Towards an Effective Evolutionary Approach for Binary Lennard-Jones Clusters

Francisco B. Pereira, *Member, IEEE* and Jorge M. C. Marques

***Abstract*— We present a hybrid approach for the optimization of the geometry of heterogeneous chemical aggregates. The method combines a steady-state evolutionary algorithm with a local search procedure and is able to deal with an optimization situation where both the composition of the aggregate and the spatial distribution of the particles must be determined. Specifically, we address the problem of discovering the optimal configuration for several binary Lennard-Jones instances with different atomic sizes. Results show that the hybrid approach is effective in determining the putative global optima of clusters up to 55 atoms.**

I. Introduction

A chemical cluster is an aggregate of between a few and millions of atoms or molecules. The study of the properties of these structures is an important topic of research, as it has immediate relevance in many different areas, from protein structure prediction to the field of nanotechnology. The interactions among the particles of a cluster are established by the potential energy surface (PES), a multidimensional function of the internal coordinates that defines the energetic landscape of the system. The PES incorporates information about the most stable structure of the cluster, which corresponds to the global minimum of that function [1]. In simple terms, when solving a cluster geometry optimization problem, one aims to determine the distribution of particles in the 3D space that corresponds to the lowest potential energy. In the optimal configuration, atoms can neither be too close (if this happens the potential energy will tend to infinity), nor to distant (in this situation there will be no interaction between particles and the potential will be close to zero). Due to the large number of interactions that occur inside a cluster, the task of discovering the conformation with the lowest potential energy is extremely difficult even for medium size aggregates and, indeed, it is considered a NP-Hard problem [2].

Many stochastic methods have been proposed for cluster geometry optimization. Basin-hopping, population basin-hopping, dynamic lattice searching and evolutionary algorithms are state-of-the-art methods for discovering low-energy configurations for clusters [3], [4], [5], [6], [7], [8]. Optimization experiments are usually performed on model potentials based on the sum of all pair potentials (*i.e.*, the PES is a function that depends only on the distance between every pair of particles that compose the cluster); for a recent application to argon clusters using more complicated potential functions (e.g., using three-body terms), see ref. [9].

Most of the work on global minimum search has been done with homogeneous clusters, *i.e.*, aggregates where all particles are identical. Conversely, studies that address heterogeneous clusters are rare, although they are relevant in the material science domain. Moreover, due to the higher number of local minima, heterogeneous clusters are very challenging for global optimization algorithms and, hence, provide excellent benchmarks to access their effectiveness [10]. It is worth noting that part of the local minima arising for heterogeneous clusters result from the presence of homotops, *i.e.*, aggregates with the same geometry but where, at least, two atoms of different types were exchanged in their positions. Also, the composition of the aggregate (*i.e.,* how many atoms of each type should appear) introduces a new variable that adds to the complexity of the optimization problem [10].

In this paper we present a steady-state hybrid evolutionary algorithm (EA) that can effectively search for good quality solutions for binary atomic clusters. The optimization method is based on state-of-the-art EA versions for homogeneous clusters [8], [11], [12], although some components are adapted, so that the algorithm can be applied to the new situations. The EA is applied to binary Lennard-Jones (BLJ) clusters. The BLJ model is an extension of the original Lennard-Jones potential, where atoms of two different types (and with different sizes) compose the cluster. For an aggregate of a given size, one must both determine how many atoms of each type should be present and what is the configuration that leads to the minimum potential energy [10].

The paper is structured as follows: in section II we present the binary Lennard-Jones potential. Section III comprises a detailed description of the hybrid optimization algorithm, whilst the main results are presented and analyzed in section IV. Finally, section V gathers the main conclusions.

II. Binary Lennard-Jones Clusters

BLJ clusters are composed by two types of atoms (hereafter denoted as A-type and B-type) corresponding to different sizes. The pairwise interaction between atoms i and j is given by:

$$V(R_{i,j}) = 4\epsilon_{i,j}\left[\left(\frac{\sigma_{i,j}}{R_{i,j}}\right)^{12} - \left(\frac{\sigma_{i,j}}{R_{i,j}}\right)^{6}\right] \qquad (1)$$

where $R_{i,j}$ denotes the distance between the two particles and $\epsilon_{i,j}$ is the minimum value of each pair-potential, which

Francisco B. Pereira is with Instituto Superior de Engenharia de Coimbra, Quinta da Nora, 3030-199, Coimbra, Portugal, and also with Centro de Informática e Sistemas da Universidade de Coimbra (CISUC), 3030-290 Coimbra, Portugal (email: xico@dei.uc.pt).

