) < D_{\text{cut}}$, α is considered as similar to A . In this case, the configuration with lower energy among α and A is kept in the bank, and the other one is discarded. However, if $D(\alpha, A) > D_{\text{cut}}$, α is regarded as distinct from all configurations in the bank. In this case, the configuration with the highest energy among the bank configurations plus α is discarded, and the rest are kept in the bank. We perform this operation for all trial configurations.

This process of generating new conformations by perturbation and subsequent local minimizations, and updating the bank, can be visualized as follows (Fig. 1). Each of the bank configurations can be considered to represent all local minima contained in the sphere with radius D_{cut} centered on it. To improve a bank configuration A , we first select A as a seed. We perturb A and subsequently energy minimize it to generate a trial configuration α . When α originates from A by small perturbation, it is likely that α is contained in a sphere centered on A . If α replaces A , the center of the sphere moves from A to α . If α belongs to a different sphere centered on B , α can replace B in a similar manner. When α is outside of all existing spheres, a new sphere centered on α is generated. In this case, to keep the number of spheres fixed, we remove the sphere represented by the highest-energy configuration. Obviously, the former two cases are more likely to happen when the spheres are large, and the latter when spheres are small. A larger value of D_{cut} produces more diverse sampling, whereas a smaller value results in quicker search of low-energy configurations at the expense of getting trapped in a basin probably far away from the GM.

Therefore, for efficient sampling of the phase space, it is necessary to maintain the diversity of sampling in the

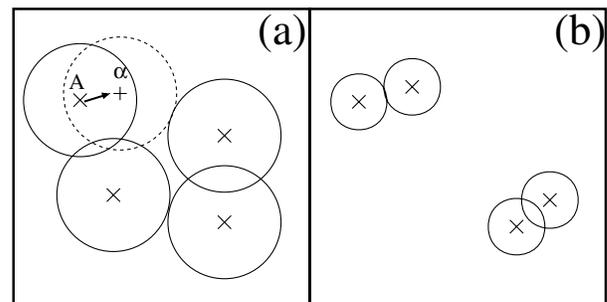


FIG. 1. Schematic figure showing the search procedure of CSA. The boxes represent the identical phase space. (a) Initially, we cover the phase space by large spheres centered on randomly chosen local minima denoted by \times and replace the centers with lower-energy local minima. When A is replaced by α , the sphere moves in the direction of the arrow. (b) As the algorithm proceeds and the energies of the representative configurations at the centers of the spheres are lowered, the size of the spheres is reduced and the search space is narrowed down to small basins of low-lying local minima.

early stages and then gradually shift the emphasis toward obtaining low-energy configurations, by slowly reducing D_{cut} . In practice, the D_{cut} is reduced by a fixed ratio after the bank update has been attempted by all the newly generated trial configurations, in such a way that D_{cut} reaches $D_{\text{ave}}/5$ after 10 000 local minimizations. Then seeds are selected again from the bank configurations which have not been used as seeds yet, to repeat the aforementioned procedure. The value of D_{cut} is kept constant after it reaches the final value.

When the energy of a seed configuration does not improve after a fixed number of perturbations, we stop perturbing it. To validate this judgment, it is important that typical perturbations are kept small, so that the perturbed configurations are close to their original seeds. However, large perturbations are also performed, in order to efficiently sample various regions of the search space.

It should be noted that in the early stages of CSA the seed configurations are continuously being replaced by low-energy local minima close to it. Therefore, when all of the bank configurations are used as seeds (one iteration completed), usually after tens of thousands of local minimizations, this implies that the procedure of updating the bank might have reached a deadlock. However, we give these configurations another chance by resetting them to be eligible for seeds and repeat another iteration of search. After a preset number of iterations, we conclude that our procedure has reached a deadlock. When this happens, we enlarge the search space by adding more random configurations into the bank and repeat the whole procedure until the stopping criterion is met. In this work, after three iterations are completed, we increase the number of bank configurations by adding 50 randomly generated and minimized configurations into the bank (and also into the first bank) and reset D_{cut} to $D_{\text{ave}}/2$. The algorithm stops when the known GM [38] is found, which is examined after all the new trial configurations are used for possible bank updates.

It should be noted that since one iteration is completed only after all bank configurations have been used as seeds, and we add random configurations whenever our search has reached a deadlock, there is no loss of generality for using particular values for the number of seeds, the number of bank configurations, etc.

