# Improved real-space genetic algorithm for crystal structure and polymorph prediction

N. L. Abraham and M. I. J. Probert
*Department of Physics, University of York, Heslington, York YO10 5DD, United Kingdom*
(Received 21 January 2008; revised manuscript received 24 March 2008; published 30 April 2008)

Existing genetic algorithms for crystal structure and polymorph prediction can suffer from stagnation during evolution, with a consequent loss of efficiency and accuracy. An improved genetic algorithm is introduced herein which penalizes similar structures and so enhances structural diversity in the population at each generation. This is shown to improve the quality of results found for the theoretical prediction of simple model crystal structures. In particular, this method is demonstrated to find three new zero-temperature phases of the Dzugutov potential that have not been previously reported.

DOI: 10.1103/PhysRevB.77.134117 PACS number(s): 61.50.Ah, 02.70.−c

## I. INTRODUCTION

Genetic algorithms (GAs) are emerging as a useful tool in the theoretical prediction of crystal structures (see Ref. 1 and references therein).[2–4] During a GA calculation, it is possible that the system will *stagnate*. When stagnation occurs, one or more local minima dominate the search and the method is unable to find the global minimum solution. In this paper, we improve the convergence to the global minimum solution of the CASTEP-GA[1] through the use of a fitness function that is able to differentiate structures during the course of a GA minimization.

Binary-encoded GAs such as the method of Hart *et al.*[5] and Blum *et al.*[6] are able to directly compare the binary strings that make up their population members and determine if two population members are the same. In this way, it is possible to remove any highly prevalent local minimum from the population and prevent its creation in future mating operations. While this method is not possible in the framework of the CASTEP-GA, we have developed an alternative approach that significantly reduces the stagnation rate and also forces the system to explore new minima. This alternative approach is also broadly transferable to a wide range of other GAs.

## II. METHOD

Our GA method[1] is a real-space encoded technique for crystal structure prediction that takes advantage of the periodicity of the simulation supercell to improve the speed and accuracy of convergence to the global minimum free-energy crystal structure. There is a *population* of structures (or *members*) which are bred together to produce new members, such that with each subsequent *generation*, the population evolves in an attempt to determine the global minimum structure. The *fitness function* of the GA is used to determine how good ("fit") a structure is and this is then used to weight the probability of survival of that structure and its probability to produce offspring.

While this method has been very successful in the past, we wanted to reduce the stagnation rate and thereby improve the quality of the solutions produced during a GA structure search. Since this is a real-space based approach, it is not possible to directly compare the atomic coordinates of two population members to determine if they are the same structure. In our previous work,[1] the enthalpy of the structure was used to calculate the fitness. In this work, we propose augmenting this fitness function with an additional function which is able to determine the similarity of the two structures. We shall illustrate the effectiveness of this approach by first studying the Lennard-Jones crystals in comparison with our previous results and then the high pressure phases of the Dzugutov potential.[7]

The enthalpy-based fitness function is

$$f_i = \frac{[1 - \tanh(2\rho_i - 1)]}{2}, \tag{1}$$

with the variable $\rho_i$ being defined by

$$\rho_i = \frac{V_i - V_{\min}}{V_{\max} - V_{\min}}, \tag{2}$$

where $V_{\max}$ is the enthalpy of the highest-enthalpy member of the population, $V_{\min}$ is the enthalpy of the lowest-enthalpy member, and $V_i$ is the enthalpy of the member $i$ being considered. The fitness of each member $i$ is $f_i$ and this is a function that varies between 0 and 1. Population members with a fitness close to 0 are less fit, and members with a fitness close to 1 are more fit. Population members are then selected (using roulette-wheel selection) for reproduction or are removed from the population based on this fitness value.

This should mean that only fit members are selected to remain in the population or are allowed to breed (*crossover*). It is often very likely that during the course of a calculation, multiple copies of population members are made. In a bit-string represented GA, duplicate members are very easy to spot, but in a real-space encoded GA, it is very hard to tell if two members are the same during the course of a calculation, since the crystal structure may be orientated or translated in any way within the simulation cell (due to use of periodic boundary conditions). This is even harder if the simulation cell parameters are also allowed to vary during the course of a calculation.