Jorge M. C. Marques is with Departamento de Química, Universidade de Coimbra, 3004-535 Coimbra, Portugal (email: qtmarque@ci.uc.pt).

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

is located at the equilibrium distance defined as $R_e = \sigma_{i,j}2^{1/6}$; both $\epsilon_{i,j}$ and $\sigma_{i,j}$ depend only on the type of atoms considered. In this work, we follow the conventions proposed in previous studies [10], [13]: $\epsilon_{AA} = \epsilon_{BB} = \epsilon_{AB} = 1$ and $\sigma_{AA} = 1$, while σ_{AB} is calculated by the usual combination rule $(\sigma_{AB} + \sigma_{BB})/2$, once σ_{BB} has been defined. We have then just one parameter on the potential defined by equation 1 left to be specified. In our study, we set $\sigma_{BB} = \{1.05, 1.1, 1.15, 1.2, 1.25, 1.3\}$, which allows to assess the effects of the size ratio on the stability of the cluster.

The total potential energy of a cluster C with N atoms is given summing up all pairwise contributions:

$$E_{BLJ}(C) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} V(R_{i,j}) \quad (2)$$

The goal of BLJ cluster geometry optimization is to find (for a given number of atoms N) the optimal distribution of particles in the 3D space that corresponds to the global minimum of equation 2. When trying to find the best configuration for a BLJ cluster with N atoms, it is not known *a priori* how many particles of each type should be considered. The optimization algorithm must simultaneously determine the best proportion of atoms and the geometry of the cluster. It is worth noting that this defines an interesting problem combining both a combinatorial optimization component and a numerical part: for each atom, one must specify whether it is from A-type or B-type (this is the discrete component of the problem). At the same time, a structural distribution of the atoms in the 3D space must be carried out (this is the numerical numerical component).

III. BLJ Clusters Optimization

Hybrid approaches combining steady-state EAs with gradient driven local search procedures are state-of-the-art methods for cluster geometry optimization. In the next section we briefly review some relevant proposals. Afterwards, we present the main components of the algorithm adopted in this work.

A. Related Work

Most efforts related to the application of EAs to cluster geometry optimization deal with homogeneous clusters. The two most widely used models are Lennard-Jones [14] and Morse potentials [15]. The second one is considered to be much more difficult, particularly when a short-range interaction between particles is adopted (it defines very rough search landscapes with an extremely high number of local optima) [16]. Therefore, recent efforts have focused on this variant of Morse clusters. Doye and Wales pioneered the application of an optimization algorithm (a basin-hopping method, which is a combination of a Monte Carlo procedure with local search) to this problem [3]. Roberts et al. proposed the first hybrid EA in 2000 [17] and, in 2007, Grosso et al. presented a population basin-hopping (PBH) approach with a two-phase local optimization [6]. PBH is an evolutionary approach without crossover and proved to be very effective in discovering the putative global optima of a large number of Morse instances, although it requires the specification of a set of parameters that are problem specific. Recently, Pereira and Marques presented the first completely unbiased EA that can discover the putative global optimum of all short-ranged Morse instances until 80 atoms [8].

In what concerns heterogeneous clusters, there are several reports describing the application of optimization algorithms to specific materials (see, *e.g.*, the application of a basin-hopping variant to mixed argon-xenon clusters [18] or the search for stable structures of copper and silicon clusters performed by a genetic algorithm [19]). In any case, BLJ clusters can be considered as a benchmark model for heterogeneous cluster optimization. The corresponding PES is easily defined, it creates a rough search landscape and its single free parameter (σ_{BB}) allows for an accurate modeling of different materials. Moreover, putative global minima are available from the Cambridge Cluster Database (CCD) [20] and, hence, it allows a fair comparison between alternative optimization methods.

