Predicting Structural Models for Silicon Clusters

CARLOS RENATO ZACHARIAS,1 MAURÍCIO RUV LEMES,2 ARNALDO DAL PINO JÚNIOR,2 DAVID SANTO ORCERO3
1Av. Ariberto Pereira da Cunha, 333-Pedregulho, 12516-410-Guaratinguetá, SP, Brasil, UNESP-Guaratinguetá, Brazil
2CTA/ITA-São José dos Campos, Brazil
3Universidad de Málaga, Spain

Received 26 November 2001; Accepted 27 August 2002

Abstract: This article introduces an efficient method to generate structural models for medium-sized silicon clusters. Geometrical information obtained from previous investigations of small clusters is initially sorted and then introduced into our predictor algorithm in order to generate structural models for large clusters. The method predicts geometries whose binding energies are close (95%) to the corresponding value for the ground-state with very low computational cost. These predictions can be used as a very good initial guess for any global optimization algorithm. As a test case, information from clusters up to 14 atoms was used to predict good models for silicon clusters up to 20 atoms. We believe that the new algorithm may enhance the performance of most optimization methods whenever some previous information is available.


Key words: classifier system; optimization; cluster; structural models; genetic algorithm

Introduction

Traditionally, relevant problems in science and engineering that defy simple direct solution have been formulated as a search for the global minimum of a suitably chosen cost function. The determination of the spatial conformation of molecules,1–4 interplanetary trajectories,5 docking processes,6 and the project of optical systems7 are just a few examples of problems that are usually transformed into optimization procedures. In the literature, there exist many different iterative methods for solving optimization problems of linear and nonlinear systems of equations.8 Genetic algorithms (GA)9 and simulated annealing (SA)10 are two of the most successful strategies in this field. Such algorithms need one or more initial candidate solutions to apply either a derivative-based or a stochastic strategy11 to obtain the desired solution. Their performances, that is, time required to find the solution, are strongly dependent on the initial guess and other characteristic parameters.12,13 Poor selection of starting point(s) often results in the solution process getting stuck in a local minimum, or taking a long time to achieve convergence, or even not converging at all. There are no obvious rules to determine these starting points, and they are certainly problem dependent. Normally, these guesses are randomly chosen or they are set according to investigator’s personal experience. It is very important to be able to select good starting points to reduce computational cost.

As the dimension and complexity of the problem increase, the need for good initial candidate solutions becomes a dramatic issue. Knowledge can be acquired from simpler problems but there is no guarantee that it can be transferred to more complex ones. For instance, problems related to shape transitions may be difficult to deal with. Consider, for example, the well-known hard case LJ38 in the Lennard-Jones (LJ) cluster system. Almost all LJn clusters have icosahedral forms, but LJ38 is nonicosahedral. One could expect that a good guess based upon global minima of smaller sizes would favor only icosahedral forms and could lead the optimization procedure necessarily to a local icosahedral minimum. To avoid this inconvenient situation, it is essential not to restrict the previous knowledge injected into a predictor procedure to global minimum. One must include information of local minima as well. Our procedure is designed to overcome this kind of problem.

This article presents an efficient approach to generate good-quality candidate solutions for medium-sized atomic clusters. The first step is to classify geometric information obtained from previous investigations of small clusters. Next, the method predicts structural models for larger ones, with very low computational cost. Finally, predicted models are used as initial candidate solutions for global optimization algorithms such as GA and SA. This

Correspondence to: C. R. Zacharias; e-mail: Zacha@feg.unesp.br

© 2003 Wiley Periodicals, Inc.
procedure increases the probability of finding the global minimum (ground-state) conformation, and it reduces the total computational time required to find the desired geometry as well. This concept can be extended to other optimization problems and fields of interest.

Here, the following reasons led us to apply this new method to investigating silicon clusters:

1. It is interesting to ask to what extent structural elements of crystalline solids are already apparent in cluster, because clusters are bridges between atoms and crystals.

