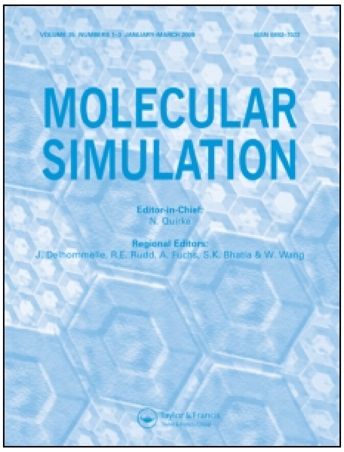


Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Applications of Genetic Algorithms in Cluster Optimisation

Steven Hobday[a]; Roger Smith[a]
[a] School of Mathematics and Physics, University of Loughborough, Loughborough, UK

To cite this Article Hobday, Steven and Smith, Roger(2000) 'Applications of Genetic Algorithms in Cluster Optimisation', Molecular Simulation, 25: 1, 93 — 120

To link to this Article: DOI: 10.1080/08927020008044115

URL: http://dx.doi.org/10.1080/08927020008044115

Molecular Simulation, 2000, Vol. 25, pp. 93–120
Reprints available directly from the publisher
Photocopying permitted by license only

APPLICATIONS OF GENETIC ALGORITHMS IN CLUSTER OPTIMISATION

STEVEN HOBDAY and ROGER SMITH*

*School of Mathematics and Physics, University of Loughborough,
Loughborough LE11 3TU, UK*

(Received May 1999; accepted June 1999)

Applications of Genetic Algorithms for optimisation of atomic clusters are reported. It is shown that the genetic algorithms are very useful tools for determining the minimum energy structures of clusters of atoms described by interatomic potential functions containing up to a few hundred atoms. The algorithm generally outperforms other optimisation methods for this task. A number of applications are given including covalent carbon and silicon clusters, close-packed structures such as argon and silver and the two-component C—H system.

Keywords: Genetic algorithm; atomic clusters

INTRODUCTION

This paper addresses the problem of determining the lowest energy configuration of a system of particles. This is a problem for which genetic algorithms (GA's) have been shown to be especially useful [1–3]. We also address this problem by the use of (GA's) and concentrate on applications involving many-body potentials as well as the extension to multi-component systems and comparison with other techniques such as Molecular Dynamics.

The determination of the minimum energy configuration of a system of particles from a potential energy function is clearly important since it determines the structural configuration of the cluster. Clusters are being increasingly used practically, for example in nanotechnology. One is often interested in simulating the interaction of clusters with surfaces and so before studying

---

*Corresponding author.

this interaction it is necessary to know the minimum energy structure for the cluster, as it approaches the surface otherwise it could be unstable. However the structure of clusters containing from tens up to several hundred atoms is also interesting in its own right and it can help throw light on the formation of the bulk material. In complex chemical systems the calculation of minimum energy structures might also point the way to how new molecules might be constructed. In this paper a method for calculating the minimum energy configuration of a system of particles is developed using a genetic algorithm (GA). No prior knowledge of the atomic geometry is assumed and the algorithm evolves the minimum energy arrangement by means of natural selection. The method is equally applicable to problems involving pairwise and many-body interaction potentials and can also be adapted for multi-component systems.

A GENETIC ALGORITHM FOR ATOMISTIC STRUCTURE CALCULATIONS

General Description

A GA is a global optimisation procedure that uses an analogy of the genetic evolution of biological organisms [1 – 3]. It is an heuristic search procedure that modifies function values encoded as binary strings, through the application of predefined reproduction operators in a stochastic manner. The binary string, referred to as a chromosome, is divided into individual sections called genes. Pseudocode for the genetic algorithm is shown in Figure 1.

```
initialize_population()                //Generate random initial population
                                       //of size npopsiz.
while{ not_converged } do

   assign_population_fitness()         //Assign Fitness to each individual.

   do from 1 to npopsiz
      select_parents(p1,p2)            //Select two parents, p1 and p2.
      reproduction(p1,p2,child)        //Reproduction produces a child.
   enddo

   select_next_generation()            //Choose from adults & off-spring

endwhile
```

FIGURE 1 Pseudocode for the genetic algorithm.

There is an absolute measure of quality also assigned to an individual called the fitness f that is a function of the genes. When designing a GA to search for the molecular structure with minimum energy an individual x from the population corresponds to a molecular configuration and a gene is the atomic co-ordinates of an atom, binary encoded. The chromosome of x is the string of genes defining the molecular structure and the fitness measure of x is the potential energy.

Following the initial random generation of a population, the reproductive phase of the GA starts by selecting two parents p_1 and p_2. This choice is weighted depending on the fitness of the parents. With two parents selected the binary digits which encode the string are changed in part by crossover and mutation (*i.e.*, genes are interchanged).

Crossover in its simplest form replaces some of the digits that encode parent p_1 by those of parent p_2. Mutation also changes some of the replaced digits with a small probability and ensures that the likelihood of exploring any subset of the search space is always non-zero. The next population is selected when there is an equal number of 'adult' and 'child' individuals.

Parent Selection

In a population P of N parents, $P = \{x_1, \ldots, x_N\}$, let the individuals be ordered in descending fitness so that $f(x_i) < f(x_{i+1})$ for $i = 1, \ldots, N-1$. Further, let p_i denote the probability that a parent x_i is selected as one of the two parents to which the reproductive operators are applied. The probabilities $\{p_1, \ldots, p_N\}$ such that $\sum^N p_i = 1$, denotes the set of parent selection probabilities which are assigned by a parent selection scheme. Our procedure is to use a binary-tournament selection, first introduced by De Jong [4]. This method takes a random choice of 2 individuals from the population P and selects the fitter as the parent. The selection probabilities p_i, $i = 1, \ldots, N$ are given by $p_i = (2(N-i) + 1)/N^2$, for $i = 1, \ldots, N$ where $\sum^N p_i = 1$. This scheme assigns a selection probability to an individual according to its relative fitness, so that the most fit individual has a probability $p_1 = (2(N-1) + 1)/N^2$, the second most fit $p_2 = (2(N-3) + 1)/N^2$ and so on down to $p_N = 1/N^2$ but without requiring that the individuals be sorted by fitness.

Reproduction

To apply the GA in the optimisation of single element molecular structures it is necessary to give a definite meaning to each of the terms defined in the general description. The population refers to the set of isolated n atom molecules. The fitness of an individual is the molecule's total potential energy and the genes define the molecular structure. It is also necessary to assign a potential energy per atom and the genetic encoding procedure should be such that only those atoms j which most influence the potential energy of atom i are located in proximity on the chromosome. The reproduction process continues in four separate stages (a) parameter encoding, (b) crossover, (c) mutation and (d) structural relaxation.

Parameter Encoding

The optimisation scheme replaces a high energy region, normally less than half of the molecule, (referred to as h) of parent P_1 with a low energy region (referred to as l) of parent P_2. In this selection process both h and l must contain the same number of atoms. Next a plane is chosen with randomly selected position and orientation but such that h region of parent P_1 is on one side of the plane and the centre of mass $\mathbf{c}$ of P_1 lies on the other side. The second molecule P_2 is now rotated so that it also is similarly orientated with l on one side of the plane and its centre of mass on the other. The genes corresponding to the h and l regions are then interchanged. The practical implementation of this step entails first choosing a location on a molecule biased towards a high or low energy region. The vector $(\mathbf{h}-\mathbf{c})/|\mathbf{h}-\mathbf{c}|$, where $\mathbf{h}$ is usually chosen as the atomic co-ordinates of the highest energy atom in the region h, forms the unit normal for the dividing plane. This dividing plane is then located between $\mathbf{c}$ and $\mathbf{l}$, where $\mathbf{l}$ is a vector located in region l, on the other molecule in such a way as to preserve the number of atoms. This number is not fixed but varies between 1 and $n/2$ for n-atom molecules. The dividing plane being located at $\mathbf{c}+\rho(\mathbf{l}-\mathbf{c})$ where ρ is a uniform random variate. Parent P_2 is rotated to align $(\mathbf{h}-\mathbf{c})$ with $(\mathbf{h}-\mathbf{l})$ and the region h is replaced by l by a simple translation.

For a two component system, the procedure is a little more complicated as the number of atoms of both species has to be preserved. Consider the first parent P_1 and let there be n_1 atoms of species 1 and n_2 atoms of species 2 in region h. The low energy region l of parent P_2 is chosen as before to contain n_1+n_2 atoms. However there is no guarantee that n_1 and n_2 are the same in both h and l. Thus the random choice of l in parent P_2 is repeated up to a maximum of $n_{\max}$ times. If a configuration does not arise with n_1 and n_2 the same in both h and l in these $n_{\max}$ attempts then a strategy is implemented to change the atoms types in l. It was found that when applying the method to C—H systems with $n_{\max} = 10$ an atomic arrangement could be found on 70% of occasions where n_1 and n_2 were the same in both h and l. In the 30% of occasions where this was not the case, the book-keeping strategy adopted was randomly to replace additional H atoms, preferentially those with the highest co-ordination, with a C atom but in the case of additional C atoms to replace with H atoms preferentially those of lowest co-ordination.

Single Point Crossover

Crossover assembles a new child geometry from the low energy parts of both regions. This process is shown in Figure 2. The crossover takes place at the point separating the genes of δ_1 and δ_2. The child chromosome takes the δ_1 genes of P_2 with the δ_2 genes of P_1. In terms of atom co-ordinates, the child molecule takes the geometry of its first parent except for a high energy region which is replaced with a low energy region from the other parent. The number of atoms replaced by this operation varies from 1 to n_2 depending on the value of ρ.