Doye and Meyer were the first to apply an optimization algorithm to BLJ clusters. They used basin-hopping and established upper bounds on the potential energy for BLJ clusters up to 100 atoms (for all possible 6 size ratios) [10]. In 2008, Cassioli et. al, applied PBH to the same set of instances and established 95 new upper bounds, while missing 11 of the best solutions previously discovered [13]. Very recently, our research group developed a specific version of a hybrid EA for the optimization of binary atomic clusters, which was applied, for the first time, to mixed argon-krypton clusters [21]. This algorithm was also applied to some BLJ instances up to $N = 50$ and discovered all putative global optima, including a new upper bound for $N = 38$ with $\sigma_{BB} = 1.05$.

B. Hybrid Evolutionary Algorithm

The EA described in this paper is an improved version of the one we developed for the optimization of both BLJ and mixed argon-krypton clusters [21]. Its effectiveness will be tested with BLJ instances, but we endeavor the development of a general framework that can be applied to any heterogeneous cluster. It combines a steady-state EA and a local search procedure, as many previous studies confirm that hybrid approaches are essential to increase the likelihood of discovering high quality solutions [6], [8], [17]. We rely on the Broyden-Fletcher-Goldfarb-Shanno method (L-BFGS), a quasi-Newton procedure that efficiently guides search into the nearest local optimum [22]. Local search is applied to every individual generated by the EA.

1) Population Model: The EA follows a simple steady-state model. In each iteration, a sequence of steps are performed: i) tournament selection chooses two parents from the current population; ii) genetic operators (described below) are applied to the parents leading to the creation of two descendants; iii) offspring are locally optimized and evaluated; iv) replacement rules decide if the descendants

Algorithm 1 Replacement strategy

Find $X \in Pop$ such that $d(X, D)$ is minimum
if $[d(X, D) \leq d_{min}$ and $E_{BLJ}(D) < E_{BLJ}(X)]$ **then**
  D replaces X in Pop
else if $[d(X, D) > d_{min}]$ **then**
  Select $Y \in Pop$, such that $E_{BLJ}(Y)$ is maximum
  if $[E_{BLJ}(D) < E_{BLJ}(Y)]$ **then**
    D replaces Y in Pop
  end if
else
  D is discarded
end if

are allowed to join the population (and, if this happens, who will be replaced).

Quality and diversity are two main driving forces controlling the replacement strategy [6], [8]. Diversity ensures that the population is composed by distinct solutions, therefore preventing (or, at the least, delaying) convergence. It has been shown in [8] that diversity is a key issue to enhance the effectiveness of cluster geometry optimization. The application of the replacement strategy requires the specification of a distance measure to estimate the dissimilarity between two solutions. Additionally, one must also define a parameter d_{min}, specifying the minimum allowed distance between two solutions that simultaneously belong to the population. The rules that determine if a descendant D is accepted in the population are detailed in Algorithm 1. If an offspring is similar to a solution already in the population, then only the best of the two is kept. Otherwise, if the descendant is distinct from all current solutions, then it has the chance to replace the worst individual in the population (providing that its fitness is better).

In this work we adopt the center of mass distance [6], a dissimilarity measure that proved to be suitable for BLJ clusters [13]. It is based on the distance of atoms to the cluster's center of mass, and, given two solutions X and Y with N atoms, is defined as follows:

$$d^3_{ord}(X, Y) = \sum_{i=1}^{N} |Ord_X(i) - Ord_Y(i)|^3 \qquad (3)$$

where Ord_X (likewise Ord_Y) denotes a sorted vector containing the distances of atoms from X (likewise, Y) to the cluster's center of mass. The value for d_{min} is obtained with the following expression [8]: $d_{min} = 0.25 \times d_{avg}$. Here, d_{avg} stands for a rough estimate of the distance between any pair of solutions that belong to the search space. The individuals from the randomly generated initial population are used as sample to estimate d_{avg}.

2) Representation and Genetic Operators: A solution for an instance with N atoms must specify the location and type of each particle. Then, a chromosome encodes two components: a vector with $3 \times N$ real values specifying the Cartesian coordinates of the atoms and a binary vector with N positions identifying the type of every particle. The coordinates values range between $[0, Max_C]$. Max_C is set to $\sigma_{BB} \times 2^{(1/6)} \times N^{(1/3)}$, as we verified that this value enables the cluster volume to scale correctly with both N and σ_{BB}. Only inter-atomic distances larger than 0.5 are allowed, as the potential becomes too repulsive if two atoms approach too much. The minimum distance constraint is enforced in the creation of the initial population and during the application of genetic operators.

