Enhancing Cluster Geometry Optimization with Island Models

António Leitão
CISUC
Department of Informatics Engineering
University of Coimbra
3030-290 Coimbra, Portugal
Email: apleitao@dei.uc.pt

Francisco B. Pereira
CISUC
Instituto Superior de Engenharia de Coimbra
3030-199 Coimbra, Portugal
Email: xico@dei.uc.pt

Penousal Machado
CISUC
Department of Informatics Engineering
University of Coimbra
3030-290 Coimbra, Portugal
Email: machado@dei.uc.pt

***Abstract*—Island Models are parallel approaches to Evolutionary Algorithms that not only offer the benefits of parallelization but are also regarded as models with an extensively distinct behaviour. This study applies for the first time an Island Model to the optimization of short-ranged Morse clusters, combined with a hybrid steady-state evolutionary algorithm and a local optimization method. Different migration parameters are experimented and the resulting behaviours are extensively analysed. Results are compared to a state-of-the-art sequential approach, showing slight improvements. Differences in behaviour between the Island Model and the sequential approach are comprehensively discussed. This study shows that Island Models are a competitive parallel approach with promising results on cluster geometry optimization problems.**

I. Introduction

Cluster geometry optimization (CGO) problems play an important role in the benchmarking of algorithms. The goal of such problems is to determine the organization of atomic and molecular clusters so that the total potential energy is minimized. Lennard-Jones clusters [1], [2] and Morse clusters [3] are the two most broadly used pair-wise models [4] due to their simplicity and practical relevance. In particular, Morse clusters are considered a tough test system and a highly relevant problem to benchmark global search algorithms, having been adopted by many research groups [4]. Functions that model the interactions inside a cluster are often denominated potential energy surfaces (PES) and depend solely on the distance between particles that compose the cluster. The Morse function can be used to estimate both long-range and short-range interactions. Short-ranged Morse clusters are regarded as particularly tough [5], [6].

PES are multidimensional functions that result in highly roughed landscapes with a large number of local minima, which increases exponentially with the size of the cluster, and a deep multiple-funnel character [7]–[9]. Evolutionary Algorithms (EAs) were first applied to CGO problems in the early 1990s [10], [11] and are currently considered a state-of-the-art method. The most successful approaches represent particles using Cartesian coordinates and rely on problem specific operators. Also, local search methods that depend on first-order derivative information to guide individuals into the nearest local optimum have been integrated and contributed to better results. Such hybrid architectures have been first proposed in [12] and have been used extensively [6], [13], [14].

Population-basin hopping (PBH) [13] was the first population-based approach to find the putative global optima of short-ranged Morse clusters up to 80 atoms. However, it needs to be provided a set of problem-dependent parameters, making it difficult to use in situations where the global minima is unknown. Two approaches were later proposed to overcome this problem [15], [16]. Pereira and Marques [16] suggested that the success of their approach was largely a result of an increase in diversity induced by the adaption of parameters. In [4] an unbiased hybrid steady-state EA with a single-phase local optimization method based on a general quasi-Newton method is presented. This EA includes mechanisms to ensure that diversity is maintained and shows that such an approach can efficiently tackle hard cluster optimization problems and that preserving diversity is absolutely necessary to achieve efficiency.

Islands Models (IMs) behave differently from standard EAs. These distributed models divide individuals into subpopulations that communicate only through migration, therefore limiting selection and mating of individuals to those on the same island. The creation of such borders may on the one hand prevent the successful mixing of individuals but on the other hand keep temporarily best solutions from dominating the population and promote diversity on the global level [17]. IMs have been strongly influenced by the theories of shifting balance [18] and punctuated equilibria [19].

Considering the characteristics of the PES, IMs seem adequate to tackle CGO problems. In this paper we study the design and implementation of an IM approach to the unbiased hybrid steady-state EA introduced in [4]. We assess the influence of migration parameters in the behaviour of the algorithm and make a comparison with published results for the sequential approach. We show that the IM is able to find the putative optima for Morse clusters ranging from 41 to 80 atoms and that despite the differences in behaviour, the number of successful runs are competitive with those obtained by a state of the art approach thus making it an adequate parallelization model for these problems.

In Sect. II, CGO is addressed, followed by a description of a steady-state EA and a quasi-Newton local search procedure. Sect III introduces IMs, covers related work, and introduces the design of IMs for the proposed hybrid steady-state EA. Sect. IV describes the experimental setup and includes the discussion of a parametric study on IMs as well as a comparison of the obtained results to those achieved by a sequential version. Conclusions are drawn in Sect. V.

II. Morse clusters optimization

In this section we will introduce the concept of Morse clusters and make a brief overview on related work on CGO. Also, we will present a hybrid steady-state EA for the optimization of Morse clusters.

