TÉCNICO LISBOA

INSTITUTO SUPERIOR TÉCNICO

ADVANCED TOPICS IN COMPUTATIONAL PHYSICS

Structural optimization of Lennard-Jones clusters using genetic algorithm

Author:
Inês Pinto Juna Santos Rodrigo Pereira

Supervisor:
Prof. Carlos Loia Reis

Abstract

The aim of this study is to implement a Genetic Algorithm (GA) to optimize the Lennard Jones potential for atomic clusters, with the ultimate goal of finding the global minima while minimizing the number of times the potential is calculated. The GA consists of initializing a generation of molecules within a box, followed by mutations and mating in the population, and finally substituting molecules with the worst potential for those with the best. The expected cluster configurations were obtained for cluster sizes $3 \leq N \leq 25$, but as the cluster size increased, it became more challenging to find the global minima, and the number of function evaluations required also increased. We also found that for optimal results the mutation rate should be kept between 1% and 5%. For bigger cluster sizes the Genetic Algorithm was combined with the conjugate gradient local minimization. This along with initializing the atoms in spherical surfaces produced the better results but required significantly more function evaluations. A study of the influence of the parameters was also performed for a molecule with thirteen atoms that concluded that the size of the box, the survival rate along with the maximum mutation translation proved to be the most significant factors for arriving at the global minimum with the least fitness evaluations. Finally, we obtained highly symmetric structures of 38 and 55 atom Lennard Jones clusters.

APRIL 2023

1 INTRODUCTION

For a very long time, Nature has taken care of the evolution of all living organisms, using processes such as mutations and natural selection. Based on this concept, Genetic Algorithms (GA) were developed to optimize solutions for a given problem. In a genetic algorithm, a population of candidate solutions (individuals) evolves toward better solutions. This is an iterative process, where each individual in the initial population is generated with random characteristics (genes). In each iteration (generation) we compute every individual's fitness, which corresponds to the value of the objective function that we aim to optimize, and we give a small probability of mutation to all members of the population, which corresponds to randomly changing their genes. Furthermore, we can also combine genes from different individuals through sexual reproduction, which helps to maintain a high diversity of the population. Finally, we select the fittest individuals in the population and substitute the less fit by them, this way forming the next generation's population. A schematic overview of this algorithm can be seen in Fig. 1

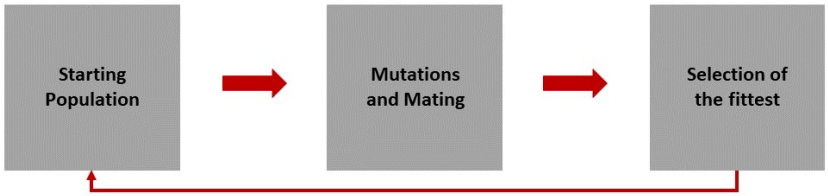


Figure 1: Schematic view of the Genetic Algorithm.

This is known to be an extremely powerful and easy to implement algorithm, capable of finding the global minimum or maximum of high dimensionality problems, whilst maintaining a reduced number of calls to the objective function.

1.1 Our Optimization Problem

The aim of this project was to find the positions of a cluster of N atoms that minimized the pair-wise Lennard-Jones potential, which can be computed as

$$\mathrm{V_{LJ}} = 4\epsilon \sum_{i=1}^{N-1} \sum_{j>i} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right].$$

However, since both ϵ and σ are usually very small numbers, we need to change variables to make this calculation more computationally stable. Thus, the above expression becomes

$$\mathrm{V^*_{LJ}} = 4 \sum_{i=1}^{N-1} \sum_{j>i} \left[\left(\frac{1}{r^*_{ij}} \right)^{12} - \left(\frac{1}{r^*_{ij}} \right)^{6} \right], \tag{1.1}$$

where $\mathrm{V^*_{LJ}} = \mathrm{V_{LJ}}/\epsilon$ and $r^*_{ij} = r_{ij}/\sigma$. The expression of Eq. (1.1) is the objective function that we aim to minimize. It is important to mention that we not only want to obtain the structure of the cluster that minimizes the Lennard-Jones potential but are also focused on doing so by calculating the expression (1.1) as few times as possible.

Thus, in the first generation, a population of molecules is randomly formed as described in Section 2. The following generations emerge from a cyclical evolutionary process:

- Reproduction - either asexual (as described in Section 3) or sexual (as described in Section 4).
- Selection - the population is ordered according to each individual's fitness, and the highest ranked move on to the next generation, and the lowest ranked are replaced by copies of them.