Cut and Splice crossovers are the most effective recombination operators for cluster geometry optimization [8], [11]. They act in the 3D space and, therefore, are able to recognize and maintain some semantic properties of the individuals. The original operators were developed for homogeneous clusters. In this paper, we describe $H - C\&S$, a variant of Generalized Cut and Splice [23] specifically designed to handle heterogeneous materials. The operator selects a sub-cluster from each parent and merges them to create a new solution. A sub-cluster is defined as a subset of atoms that are neighbors in the 3D space and, therefore, tend to have a low potential energy. When applied to $P1$ and $P2$, $H-C\&S$ generates two descendants $D1$ and $D2$, where the number of A-type atoms in $D1$ (respectively $D2$) is the same as that in $P1$ ($P2$). In concrete, $D1$ is created through the following steps ($D2$ is obtained by switching the role played by the parents):

1) Select a random atom CP (the cut point) in $P1$
2) Select a random number $S \in [1, N-1]$
3) Create a list L_{P1} with the N atoms from $P1$ sorted in an increasing distance to the location of CP
4) Copy the first S atoms from L_{P1} to $D1$
5) Create a list L_{P2} with the N atoms from $P2$ sorted in an increasing distance to the location of CP
6) Remove from L_{P2} atoms that are too close (*i.e.*, at a distance < 0.5) to particles already copied to $D1$
7) Copy the first $(N - S)$ atoms from L_{P2} to $D1$. Skip atoms from a given type if $D1$ already contains all required particles from that category
8) If less than N atoms were copied, $D1$ is completed with particles placed at random locations

Two mutation operators are considered, one for each component of the chromosome. Sigma mutation is used to modify the location of a particle. The new location is obtained by perturbing each coordinate with a random value sampled from a Gaussian distribution with mean 0 and standard deviation σ_{mut}. As for switch mutation, it modifies the type of an atom (from A-type to B-type or the other way around).

IV. Experimental Results

Results described in this section allow us to gain insight into the behavior of the hybrid EA. Our aim is twofold: to access its efficacy and to understand how different components help the EA to discover good quality solutions.

A. General Trends

We applied the EA to search for the optimal configuration of BLJ clusters between 5 and 55 atoms, for the six possible

values of σ_{BB}. In this section we do not present the outcomes for $5 \leq N \leq 37$, as these are rather easy instances. The optimization algorithm consistently discovered all global optima in this range and the results do not provide any relevant contribution to the analysis. The settings of the algorithm are the following: Number of runs: 30; Population size: 100; Evaluations: 5,000,000; Tournament size: 5; Crossover rate: 0.7; Sigma mutation rate: 0.05; $\sigma_{mut} : 0.05 \times Max_C$; Switch mutation rate: 0.025. Each iteration performed by L-BFGS counts as one evaluation. Initial populations are randomly generated: both the location and the type of each one of the atoms that compose a solution are arbitrarily determined. The analysis of the results focuses on the ability of the hybrid EA to discover the putative global optima for different BLJ instances (or to establish new upper bounds). This is the most widely adopted performance measure when evaluating state-of-the-art algorithms for cluster geometry optimization [6], [8], [17] .

A brief overview of the results presented in table I reveals that the EA is effective in discovering the best solutions for BLJ clusters. Columns 2 to 7 display, for every σ_{BB} value considered, the number of runs where the putative global optimum was discovered (*i.e.*, the success rate). The symbol $'-'$ indicates an instance where the known best solution was not found. Each line contains results obtained with a cluster of a given size. From the 108 instances reported in the table (18 cluster sizes $\times$ 6 σ_{BB} values), the optimization algorithm failed the putative optimum just once ($N = 55$, $\sigma_{BB} = 1.15$). On the other hand, it confirms the existence of a new upper bound for the instance $N = 38$, $\sigma_{BB} = 1.05$. The potential energy of the new solution discovered is -177.260695 for a cluster belonging to T_d symmetry point group (*i.e.*, with the same properties of symmetry of a tetrahedron), whilst the previous known best solution stored at the CCD [20] has a minimum energy of -177.209085 corresponding to a cluster with C_s symmetry (*i.e.*, it has only a plane of symmetry).