2. The determination of the ground-state geometry of atomic clusters is an extremely difficult optimization problem. Realistic total-energy calculations are very demanding of computational resources. The number of local minima increases dramatically with the number of atoms. For instance, it has been estimated that a 150-atom LJ cluster has local minima. At least a similar pattern is expected for silicon clusters.

3. It is known that silicon clusters—Siₙ, 18 < n < 40—can coexist in two structural patterns (prolate and compact). Even though the transition mechanisms are not well understood, some structural models, particularly for Si₂₀, have been proposed.

4. Our previous experience with silicon clusters and optimization methods.

**Basic Ideas**

Unbiased search for the ground-state geometry of silicon clusters is a rather cumbersome task. The directional nature of covalent bonds leads to a number of local minima that may exceed those of LJ clusters. Electronic effects can hardly be disregarded, thus quantum-mechanical treatment of the total energy needs to be considered. Due to the importance of silicon and its potential applications in nano-electronics, the aforementioned difficulties have not kept researchers from studying silicon clusters. Some investigators have tried to use previous knowledge to reduce the number of candidate structures. Kaxiras and Jackson reasoned that most atoms in clusters are on its surface, then they generated models of Siₙ based on well-known terminations of silicon surfaces. On the other hand, Grossman and Mitás proposed structural models for prolate Siₙ clusters that can be described as a superposition of layers of atoms. In each layer, three atoms were arranged in a triangular shape. These layers were stacked along a symmetry axis, with atoms at the ends. Obviously, interpreting an arrangement of atoms as a superposition of atomic layers is an old concept. It has been used for decades to describe crystal lattices and polypolytypes. Recently, knowledge of medium-sized silicon clusters has improved a lot thanks to the combined efforts of several researchers. Several experiments have been developed to elucidate the dependence of the cluster shape on the number of atoms. Theoretical calculations have lead to a new picture with the ground-state geometry of larger clusters? It is possible, even if the amount of available information is very limited? It is also worth saying that this is a presentation of a new method and not an investigation into particular cluster structures or into the quality of the Tight-Binding approach.

Not long ago, Hartke tried to transfer structural knowledge from smaller to larger clusters. That author introduced a very appealing idea: growth strategy without subjecting the cluster to a global optimization. A cluster with n + 1 atoms was created by adding a new atom to a previously optimized cluster with n atoms. Several assembly routines were optimized. However, his scheme did not allow for a nonsequential type of growth. This presented two drawbacks:

1. To search for the geometry of the cluster with (n + 1) atoms, one must know the ground-state geometry of the cluster with N atoms. The larger cluster (n + 1 atoms) was submitted only to a local optimization procedure.

2. It would be very difficult to recognize shape transitions because the shape of the larger cluster would be totally different from its predecessor. This difficulty was already mentioned by that author.

Here, we shall use our previous knowledge of total energy and geometry of several local and global minima of small silicon clusters. Unlike other authors, we will feed information into an automatic predictor/classifier system. Similarly to other authors, we will represent a cluster as a superposition of layers, each one representing a planar polyatomic geometric element. A set of 13 planar elements inspired in the geometric features of small clusters Siₙ (n < 15) is used (Fig. 1). This choice proved to be effective, but it is certainly not unique.

Now, we can describe Siₙ clusters as stacked sequences of polyatomic layers. It is worth noticing that a cluster may be described by more than one distinct sequence of layers. This is due to different choices of the symmetry (stacking) axis. Sequences of layers that represent the same cluster will be referred to as degenerate sequences. Figure 2 shows the sequences related to two Siₙ geometries corresponding to local energy minima with respective degenerated sequences.