2 INITIALIZATION

Several methods for initializing the positions of the first generation of molecules were implemented and tested:

- Randomly uniformly generating each coordinate of each atom in a box from 0 to L, with L being a free parameter. This method proved to be efficient, especially for molecules with a low number of atoms and small boxes.
- Randomly generating each atom in a spherical surface of radius $L/2$.
- Setting the first atom at the center and the remaining atoms in two concentric spherical surfaces. This helped when optimizing molecules with a high number of atoms.

3 MUTATIONS

The genes of a molecule can be interpreted as the positions of every atom composing it. In this sense, a mutation corresponds to randomly translating atoms in a molecule, with a small probability of occurrence, which can be varied according to the number of atoms being considered. This translation can be computed by generating three random numbers, each for each spacial coordinate, in the range $[-m_0 L, m_0 L]$, where m_0 is a free parameter. The effect of the following types of mutations on the global optimisation was tested:

1. Randomly select one and only atom to translate.
2. Each atom has a fixed probability of being translated.

It was verified that the most effective way of mutating the population was the first method. Furthermore, we also considered the case where the value of m_0 could decrease for a higher number of iterations, adopting an inverse cooling scheme, where our m_0 now becomes $\frac{m_0}{1+\alpha N_{it}}$, where α is a free parameter, usually small, to choose accordingly to the problem. This proved to be a very efficient method to find the global minimum of the objective function.

Finally, for molecules with many atoms, a local minimization algorithm - the conjugate gradient method for multi-dimensions [1] - method was implemented for each molecule that suffered a mutation. The goal here was to ensure that we had sufficient genetic diversity so that all the minima of the fitness function were explored. Thus, a high number of molecules and a high survival rate were chosen. In addition, a high m_0 was selected to allow a molecule that is stuck in a valley to explore others, increasing the chance of encountering the global minimum. Of course, the more individuals constitute the population and the higher the mutation probability, the more computationally expensive any method, and this method in particular become, due to the local minimization.

4 MATING

A certain percentage (survival rate) of the fittest population is selected as parents. The remaining population has a probability (sex probability) of undergoing mating. The following methods were implemented and their pertinence for the global optimization problem was tested:

1. The child's positions are a combination of the positions of both parents.
2. The father and mother molecules are divided by a plane, with the mother's atoms above the plane passed to the child, and the father's atoms below the plane passed to the child. If these conditions haven't yet formed a molecule with the original number of atoms, the opposing conditions are imposed [2].
3. Similar mating process to the previous one, but the parents are split by a spherical surface rather than a plane [2].

Implementing the first and second mating processes (with a small probability for the second one to occur) proved to be the most efficient method.

5 PARALLELIZATION

We used *OpenMP* to perform the parallelization of our code to accelerate the optimization process. The chosen approach is called Data Parallelism and it consists in dividing the population into several smaller sub-populations. Then, each sub-population evolves independently in separate threads. It is important to notice that individuals of different sub-populations are allowed to reproduce with each other. This type of parallelization is especially useful when the fitness function is computationally expensive or when we need to calculate it a lot of times, which is the case of molecules with a high mutation probability or a number of atoms.

6 RESULTS FOR 3-25 ATOM CLUSTERS

6.1 Potential as a function of Function Evaluations

We not only want to find the global minimum of the system, but we also seek to find it by calculating the Lennard-Jones potential as few times as possible. In this sense, it becomes important to study how the potential of the best individual in the population evolves with the number of times we call the objective function. Figs. 2a-2c show this evolution for the twenty atom cluster.