The results from table I show that, as a rule, the EA is fairly robust. In approximately 82% of the instances (89 out of 108), its success rate was higher than 50%. This is a relevant result, given the size and, particularly, the challenges posed by the rugged structure of the search landscape [10]. If we consider all instances with $\sigma_{BB} > 1.15$, the average success rate is around 81% and never drops below 33%. As for the instances with $\sigma_{BB} \leq 1.15$, the average success rate is approximately 68%. This lower value is mainly a consequence of 5 instances where the putative optimum was discovered in less than 10% of the runs. In the next section we identify the obstacles found by the EA and verify if small changes in the components of the algorithm contribute to enhance its efficacy in these difficult instances. Anyway, the trend for the enhanced performance in instances with larger values for σ_{BB} (the size ratio) suggests that the optimization algorithm is better equipped to handle heterogeneous clusters where the properties of different particles are clearly distinct. In BLJ clusters, the only difference between the two types of atoms is characterized by the size ratio.

A comparison with other methods that were also applied to BLJ clusters is not straightforward, as neither Doye and Meyer [10] nor Cassioli et. al [13] provide details concerning the computational effort or the success rate. Focusing our attention on the entries from the CCD [20], one can see that basin-hopping [10] failed to discover the putative optimum in 10 instances up to N=55. Cassioli et. al [13] refer that PBH missed the optimum 11 times (considering all instances up to 100 atoms). Although they do not provide details, it is likely that failures occurred mainly for large clusters. Anyway, they also missed the new best solution for N=38 with $\sigma_{BB} = 1.05$. For completeness, the EA previously developed for the specific optimization of argon krypton clusters [21], was also applied to all instances considered in this study. Results obtained show that its effectiveness is lower than that of the improved hybrid EA proposed in this work (section III-B). For instances with $\sigma_{BB} > 1.15$, the average success rate is around 58% and, in 4 situations, is below or equal to 10%. When $\sigma_{BB} \leq 1.15$, the average success rate is approximately 53% and, in 5 instances, the putative optimum was not found. The key difference between the EA of our previous study and the approach proposed here is that the former contains specific local search operators for the discrete component of the chromosome (*i.e.*, they are able to switch the type of some atoms). Changes are accepted if they lead to an improvement in fitness. Results show that discrete local search does not help to effectively explore the search landscape defined by BLJ clusters.

B. Analysis of Difficult Instances

Five instances proved to be particularly difficult to optimize. Since the problem is composed by two components, it is important to determine if the EA failed to discover the appropriate proportion of atoms or was unable to determine the optimal 3D configuration (or both). An inspection of the best solution found by the EA in each one of the 30 runs for these 5 instances provides useful information. Column 2 from table II displays the number of runs where the best solution discovered has a proportion of atoms that matches the composition of the putative optimum for that instance (in column 3 we recall the number of runs where the known best solution was found). An inspection of both columns reveals no clear pattern. In some instances it was difficult for the EA to discover promising solutions with the appropriate composition, while, in others, many runs discovered good quality solutions with the right proportion. This outcome was predictable, as different instances define search landscapes with specific properties. It might happen that, for some cases, good quality local optima with the right proportion of atoms are easily discovered. However, finding these solutions does not imply that the putative global optimum is achieved (see instance $N = 51$; $\sigma_{BB} = 1.15$).

As the optimization advances, there is a trend for individuals that belong to the population to have a similar composition (or a limited subset of compositions). The chart from figure 1 shows the variation on the number of

TABLE I

NUMBER OF RUNS WHERE THE PUTATIVE GLOBAL OPTIMUM WAS FOUND FOR BLJ CLUSTERS BETWEEN 38 AND 55 ATOMS (FOR EVERY σ_{BB} CONSIDERED). THE SYMBOL '-' INDICATES INSTANCES WHERE THE KNOWN BEST SOLUTION WAS NOT DISCOVERED.

	σ_{BB}					
N	1.05	1.1	1.15	1.2	1.25	1.3
38	15	29	30	30	30	30
39	29	29	30	27	29	30
40	28	30	29	29	29	28
41	26	27	29	30	30	30
42	29	30	30	28	28	23
43	3	30	30	30	30	17
44	2	28	30	29	30	22
45	11	14	25	15	19	14
46	18	17	24	23	19	20
47	26	8	22	25	22	24
48	22	12	24	25	24	29
49	28	9	22	23	23	25
50	23	23	26	27	25	23
51	24	4	2	24	26	22
52	16	24	20	23	25	25
53	26	8	17	19	26	27
54	26	1	16	26	10	12
55	19	8	-	14	17	17

TABLE II

SUCCESS RATE OF DIFFERENT EA VARIANTS IN THE OPTIMIZATION OF DIFFICULT BLJ INSTANCES.