Our next step is to learn from previously available data if a certain sequence of planar elements is energetically favored or not. To this end, we will create coefficients that measure the affinity of pairs and triplets of consecutive layers. Combining these affinity coefficients, our algorithm is able to estimate an index closely related to the binding energy of any selected cluster. We have named this index expressivity [see eq. (1)]. Now, we are in a position to develop a procedure to generate structural models for larger clusters. Those stacking sequences with high values of expressivity will be submitted to total energy calculations, while those with poor expressivity values can be discarded. It is important to notice that our scheme can be associated with any method of total energy calculation. In this article, for practical reasons, we will be calculating binding energies per atom using the Tight-Binding Method described elsewhere.
Classifier and Predictor Systems

Classifier Rules

In order to create a classifier and predictor system (CPS), one generally has to follow three steps:

1. Collect relevant information. Here we have included geometries and binding energies of local and global minima of small silicon clusters. Inclusion of the local minima is essential to allow the code to deal with shape transitions.

2. Select and classify this information to learn the general rules.

3. Make predictions as accurate as possible. If the amount of information is insufficient, the comprehension about the theme can be poor, while too much information can lead to redundancy.

Generally, one does not know, a priori, if the available information forms a complete descriptive set. Selection and classification depend on previous knowledge and will determine the quality of predictions. The prediction consists of processing the available information in order to obtain new results. In practice, these steps are correlated, consisting of an iterative procedure, converging to the reproduction of known results, and after that, the prediction of new ones.

Unfortunately, one generally doesn’t know the best procedure to classify the information. Neural networks (NN),\(^\text{30,31}\) classifier systems (CS),\(^\text{8}\) or adaptive algorithms (AA)\(^\text{32}\) can be useful in these situations. We have developed a classifier system able to correlate geometric information with binding energy of a silicon cluster. The main reasons to develop a CPS were: it is low cost from the computational point of view, and it works well in spite of the reduced quantity of available geometric information related to silicon clusters.

The option for a polyatomic multilayered description for clusters permits us to associate an affinity coefficient for pairs \([c(i, j)]\) and triplets \([c(i, j, k)]\) of consecutive layers, that is, the rules of the CPS are defined as:

\[
\begin{align*}
\text{if} \ (\text{layer, and layer}) & \text{ then } c(i, j); \\
\text{if} \ (\text{layer, and layer, and layer}) & \text{ then } c(i, j, k)
\end{align*}
\]

From the coefficients \(c\)'s, we can define a parameter associated with each cluster description (defined in the next section), related to the cluster’s binding energy. Applying some properties to these parameters, a suitable CPS can be implemented for our proposes.

Classifier Formalism

We chose a finite basis set—\(B\)—with \(q\) geometric layer elements:

\[B = \{b_1, b_2, \ldots, b_q\}\]

Any cluster \(k\) can be represented by a sequence \(S^k\) of geometric elements from \(B\)

\[S^i = \{s_1^i, s_2^i, \ldots, s_m^i\} \text{ with } (s_i^i \in B; i = 1, m_i)\]

We will assume that the energy of cluster \(k\) may be approximately obtained as a sum of affinities over pairs and triplets of layers. We named this quantity expressivity \(P(S^k)\); defined as

\[
P(S^k) = \sum_{j=1}^{m_i-1} \frac{1}{m_i - 1} c(s_i^j, s_{j+1}^k) + \sum_{j=2}^{m_i-1} \frac{1}{m_i - 2} c(s_{j-1}^i, s_j^k, s_{j+1}^k)
\]

where,

\[
\text{if} \ (\text{layer, and layer}) \text{ then } c(i, j); \\
\text{if} \ (\text{layer, and layer, and layer}) \text{ then } c(i, j, k)
\]

Figure 1. Planar elements used to represent the geometry of 3-dimensional clusters. They range from a zero-atom layer, which corresponds to empty space and is used to represent the end of a cluster, up to a five-atom layer. Notice that some elements differ only by their spatial orientation.