Mutation

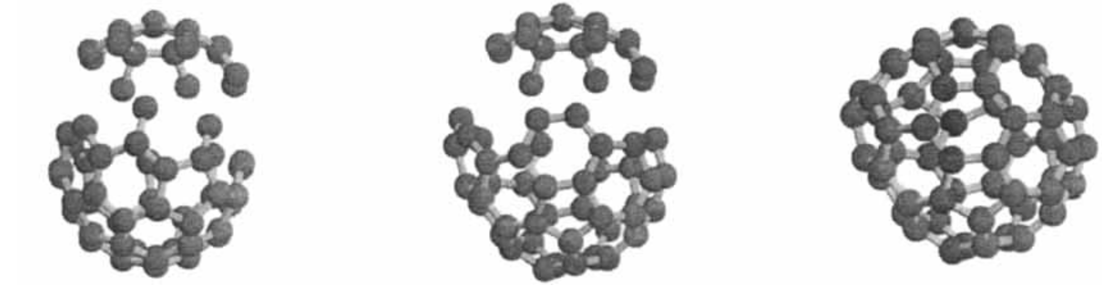
Mutation occurs by replacing typically one or two atoms of high energy (poor placement) whilst maintaining the total number of atoms in the molecule. The replaced atoms are not already in the high energy region and are removed completely from the low energy region of parent P_1. This means that the number of atoms added in the low energy region during the crossover must be increased proportionately. We choose an effective mutation operator that randomly replaces the genes that encode the placements of up to κ atoms of the child chromosome with those from the parent P_2. The procedure is as follows. On the new child chromosome, locate $i = 1$ or 2 or .... κ atoms with high energies $h_1, h_2, \ldots, h_\kappa$. Mutation occurs when $i > 0$ and then the genes corresponding to the atoms $1, \ldots, \kappa$ are deleted. New genes are added by effectively translating the second parent normal to the separation plane until κ more atoms lie within the δ_1 region. Typically $\kappa = 0$, 1 or 2.

For the two component C—H system, the strategy is a little different. Locate only $i = 1$ or 2 or .... λ, H atoms with poor placement. Locate a similar set of C atoms that have a low number of neighbours and reassign the λH atoms randomly to points in the neighbourhood of these C atoms.

Parent p1

Parent p1 following plane separation.

Parent p2



p2 following rotation and plane separation.

Replace high energy region with low.

Relaxed child Configuration

FIGURE 2 The GA implementation for molecules.

Child Structure Relaxation

The child structures are relaxed using a local optimisation routine BFGS developed by Liu and Nocedal [5]. This has the advantage that the memory requirements and iteration time scale linearly with the dimension of the search space and in addition any second derivative discontinuities in the potential energy surface do not present problems. A fixed tolerance is set and this algorithm is iterated until the sum of squared forces is less than this tolerance.

Population Replacement Strategy

The pseudocode of Figure 1 suggests that every generation creates N offspring from the current population of N adults. The question that arises is how to select the replacement generation from this set of $2N$ parent and child structures. The strategy adopted here is to select the subset of the N most fit individuals from the current set of adults and offspring. This strategy ensures that the increase in $f(x_1)$ is monotonic, where x_1 is the most fit individual in the population. To counteract premature convergence a restarting procedure was employed. If the chromosomes of the population become identical to within a small Hamming distance (typically 5%) in insufficient iterations to expect global convergence to have occurred, then the population is reinitialised but keeping the chromosome of the fittest individual.

APPLICATION OF THE GENETIC ALGORITHM

Carbon Clusters

Table I gives a list of the optimum carbon structures calculated using the genetic algorithm for the Brenner carbon potential [6, 7]. This is a modification of the short-ranged Tersoff [8] many-body potential which has been used with some success in large scale computer simulation problems involving carbon [9–11].

Figure 3 shows also the geometries of some of the small cluster structures including fullerenes. The transition from linear chains to rings occurs between 4 and 5 atoms and the transition from rings to cages between 17 and 18 atoms. In all cases the Brenner potential gave fullerene or fullerene-like cages as the minimum energy configurations for all molecules containing between 18 and 60 atoms. More details of these structures are given in [12].

TABLE I Energetics of C clusters calculated using the GA algorithm for clusters containing between 3 and 60 atoms. The geometry column is labelled l for linear chain, *m* for monocyclic ring, *f* for fullerene cage and *c* for fullerene-like cage

	Potential energy of lowest minima (Brenner)		
Number of atoms	*Potential energy of molecule (eV)*	*Energy per atom (eV)*	*Geometry*
3	− 12.40	− 4.1329	l
4	− 18.58	− 4.6439	l
5	− 26.33	− 5.2651	m
6	− 33.96	− 5.6600	m
7	− 41.10	− 5.8714	m
8	− 47.91	− 5.9883	m
9	− 54.50	− 6.0551	m
10	− 60.95	− 6.0950	m
11	− 67.32	− 6.1199	m
12	− 73.61	− 6.1342	m
13	− 79.90	− 6.1416	m
14	− 86.11	− 6.1507	m
15	− 92.34	− 6.1560	m
16	− 98.61	− 6.1631	m
17	− 104.80	− 6.1647	m
18	− 112.29	− 6.2386	c
19	− 118.83	− 6.2543	c
20	− 128.39	− 6.4195	f
21	− 134.52	− 6.4058	c
22	− 142.09	− 6.4586	c
23	− 148.41	− 6.4528	c
24	− 157.16	− 6.5483	f
25	− 163.13	− 6.5252	c
26	− 171.98	− 6.6146	f
27	− 178.01	− 6.5930	c
28	− 186.88	− 6.6743	f
29	− 192.67	− 6.6438	c
30	− 200.44	− 6.6813	f
31	− 207.00	− 6.6774	c
32	− 216.66	− 6.7706	f
33	− 221.78	− 6.7206	f
34	− 230.30	− 6.7735	f
35	− 236.38	− 6.7537	f
36	− 245.45	− 6.8181	f
37	− 251.23	− 6.7900	f
38	− 259.97	− 6.8413	f
39	− 265.69	− 6.8126	f
40	− 274.64	− 6.8660	f
41	− 280.38	− 6.8385	f
42	− 289.34	− 6.8890	f
43	− 295.12	− 6.8633	f
44	− 304.09	− 6.9111	f
45	− 309.02	− 6.8671	f
46	− 318.54	− 6.9248	f
47	− 324.06	− 6.8949	f
48	− 333.38	− 6.9454	f
49	− 339.03	− 6.9190	f
50	− 348.37	− 6.9674	f
51	− 353.88	− 6.9388	f
52	− 362.83	− 6.9742	f

TABLE I (Continued)

	Potential energy of lowest minima (Brenner)		
Number of atoms	*Potential energy of molecule (eV)*	*Energy per atom (eV)*	*Geometry*
53	−367.92	−6.9419	f
54	−377.69	−6.9943	f
55	−383.08	−6.9789	f
56	−392.37	−7.0066	f
57	−397.73	−6.9777	f
58	−407.25	−7.0216	f
59	−412.74	−6.9956	f
60	−422.55	−7.0425	f

An interesting point is that the evolution of clusters containing 18−22 atoms, unlike other sizes of cluster, is critically dependent on the dimensions of the region in which the atoms are confined. If the region is too large as in Figure 4(a) then a different structure is obtained than if the region is smaller as in Figure 4(b). This indicates that there must be some large energy barriers to be overcome at these nuclearities and that if the Brenner potential can be regarded as a true description of reality then the minimum energy configurations for such cluster sizes might only be formed under high pressure.

Hydrocarbon Clusters

The modified GA for two component systems is applied to determine low energy hydrocarbon structures. The Brenner Hydrocarbon potential has been fitted to a number of small alkane, alkene, alkyne, radical and aromatic molecules and additionally has been shown to reproduce the correct geometry and binding energies of larger hydrocarbon isomers. In this section we test whether the potential identifies the ground state configurations of C_nH_m for specific values of m and n. The optimised structures are compared with molecular data available from the pdb databank at Okanagan University [12]. The GA identifies ground-state isomers with energies in every case that are less than or equal to those of the known isomers from the pdb database. The results for a number of alkanes, alkenes, alkynes and aromatic molecules are shown in Figures 5−8. Figure 5 shows the application of the algorithm to the Brenner hydrocarbon potential for the alkane hydrocarbon isomers containing relatively small numbers of atoms, Figure 6 gives the alkenes, Figure 7 the alkanes and Figure 8 the aromatics.