Instance	Number of runs				
(N, σ_{BB})	Proportion	Std-EA	Fix-EA	Mod-EA	Ext-EA
(43, 1.05)	3	3	2	1	9
(44, 1.05)	6	2	-	1	9
(54, 1.10)	8	1	1	3	2
(51, 1.15)	26	2	1	1	3
(55, 1.15)	8	-	-	1	2

different compositions that appear in the population. For clarity, we provide results only from two instances ($N = 43$, $\sigma_{BB} = 1.05$ and $N = 51$, $\sigma_{BB} = 1.15$), but the same trend is visible for all other situations. Results confirm that, as the optimization advances, the individuals from the population converge to similar compositions. Even though the replacement strategy and switch mutation help to fight against this convergence, the search ability of the EA might be compromised if the correct proportions of atoms (*i.e.*, the composition of the putative global optimum) disappears from the population.

We performed two straightforward modifications in the hybrid EA that help us to gain insight into how important it is to breed the population with solutions containing the correct proportion of atoms. The first option, identified as $Mod - EA$, aims to prevent that a small subset of cluster compositions takes over the search. The maximum number of individuals with the same number of A-type atoms that can simultaneously belong to the population is set to 10% of the population size (this is an arbitrary value, but hopefully it will provide an educated guess on the relevance of this modification). When $Mod - EA$ checks if a descendant D is allowed to join the population, two situations might happen: if less than 10% of the current individuals have the same composition as D, the original replacement rules apply; otherwise, D can only replace a solution with the same proportion of atoms. The other option, labelled $Fix - EA$, fixes the composition of the individuals. When solving a specific instance, all solutions have the same proportion of atoms as that of the putative global optimum and, therefore, the EA only needs to discover the best possible configuration. Columns $Mod - EA$ and $Fix - EA$ from table II show the success rate of the two variants. Clearly, none of these variants improves the performance of the original EA. The outcomes obtained by $Fix - EA$ show that removing the discrete component does not lead to a simplified problem. Results seem to be even slightly worse than those achieved by $Std - EA$ (even though differences are not statistically

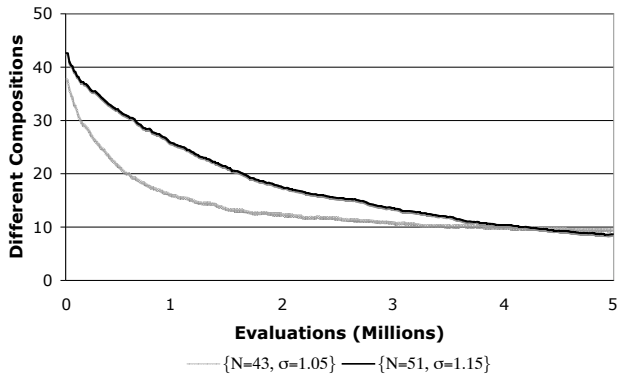

Fig. 1. Variation on the number of different compositions that appear in the population. Results were obtained by the $Std-EA$ variant in instances $\{N = 43, \sigma_{BB} = 1.05\}$ and $\{N = 51, \sigma_{BB} = 1.15\}$ and are averages of 30 runs.

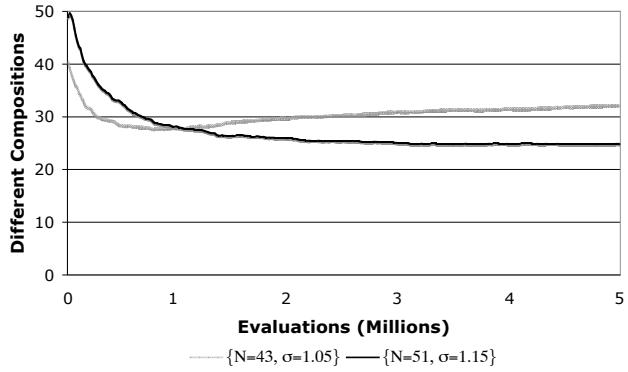

Fig. 2. Variation on the number of different compositions that appear in the population. Results were obtained by the $Mod-EA$ variant in instances $\{N = 43, \sigma_{BB} = 1.05\}$ and $\{N = 51, \sigma_{BB} = 1.15\}$ and are averages of 30 runs.

significant), suggesting that the EA can benefit from visiting and maintaining solutions with different components.