Hence, we need a simple measure of structural similarity so that we can detect when duplicate structures exist within a population. While this is encouraging from the point of view of ultimate structural convergence, in the early stages of the GA minimization, we want to ensure as much structural di-

 

versity as possible to enable a broad search of possible solutions and so we want to penalize similar structures.

Since we are using this routine to differentiate between like and unlike structures, rather than any form of comprehensive structural analysis, we can simplify this comparison somewhat. If we are performing a calculation in which we allow the number of atoms to vary, then we can make an educated guess that two structures with different numbers of atoms are different (or rather in this case any offspring produced in the crossover procedure will have a greater number of degrees of freedom to explore the potential energy surface), so we will have no need to compare these structures. We also do not need to compare each structure with all other structures, since we are merely trying to prevent stagnation rather than give a definitive structural comparison, and so we can simply compare all structures with the minimum-enthalpy structure that has the same number of atoms as itself that exist in the current generation. We will define a comparison function between structures which returns zero if the structures are the same and one if the structures are suitably dissimilar. We also want to keep the fact that lower-enthalpy structures are "better" than higher-enthalpy ones, so we further weigh the value that any given structure has by the value of $f_i$ of the fittest member in that "set" which is made up of members with the same number of atoms. Here, we define our improved fitness function as

$$ {}^{j}f_i' = (1-\omega)\,{}^{j}f_i + \omega\,{}^{j}f_{\mathrm{fit}} \star \begin{cases} 1, & i \equiv \mathrm{fit} \\ R[\Lambda(k_r)], & i \not\equiv \mathrm{fit} \end{cases}, \qquad (3) $$

where the left superscript $j$ above denotes the comparison between groups with the same number of atoms only, $f_i$ is as defined in Eq. (1), $w$ is a weighting value between zero and one, and $R[\Lambda(k_r)]$ is a function which compares member $i$ of the set of atoms $j$ with the fittest member in that set [as defined by Eq. (1)]. This means that the fitness of the fittest member of each group $({}^{j}f_{fit})$ will be unchanged from its enthalpy value, and all other values in the group will be accordingly scaled. If the value of the fitness weight $w$ is set to 1, then the maximum value of ${}^{j}f_i'$ that any member could have is the same value of the fittest member of the group ${}^{j}f_{fit}$. If $w$ is set to zero, then this function reduces to that given in Eq. (1). The comparison function $R[\Lambda(k_r)]$ is

$$ R[\Lambda(k_r)] = \frac{\sum_{k_r} |\Lambda'(k_r) - \Lambda(k_r)|}{\sum_{k_r} \Lambda'(k_r)}, \qquad (4) $$

where consideration of the spherically averaged scattering intensity leads to

TABLE I. Table of parameters used in the Dzugutov potential (Ref. 7) [Eq. (6)].

| $m$ | $A$ | $c$ | $a$ | $B$ | $d$ | $b$ |
|---|---|---|---|---|---|---|
| 16 | 5.82 | 1.1 | 1.87 | 1.28 | 0.27 | 1.94 |

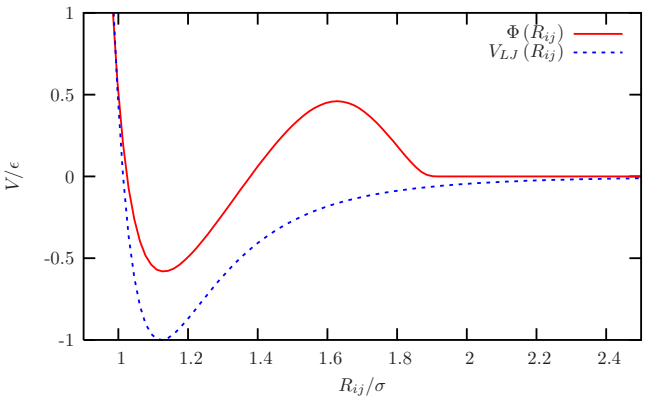


FIG. 1. (Color online) Comparison of the Lennard-Jones and Dzugutov pair potentials.