Data for clusters containing a larger number n of atoms is given in Table II. When n is greater than 5, the GA determined global geometries that seldom

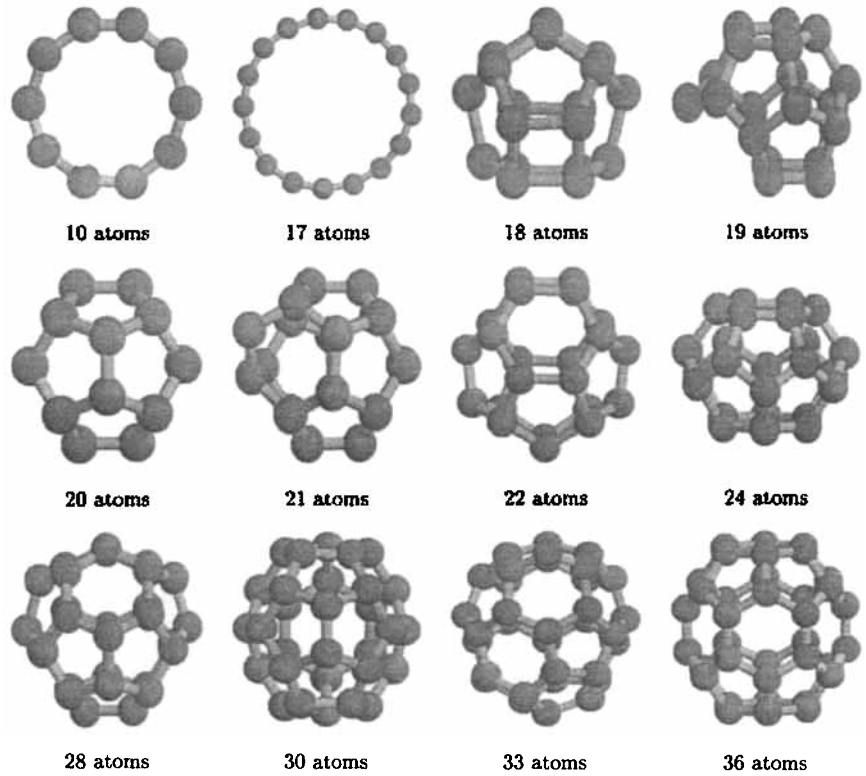
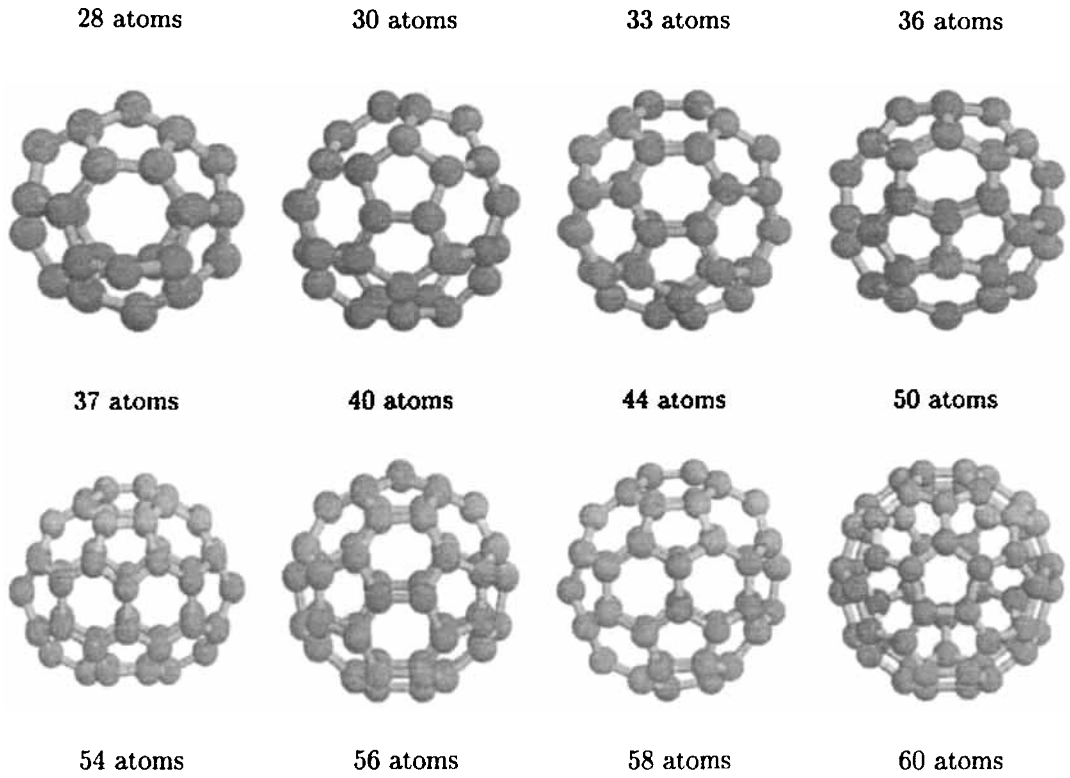


FIGURE 3 Small carbon molecular structures containing between 10 and 60 atoms calculated using the GA and the Brenner potential.



28 atoms
30 atoms
33 atoms
36 atoms
37 atoms
40 atoms
44 atoms
50 atoms
54 atoms
56 atoms
58 atoms
60 atoms

FIGURE 3 (Continued).

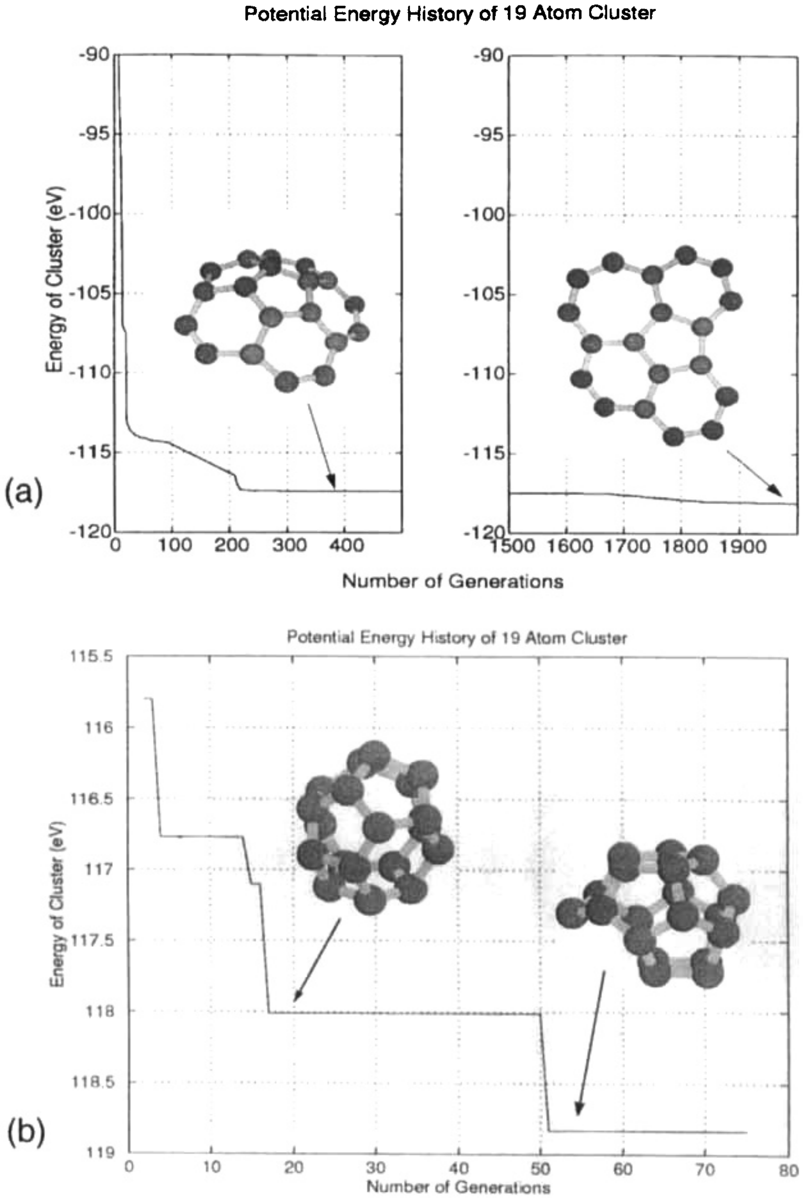


FIGURE 4 Evolution of the C_{19} cluster using the genetic algorithm. The initial population size was 30. In (a) the size of the box was a 1 nm cube. In (b) the size of the box was a 0.6 nm cube. In (a) after 210 generations, the configuration gas evolved to a cap polycyclic structure with an energy of − 117.45 eV. It takes a further 1600 generations to find a planar structure with energy − 118.11 eV. In (b) the minimum energy configuration is that of a complex cage with energy − 118.83 eV, located in only 50 generations.