Figure 2. Two different views of the Si\(_9\) tricapped trigonal prism, with respective stacked layer description. Note that this is the subunit\(^\text{19}\) that has been proposed as a building-block for larger silicon clusters.
m_k \rightarrow \text{number of layers required to describe the cluster } k \text{ according to the sequence } S_k;
\begin{align*}
c(s^k_{i,j}, s^k_{i,j+1}) & \rightarrow \text{affinity coefficients for a pair of consecutive layers;} \\
c(s^k_{i-1,j}, s^k_{j}, s^k_{j+1}) & \rightarrow \text{affinity coefficients for a triplet of consecutive layers.}
\end{align*}

To determine the affinity coefficients, c’s, we need a set of training data. Here, we have used previous knowledge of the binding energy per atom and geometry of local and global minima of small Si clusters (Si_{n}, 2 \leq n \leq 14). The ideal set of c’s is obtained when the following requirements are satisfied:

1. Degenerated sequences have exactly the same expressivity value [eq. (1)].
2. The expressivity of any cluster belonging to the training set will be numerically equivalent to its binding energy per atom.

Next, we define a cost-function that equals the square of the differences between expected and calculated values for expressivity. A least-square fit is performed to determine the values of the coefficients c’s.

Thus, the determination of affinity coefficients is also transformed into an optimization problem, where we search for c’s that minimize the cost-function. These coefficients are optimized using a hybrid version of a GA. Obviously, this is much easier than the optimization based on total-energy calculations.

**Predictor Algorithm**

The classifier algorithm just described is the first step to predict a set of clusters with n atoms that may be used as an initial population for a GA optimization. A predictor algorithm was conceived reasoning that the description of a good structural model for a silicon cluster with n atoms must have a high expressivity [eq. (1)]. No restrictions, such as symmetry, crystal fragments, or surface similarities, were enforced. Notice that expressivity is an artificial parameter and a high expressivity value does not assure that the geometry is physically relevant. Obviously, a realistic evaluation of a predicted cluster must be done using a total energy calculation. Next, we implemented a new GA code to search for high expressivity descriptions with n atoms. Then, these solutions were decoded from the stacked multilayered description to Cartesian coordinates. Finally, a local optimization procedure was performed and those preoptimized clusters constitute the initial population (starting points) for a regular optimization procedure.13

**Numerical Implementation**

We have previously studied13 small silicon clusters (Si_{n}, n \leq 14) whose binding energy was calculated within a Tight-Binding approach described elsewhere.29 For clusters, ranging from Si_{3} to Si_{14}, we had 21 conformations, corresponding either to local or global energy minima. Each one was described according to the polyatomic stacked layer concept. As the same cluster can be described by different sequences of layers, we had 51 different S^k descriptions with known binding energy per atom. For convenience, they were sorted in decreasing order of binding energy. Next, we minimized the cost function with a GA, that is, we adjusted the affinity coefficients in order to make expressivity values, P(S), proportional to those of energy. To perform the GA optimization, each individual was represented by a rectangular matrix whose elements were c(i, i + 1) and c(j - 1, j, j + 1).

An initial population with 100 individuals (matrices) with elements randomly set in the range [−1000, +1000] evolved until it reached convergence. With the optimized coefficients, a new GA-SA code searched for the top 10 descriptions (clusters), with n atoms in terms of expressivity. The possible presence of unrealistic models with high expressivity values contributes to maintain genetic diversity in the gene pool of GA. Thus, these structures do not deteriorate an algorithm’s performance. For each n, the GA-SA engine converges to a population with 10 polyatomic stacked layer descriptions. These descriptions were transformed back into Cartesian coordinates and submitted to a quick preoptimization calculation. This approach produces physically relevant structural models with very low computational cost. These models can be used as an initial population for a global optimization algorithm such as GA or SA.13

**Features and Results**

**Merits and Limitations**

Many factors contribute to make the prediction of the ground-state geometries of Si clusters a stringent test for our classifier-predictor system. First, the classifier function, relating geometric features to energy, was not (see the next section) uniquely defined. Second, the amount of input data, used to determine the affinity coefficients, was very small. The information on triplets proved to be very important because it solved conflicts among some sequences of pairs (see the next section).

The initial affinity coefficients were randomly generated. Figure 3 shows a typical initial distribution. After optimization, the correlation between real and classified energies (normalized expressivity) is shown in Figure 4.