We define the distance measure $D(k, k')$ as follows. Given a configuration k , we define the first and second shells centered on each atom as the spheres of radii 1.35 and 1.70. These values of the radii are obtained from the radial distribution function, which is approximately estimated from randomly generated local minimum configurations with $N = 201$, and do not depend much on N . We then obtain the histogram $H^k(1, n)$ [$H^k(2, n)$] which is the number of atoms having n atoms in the first [second] shell. The first and second coordination numbers contain the geometrical information of the local environment of each atom. Therefore, the histograms provide collective information of such local structures in a cluster. We put larger

weight for the first coordination number than the second one since the former is more important than the latter. In addition, larger weights are assigned to the histograms with larger n , since they correspond to the core part of a cluster. This motivates us to define $D(k, k')$ as

$$D(k, k') = \sum_n n [2|H^k(1, n) - H^{k'}(1, n)| + |H^k(2, n) - H^{k'}(2, n)|]. \quad (2)$$

It should be noted that the details of the definition of $D(k, k')$ do not affect the overall performance.

The method of perturbing a seed configuration s to generate a new configuration is as follows [25]. We first generate random planes passing through the center of mass of the seed s and another bank configuration k . We then choose from s a certain fraction (25%–50%) of atoms which are farthest from the plane and replace them by the corresponding counterpart in k . We also make a random rotation perpendicular to the face of the cut surface when we put together the two parts. The main difference from the method used in Ref. [25] is that we are making a fractional replacement of a seed, for the reason discussed earlier. In addition, we also generate new configurations by choosing an atom in s with the lowest coordination number and move it to the neighborhood of an atom with the second lowest coordination number.

We could reproduce all published GM configurations [38] for $N \leq 201$. To measure the overall efficiency of our algorithm, we performed exhaustive systematic runs. We performed ten independent CSA runs for each cluster size $N \leq 183$ and found all known GMs for all ten cases, *without an exception*, which demonstrates the robustness of the algorithm with respect to the randomness of the initial conditions. The same test was performed for $184 \leq N \leq 201$, but for $N = 184, 188\text{--}192, 198, 199$ the optimization was only partially successful (four through nine out of ten). The average values and fluctuations of CPU times for these runs are shown in Fig. 2. The computations were carried out on Athlon processors (1.667 GHz), and the limited-memory quasi-Newton method [39] was used for local minimization. The minimization stopped whenever $|G|/\sqrt{3N} \leq 0.001$, where G is the gradient. A similar plot was presented in Ref. [12] for $N \leq 150$, where the identical local minimizer was used. These results are also included in Fig. 2 for comparison, after the rescaling of the CPU clocks. The result suggests that our algorithm is faster on average for the sizes where the data are available, although a rigorous conclusion is difficult to be drawn due to the possible technical differences such as the stopping criteria of the local minimizer, and the fact that our result is the average of ten independent runs.

The CSA method is now also being applied to quite different kinds of problems such as the traveling salesman problems, spin glasses, not to mention complex molecular clusters. For example, the shortest path of ATT532 [40]

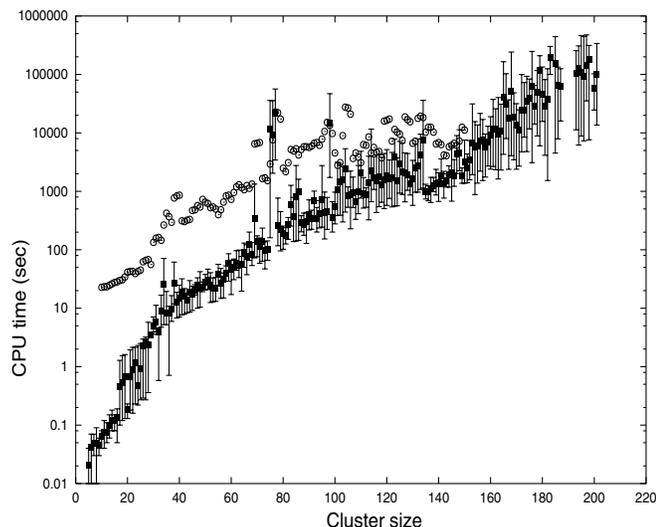


FIG. 2. CPU time (sec) of ten separate runs for cluster size $N \leq 201$. The square boxes denote the average values and the error bars indicate the ranges of the values. The partially successful runs ($N = 184, 188-192, 198, 199$) are excluded (see text). The circles denote the result for $N \leq 150$ reported in Ref. [12] after rescaling of the CPU clocks.

was found for all 100 independent runs, whose results will be reported elsewhere. This suggests that the CSA method is a general and yet efficient global optimization algorithm applicable to many systems. As is the case for GA, the CSA can also be easily adapted for efficient parallel computation [4,9].

*Current address: Department of Bioinformatics and Life Science, Soongsil University, Seoul 156-743, Korea.

†To whom all correspondence should be addressed.