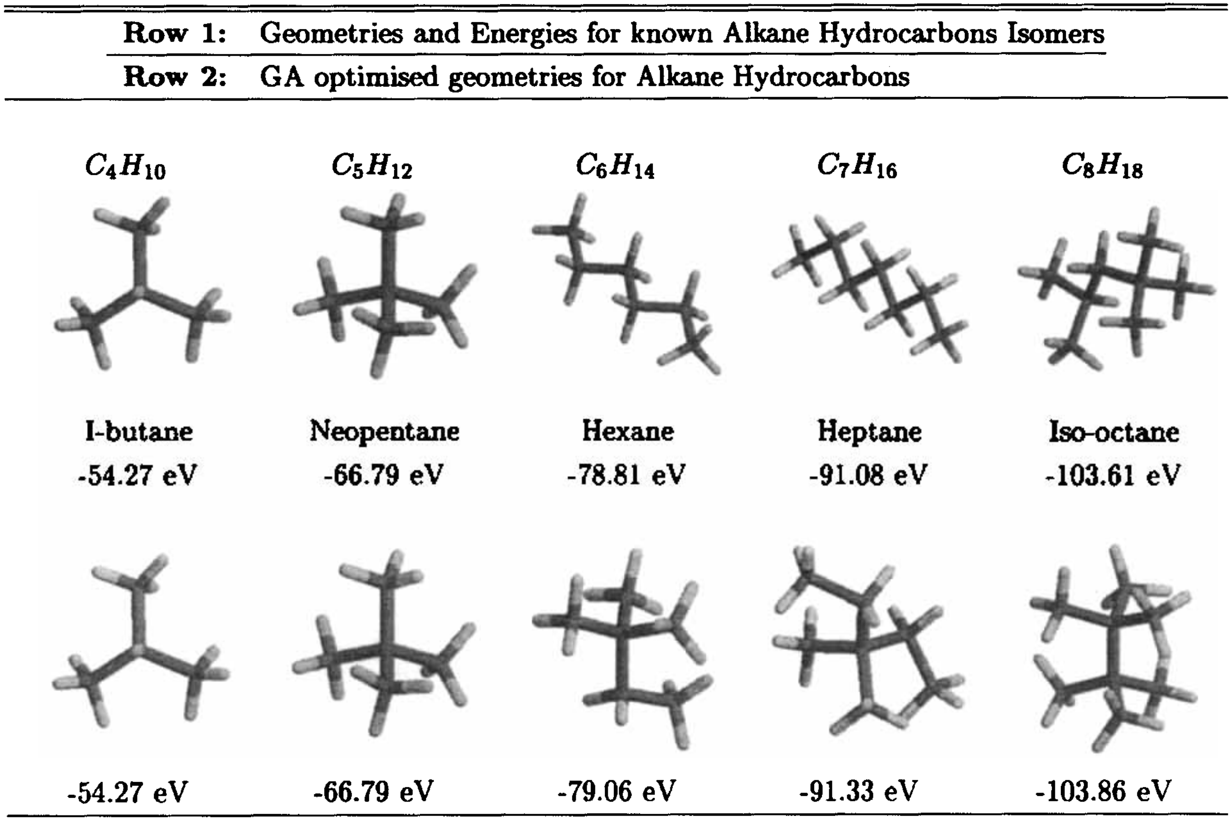


FIGURE 5 Alkane hydrocarbon structures calculated using the GA and the Brenner hydrocarbon potential.

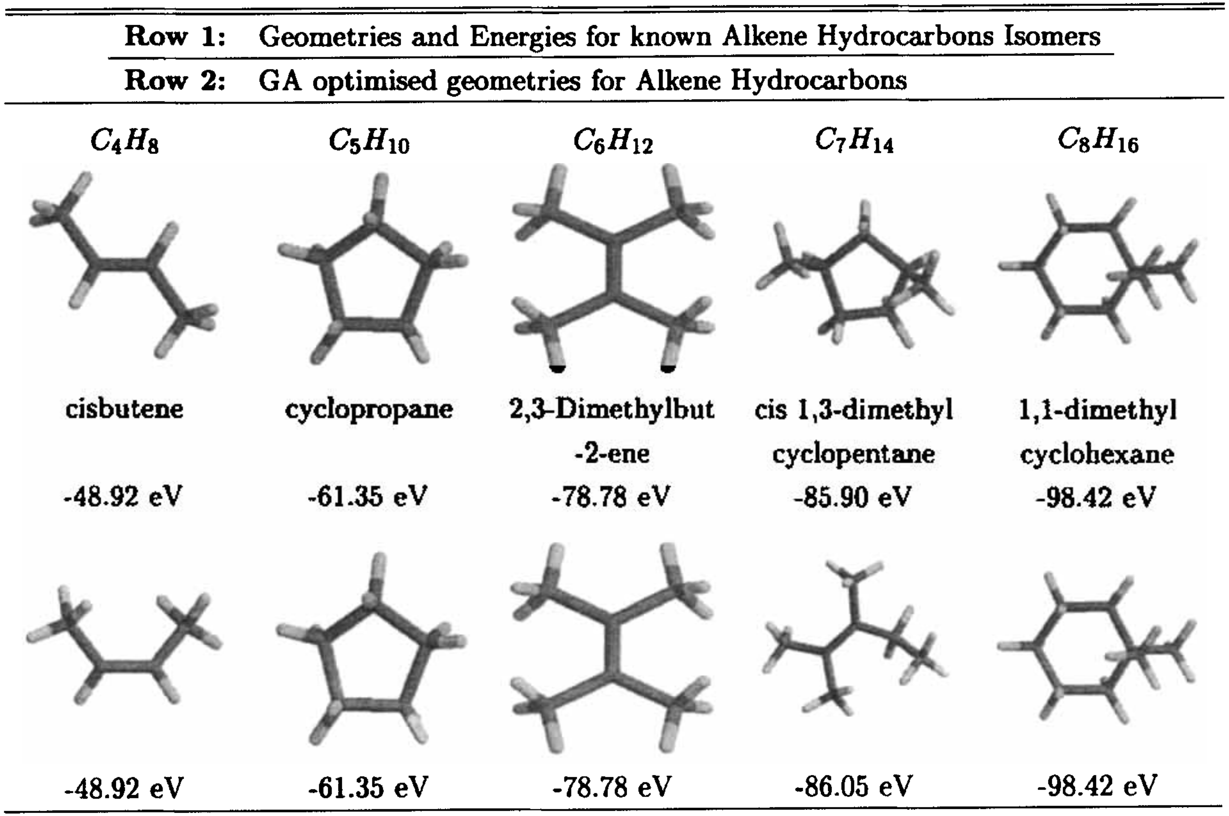
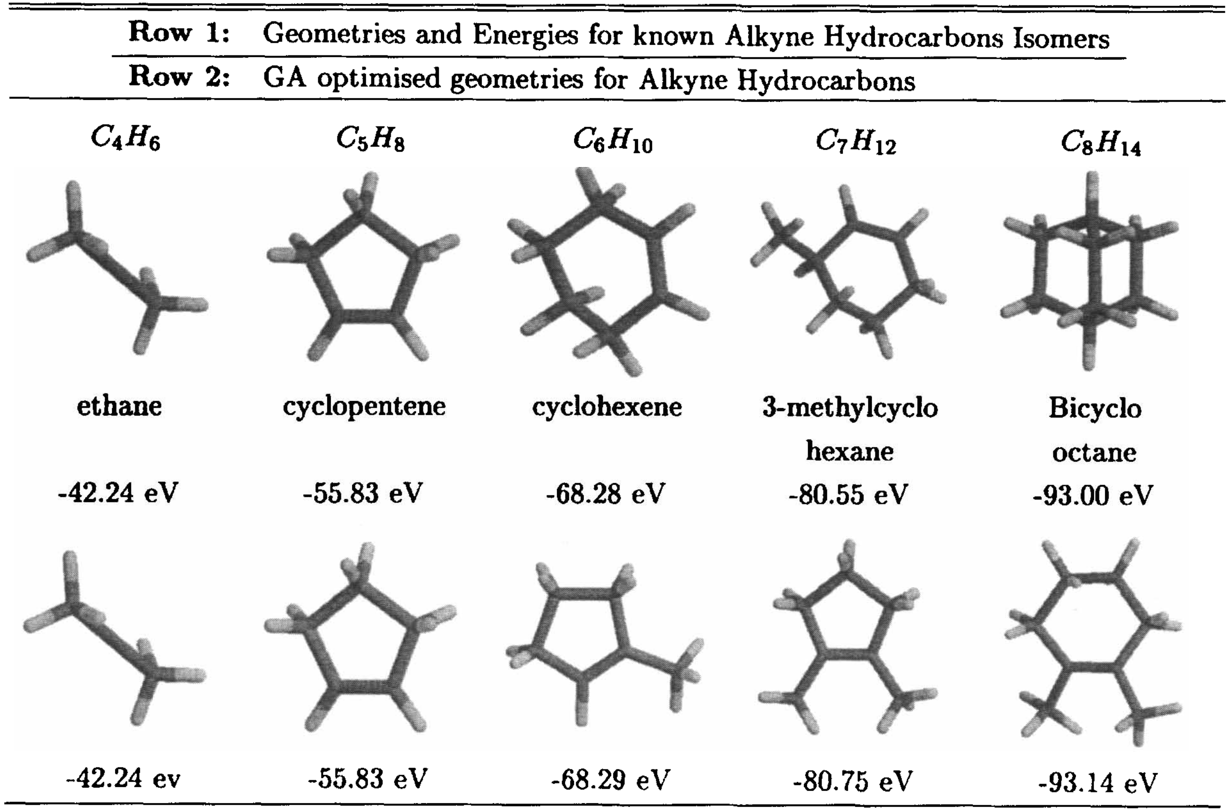


FIGURE 6 Alkene hydrocarbon structures calculated using the GA and the Brenner hydrocarbon potential.



Row 1: Geometries and Energies for known Alkyne Hydrocarbons Isomers

Row 2: GA optimised geometries for Alkyne Hydrocarbons

C_4H_6	C_5H_8	C_6H_{10}	C_7H_{12}	C_8H_{14}
ethane	**cyclopentene**	**cyclohexene**	**3-methylcyclo hexane**	**Bicyclo octane**
-42.24 eV	-55.83 eV	-68.28 eV	-80.55 eV	-93.00 eV
-42.24 ev	-55.83 eV	-68.29 eV	-80.75 eV	-93.14 eV

FIGURE 7 Alkyne hydrocarbon structures calculated using the GA and the Brenner hydrocarbon potential.