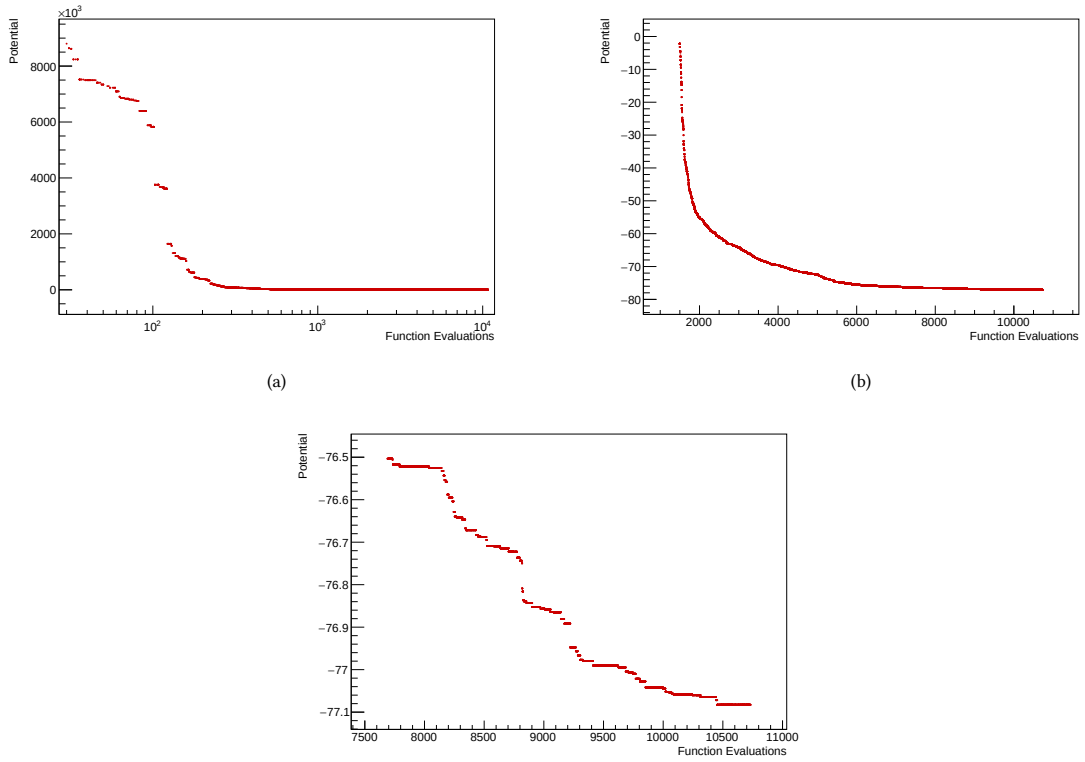


Figure 2: Structure of the thirteen atom cluster. (a): Side view of the cluster. (b): Top view of the cluster.

In the beginning, since the total potential is really high, due to randomly generating atoms in a small volume, almost every time a favorable mutation happens in an individual its potential drops by a lot, as can be seen in the first two hundred function evaluations in Fig. 2a. If we now look a little closer at what happens later, in Fig. 2b, we see that almost every mutation leads to a decrease in the total potential, however, their effect is much smaller than in the case before, since the cluster is becoming more stable every time a mutation occurs. Finally, in Fig.

2c we observe the evolution curve in the final stages where the potential starts to stabilize. Here, the algorithm arrives at a lot of local minima, thus it needs more time until a mutation becomes favorable. Thus, our choice of parameters and mutation type must be adjusted to pull out of the local minima faster.

6.2 Global minimum with two decimals precision

Table 1 presents the results obtained from the implementation of our algorithm. As the cluster size increases, obtaining the fitness value becomes more challenging, leading to an increase in the number of function evaluations. For optimal results, we found that it is best to keep a low mutation probability, ideally between 1% to 5%.

We also found that while mating did help us to converge to the correct value more frequently, it also resulted in a significant increase in the number of function evaluations. As a result, the values found in table 1 did not take mating into account.

To find suitable parameters, we performed our algorithm 10 times and considered that if more than half of the runs successfully converged to the global minimum with two decimals places accuracy then the parameters were acceptable.

Cluster Size	Fitness	Reference Fitness	Function Evaluations (Avg)	Mutation Probability	Population Size
3	-3.00	-3.000000	1125	0.01	4
4	-6.00	-6.000000	891	0.01	6
5	-9.10	-9.103852	3773	0.01	6
6	-12.71	-12.712062	4926	0.01	20
7	-16.50	16.505384	8807	0.01	15
8	-19.82	-19.821489	18734	0.01	20
9	-24.11	-24.113360	23584	0.01	18
10	-28.42	-28.422532	31956	0.01	18
11	-32.76	-32.765970	33948	0.01	18
12	-37.96	-37.967600	35204	0.01	18
13	-44.32	-44.326801	29954	0.01	30
14	-47.84	-47.845157	17885	0.01	15
15	-52.32	-52.322627	54965	0.02	30
16	-56.81	-56.815742	63807	0.02	20

Table 1: Fitness Results obtained for cluster sizes between 3 and 16.