On the other hand, the chart from figure 2 reveals that $Mod-EA$ indeed increases the diversity of the population (in what concerns the cluster compositions). It displays results obtained with the same instances previously selected for figure 1. Differences between the two charts are clear and confirm that the modified EA is able to promote composition diversity at the population level. Anyway, the optimization outcomes of $Mod-EA$ presented in table II reveal that forcing the population to have individuals with distinct compositions is not, by itself, a guarantee of obtaining improved results. In what concerns the strategy adopted by $Mod-EA$, it might not be the most suitable for this problem. There are many other options that can be considered, like: modifying the dissimilarity measure in such a way that it also considers the clusters composition or adjusting the crossover operator so that it is able to create descendants with a proportion of atoms different from that of their parents. These and other alternatives will be addressed in future research.

There is another possible explanation for the decreased performance in these specific instances: they might define particularly rugged search landscapes and the limited number of evaluations is not enough to allow an appropriate exploration (a similar effect for Morse clusters is described in [8]). To check if this is happening we let the original hybrid EA to run for 10,000,000 evaluations in each one of the difficult instances (twice as much as the original tests). The success rate is presented in column $Ext-EA$ from table II. A negligible increase is visible for instances with $\sigma_{BB} = \{1.1, 1.15\}$. On the other hand, the two smaller instances with $\sigma_{BB} = 1.05$ exhibit a statistical significant increase over the success rate obtained by experiments that were allowed to run only for 5,000,000 evaluations ($p = 0.016$ for $N = 43$ and $p = 0.006$ for $N = 44$ obtained with the Taillard statistical test for comparing proportions [24]).

V. Conclusions

In this paper we presented a hybrid EA for heterogeneous cluster geometry optimization. The framework is based on state-of-the-art approaches for homogeneous clusters, whilst some components were adapted for dealing with aggregates containing distinct particles. The effectiveness of the approach was tested with several BLJ instances. Results confirm that the hybrid EA is able to successfully handle both the discrete and numerical components of the problem. It was effective in discovering the putative global optima of nearly all instances addressed and it exhibited signs of reasonable scalability and robustness.

Since we endeavor to build an effective EA for global optimization of binary clusters, we selected a few instances where the results were worse and tried to gain insight into the difficulties of the EA. A preliminary analysis reveals that the biased components used to maintain distinct compositions in the population do not contribute for a meaningful exploration of the search space. The effectiveness of the standard EA that relies solely in switch mutation is similar to that of a version that explicitly keeps diversity at this level.

In addition, the results obtained with extended runs demonstrate that, for the difficult instances, the reliability of the algorithm can be improved (its efficacy does not depend on the time that we grant it to explore the search space). Also, this inconsistency is likely to be amplified in BLJ instances with an increasing number of atoms. Therefore, our current efforts focus on the development of new components that might help the EA to effectively tackle these hard cases. An important issue that we will address in the near future is the study of alternative distance measures to estimate the similarity between two solutions. A recent proposal that we will consider in our work is the ultrafast shape recognition [25], a fast and highly accurate measure that creates a signature vector with 12 values to describe the shape of a molecule.

ACKNOWLEDGMENTS

This work was supported by Fundação para a Ciência e a Tecnologia (FCT), Portugal, under grant PTDC/QUI/69422/2006. We are grateful to the John von Neumann Institut für Computing, Jülich, for the provision of supercomputer time on the IBM Regatta p690+ (Project EPG01).

REFERENCES

[1] Stillinger, F. Exponential Multiplicity of Inherent Structures. Phys. Rev. E. 59, 48–51 (1999).

[2] Wille, L., Vennik, J. Computational Complexity of the Ground-State Determination of Atomic Clusters. J. Phys. A: Math. Gen. 419–422 (1985).

[3] Doye, J., Wales, D. Structural Consequences of the Range of the Interatomic Potential. A Menagerie of Clusters. J. Chem. Soc. Faraday Trans. 93, 4233-4243, (1997).