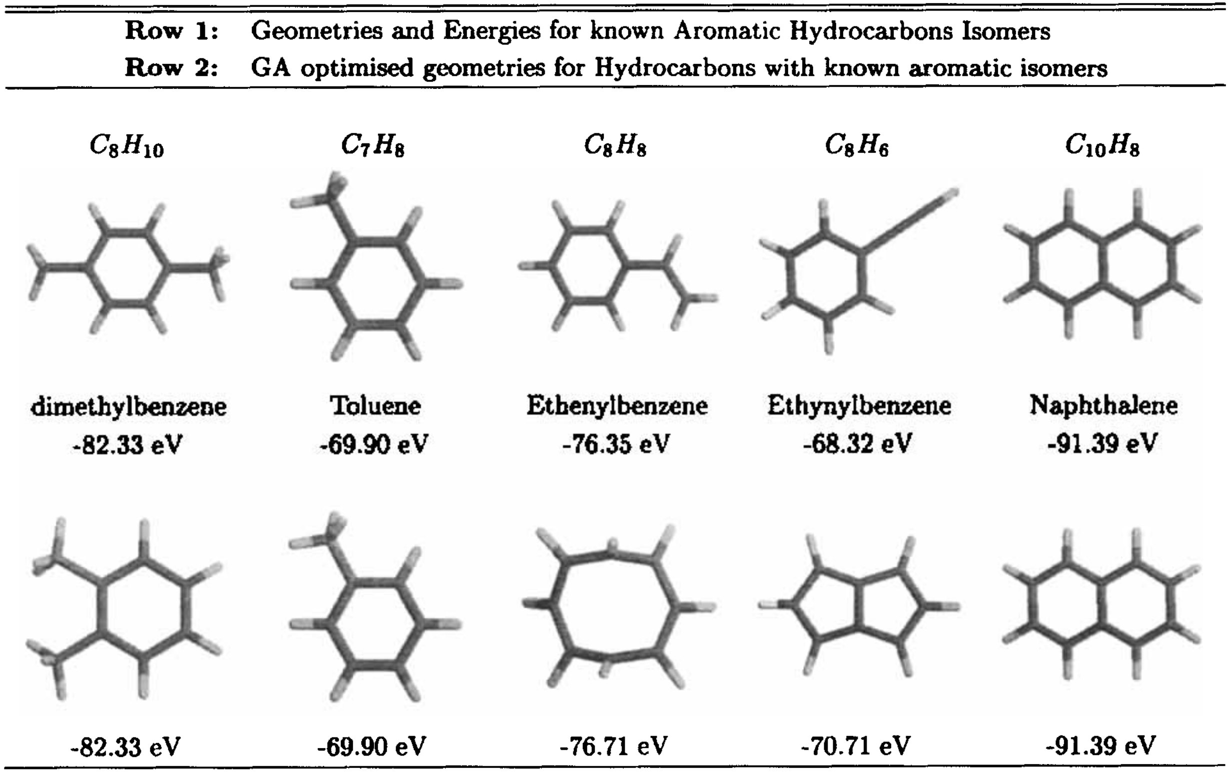


FIGURE 8 Aromatic hydrocarbon structures calculated using the GA and the Brenner hydrocarbon potential.

TABLE II This table compares the energy of database molecules with those evolved by the genetic algorithm. All configurations are locally optimised with the LBFGS routine according to the Brenner hydrocarbon interatomic potential energy function. The Database ID column provides the filename for the pdb format data base molecule. Note that although the high level of significant figures in the energy table is physically meaningless, it is necessary as two molecules with completely different topologies may have energies that are similar to within as little as 0.005 eV

Lowest energy hydrocarbon chart

		No. of		*Energy, eV*		
Series	*Hydrocarbon molecule*	*entries in database*	*Database pdb ID*	*Database minimum*	*GA minimum*	*Identical geometries*
	C9H20	3	nonane	-115.620	-116.127	*x*
	C10H22	3	decane	-127.891	-128.144	*x*
A	C11H24	3	34e-3m-6	-140.419	-140.923	*x*
L	C12H26	1	c-7ane	-152.448	-152.935	*x*
K	C13H28	2	b-8ane	-164.719	-165.462	*x*
A	C14H30	2	224477m8	-177.736	-177.742	*x*
N	C15H32	3	b-7ane	-190.013	-190.254	*x*
E	C16H34	3	c-9ane	-201.532	-202.446	*x*
	C17H36	1	d-10ane	-214.048	-214.802	*x*
	C18H38	1	c-10ane	-226.073	-227.164	*x*
A	C9H18	6	f-cy5ane	-110.689	-110.689	Yes
L	C10H20	5	bucyclo6	-122.968	-123.219	*x*
K	C11H22	3	g-cy6ane	-135.241	-135.482	*x*
E	C12H24	1	al0557	-147.406	-147.760	*x*
N	C13H26	3	b-6ane	-159.530	-159.935	*x*
E	C24H48	1	al3063	-294.422	-294.633	*x*
	C9H16	14	spiro-44	-105.504	-105.544	*x*
A	C10H18	6	s45ane	-117.779	-117.780	*x*
L	C11H20	5	s55ane	-130.055	-130.297	*x*
K	C12H22	4	a-s45ane	-142.328	-142.761	*x*
Y	C13H24	3	b-s45ane	-154.598	-154.840	*x*
N	C14H26	1	al0066	-167.130	-167.152	*x*
E	C16H30	1	a-s56ane	-191.417	-192.094	*x*
	C17H32	2	al0524	-202.017	-203.821	*x*

coincide with any contained in the pdb database. The data shows that the stereoisomers are equally energetic with the molecules in the database see for example, C_4H_8.

For the aromatic isomers, the GA finds the minimum energy geometries for benzene, toluene, dimethylbenzene and napthalene, although for the case of C_8H_{10}, a configuration with adjacent as opposed to opposite methyl units has also the same energy. However C_8H_6 and C_8H_8 have different molecular structures, with fused pentagonal rings and an eight-membered ring exhibiting preferential stability to ethenylbenzene and ethynylbenzene.

Although the GA finds that the Brenner potential accurately predicts the small alkene hydrocarbon structures, different molecular structures for the alkanes are predicted when the number of carbon atoms exceed 6. This trend is continued for the larger molecules. Whereas this might highlight the inadequacies of the Brenner potential it raises the intriguing possibility that new molecules could be discovered computationally by the application of the GA with accurate interatomic potential energy functions for multi-component systems.

Silicon Clusters

The GA can also be successfully applied to the Tersoff silicon potential [14, 15]. There are two parameterisations of this potential referred to as Ter2 and Ter3. Figure 9 shows the geometry of small Si clusters calculated using Ter2. In this case cages are predicted for upwards of 4 atom clusters. The 6 and 7 atom structures predicted are in agreement with recent

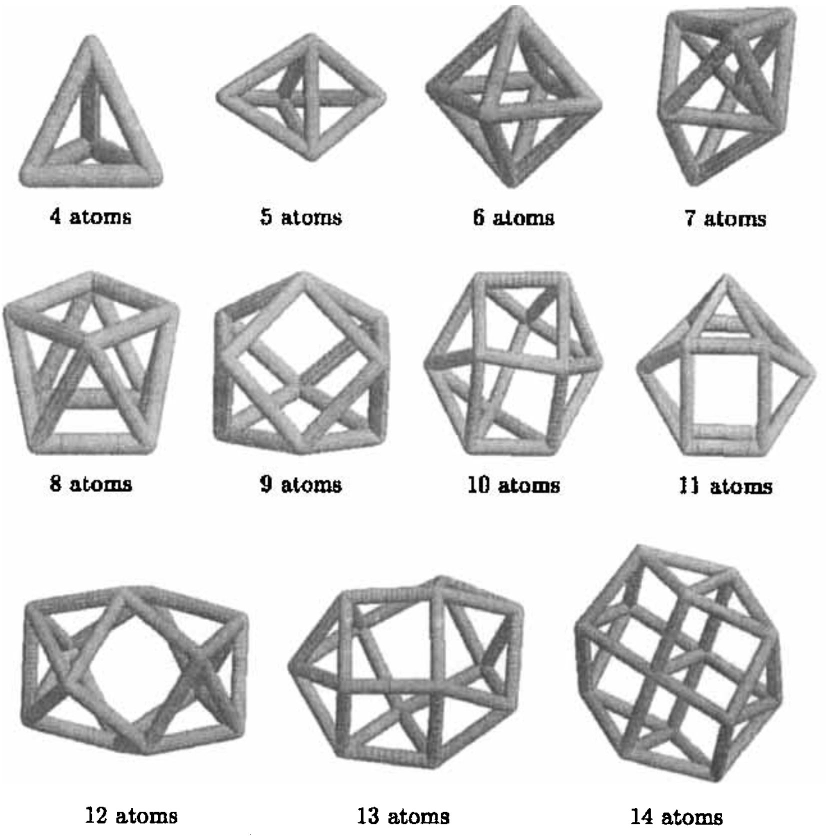

4 atoms 5 atoms 6 atoms 7 atoms
8 atoms 9 atoms 10 atoms 11 atoms
12 atoms 13 atoms 14 atoms

FIGURE 9 Small Si clusters calculated using Ter2.

experimental results [16, 17]. An interesting observation concerning the evolution of the Si_{18} atom cluster is that two isolated Si_8 and Si_{10} atom structures are predicted to have a lower energy with Ter2.

The evolution of this cluster shown in Figure 10. Larger clusters containing between 50 and 58 atoms are shown in Figure 11, calculated using Ter3. Here there are internal sp^3 bonds unlike the carbon potential. The 60 atom Si structure is about 3.3 eV more energetically favourable than the corresponding fullerene cage. These structures are very similar to recent calculations using *ab-initio* potentials [18].