6.3 Global Minimum with maximum precision

For our algorithm to achieve the same level of precision as the already established values for the Lennard-Jones potential [3] we added a local minimization. This method consists in reducing the mutation translation, allowing for mutations to occur in only one decimal place at a time. The results can be observed in the table 2.

When working with cluster sizes smaller than 20, we are able to obtain values with full precision. However, as the size of the cluster increases, achieving the same level of accuracy becomes more difficult.

The disadvantage of the local minimization is that the algorithm is less likely to converge to the global minimum as it gets stuck in other local minima, this will increase the number of trials required to reach the optimal solution.

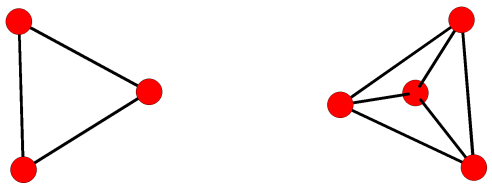
Cluster Size	Fitness	Reference Fitness	Function Evaluations
3	-3.000000	-3.000000	1355
4	-6.000000	-6.000000	1905
5	-9.1038524	-9.103852	4999
6	-12.712062	-12.712062	4990
7	-16.505384	-16.505384	4782
8	-19.821489	-19.821489	7507
9	-24.113360	-24.113360	10143
10	-28.422532	-28.422532	10153
11	-32.765970	-32.765970	19965
12	-37.967600	-37.967600	19892
13	-44.326801	-44.326801	12683
14	-47.845157	-47.845157	17885
15	-52.322627	-52.322627	15818
16	-56.815742	-56.815742	72830
17	-61.317995	-61.317995	499920
18	-66.530949	-66.530949	143737
19	-72.659757	-72.659782	381769
20	-77.1770260	-77.177043	384455
25	-102.37266	-102.372663	438458

Table 2: Best precision fitness results for cluster sizes ranging from 3 to 25.

6.4 MOLECULAR STRUCTURE OF small CLUSTERS

In the previous sections, we have studied whether or not our algorithm can reach the global minimum of the Lennard-Jones potential for clusters with several numbers of atoms. However, it is also interesting to study the molecular structure of these clusters. For that, we used the software *xbs - ball and stick molecule modeling* [4] , which reads the position of each atom and converts it into an image. Atomic bonds can also be added to provide better visualization.

We start by considering the two simpler cases, three and four atom clusters. It is expected that in the equilibrium structures of these cases, all the atoms are equally spaced, forming an equilateral triangle and a tetrahedron, respectively. Figs. 3a and 3b show that our algorithm correctly finds the most stable structure, as expected.

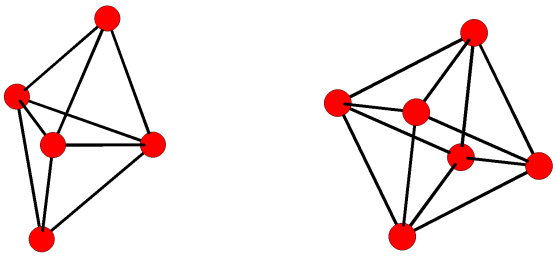


(a) (b)

Figure 3: (a): Structure of the three atom cluster. (b): Structure of the four atom cluster.

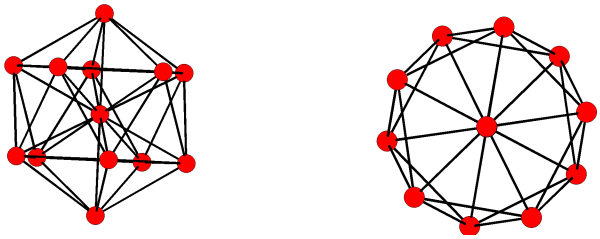
If we now study the structure of the five and six atom clusters, we find that they form hexahedron and octahedron structures, respectively, as can be seen in Figs. 4a and 4b.

The next interesting structure occurs for the thirteen atom cluster. In this case, the atoms form an icosahedron-like structure with an atom in the center as can be seen in Fig. 5a. The top view is shown in Fig. 5b where we can easily see the symmetry of this structure.



(a) (b)

Figure 4: (a): Structure of the five atom cluster. (b): Structure of the six atom cluster.



(a) (b)

Figure 5: Structure of the thirteen atom cluster. (a): Side view of the cluster. (b): Top view of the cluster.

7 USING `OPTUNA` TO FURTHER OPTIMIZE THE ALGORITHM

The hyperparameter optimization software `optuna` [5], commonly used in the search for the set of hyperparameters of a machine learning model that maximize its performance, was used with two different purposes.