Electronic address: jlee@kias.re.kr

- [1] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).
- [2] J. Lee, H. A. Scheraga, and S. Rackovsky, *J. Comput. Chem.* **18**, 1222 (1997).
- [3] J. Lee, H. A. Scheraga, and S. Rackovsky, *Biopolymers* **46**, 103 (1998).
- [4] J. Lee, H. A. Scheraga, and S. Rackovsky, *Int. J. Quantum Chem.* **75**, 255 (1999).
- [5] J. Lee, A. Liwo, D. R. Ripoll, J. Pillardy, J. A. Saunders, K. D. Gibson, and H. A. Scheraga, *Int. J. Quantum Chem.* **77**, 90 (2000).
- [6] J. Lee, A. Liwo, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2025 (1999).
- [7] A. Liwo, J. Lee, D. R. Ripoll, J. Pillardy, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5482 (1999).
- [8] J. Lee, A. Liwo, D. R. Ripoll, J. Pillardy, and H. A. Scheraga, *Proteins: Struct. Funct. Genet.* **Suppl. 3**, 204 (1999).
- [9] J. Lee, J. Pillardy, C. Czaplowski, Y. Arnautova, D. R. Ripoll, A. Liwo, K. D. Gibson, R. J. Wawak, and H. A. Scheraga, *Comput. Phys. Commun.* **128**, 399 (2000).

- [10] Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6611 (1987).
- [11] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, MA, 1989).
- [12] B. Hartke, *J. Comput. Chem.* **20**, 1752 (1999).
- [13] J. A. Northby, *J. Chem. Phys.* **87**, 6166 (1987).
- [14] D. L. Freeman and J. D. Doll, *J. Chem. Phys.* **82**, 462 (1985).
- [15] J. Farges, M. F. de Feraudy, B. Raoult, and G. Torchet, *Surf. Sci.* **156**, 370 (1985).
- [16] L. T. Wille, *Chem. Phys. Lett.* **133**, 405 (1987).
- [17] C. D. Maranas and C. A. Floudas, *J. Chem. Phys.* **97**, 7667 (1992).
- [18] S. Gomez and D. Romero, in *Proceedings of the First European Congress of Mathematics, Paris, 1992* (Birkhauser, Paris, 1994), Vol. III, p. 503.
- [19] T. Coleman, D. Shalloway, and Z. J. Wu, *J. Global Optim.* **4**, 171 (1994).
- [20] G. L. Xue, *J. Global Optim.* **4**, 425 (1994).
- [21] J. Pillardy and L. Piela, *J. Phys. Chem.* **99**, 11805 (1995).
- [22] J. P. K. Doye, D. J. Wales, and R. S. Berry, *J. Chem. Phys.* **103**, 4234 (1995).
- [23] J. P. K. Doye and D. J. Wales, *Chem. Phys. Lett.* **247**, 339 (1995).
- [24] S. K. Gregurick, M. H. Alexander, and B. Hartke, *J. Chem. Phys.* **104**, 2684 (1996).
- [25] D. M. Deaven, N. Tit, J. R. Morris, and K. M. Ho, *Chem. Phys. Lett.* **256**, 195 (1996).
- [26] R. H. Leary, *J. Global Optim.* **11**, 35 (1997).
- [27] C. Barrón, S. Gómez, and D. Romero, *Appl. Math. Lett.* **10**, 25 (1997).
- [28] D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A* **101**, 5111 (1997).
- [29] M. Wolf and U. Landman, *J. Phys. Chem. A* **102**, 6129 (1998).
- [30] R. P. White and H. R. Mayne, *Chem. Phys. Lett.* **289**, 463 (1998).
- [31] R. H. Leary and J. P. K. Doye, *Phys. Rev. E* **60**, R6320 (1999).
- [32] R. Pillardy, A. Liwo, and A. Scheraga, *J. Phys. Chem. A* **103**, 9370 (1999).
- [33] C. Barrón, S. Gómez, D. Romero, and A. Saavedra, *Appl. Math. Lett.* **12**, 85 (1999).
- [34] D. Romero, C. Barrón, and S. Gómez, *Comput. Phys. Commun.* **123**, 87 (1999).
- [35] R. H. Leary, *J. Global Optim.* **18**, 367 (2000).
- [36] S. V. Krivov, *Phys. Rev. E* **66**, 025701(R) (2002).
- [37] B. Hartke, in *Efficient Global Geometry Optimization of Atomic and Molecular Clusters*, edited by J. D. Pinter (Kluwer, Dordrecht, 2003).
- [38] The complete up-to-date list of the global minima of the LJ cluster system can be found at <http://brian.ch.cam.ac.uk/> for $N \leq 150$ and at http://www.vcl.uh.edu/~cbarron/LJ_cluster/researchpot.html for $147 \leq N \leq 309$. In this Letter, published values of global minima refer to the values listed in these sites.
- [39] D. C. Liu and J. Nocedal, *Math. Programming B* **45**, 503 (1989).
- [40] M. Padberg and G. Rinaldi, *Oper. Res. Lett.* **6**, 1 (1987).