Close-packed Clusters

The genetic algorithm can equally well deal with atoms described by either pair potentials or embedded-atom/Finnis-Sinclair [19, 20] potentials. Despite the simplicity of the pair potential, obtaining the ground state configuration for systems of atoms is a difficult problem. A study of the

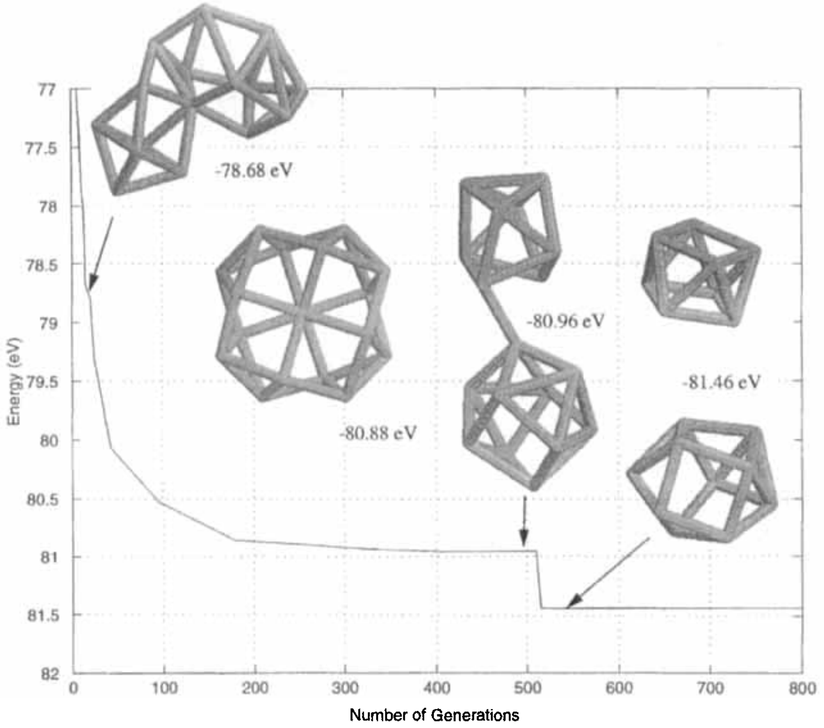


FIGURE 10 Evolution stages of 18 Si atoms. After 500 generations the cluster dissociates into Si_8 and Si_{10} geometries. Two Si_9, clusters have a higher energy at -81.316 eV.

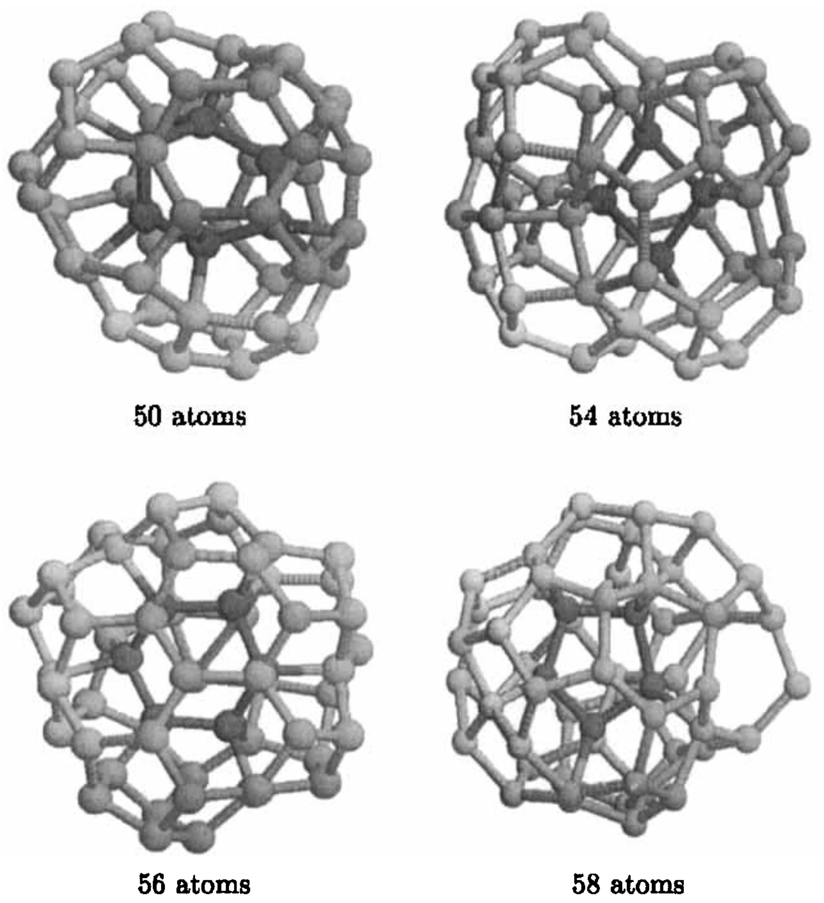


FIGURE 11 Four examples of Si cage formations with an internally bonded four-fold co-ordinated core calculated using Ter3. Within the core it is possible to see atoms arranged in the approximate forms of a square Si_{54} pentagon Si_{58} or hexagon Si_{50} and Si_{56}. For the cages containing between 48 and 60 atoms, two-thirds of the atoms are three-fold co-ordinated with the majority of the rest having four-fold co-ordination. For 60 Si atoms the cage with the internal core is 3.3 eV more stable than the fullerene.

number of local minima on a surface of the Lennard Jones potential found a lower bound of 988 local minima for a cluster with only 13 atoms. It is not the intention here to categorise all the geometrical structures predicted by the GA with the Lennard-Jones potential, see for example Wales and Doles [21], however it is worth noting the trends.

The use of the GA with pair potentials together with smoothing algorithms has also been found to be beneficial in terms of speed [22] but the calculations reported here do not include smoothing. For most clusters in the region $14 < n < 30$, a face-centred stacking arrangement occurs. The 19 atom cluster take a double icosahedral barrel geometry and larger clusters up to 30 atoms are mostly built around this core. The icosahedral lattice is optimal for most nuclearities above $n = 30$ until the completion of the outer shell at $n = 55$. Figure 12 shows some especially symmetric close-packed structures of 7, 38, 75 (sub-optimal) and 147 atoms as calculated with the

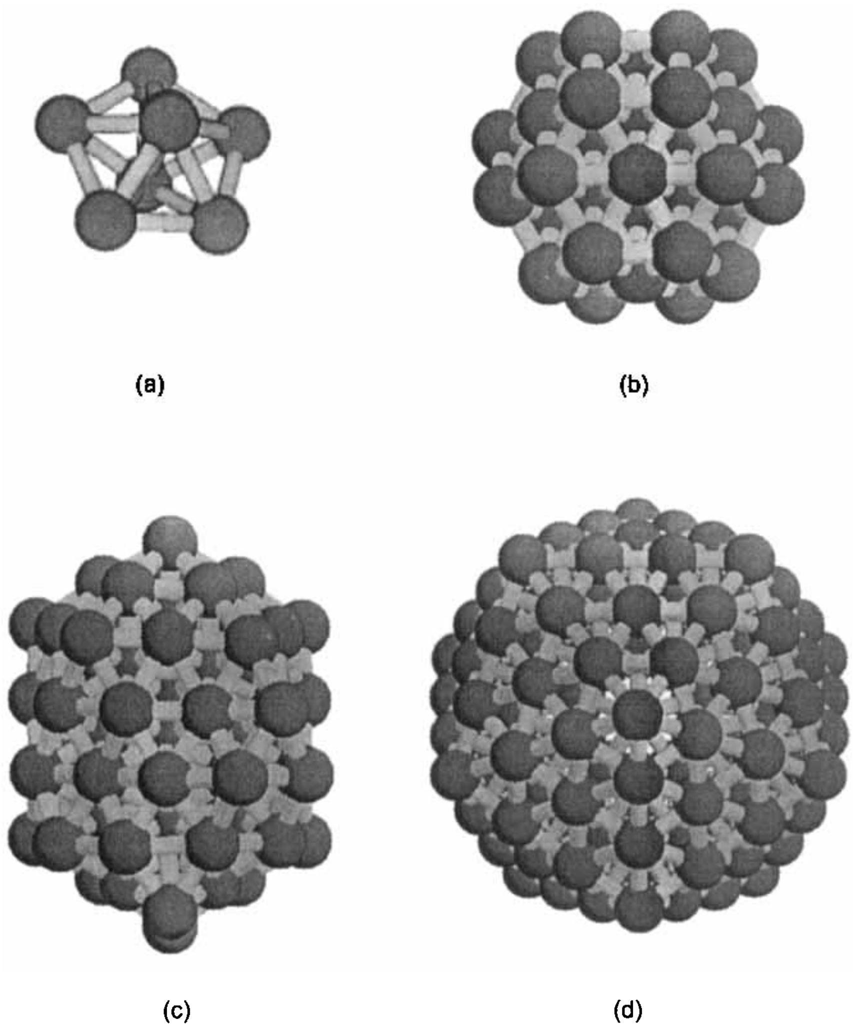


FIGURE 12 Clusters containing (a) 7, (b) 38, (c) 75 and (d) 147 atoms calculated from a Lennard Jones potential. The 38 atom cluster is an fcc truncated octahedron, the 75 atom and 147 atom clusters have icosahedral geometry.

GA using a standard Lennard-Jones 6 – 12 potential. The GA located the 38 atom cluster in less than 20 generations, the 75 atom cluster in 70 generations, 147 atoms in 180 generations. The 75 atom cluster is believed to be the lowest energy icosahedral geometry, however a lower energy configuration with dodecahedral geometry has also been found.