Firstly, we aim to find the combination of parameters (in a specified range) that minimized the number of calls to the function that calculates the Lennard-Jones potential. This was performed specifically for the thirteen atom cluster in Section 7.1.

Finally, since the number of atoms in the cluster is increased the more complex the potential function is, `optuna` was implemented in order to find the combination of parameters that were capable of achieving the global minimum. This was computed for the typical clusters of 38 and 55 atoms as can be seen in Section 7.2.

7.1 The influence of parameters in the convergence of the 13 atom cluster

A study of the influence of different parameters (such as the population size, survival rate, mutation rate, mutation maximum translation, and the size of the box where the atoms are originally generated) on the convergence (number of function calls required to reach the global minimum) of the thirteen atom cluster was conducted.

To that end, the number of maximum iterations was increased and an early stopping method, that terminates the algorithm when little to no variation is shown in the best molecule's fitness value for an appropriate number of iterations, was implemented so that the number of function calls was only the necessary.

At each trial, the algorithm was executed five times for the same set of parameters to reduce the randomness of the results. For this study, the set of parameters that generated an evolutionary system that didn't reach the global minimum in neither attempt was set with a high number of calls to avoid loss of information.

The size of the box, the survival rate and the maximum translation m_0 proved to be the most relevant features for arriving at the global minimum of the thirteen atom cluster with the least fitness evaluations, while the mutation probability and the size of the population were the least as it can be seen in Figure 6. The convergence benefited from generating molecules in a box with $L < 1$, with survival rate $> 65\%$ and $m_0 < 1.8$ as it can be seen in Figure 7.

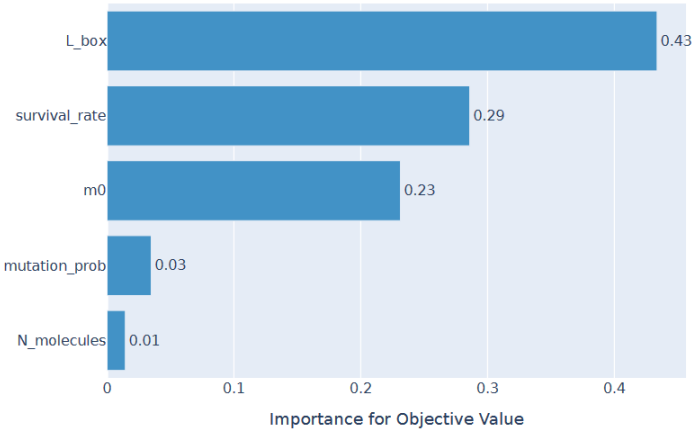


Figure 6: Parameter importance in the convergence of the 13 atom cluster.

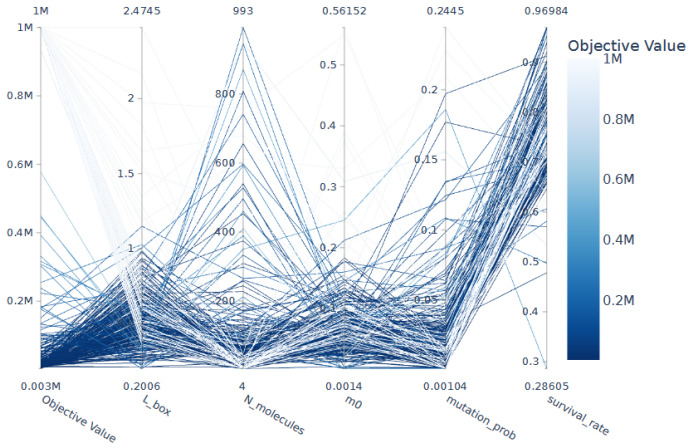


Figure 7: Parameter importance in the convergence of the 13 atom cluster.

The best combinations of parameters required only about 3000 iterations, which proves the high efficiency of the genetic algorithm in finding the global minimum of a 39-dimensional problem.

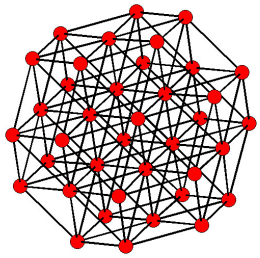
7.2 Many Atoms Cluster

Here, the goal was to find the set of parameters that allowed the algorithm to find the global minima of bigger clusters.