$$ \Lambda(k_r) = \Omega^2 \left[ N \sum_{n=1}^{N} \rho'^2(n) + 2 \sum_{n=1}^{N} \sum_{m>n}^{N} \rho'^2(n)\rho'^2(m) J_0 \times (\sqrt{3}\pi k_r |\mathbf{r}_n - \mathbf{r}_m|) \right], \qquad (5) $$

which is positive definite (and is based on the Debye scattering formula[8]). In Eq. (5), there are $N$ ions within the simulation cell which has a volume $\Omega$, $\rho'(n)$ is the scattering factor of ion $n$ which has the atomic real-space coordinate of $\mathbf{r}_n$, and $J_0(r)$ is a Bessel function. The function $\Lambda(k_r)$ of a population member $i$ of each group $j$ is tested against the function $\Lambda'(k_r)$ of the fittest member in the group $j$ containing the same number of atoms as member $i$. Equation (4) is then used to compare these two functions and returns a single number between 0 and 1. In a variable-supercell calculation, it is possible for this function to become greater than one when the structures are highly dissimilar, in which case we set the value of $R[\Lambda(k_r)]$ to 1.

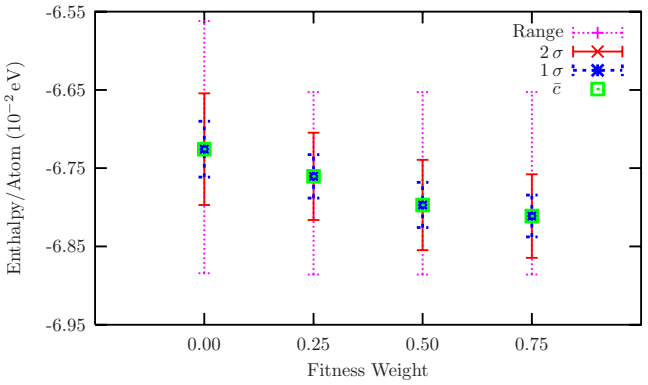


FIG. 2. (Color online) Summary of the enthalpies of the different Lennard-Jones structures found for different fitness weights, which controls how much the comparison factor is considered during selection for update and crossover. The values for $w$=0.0 are those from Abraham and Probert (Ref. 1). All points are averaged over 15 independent calculations.

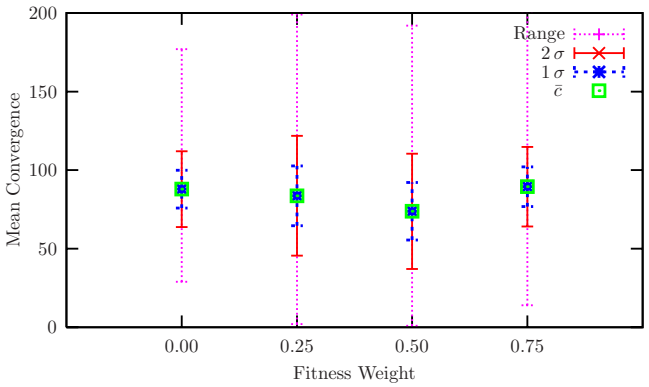


FIG. 3. (Color online) Summary of the convergence times for the results shown in Fig. 2. The values for $w=0.0$ are those from Abraham and Probert (Ref. 1). All points are averaged over 15 independent calculations.

## III. RESULTS

The results presented here will use two different empirical potentials, the Lennard-Jones potential[9,10] and the Dzugutov potential[7] which is defined as

$$\Phi(R_{ij}) = \Phi_1(R_{ij}) + \Phi_2(R_{ij}), \tag{6}$$

where

$$\Phi_1 = \begin{cases} A(R_{ij}^{-m} - B)\exp\left(\frac{c}{R_{ij}-a}\right), & R_{ij} < a \\ 0, & R_{ij} \geq a, \end{cases} \tag{7}$$

$$\Phi_2 = \begin{cases} B\exp\left(\frac{d}{R_{ij}-b}\right), & R_{ij} < b \\ 0, & R_{ij} \geq b, \end{cases} \tag{8}$$

with the constants defined in Table I. A comparison of these two potentials is shown in Fig. 1. The Dzugutov potential[7]