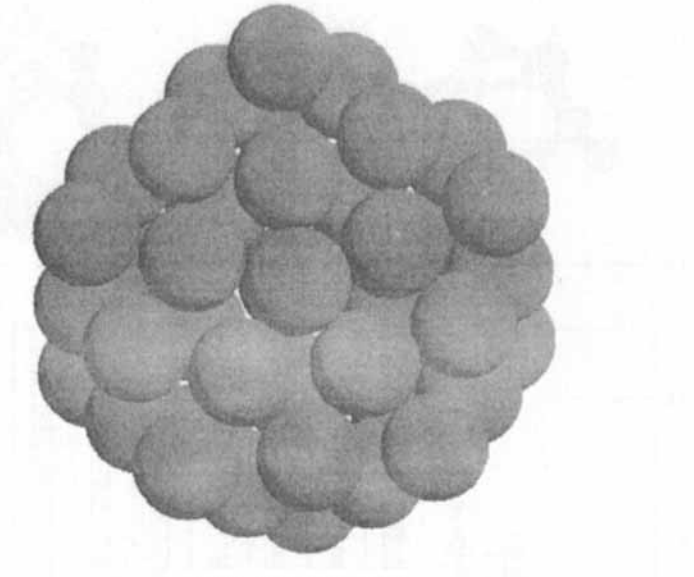
Close-packed clusters are now beginning to be routinely used in nanotechnology and thin film deposition and so a knowledge of the structural configuration of the clusters before deposition is important, not only practically but also in carrying out MD simulations. We have carried out MD simulations of the interaction of Ag clusters with graphite using an embedded-atom potential [19] for Ag in which the structure of the clusters was determined before impact using the GA [23]. Figure 13 shows the optimised Ag cluster structures containing 50 and 400 atoms. Note that the 400 atom cluster is beginning to take the form of the bulk fcc lattice.

Comparison with other Methods

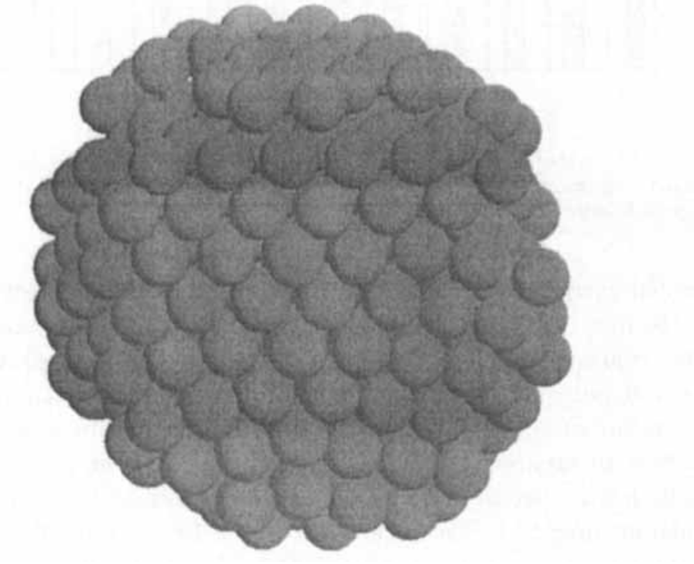
The minimum structure of small carbon clusters is an interesting test for the algorithm because this problem has been addressed by other authors *e.g.* [24, 25]. The potential used to describe the atomic interactions has two and three body terms and follows the so-called Murrell – Mottram model [26] Eggen *et al.* [25] optimised clusters with up to 20 atoms using steepest descent and conjugate gradient local optimisation routines in a controlled random search whereby the control also involved some physical intuition as to what candidate structures should be. For cluster containing between 10 and 22 atoms, the genetic algorithm found lower ground state structures for 11 – 18 atoms and also the 22 atoms cluster with no prior assumptions about geometry. The comparative results are shown in Figure 14.

The simulations involving carbon atoms described above involved the Brenner potential [6, 7]. Two separate methods are used to compare with the genetic algorithm, a quadratic approximation controlled random search method (CRS) [22, 25] and molecular dynamics (MD). On tests of C clusters containing up to 20 atoms, the CRS algorithm found the same minimum energy structures for carbon clusters containing up to 17 atoms as with the GA and in comparative CPU time, but failed to find minimum energy structures for C_{18}, C_{19} and C_{20}. The computing time required even to find the local minima at these nuclearities rapidly increases from 25 cpu secs. for 10 atoms to 3×10^4 for 20 atoms on a SUN Ultrasparc 1.

For the MD calculations the idea is as follows. The initial state of the system is determined by a stochastic process that assigns kinetic and



50 atom Ag cluster



400 atom Ag cluster

FIGURE 13 Partially optimised 50 and 400 atom Ag clusters.

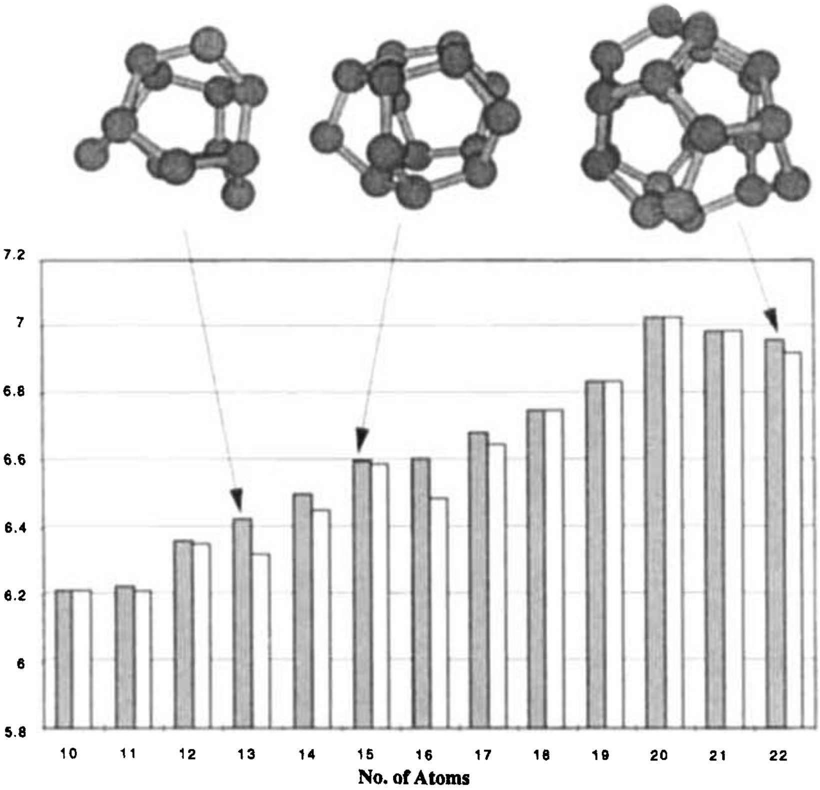
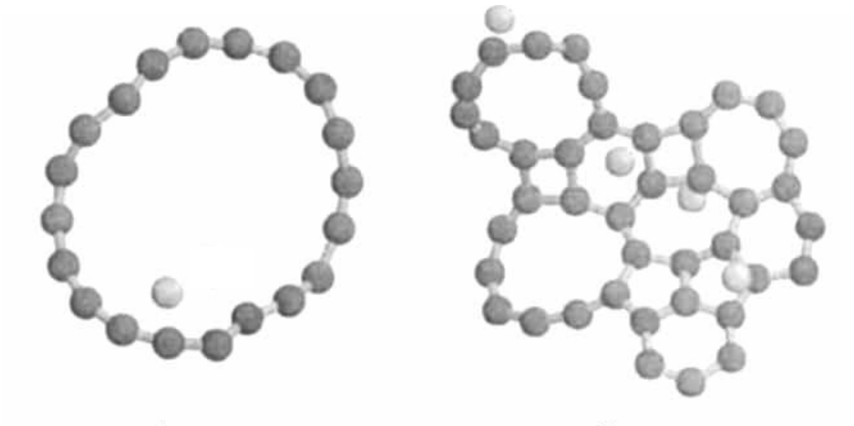


FIGURE 14 A comparison between the energetics of small ground state clusters for the Eggen potential, as predicted by the GA (left dark columns) and those calculated in the original paper of Eggen, Johnson, Li and Murrell (right).

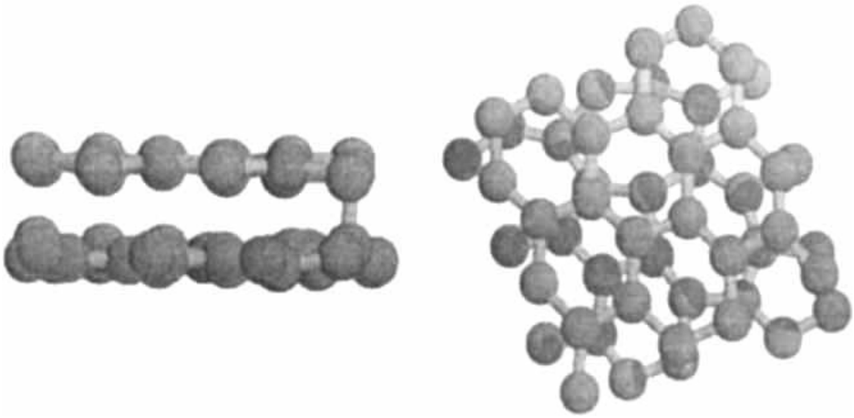
potential energy to *n* carbon atoms and typically half as many He atoms. The He atoms are connected to an external heat bath held typically between 2000–3500 K and interact with the C atoms with a repulsive screened Coulomb potential [27]. The combined initial energy is equivalent to a temperature of about 30,000 K. The simulation is run for several thousand timesteps to randomise the position and velocity components before the heat bath is connected and then the entire system is slowly annealed over a simulation time of 1–4 ns, equivalent to $5 \times 10^6 - 2 \times 10^7$ time steps. The heat bath model used was that due to Berendsen [28] with a time constant τ which varied between 4 and 10 ps.