A. Morse clusters

The potential energy of Morse Clusters is defined by the N-particle pair-wise additive Morse potential [3] given by (1), where r_{ij} is the Cartesian distance between atoms *i* and *j*, ϵ is the bond dissociation energy, r_0 the equilibrium bond length and β represents the range exponent of the potential. By setting both ϵ and r_0 to *1*, the potential of (1) becomes a scaled version of the Morse function [20] with non-atom specific interactions, leaving β as the only adjustable parameter. In this study β is set to *14* which corresponds to short-range interactions, considered specially challenging as the PES is extremely rough and a large number of local minima are present [5]. Additionally, the global minima for instances with a different number of atoms correspond to different cluster geometries [20], [21].

$$V_{Morse} = \epsilon \sum_{i}^{N-1} \sum_{j>i}^{N} \{exp[-2\beta(r_{ij} - r_0)] - 2exp[-\beta(r_{ij} - r_0)]\} \quad (1)$$

The employment of local minimization methods requires the specification of the analytical gradient of the function being optimized. The generic element *n* of the Morse cluster potential gradient is given by (2) where x_{ni} represents the difference between the Cartesian coordinates of particles *i* and *n*. Similar expressions apply for the *y* and *z* axis.

$$g_n = -2\beta\epsilon \sum_{i \neq n}^{N} \left(\frac{x_{ni}}{r_{ni}} \right) \{exp[-2\beta(r_{ni} - r_0)] - exp[-\beta(r_{ni} - r_0)]\} \quad (2)$$

B. Related work

An overview of the most important approaches of EAs for CGO can be found in [4]. We will be focusing on Morse Clusters and will briefly review some of the most relevant achievements in this section.

Morse clusters were first used as an optimization problem by Doye and Wales [20] who applied a basin-hopping algorithm. They report finding all but 12 of the putative global optima for short-ranged Morse clusters up to 80 atoms. EAs were first applied to this problem by Roberts et al. [22] and combined real-valued representation [23] with a Lamarckian local optimization method [12] and a Cut and Splice crossover operator [12]. This approach was able to find nearly all putative global optima for medium and short-range Morse clusters sized from 19 to 50 atoms. A later revisited version by the same research group successfully found all global optima for the same range [6].

Locatelli and Schoen proposed a two-phase local optimization procedure designed to increase the efficiency of methods that need to explore PESs with a multiple funnel topology [24]. A basin-hopping approach using the two-phase local optimization method was applied to Morse clusters [5] discovering almost all putative global optima in instances ranging from 41 to 80 atoms. The drawback of the two-phase local approach is that it is not completely unbiased as it requires the specification of a number of parameters which are dependent on the structure of the optimal solution for each instance to be optimized. Two research groups recently proposed self-adaptive approaches to overcome this limitation [15], [16].

The study presented in [16] suggested that the success of the proposed approach was for the most part due to an increased diversity introduced by the adaptive settings. A recent study from the same research group confirmed this hypothesis [4]. This paper proposed a straightforward hybrid steady-state EA relying on a single-phase local optimization procedure based on a general quasi-Newton method and showed how CGO problems can be tackled efficiently by an unbiased hybrid EA.

C. A hybrid steady-state EA for Morse clusters optimization

In the remainder of this section we present the main components that compose the algorithm used in the experiments. They are all similar to other state-of-the-art EAs and have been thoroughly discussed in Pereira's study on diversity on CGO [4], [25]. The study proposes and compares different operators and parameters that we have selected based on the results obtained and the discussion on their behaviour. Pereira gives special attention to the mechanisms responsible for the maintenance of diversity which are therefore linked the success of the proposed approach. Various studies support that diversity is a key issue for the success of optimization algorithms [4], [13], [15], [16], [26].

1) Local optimization: Various studies show that hybrid approaches to CGO achieve striking effects regarding search efficacy [6], [12], [14]. In this study, it is performed with the Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS), a quasi-Newton conjugate gradient method [27]. L-BFGS is applied to every generated individual. During local search the method is applied until a local optimum is found or until a maximum number of iterations, the Local Search Length (LSL), is reached. The accuracy of local search is set to $1.0E-8$.

2) Representation and genetic operators: A solution represents the position of all the particles that compose the

cluster, which means that for *N* atoms, an individual encodes $3 \times N$ real values specifying the Cartesian coordinates of each particle. The gene values range from *0* to $N^{1/3}$. It is widely accepted that this range allows the cluster volume to scale correctly with *N* despite the lack of formal proof [6], [23]. Inter-atomic distances are kept larger than a pre-specified parameter δ since atoms being too close results on the potential becoming too repulsive. This restriction is enforced during initialization of the population and when the genetic operators are applied. In this work, δ is set to *0.5*. Fitness is assigned by applying L-BFGS on each individual before using (1) to calculate its potential energy.

Crossover is performed using the Generalized Cut and Splice (GenC&S) operator [25] which was designed for CGO problems with the aim of preserving some semantic properties of the parents involved in crossover. GenC&S tries to ensure that each parent contributes for the generation of offspring with a set of atoms that are close together. Briefly, the operator works by selecting a random number $X \in [1, N-1]$, where *N* is the size of the cluster. The first parent contributes with *X* atoms while the second parent contributes with $N - X$ atoms. Sub-clusters are selected from each parent so that their atoms are neighbours in the 3D space and are superimposed to create new individuals (consult [25] for implementation details regarding the GenC&S operator).

Sigma mutation is applied to modify the location of an atom inside a cluster. This is accomplished by changing the three coordinates that determines the position of an atom in the 3D space. The new location is obtained by perturbing each coordinate with a random value sampled from a Normal distribution with mean *0* and standard deviation σ.