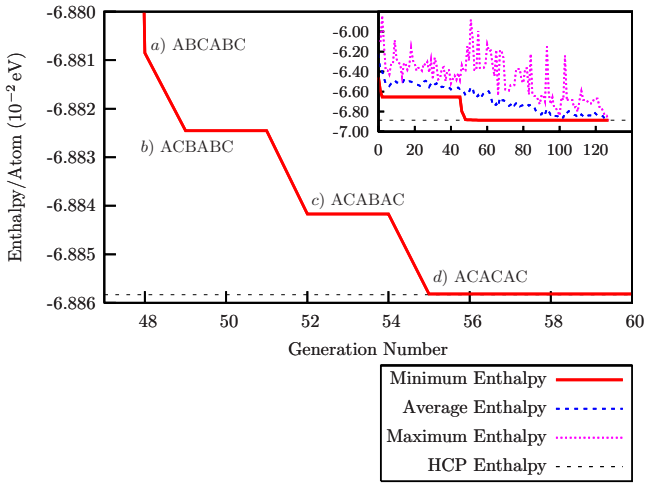


FIG. 4. (Color online) Plot showing convergence to hcp minimum structure for a Lennard-Jones calculation with $w=0.75$. The stacking patterns of the minimum-enthalpy solutions are shown next to their appearance during the course of the simulation. The system converged to a hcp structure in 55 generations, and by the 127th generation, all members were the same.

TABLE II. Comparison of the number of each ordered structure type of the lowest-enthalpy structure found (i.e., ignoring higher-enthalpy structures found during the course of a GA minimization) for different values of the fitness weighting factor $w$. Numbers given are out of a total of 15 calculations.

| Fitness weight | Pure hcp | Intermediate hcp-fcc | Pure fcc |
|---|---|---|---|
| 0.00 | 0 | 6 | 0 |
| 0.25 | 3 | 3 | 0 |
| 0.50 | 3 | 6 | 0 |
| 0.75 | 6 | 3 | 0 |

was originally formulated to simulate liquid systems; however, it has also been shown to have some interesting solid phases[11] and can also be used to form quasicrystals.[7]

The Dzugutov potential[7] is designed such that the force on, and energy of, an atom moving within the potential goes to zero at $b/\sigma$. As is reported by Roth and Denton,[11] the Dzugutov potential[7] has three known stable phases at varying pressures: bcc, the $\sigma$ phase, and fcc.

### A. Results from the Lennard-Jones potential

The use of the comparison factor in the selection procedure has a marked effect on the quality of the results produced, as shown in Fig. 2. While the global minimum structure is hexagonal close packed (hcp), this structure is nearly degenerate with the face-centered cubic structure (fcc) [with an energy difference of less than 0.1% (Ref. 12)]. There are also a number of other stacking-fault structures that exist in-between fcc and hcp. The use of the comparison factor encourages the system to explore and hence escape from

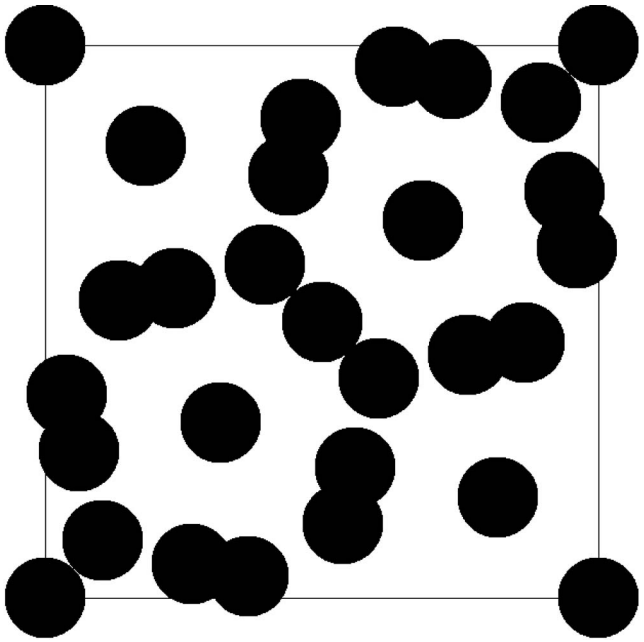

FIG. 5. The unit cell of the Dzugutov potential (Ref. 7) $\sigma$ phase looking down the $[00\bar{1}]$ direction.