[4] Hartke, B. Global Geometry Optimization of Atomic and Molecular Clusters by Genetic Algorithms. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001), pp. 1284-1291, Morgan-Kaufmann, 2001.

[5] Pullan, W. An Unbiased Population-based Search for the Geometry Optimization of Lennard-Jones Clsuters: $2 \leq n \leq 372$. J. Comput. Chem., 26, 899-906, 2005.

[6] Grosso, A., Locatelli, M., Schoen, F. A Population-based Approach for Hard Global Optimization Problems Based on Dissimilarity Measures. Math. Program. Ser. A. 110, 373–404, (2007).

[7] Cheng, L., Yang, J. Global Minimum Structures of Morse Clusters as a Function of the Range of the Potential: $81 \leq N \leq 160$. J. Phys. Chem. A. 111, 5287–5293, (2007).

[8] Pereira, F. B., Marques, J. M. C. A Study on Diversity for Cluster Geometry Optimization. Evolutionary Intelligence, 2, 121–140, (2009).

[9] Marques, J. M. C., Pereira, F. B., Leitão, T. On the Use of Different Potential Energy Functions in Rare-Gas Cluster Optimization by Genetic Algorithms: Application to Argon Clusters. J. Phys. Chem. A. 112, 6079–6089, (2008).

[10] Doye, J., Meyer, L. Mapping the Magic Numbers in Binary Lennard-Jones Clusters. Phys. Rev. Lett. 95, 063401 (2005).

[11] Deaven, D., Ho, K. Molecular Geometry Optimization with a Genetic Algorithm. Phys. Rev. Lett. 75, 288–291, (1995).

[12] Zeiri, Y. Prediction of the Lowest Energy Structure of Clusters Using a Genetic Algorithm. Phys. Rev. 51, 2769-2772, (1995).

[13] Cassioli, A., Locatelli, M., Schoen, F. Global Optimization of Binary Lennard-Jones Clusters. Optimization Methods and Software, 24, 819–835, (2009).

[14] Lennard-Jones, J. E. Cohesion. Proc. Phys. Soc., 43, 461–482, (1931).

[15] Morse, P. Diatomic molecules according to the wave mechanics. ii. vibrational levels. Phys. Rev., 34, 57–64, (1929).

[16] Doye, J., Leary, R., Locatelli, M., Schoen, F. Global Optimization of Morse Clusters by Potential Energy Transformations. Informs Journal on Computing, 16, 371-379, (2004).

[17] Roberts, C., Johnston, R. L., Wilson, N.: A Genetic Algorithm for the Structural Optimization of Morse Clusters. Theor. Chem. Acc. 104, 123–130, (2000).

[18] Munro, L., Tharrington, A., Jordan, K. Global Optimization and Finite Temperature Simulations of Atomic CLusters: Use of Xe_nAr_m Clusters as Test Systems. Comput. Phys. Comm. 145, 1–23 (2002).

[19] Ona, O., Bazterra, V., Caputo, M., Ferraro, M., Fuentealba, P., Facelli, J. Modified Genetic Algorithms to Model Atomic Cluster Structures: CuSi Clusters. J. Mol. Struct. 681, 149–155 (2004).

[20] The Cambridge Cluster Database, Available at: *http://www-wales.ch.cam.ac.uk/CCD.html*

[21] Marques, J. M. C., Pereira, F. B. An Evolutionary Algorithm for Global Minimum Search of Binary Atomic Clusters. Chemical Physics Letters, 485, 211–216, (2010).

[22] Liu, D., Nocedal, J. On the Limited Memory Method for Large Scale Optimization. Mathematical Programming B. 45, 503-528, (1989).

[23] Pereira, F. B., Marques, J. M. C., Leitao, T., Tavares, J. Efficient Evolutionary Algorithms for Cluster Optimization: A Study on Locality. In: Advances in Metaheuristics for Hard Optimization, pp. 223–250. Springer-Verlag (2008).

[24] Taillard, E., Waelti, P., Zuber, J. Few Statistical Tests for Proportions Comparison. European Journal of Operational Research. 185, 1336-1350, (2008).

[25] Ballester, P., Richards, W. Ultrafast Shape Recognition to Search Compound Databases for Similar Molecular Shapes. J. Comput. Chem. 28, 1711–1723, (2007).