3) Population model: Unlike generational models, in steady-state EAs parents compete for survival with their offspring. Usually, in these models only one or two individuals are generated at each step. After their creation it must be decided if their are to be part of the population and if so, which individuals should be replaced. The replacement strategy plays a major role on the performance of steady-state algorithms [28]–[30]. One well known strategy is to replace the worst individual in the population with a newly generated one. This method is known as the GENITOR [31] and is bound to put the EA under high selective pressure, guiding it to premature convergence [30], [32]. GENITOR is a simple example of replacement strategies, other methods that handle selective pressure in a more appropriate manner have been used. A detailed characterization of selection pressure in different replacement strategies can be found in [30] along with analysis on takeover time and loss of diversity. The crucial issue raised by this and other analysis is the need to maintain diversity during the search. The search landscape that results from (1) presents a number of local minima that increases exponentially with the size of the cluster [13], therefore, it is imperative to provide the EA with mechanisms that help prevent it from getting stuck. A number of interesting approaches can be found in the literature [33]–[35].

The replacement strategy chosen for this study was proposed by Lee et al. [36] and later incorporated in the PBH algorithm [13]. It includes mechanisms that help maintain an appropriate level of diversity. The algorithm in Fig. 1 shows the steps that determine if an individual *D* will be inserted in the population *Pop*. A brief explanation of the strategy follows: In each iteration, two parents are selected and two descendants are created. Considering a newly generated individual *D* already locally optimized and evaluated, if the population contains a solution that is close to *D*, then only the best of the two is kept. Otherwise, *D* is different from all the solutions in the population and it will replace the worst one if *D* has a better fitness. In order to perceive if individuals are close, a distance measure *d(X,Y)* that captures the dissimilarity between individuals *X* and *Y* is required. Also, a parameter d_{min} must be specified, representing the minimum allowed distance between two individuals from the population.

```
Find X ∈ Pop such that d(X, D) is minimum
if [d(X, D) ≤ d_min and V_Morse(D) < V_Morse(X)] then
    D replaces X in Pop
else if [d(X, D) > d_min] then
    Select Y ∈ Pop, such that V_Morse(Y) is maximum
    if [V_Morse(D) < V_Morse(Y)] then
        D replaces Y in Pop
    end if
else
    D is discarded
end if
```

Fig. 1. Replacement strategy adopted by the EA

4) Diversity measure: A large number of measures are available in the literature. In [4] a review of different measures for CGO can be found. In [26] a different classification methodology is also available.

In this study, a measure based on distances of atoms to the center of mass is used, as proposed in [13]. This choice was based on the results and analysis presented in [4]. This measure works as follows: First, for a given cluster, the distance of each atom to the center of mass is calculated and the values are stored in a vector in a non-decreasing way. Then, given two clusters *A* and *B* and the matching ordered sets Ord_A and Ord_B, the dissimilarity between the clusters is calculated using (3).

$$d_{ord}(A, B) = \frac{1}{10} \sum_{i=1}^{N} |(Ord_A(i) - Ord_B(i))|^3 \qquad (3)$$

This measure has been incorporated in two-phase local optimization PBH and applied to Morse clusters up to 80 atoms, showing excellent results.

In order to estimate d_{min} we rely on information provided by the randomly generated population using (4) [13]. The parameter ζ specifies the proportion among the average distance calculated with a set of individuals randomly generated and the minimum allowed distance that may exist between two solutions that simultaneously belong to the population. ζ has

been set to 0.25 in this study. Pop represents the size of the population.

$$d_{min} = \zeta \times \frac{\sum_{A=1}^{Pop-1} \sum_{B=A+1}^{Pop} d(A,B)}{\frac{1}{2}(Pop^2 - Pop)} \quad (4)$$

III. Island models for Morse clusters optimization

In this section we will introduce IMs and their distinct characteristics as well as related work. We will also discuss an approach to hybrid steady-state EAs.

A. Island models

Island models are models of EAs that use a number of subpopulations (islands) as opposed to standard EAs (Panmictic) [37] which evolve on their own and from time to time exchange individuals through migration. IMs have the appealing advantage of allowing the distribution of computation effort but more interesting is the fact that IMs result on a behaviour distinct from standard EAs. While the separated evolution processes of an IM make it different from standard EAs, migration makes it different from isolated runs of EAs [37].

The most distinctive aspect of IMs is that individuals are restricted to mate with individuals of the same island since selection is done independently. Regarding steady-state EAs, IMs also influence the replacement strategy by restricting it to the island of the emerging individual. These restrictions are therefore borders that separate individuals and make different evolutionary processes possible. Such an approach may on the one hand prevent successful mixing of individuals but on the other hand it helps keep temporarily better individuals from dominating the whole population, resulting on blocks of solutions that occupy different regions of the search space [37].

The later characteristics support the maintenance of diversity on the global population and provide ecological opportunity for weaker individuals to participate in the evolution process and hopefully result in better individuals [38]. Regarding CGO problems, these characteristics may be helpful in covering a large part of the search space. While intra-island evolution pushes individuals toward different local optimums, first stochasticity and later through selection, migration introduces new genetic material into the islands, ultimately helping them progress.

IMs require the specification of different components, including the number and size of the islands, a topology that states how islands are connected, operators for the selection of emigrants and for the selection of individuals to be replaced (migration policies) as well as parameters that determine migration sizes and frequency. All these components play an important role on IMs and need to be balanced in order to achieve the desired effect.

B. Related work

Studies on IMs go back as far as Grefenstette's study [39] where he proposes a parallel GA. In his approach, the best individuals from each island are broadcast to all the others at every generation. Grosso was the first to observe that dividing the population into islands resulted on faster fitness improvements than on a Panmictic population. He also showed that isolated islands performed worst than single large populations [40].