TABLE III. Summary of results for 62-atom variable-cell, constrained variable-atom-number calculations. 22 independent GA calculations were performed at each pressure.

| Pressure (MPa) | Lowest enthalpy phase[a] | Higher enthalpy phase | Number of each phase found bcc | $\sigma$ | fcc | Lower enthalpy phase[b] |
|---|---|---|---|---|---|---|
| 0 | bcc | 13 | 8 | 1 | 0 | 0 |
| 50 | bcc | 2 | 16 | 1 | 0 | 3 |
| 100 | $\sigma$ | 1 | 9 | 11 | 0 | 1 |
| 150 | fcc | 1 | 0 | 0 | 15 | 6 |

[a]Data taken from Roth and Denton (Ref. 11).
[b]Where the term "lower enthalpy" refers to having lower enthalpy than the phase in column 2.

these local minima and find the hcp structure. With a fitness weight of $w=0.75$, finding a hcp structure is much more likely.

The effect on convergence is interesting, as shown in Fig. 3. There is little increase in the mean number of generations required for convergence, although there is a greater spread in the values.

Figure 4 shows the results from a calculation performed with $w=0.75$. We have included these results, in particular, because it shows the system going from an fcc structure to a hcp structure through two intermediate stacking-fault structures.

These results are summarized in Table II. The results from $w=0.00$ are those presented by Abraham and Probert.[1]

### B. Results from the Dzugutov potential

For results obtained by using this potential, an additional modification was made to the GA in the crossover step. Previously, the atom number could either be kept fixed or be allowed to vary in an unconstrained manner. For these Dzugutov calculations, a third option was added, which is to allow the atom number to vary within an allowed percentage of the original number of atoms within the simulation supercell.

While this is not necessary in a fixed-cell size and shape calculation, for a variable-cell calculation, it is essential. Without this constraint, it would be possible for the number of atoms to keep decreasing with the cell getting smaller and smaller until the minimum image convention is violated, at which point the calculation will stop. It might also allow a calculation to keep adding atoms at the crossover stage and then allow the cell to grow to accommodate them. In this way, the calculation would increase in size and take a longer and longer time for each minimization step. This percentage cutoff keeps the advantages of a variable-atom-number calculation without these problems.

It is already known that the Dzugutov $\sigma$ phase has a complicated 30-atom unit cell (see Fig. 5) and so all calculations had to have at least this many atoms. To prevent any bias of the final results, we started each run with 62 atoms in the unit cell and allowed the number of atoms to vary, in order to have an unbiased search of a large enough phase space.

A summary of the Dzugatov results is given in Table III. Calculations were performed at four pressures, 0, 50, 100, and 150 MPa, which allows each of the three structures suggested by Roth and Denton[11] to be the most stable at least one point during the experiment.

As can be seen in Table III, a number of GA minimizations found structures with a lower enthalpy than the previ-

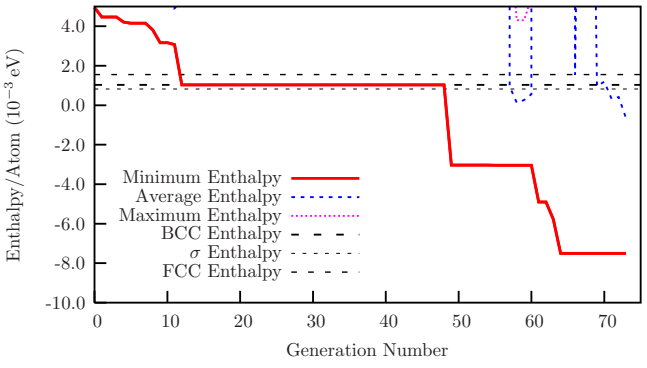


FIG. 6. (Color online) Convergence plot of a variable-atom, variable-cell calculation, starting from 62 atoms. This gives rise to a previously unknown phase (labeled as phase $a$ in Fig. 8). The inset shows the complete calculation. The minimum-enthalpy structure found has 65 atoms and is shown in Fig. 7.