These results indicate that the implemented classifier system was successful in classifying structural geometric relations obtained from small clusters. As we have previously discussed, our predictor procedure must be coupled to a global optimization method. Our choice was the GA due to its success in similar problems.13,19,33 Thus, we used our predictor algorithm to generate “good” initial populations of clusters Si_{n} for 13 \leq n \leq 20.

To determine how well the predictor performs, we calculated the average binding energy of the initial population for each size of the clusters, \( E_{n_{av}} \). We named relative energy, the ratio \( E_{n_{av}} \) over the energy of the ground-state for Si_{n}, according to the Tight-Binding Method. These values give us an indication of the predictor performance, that is, they show how good the initial population generated by the predictor algorithm really is. These numbers are presented in Table 1.
For sizes in the range from 13 to 20, Table 1 shows that predicted initial population reaches a relative energy around 90–95%. It seems to be a very good initial guess for the global optimization procedure. It is important to remark that the affinity coefficients were optimized using geometric information from silicon clusters up to 14 atoms. However, these coefficients could be used successfully to predict clusters up to 20 atoms. Thus, this suggests that the CPS was able to acquire useful geometric information from small clusters.

**Interpretations**

It is well known that a covalent system cannot be suitably described by a two-body potential. Angular terms must be included, in order to reproduce more realistic potentials.\(^34\)–\(^36\) If one expands the potential energy function in many-body terms, one would get:

$$U_{\text{total}} = \sum_{i,j} A_{i,j} U_{i,j}^{(2)} + \sum_{i,j,k} B_{i,j,k} U_{i,j,k}^{(3)} + \sum_{i,j,k,l} C_{i,j,k,l} U_{i,j,k,l}^{(4)} + \cdots$$

where the first summation represents the two-body interactions, the second, three-body terms, and so on;

$$A_{i,j}, B_{i,j,k}, C_{i,j,k,l} \rightarrow \text{interaction parameters;}$$

$$U_{i,j,k,l}^{(n)} \rightarrow n\text{-body potentials}$$

One may be tempted to associate a physical interpretation of the parameters (c’s) by comparing eqs. (1) and (2). Such reasoning would lead to the (incorrect) conclusion that the affinity coefficients represent the effective potential between a pair (or triplet) of polyatomic layers. This is not true, and the investigation of two degenerated description supports this argument. For example, the two distinct descriptions of tricapped trigonal prisms (TTP) shown in Figure 2 have exactly the same total expressivity, but the contribution of pairs and triplets is different in each case. However, their sums are the same as those required for degenerated descriptions. A simple rotation can make the n-body terms different. Consequently, the coefficients, c’s, have no obvious physical meaning. However, as the expressivity values were adjusted to reflect an energy scale, they can be seen as a crude approximation to the cluster energy.

**General Conclusions**

Researchers\(^37\)–\(^38\) have established that the solution to a hard optimization problem can become faster and more efficient if information acquired from earlier studies can be conveniently used. Such information can be used to find a good initial guess or even to guide the optimization process.

The search for the ground-state geometry of medium-sized Si clusters has frequently been restricted by crystal-inspired insights\(^17\),\(^18\) or by other external constraints based on investigator’s experience. Only recently has the combined effort of several groups led to a better understanding of silicon clusters with up to 23 atoms. There is evidence\(^19\),\(^39\)–\(^42\) that clusters with 12–18

---

**Table 1.** Percentage of the Binding Energy per Atom, Averaged over the Initial Population Selected by Our Method, for Silicon Clusters Ranging from 13 to 20 Atoms.