Tanese was the first to perform systematic experiments with migration and its effects on IM efficiency [41]–[43]. She varied migration sizes and migration intervals and studied their effects on the maintenance of allele diversity. She showed that both need to be balanced and that too high or too small rates result on worst solutions comparing with a Panmictic approach. Also, she suggested using different GA parameters in each island in order to prevent premature loss of diversity.

Arguably, migration sizes and intervals are the most important parameters in IM's. Different studies are available in the literature on the effects of these parameters. The most popular approach is to use fixed sizes and intervals between migrations. Effects on Genetic Algorithms [38], [44] and Genetic Programming [45], [46] are available. Variation of island numbers and sizes have also been regarded as important. A large number of islands is usually desirable, helping the system cover a larger area of the search space. However, for a fixed population size, the size of each island reduces as the number of islands increases. In [38] it is suggested that increasing the number of islands is beneficial as long as the island size doesn't fall under a given threshold.

Migration policies have also been thoroughly studied. Cantu-Paz, in [47], shows how different migration policies affect selection pressure and speed of convergence. He shows that choosing migrants and individuals to be replaced according to their fitness increases selection pressure and guides the process to convergence faster. Similar conclusions are obtained in [38]. Migration topologies help control how information spreads through the islands. The sparser the topology is, the slower information spreads in the system. Popular topologies include an array, a ring, a grid, a star or fully connected topologies. A study on different topologies can be found in [48]. Fernandez [45] has proposed a dynamic topology where connections between islands represent potential migrations and islands choose a target at each migration step.

C. An approach to hybrid steady-state EAs

IMs require a population of size P to be divided in N islands of size M so that $P = N \times M$. The steady-state approach used in this study has a strong effect on the behaviour of the islands, helping reduce premature convergence and maintaining diversity internally, thus allowing the use of a large number of islands without compromising their performance. In this study we have split a population of *100* individuals into *10* islands of *10* individuals each. Despite this effort, a system consisting of *10* islands is quite small, promoting the fast spread of individuals through migration and the consequent loss of diversity, specially when considering the size of the islands. In order to reduce this effect we have used a dynamic topology [45] as described in Sect. III-B. We

```
Consider a candidate immigrant (I) to a subpopulation (Pop)
for each individual (P) in Pop do
    if d(I, P) == 0 and V_Morse(I) == V_Morse(P) then
        Discard I
        Return
    end if
end for
sort Pop
Replace a random individual except Pop[0]
```

Fig. 2. Replacement strategy used by the migration policy

think of our topology as a bidirectional fully connected graph where each edge represents potential migrations. This choice makes each island able to communicate with every other but with a probability of *10%* in each migration step.

The migration policy applied was *random-random* as it reportedly causes less impact on the algorithms convergence. Therefore, individuals will be randomly selected from their island of origin and will replace random individuals at the target island. However, in order to adapt this strategy to the steady-state model, restrictions were imposed so that migrants arriving at an island are discarded if an equal individual is already present in the population as shown by the algorithm in Fig. 2. Comparison between individuals relies on their fitness values and distance between them as described by (3). This strategy will keep migration from causing duplication of individuals which would increase their influence on selection for reproduction. We also protect the best individual from being replaced.

Despite these mechanisms to maintain diversity, migration, on a global level, still causes duplication of some genes and deletion of others. Therefore, the number of migrants sent by each island to their target should on the one hand be enough to promote innovation and progress but on the other hand small enough to prevent flooding, fast duplication and deletion of genes. In this study we have experimented with migration rates of *1*, *2* and *4* individuals.

We have also experimented with intervals of *5* and *10* generations between migrations. The interval should be long enough to allow for the exploitation of new genetic material that is injected into each island through migration. This will allow useful genes to spread through the subpopulations, guiding them to new areas of the search space and generating locally optimized individuals. Spreading these individuals allows the recombination of optimized solutions. However, increasing the size of the intervals will cause already optimized islands to remain in stasis for longer periods of time, wasting resources and causing them to be less productive until the stop condition is reached.

IV. Experimental Results

In this section we will cover the experimental setup and analyse the results obtained. A parametric study on the IM is first discussed followed by a study on the results of the best setting.

A. Experimental settings

1) Panmictic approach: Each experiment evolves a population of 100 individuals until 5,000,000 evaluations are performed. Each iteration of the L-BFGS algorithm counts as 1 evaluation. LSL is set to 1,000. Selection is done through tournament of 5 individuals, followed by crossover at the rate of 0.7 and mutation at the rate of 0.05. Also, $\sigma = 0.05 \times N^{1/3}$. Experiments were repeated 30 times for Morse clusters ranging from 41 to 80 atoms. For analysis purposes, a generation counter is incremented every time 100 individuals are generated.

2) IM approach: Similarly to the number of individuals, the number of allowed evaluations have also been divided by the islands. Therefore, each island will stop when 500,000 evaluations have been conducted. Also, while tournament of 5 individuals may be well suited for a population of 100 individuals, in subpopulations of 10 individuals it would cause a very high selection pressure, ultimately stopping the 4 individuals with worst fitness from sharing their genes on their present islands (40% of the global population size). While migration might at some point allow these individuals to move to islands where they have a chance of sharing their genes, there is high probability of them being deleted before doing so, therefore, we have set the tournament size to 2.