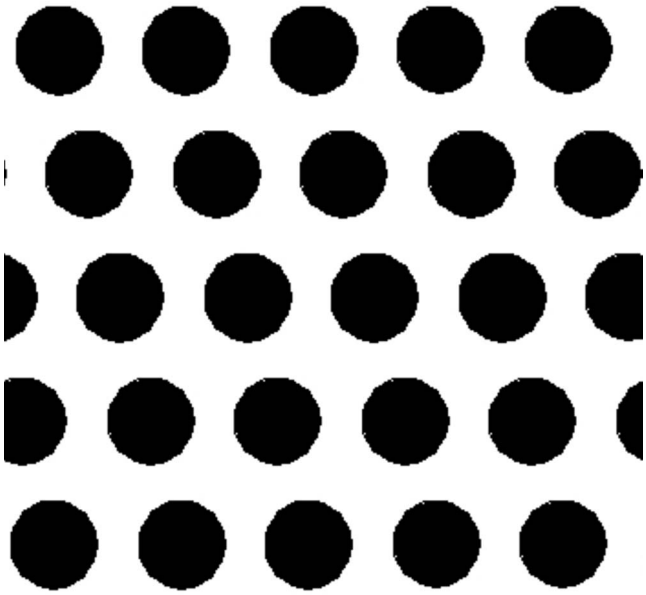

FIG. 7. The structure of the new phase $a$, a 65-atom phase found in generation 64 of the calculation shown in Fig. 6.

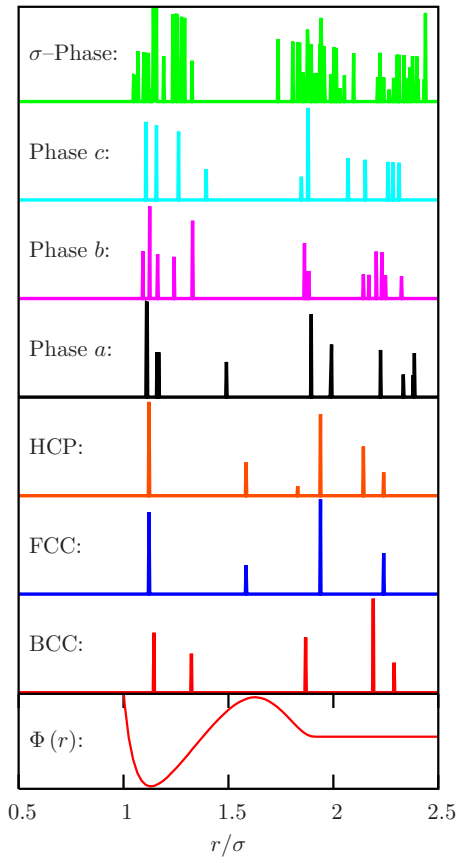


FIG. 8. (Color online) Comparison of the radial distribution function $g(r)$ for the distinct lower-enthalpy structures found with bcc, fcc, hcp, and the $\sigma$ phase. The Dzugutov potential (Ref. 7) is also shown.

ously reported minimum-enthalpy structure. In total, three distinct new structures were found. A plot showing the progress of a GA minimization down to the new lowest-enthalpy structure found is shown in Fig. 6 with the structure itself shown in Fig. 7.

A plot comparing the radial distribution functions of all the known and unknown phases of the Dzugutov potential[7] is shown in Fig. 8, as well as the hcp phase. The three new phases found are all significantly different from the established phases of this potential. Examination of these phases suggests that the simulation cells correspond to primitive cells and not to supercells, but attempts to further characterize these structures by space group that have so far been unsuccessful. The atomic coordinates of these structures are available online.[13]

A plot showing the energy-volume curves for the six phases of the Dzugutov potential[7] found in the course of this study is shown in Fig. 9. Phase $a$ is the most stable phase at all positive pressures.