<table>
<thead>
<tr>
<th>Number of Si atoms</th>
<th>Relative energy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>91.5</td>
</tr>
<tr>
<td>14</td>
<td>97.2</td>
</tr>
<tr>
<td>15</td>
<td>94.9</td>
</tr>
<tr>
<td>16</td>
<td>89.6</td>
</tr>
<tr>
<td>17</td>
<td>95.3</td>
</tr>
<tr>
<td>18</td>
<td>95.5</td>
</tr>
<tr>
<td>19</td>
<td>92.3</td>
</tr>
<tr>
<td>20</td>
<td>93.1</td>
</tr>
</tbody>
</table>

---

**Figure 3.** Randomly generated initial matrix population histogram. Smoothed curve represents a Gaussian fit averaged over 20 runs.

**Figure 4.** Expected energy versus classified energy for 51 Si\(_n\) clusters for \(n < 15\). Correlation coefficient is 0.9999.
atoms are built on a structural motif consisting of a stack of Si$_9$ TTPs. For 19 $\leq n \leq 27$, mobility experiments$^{16,43}$ indicate the coexistence of prolate and more spherical shapes.

Figure 5 shows the geometry of the ground-state of the Si$_{20}$ cluster according to our new method, as well as a more compact local minima. It’s important to observe that even though the piling of layers tends to favor elongated structures, the method can also predict more compacted structures, with internal atoms, as has been discussed in refs. 13 and 44. The more spherical shaped structure can be obtained by performing the search avoiding the presence of the elements that were too similar in the population during the application of the GA. We have introduced a resolution in energy $\delta$, in such a way that the coexistence of more than two elements whose difference in energy was less than $\delta$ (typically 0.02 eV/atom) was not permitted. This procedure allowed us to find a more spherical structure, presented in Figure 5b, as well as the ground-state geometry shown in Figure 5a. The binding energy per atom for the more spherical structure differs from the binding energy per atom of ground-state by only 0.06 eV/atom.

Finally, we have presented a new automatic method to classify geometric information, related to small silicon clusters, in order to propose structural models for larger ones. The performance of the classifier algorithm depends on the geometric data adopted as a training set and it also depends on the affinity coefficients obtained. Fortunately, the training procedure is fast and easy. We have verified that the CPS is much easier and faster than those procedures based on brute force.

CPS, using information from clusters up to 14 atoms, predicted good initial models for clusters up to 20 atoms. A first important fact is that CPS really learns from a restricted amount of geometrical information. It classifies distinct stacks of planar geometric arrangement of atoms based on a geometric basis set $B$. This set proved to be effective, but it is certainly not unique. A second point is the possibility to predict a 20-atom cluster, using information up to 14-atom clusters. Our method not only saves a lot of computational effort, it learns the geometric features of small clusters, and uses them to find new, larger candidate structures to optimization. Another important fact is that even with a small amount of information, the correlation between expected and classified energy was excellent (Fig. 4). A final interesting fact is that new information, related to new clusters, can be incorporated into the algorithm, in order to improve its learning ability. It can be done starting from the former coefficient matrix, without restarting the optimization from a random starting point.

Because the expressivity must be seen as an artificial parameter, the prediction based on it must be carefully interpreted. Some predictions can be physically inadequate. However, one of the main conditions to improve the performance of a GA is diversity. Thus, these “unrealistic” models are not harmful to the method. Indeed, they help to avoid premature convergence of GA.

The prediction of a model with 95% relative energy does not assure that it is a good candidate. The existence of many local minima requires the use of a global optimization method, in order to find the global minimum. However, the quality of the initial solution is one of the most important conditions to find the global minimum in a reasonable amount of time. Generally, to reach a 95% relative energy model requires human intervention in order to prevent the calculation from getting stuck in a local minima. The algorithm here proposed eliminates this constraint, generating a population of preclassified and optimized models.

The main limiting factor of the classifier-predictor algorithm is the geometrical basis set $B$. Other geometric bases may be necessary to address the ground-state geometry of larger systems. In short, our algorithm can be expanded to treat different problems, starting from previous well established information, apparently uncorrelated. At least, it is worth mentioning that the same structure can be easily adapted to solve other optimization problems whenever some kind of previous knowledge is available.

References

7. Shafer, D. Comput Phys 1994, 8, 188.
Predicting Structural Models for Silicon Clusters