For each combination of migration rates and intervals, experiments were repeated 30 times for a small set of Morse clusters instances. The analysis of the behaviour for different setups is presented in Sect. IV-C and leads us to believe that using a rate of *1* individual with intervals of *5* generations between migrations is the most beneficial to the evolution process. For this particular setup, experiments were repeated 30 times for Morse clusters ranging from 41 to 80 atoms.

B. Performance metrics

On a global level, the performance of the IM is based on its ability to reach the putative optima. Therefore, during experiments we registered the number of runs where the optima was found for different Morse cluster sizes. This measure is widely accepted for the evaluation of state-of-the-art approaches for CGO [4], [5], [13], [21].

In order to compare the behaviour of different parameters in IMs, other metrics were necessary. For each generation, the following elements were registered for each island: best fitness, average fitness, number of substitutions, average dissimilarity between individuals and fitness of incoming immigrants. The metrics are independent for each island. On a global level this allows for the calculation of the global best fitness, global average fitness, total number of substitutions and average inner diversity of the islands. Some of this metrics have been used for the analysis of steady-state approaches to CGO [4] while others have been used in the analysis of IM behaviour [37].

C. The influence of IM parameters

This section details the differences in behaviour between various parameter setups. First we will address the best setup followed by a comparison with other migration rates and

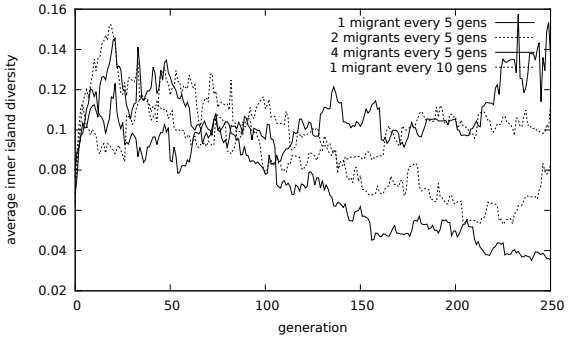

Fig. 3. Diversity measure along a run for a cluster of 51 atoms.

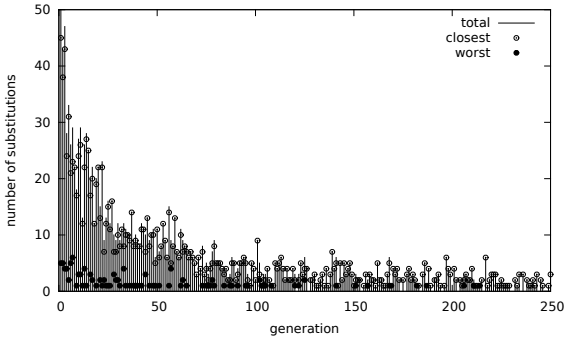

Fig. 4. Acceptance rate of the replacement strategy adopted by the EA along a run for a cluster of 51 atoms. 1 migrant every 5 generations.

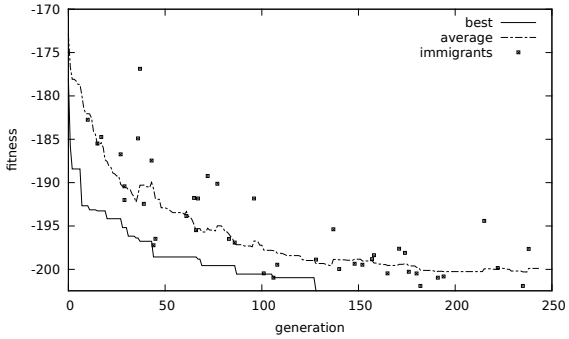

Fig. 5. Fitness evolution along a run for a cluster of 51 atoms. 1 migrant every 5 generations.

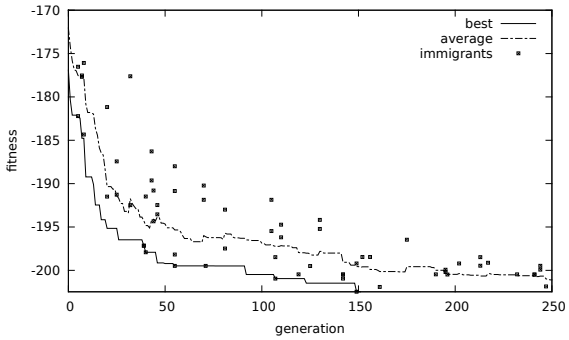

Fig. 6. Fitness evolution along a run for a cluster of 51 atoms. 2 migrants every 5 generations.

intervals. While comparing a given parameter, all other keep the value of the best setup.

Fig. 3 shows that the average inner diversity of the islands is maintained along the run, in spite of the sudden peaks that are the result of migration. Despite the small size of the islands, the replacement strategy applied by the steady-state approach is able to prevent the fast convergence of individuals. Also, Fig. 4 shows that newly bred individuals are able to integrate the subpopulations consistently along the run, suggesting that diversity, on a global level, is maintained sufficiently high to promote progress. Migration plays a delicate double role, on the one hand duplicating and deleting genes and on the other hand injecting new genetic material into islands, thus helping the emergence of new innovative individuals.

More interestingly, on an island level analysis, Fig. 5 shows that islands are able to adequately take advantage of new genetic material. It is clear that a correlation between progress in terms of best fitness and immigrants exists, either by the inclusion of new best individuals or more interestingly by the exploitation of worse ones. As soon as these individuals are available for selection, they are able to spread their genes through the population, allowing it to escape local optima.