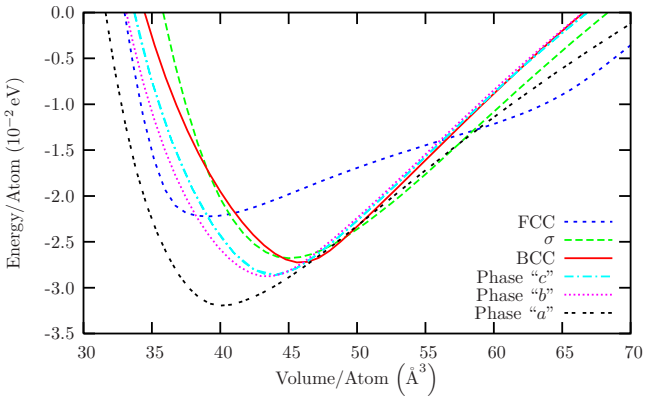


FIG. 9. (Color online) Energy-volume curve for the Dzugutov potential (Ref. 7) showing the three new phases calculated at zero pressure. The curves for the $\sigma$ phase and structures $a$, $b$, and $c$ were calculated assuming an isotropic expansion.

## IV. CONCLUSIONS

In this paper, we have developed a fitness function that combines a traditional approach to fitness based upon enthalpy, with a simple structural comparison factor to find new, more stable crystal structures within a GA for crystal structure prediction. This method penalizes the presence of similar structures within the population which prevents the GA stagnating in some local minimum. The GA method itself was also extended to allow both the simulation supercell and the number of atoms within that supercell to vary. The number of atoms must only be varied within fixed limits to prevent the system size becoming too large or too small.

Studies using the Lennard-Jones potential showed the calculation progressing through the fcc local minimum and two other stacking-fault local minima before finding the hcp global minimum-enthalpy structure. This was shown to be repeatable and efficient.

When this GA was used to study phases of the Dzugutov potential[7] at different pressures, all the previously reported zero-temperature phases were found, along with three new phases, one of which is the most stable phase at all positive pressures. These new structures are markedly different from the three previously known phases. This clearly illustrates the power of this GA to find new crystal structures that were *hitherto* unexpected.

## ACKNOWLEDGMENTS

Calculations were performed on our Departmental Beowulf Cluster, EPSRC-GB under Grant No. R47769 from the Multi-Project Research Equipment Initiative. N.L.A. is grateful to the EPSRC-GB for financial support. The authors acknowledge useful discussions with Peter Main.

[1] N. L. Abraham and M. I. J. Probert, Phys. Rev. B **73**, 224104 (2006).
[2] G. Trimarchi and A. Zunger, Phys. Rev. B **75**, 104113 (2007).
[3] C. W. Glass, A. R. Oganov, and N. Hansen, Comput. Phys. Commun. **175**, 713 (2006).
[4] R. M. Briggs and C. V. Ciobanu, Phys. Rev. B **75**, 195415 (2007).
[5] G. L. W. Hart, V. Blum, M. J. Walorski, and A. Zunger, Nat. Mater. **4**, 391 (2005).
[6] V. Blum, G. L. W. Hart, M. J. Walorski, and A. Zunger, Phys. Rev. B **72**, 165113 (2005).
[7] M. Dzugutov, Phys. Rev. Lett. **70**, 2924 (1993).
[8] P. Debye, Ann. Phys. **46**, 809 (1915).
[9] S. D. Stoddard and J. Ford, Phys. Rev. A **8**, 1504 (1973).
[10] J. E. Jones and A. E. Ingham, Proc. R. Soc. London, Ser. A **107**, 636 (1925).
[11] J. Roth and A. R. Denton, Phys. Rev. E **61**, 6845 (2000).
[12] G. Kane and M. Goeppert-Mayer, J. Chem. Phys. **8**, 642 (1940).
[13] N. L. Abraham and M. I. J. Probert, *New Dzugatov Structures* (December 2007), http://www.cmt.york.ac.uk/cmd/downloads.html/.