On tests involving between 10 and 60 C atoms, the MD method found the correct ground state structure only for 10 atoms. On many occasions the algorithm correctly predicted fullerene-like cages for nuclearities of 60 and

120 atoms, albeit not minimum energy configurations. However this was very much dependent on the initial and boundary conditions and region size. Fullerene-like cage structures 40, 50 atom clusters were predicted but only sporadically. CPU time varied from $1 \times 10^4 - 4 \times 10^4$ s for these calculations. Typical configurations found by the algorithm for 20, 40 and 50 atoms are shown in Figure 15. The formation of the C_{60} fullerene-like cage



20 atoms **40 atoms**



50 atoms

FIGURE 15 Minimum energy structures calculated using molecular dynamics and annealing. This algorithm always failed to find the minimum energy 20 atom dodecahedron, annealing instead to a monocyclic ring. For 40 atoms a polycyclic structure, rather than fullerene cage is formed. Only for 60 atoms is the fullerene cage structure formed. For 50 atoms confined to a cubic region of side 0.8 nm, a two-layer graphite sheet is formed.

structure is shown in Figure 16. The 50 atom structure was obtained using periodic boundary conditions showing a two-layered graphite lattice structure.

Fullerenes are in practice most effectively formed by He or Ar collisions C atoms emitted from a graphite electrode in a glow discharge. Thus there are some similarities between the experiment and the MD simulation. The major dissimilarity is the density of atoms which is much less in the practical case and therefore the formation time would be much longer. Nevertheless it could be argued that the fullerene formation process shown in Figure 16 whereby rings and chains wrap around to form three dimensional structures

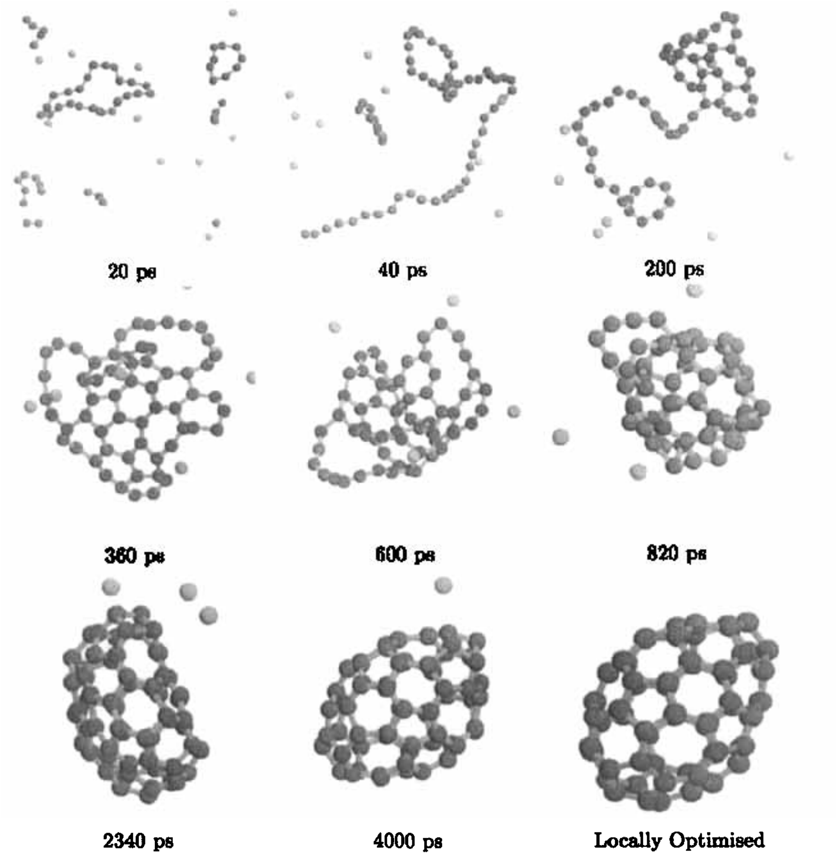


FIGURE 16 The formation of a 60 atom cluster over the course of 4 ns using molecular dynamics and annealing. Polycyclic structures first start to form after a few hundred ps as a precursor to cage structures that form after about 1 ns. Refinement of the cage structure takes place over the remaining time interval. The final optimised structure is not a perfect fullerene cage and has a potential energy of −417 eV compared to −422.6 eV for the perfect fullerene.

which are gradually annealed to a cage might be the way in which these are formed. However numerical experiment with 120 C atoms produced a 120 atoms fullerene cage not two separate C_{60}'s.

In terms of cpu time the GA method is comparative with both CRS and MD. For example, for 60 atoms cluster typical cpu time was of the order of 3×10^4 s. However it should be pointed out that the MD scheme was designed with cpu time in mind and a slower annealing schedule with a lower atom density might find the minimum energy structures more consistently. The advantage of the GA method was that it always found lower energy structures than any of the methods described above in comparable cpu time.

CONCLUSION

The application of genetic algorithms to a number of interatomic potential functions has been demonstrated. The advantage of the GA, compared to a black box global optimisation approach, is that it can be tailor-made to take into account the special features of the atomic arrangements. The efficiency of the GA is dependent on the method used to place the genes on the chromosome *i.e.*, the parameter encoding scheme. For the case of large molecules this could be an important factor in improving the performance of the method and further work on this topic might yield additional improvements to speed. The approach used for the two-component systems could also have potential application in the design of new drugs. Molecular design and candidate geometries for new drugs are increasing being investigated by using computers. The combination of the GA approach together with molecular design techniques might therefore yield substantial benefits in this developing field.

Acknowledgements

This research was supported by NATO grant CRG971170 and an EPSRC studentship.

References

[1] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison Wesley.
[2] Deaven, D. M. and Ho, K. M. (1995). *Phys. Rev. Lett.*, **75**, 288.
[3] Wolf, M. D. and Landman, U. (1998). *J. Phys. Chem. A*, **102**, 6129.

[4] De Jong, K. A. (1975). *Ph.D. Thesis*, University of Michigan.
[5] Lui, D. C. and Nocedal, J. (1989). *Math. Programming*, **45**, 503.
[6] Brenner, D. W. (1990). *Phys. Rev. B*, **42**, 9458.
[7] Brenner, D. W. (1992). *Phys. Rev. B*, **46**, 1948.
[8] Tersoff, J. (1988). *Phys. Rev. Lett.*, **61**, 2879.
[9] Gras-Marti, A., Smith, R., Beardmore, K., Jiminez-Rodriguez, J. J., Konoplev, V. and Ferron, J. (1995). *Comp. Mater. Sci.*, **3**, 413.
[10] Smith, R. and Beardmore, K. (1996). *Thin Solid Films*, **272**, 255.
[11] Hobday, S., Smith, R., Gibson, U. and Richter, A. (1997). *Rad. Eff. and Defects in Solids*, **142**, 301.
[12] Hobday, S. and Smith, R. (1997). *J. Chem. Soc., Faraday Trans.*, **93**, 3919.
[13] Woodcock, D. (1997). *http://www/sci.ouc.bc.ca/chem/programmes/modsrch.htm*, **93**, 3919, Hydrocarbon Protein Data Bank (PDB) Database, Okanagan University, Canada.
[14] Tersoff, J. (1988). *Phys. Rev. B*, **37**, 6991.
[15] Tersoff, J. (1988). *Phys. Rev. B*, **38**, 9902.
[16] Raghavachari, K. and Rohlfing, C. M. (1988). *J. Chem. Phys.*, **89**, 2219–2234.
[17] Raghavachari, K. and Rohlfing, C. M. (1992). *Chem. Phys. Lett.*, **198**, 521–525.
[18] Rothlisberger, U., Andreoni, W. and Parrinello, M. (1994). *Phys. Rev. Lett.*, **72**, 665.
[19] Finnis, M. W. and Sinclair, J. E. (1984). *Phil. Mag. A*, **50**, 45.
[20] Ackland, G. J., Tichy, G., Vitek, V. and Finnis, M. W. (1987). *Philos. Mag. A*, **56**, 753.
[21] Wales, D. J. and Doye, J. P. K. (1997). *J. Chem. Phys.*, **101**, 5111.
[22] Hobday, S. (1998). *Ph.D. Thesis*, University of Loughborough.
[23] Carroll, S., Palmer, R. and Smith, R. (1998). *Phys. Rev. Lett.*
[24] Eggen, B., Johnston, R. and Murrell, J. (1994). *J. Chem. Soc. Faraday Trans.*, **90**, 3029.
[25] Ali, M. M., Storey, C. and Torn, A. (1997). *J. Optimization Theory and Appl.*, **95**, 545.
[26] Eggen, B., Johnston, R., Li, S. and Murrell, J. (1992). *Mol. Phys.*, **76**, 619.
[27] Ziegler, J. F., Biersak, J. P. and Littmark, U., *The Stopping and Range of Ions in Solids*, Pergamon, New York, 1985.
[28] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. and Haak, J. R. (1984). *J. Chem. Phys.*, **81**, 3684.