1) Number of migrants: Fig. 5 shows that migration of *1* individual at each step is enough to help islands evolve. One may think that by increasing the number of individuals to be migrated there is a higher chance of injecting helpful genetic material into target islands, specially since the policy used may reject some individuals. However, Fig. 6 shows that there is not much gain from migrating *2* individuals instead of *1*. In fact, in Fig. 3 we can observe that the average inner island diversity is slowly decreasing which is more visible for rates of *4* migrants. The nature of this migration policy and the fact that it both duplicates and deletes genes, helped by the small size of the islands, ultimately leads to loss of global diversity which at the end disrupts both exploration and exploitation, guiding the system to premature convergence. A similar result may be expected by using a fixed topology with a high number of neighbours relative to the number of islands.

2) Migration intervals: A migration interval of *10* generations is still able to promote progress in the islands as is shown by Fig. 7 as well as maintain diversity as observable in Fig. 3. However Fig. 7 shows that islands undergo long periods of stasis, where no progress is made. This behaviour suggests that while diversity is maintained by the steady-state replacement strategy, the islands have converged to a point where no helpful reproduction is likely to occur. This behaviour is amplified by the small size of the islands. Intervals of *5* generations seem long enough for the islands to exploit new genetic material and find optimized solutions. While it is likely that, given a larger number of evaluations, results as good as the best could be achieved, periods of stasis are a waste of resources and evaluations.

D. Results

A general overview of the number of runs where the putative optimum was found is presented in Table I. Lines labeled

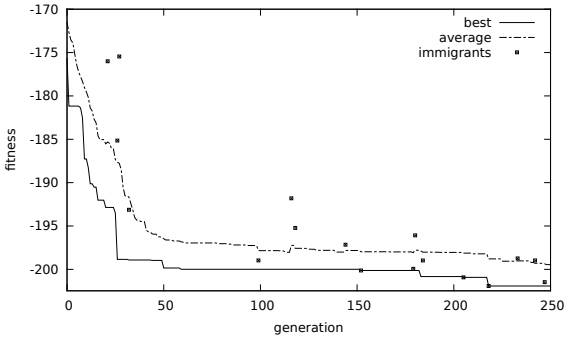

Fig. 7. Fitness evolution along a run for a cluster of 51 atoms. 1 migrant every 10 generations.

'Panmictic' show results from a previous study using the same hybrid steady-state EA as applied here [4]. Lines labeled 'IM' show the success rate of our approach. The symbol '-' was used in instances where the optimum was not found.

On a first analysis, results show that our approach was able to find all the putative optimum for Morse clusters consisting of 41 to 80 atoms. Also, the global performance seems to be slightly better than the Panmictic approach, finding the global optimum with a higher rate on 21 instances of the problem against 12 where the Panmictic found it more often and 7 ties. In order to confirm this assumption, the pairwise proportions test described in [49] was performed. Instances where significant differences regarding success rates have been found are highlighted in Table I. A significance level of 0.01 was used. Except for 5 instances, no significant differences where found. On the one hand, IM approach obtained worse results on 1 instance but on the other hand results have been enhanced for 4 other.

On a closer analysis, Table I indicates that while the IM approach performed well on some instances where the Panmictic EA had a low number of successful runs, the opposite happened as well. The instances used on this experimental setup have optimal solutions with different geometric structures as well as different properties [5], [20], [21] which suggests that IMs may be better suited for certain instances while the Panmictic may be more advantageous in others. This phenomena emphasizes the differences in behaviour between the different models, however further analysis is needed in order to confirm so.

In IMs, selection and replacement strategies behave substantially different from the Panmictic approach. The division of individuals into small islands and the consequent reduction of tournament size increases the chance of less fit individuals to be selected for reproduction. Regarding the replacement strategy, new individuals face less concurrence from existing solutions, thus having a higher chance to be included in the population. Also d_{min} is set independently in each island, meaning that different islands will ultimately have different replacement behaviours. Both of the afore mentioned characteristics help promote inter-island diversity. However, the lack of a global replacement strategy is bound to allow some loss diversity on the global population. The migration policy applied helps reduce the impact.

Despite the restrictions imposed by the policy, migration is still a duplicative and deleterious action as it is. The study performed on the rate and interval between migrations has proved to be very important and has shown that the correct choice of parameters help keep this operation valuable to the global population.

V. Conclusion

This paper presented for the first time an IM approach to the optimization of Morse clusters. A sequential hybrid steady-state model with a quasi-Newton local optimization procedure was described and later included on the IM. A comprehensive study on the influence of migration parameters on the behaviour of the model was presented and the results of the best setup were compared to the sequential approach.

The IM was able to find all putative optima for Morse clusters ranging from 41 to 80 atoms and obtained a performance slightly more robust than the sequential approach. This study has shown that IMs are a competitive alternative to sequential approaches, showing promising results. Future work on this subject may include tackling the optimization of Morse clusters up to 160 atoms as well as study how different migration topologies, policies and population sizes effect IM performance on hybrid steady-state approaches to CGO.

Acknowledgment

This work has been partially supported by the project PTDC/EIA-EIA/102212/2008, High-Performance Computing over the Large-Scale Internet.

References

[1] J. E. Jones, "On the Determination of Molecular Fields. II. From the Equation of State of a Gas," *Royal Society of London Proceedings Series A*, vol. 106, pp. 463–477, Oct. 1924.

[2] J. E. Lennard-Jones, "Cohesion," *Proceedings of the Physical Society*, vol. 43, pp. 461–482, Sep. 1931.

[3] P. M. Morse, "Diatomic Molecules According To Wave Mechanics. II. Vibrational Levels," *Phys. Rev.*, 1929.

[4] F. Pereira and J. Marques, "A study on diversity for cluster geometry optimization," *Evolutionary Intelligence*, vol. 2, pp. 121–140, 2009.

[5] J. P. K. Doye, R. H. Leary, M. Locatelli, and F. Schoen, "Global optimization of morse clusters by potential energy transformations," *INFORMS J. on Computing*, vol. 16, pp. 371–379, September 2004.

[6] R. L. Johnston, "Evolving better nanoparticles: Genetic algorithms for optimising cluster geometries," *Dalton Trans.*, pp. 4193–4207, 2003.

[7] F. H. Stillinger, "Exponential multiplicity of inherent structures," *Physical Review E Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 59, pp. 48–51, Jan. 1999.

[8] C. J. Tsai and K. D. Jordan, "Use of the histogram and jump-walking methods for overcoming slow barrier crossing behavior in Monte Carlo simulations: Applications to the phase transitions in the $(Ar)_{13}$ and $(H_2O)_8$ clusters," *The Journal of Chemical Physics*, vol. 99, pp. 6957–6970, Nov. 1993.

[9] L. Wille and J. Vennik, "Computational complexity of the ground-state determination of atomic clusters," *Journal of Physics A: Mathematical and General*, vol. 18, pp. 419–442, 1985.

[10] B. Hartke, "Global geometry optimization of clusters using genetic algorithms," *The Journal of Physical Chemistry*, vol. 97, no. 39, pp. 9973–9976, 1993.

[11] Y. Xiao and D. E. Williams, "Genetic algorithm: a new approach to the prediction of the structure of molecular clusters," *Chemical Physics Letters*, vol. 215, pp. 17–24, Nov. 1993.

TABLE I
NUMBER OF SUCCESSFUL RUNS OBTAINED BY THE IM AND PANMICTIC ON THE OPTIMIZATION OF MORSE CLUSTERS RANGING FROM 41 TO 80 ATOMS

N	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
IM	11	**20**	4	4	5	3	2	20	22	3	8	10	8	5	6	11	11	**13**	2	**11**
Panmictic	15	12	**14**	7	5	9	2	14	18	5	7	6	5	12	6	11	8	4	2	-

N	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
IM	2	8	11	6	9	10	6	4	6	10	5	10	8	**7**	5	7	4	5	1	1
Panmictic	1	12	6	11	8	8	5	4	6	8	6	7	6	1	2	4	7	3	6	5

[12] D. M. Deaven and K. M. Ho, "Molecular Geometry Optimization with a Genetic Algorithm," *Physical Review Letters*, vol. 75, pp. 288–291, Jul. 1995.

[13] A. Grosso, M. Locatelli, and F. Schoen, "A population-based approach for hard global optimization problems based on dissimilarity measures," *Mathematical Programming*, vol. 110, pp. 373–404, 2007.

[14] F. Pereira, J. Marques, T. Leitao, and J. Tavares, "Analysis of locality in hybrid evolutionary cluster optimization," in *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*, 0-0 2006, pp. 2285 –2292.

[15] A. Cassioli, M. Locatelli, and F. Schoen, "Global optimization of binary lennard-jones clusters," *Optimization Methods Software*, vol. 24, pp. 819–835, August 2009.

[16] F. Pereira and J. Marques, "A self-adaptive evolutionary algorithm for cluster geometry optimization," in *Hybrid Intelligent Systems, 2008. HIS '08. Eighth International Conference on*, sept. 2008, pp. 678 –683.

[17] Z. Skolicki, "An analysis of island models in evolutionary computation," in *Proceedings of the 2005 workshops on Genetic and evolutionary computation*, ser. GECCO '05. New York, NY, USA: ACM, 2005, pp. 386–389.

[18] S. Wright, "Stochastic processes in evolution," in *Stochastic Models in Medicine and Biology*. University of Wisconsin Press, 1964.

[19] N. Eldredge and S. J. Gould, "Punctuated equilibria: an alternative to phyletic gradualism," in *In T. J. Schopf, Models in Paleobiology*, 1972, pp. 82–115.

[20] J. P. K. Doye and D. J. Wales, "Structural consequences of the range of the interatomic potential a menagerie of clusters," *J. Chem. Soc., Faraday Trans.*, vol. 93, pp. 4233–4243, 1997.

[21] L. Cheng and J. Yang, "Global minimum structures of morse clusters as a function of the range of the potential: 81 $\leq$ n $\leq$ 160," *The Journal of Physical Chemistry A*, vol. 111, no. 24, pp. 5287–5293, 2007, pMID: 17521176.

[22] C. Roberts, R. L. Johnston, and N. T. Wilson, "A genetic algorithm for the structural optimization of morse clusters," *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, vol. 104, pp. 123–130, 2000.

[23] Y. Zeiri, "Prediction of the lowest energy structure of clusters using a genetic algorithm," *Phys. Rev. E*, vol. 51, pp. R2769–R2772, Apr 1995.

[24] M. Locatelli and F. Schoen, "Fast global optimization of difficult lennard-jones clusters," *Computational Optimization and Applications*, vol. 21, pp. 55–70, 2002.

[25] F. B. Pereira, J. Marques, T. Leitao, and J. Tavares, "Designing efficient evolutionary algorithms for cluster optimization: A study on locality," in *Advances in Metaheuristics for Hard Optimization*, ser. Natural Computing Series, P. Siarry and Z. Michalewicz, Eds. Springer Berlin Heidelberg, 2008, pp. 223–250.

[26] A. Cassioli, M. Locatelli, and F. Schoen, "Dissimilarity measures for population-based global optimization algorithms," *Computational Optimization and Applications*, vol. 45, pp. 257–281, 2010, 10.1007/s10589-008-9194-5.

[27] D. C. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical Programming*, vol. 45, pp. 503–528, 1989.

[28] K. A. D. Jong and J. Sarma, "Generation gaps revisited," in *Foundations of Genetic Algorithms 2*. Morgan Kaufmann.

[29] M. Lozano, F. Herrera, N. Krasnogor, and D. Molina, "Real-coded memetic algorithms with crossover hill-climbing," *Evol. Comput.*, vol. 12, pp. 273–302, September 2004.

[30] J. Smith, "On replacement strategies in steady state evolutionary algorithms," *Evol. Comput.*, vol. 15, pp. 29–59, March 2007.

[31] D. Whitley, "The genitor algorithm and selection pressure: Why rank-based allocation of reproductive trials is best," in *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann, 1989, pp. 116–121.

[32] D. E. Goldberg and K. Deb, "A comparative analysis of selection schemes used in genetic algorithms," in *Foundations of Genetic Algorithms*. Morgan Kaufmann, 1991, pp. 69–93.

[33] N. Krasnogor, "Recent advances in memetic algorithms, studies in fuzziness and soft computing, vol. 166, chap. towards robust memetic algorithms," 2004.

[34] D. Pelta and N. Krasnogor, "Recent advances in memetic algorithms, multimeme algorithms using fuzzy logic based," in *Memes for Protein Structure Prediction, Studies in Fuzziness and Soft Computing, Physica-Verlag, Wurzburg*, 2004.

[35] N. Krasnogor, B. P. Blackburne, E. K. Burke, and J. Hirst, "Multimeme algorithms for protein structure," in *In: Proceedings of the Parallel Problem Solving from Nature VII. Lecture Notes in Computer Science*. Springer-Verlag, 2002, pp. 769–778.

[36] J. Lee, I.-H. Lee, and J. Lee, "Unbiased global optimization of lennard-jones clusters for $n \leq 201$ using the conformational space annealing method," *Phys. Rev. Lett.*, vol. 91, p. 080201, Aug 2003.

[37] Z. Skolicki, "An analysis of island models in evolutionary computation," in *Proceedings of the 2005 workshops on Genetic and evolutionary computation*, ser. GECCO '05. New York, NY, USA: ACM, 2005, pp. 386–389.

[38] W. N. Martin, J. Lienig, and J. P. Cohoon, "Island (migration) models: evolutionary algorithms based on punctuated equilibria," in *Handbook of evolutionary Computation. IOP Publishing and*. Oxford University Press, 1997.

[39] G. J, "Parallel adaptive algorithms for function optimization," Tech. Rep., 1981.

[40] P. B. Grosso, "Computer simulations of genetic adaptation: parallel subcomponent interaction in a multilocus model," Ph.D. dissertation, Ann Arbor, MI, USA, 1985, aAI8520908.

[41] R. Tanese, "Distributed genetic algorithms for function optimization," Tech. Rep., 1989.

[42] ——, "Distributed genetic algorithm," in *Proceedings of the Third International Conference on Genetic Algorithms*, 1989, pp. 434–439.

[43] ——, "Parallel genetic algorithms for a hypercube," in *In: Proceedings of the Second International Conference on Genetic algorithms and their application, Mahwah, NJ, Lawrence Erlbaum Associates, Inc*, 1987, pp. 177–183.

[44] Z. Skolicki and K. De Jong, "The influence of migration sizes and intervals on island models," in *Proceedings of the 2005 conference on Genetic and evolutionary computation*, ser. GECCO '05. New York, NY, USA: ACM, 2005, pp. 1295–1302.

[45] F. Fernández, M. Tomassini, W. Punch, and J. Sánchez, "Experimental study of multipopulation parallel genetic programming," in *Genetic Programming*, ser. Lecture Notes in Computer Science, R. Poli, W. Banzhaf, W. Langdon, J. Miller, P. Nordin, and T. Fogarty, Eds. Springer Berlin / Heidelberg, 2000, vol. 1802, pp. 283–293.

[46] F. Fernández, M. Tomassini, and L. Vanneschi, "An empirical study of multipopulation genetic programming," *Genetic Programming and Evolvable Machines*, vol. 4, pp. 21–51, 2003.

[47] E. Cantu-Paz, "Migration policies, selection pressure, and parallel evolutionary algorithms," *Journal of Heuristics*, vol. 7, pp. 311–334, 2001.

[48] I. Sekaj, "Robust parallel genetic algorithms with re-initialisation," in *Parallel Problem Solving from Nature - PPSN VIII*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2004, vol. 3242, pp. 411–419.

[49] E. D. Taillard, P. Waelti, and J. Zuber, "Few statistical tests for proportions comparison," *European Journal of Operational Research*, vol. 185, no. 3, pp. 1336 – 1350, 2008.