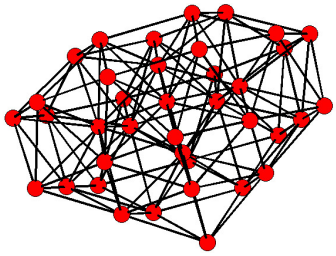
The molecules were initialized in two concentric spheres as described in Section 2, and suffered mutations with a high maximum translation, followed by local minimization (this last step was only introduced after an appropriate number of iterations due to its high computational cost) as described at the end of Section 3. These parameters were the same as before, however, their search spaces were different since both the characteristics of the molecules and approaches for the problem are distinct. Hence, higher intervals were chosen for the size of the initialization volume and of the population, the maximum translation, and the survival rate. However, the maximum mutation rate was reduced since each mutation carried a greater computational cost.

In these conditions, we were able to come extremely close to the global minimum of both the 38 atom cluster and the 55. For the 38 atom cluster, we arrived at a minimum of -173.431 (reference value of -173.828 [3]) with about 10^7 calls to the potential function and 10^5 to the function that calculates its derivative. We also arrived at a local minimum value of -172.837. These structures can be seen in Figs. 8a and 8b, respectively.

For the 55 atom cluster, we arrived at a minimum of -278.588 (reference value of -279.248 [3]) with about 10^8



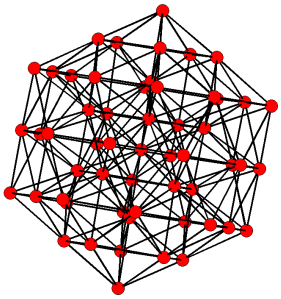
(a) LJ38 global minimum with an FCC truncated octahedron [6].



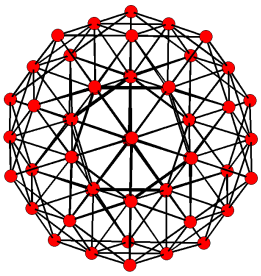
(b) LJ38 local minimum (incomplete Mackay icosahedron [6]).

Figure 8: Structure of the 38 atom cluster.

calls to the potential function and 10^6 to the function that calculates its derivative. The structure of this cluster is also an icosahedron and can be seen in Figs. 9a and 9b.



(a)



(b)

Figure 9: Structure of the 55 atom cluster. (a): Side view of the cluster. (b): Top view of the cluster.

8 CONCLUSION

In this project, we implemented Genetic Algorithms to find the configuration of a cluster that minimized the Lennard-Jones Potential. With this purpose, we implemented several strategies for the initialization and mutation of our population. First, we studied how the potential of the best molecule in the generation evolved with the number of objective function evaluations. We concluded that this behavior varies for different ranges of function evaluations, which is connected to the stability of the molecule. Afterward, we presented the global minima obtained for various cluster sizes. Firstly, we guaranteed that the algorithm could reach the global minimum the majority of times, for $N \leq 16$, with a precision of two decimal places, to study how acceptable a set of parameters were. Secondly, we found the global minimum of clusters with $N \leq 25$ with much more precision but did not try to reproduce the results consistently like before, in order to verify that our algorithm could obtain the results from [2]. We could not obtain full decimal precision for $N \geq 19$, but the results were still very positive. With the global minima found, we plotted some structures of these clusters. Subsequently, we used `optuna` to study how different sets of parameters influence the convergence of the algorithm in the thirteen atom cluster. It was found that the parameters $L < 1$, survival rate $> 65\%$, and $m_0 < 1.8$ were the most beneficial. Finally, we used this software to obtain the global minima for the clusters with $N = 38$ and $N = 55$, which were very hard to find otherwise due to their symmetry, and we could observe their structure.

With this study, we concluded that Genetic Algorithms are a powerful tool to optimize multi-dimensional functions that could not be so easily found with other methods.

9 ACKNOWLEDGMENTS

We'd like to thank our supervisor for his guidance and the weekly insightful discussions that enable us to keep improving our algorithm.

REFERENCES

[1] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, USA, 3 edition, 2007.

[2] C. Barrón, S. Gómez, D. Romero, and A. Saavedra. A genetic algorithm for lennard-jones atomic clusters. *Applied Mathematics Letters*, 12(7):85–90, 1999.

[3] Table of global minima for Lennard Jones clusters, 1998.

[4] M. Methfessel. xbs - ball and stick molecule modeling, 1998.

[5] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[6] Jonathan P. K. Doye, Mark A. Miller, and David J. Wales. The double-funnel energy landscape of the 38-atom lennard-jones cluster. *The Journal of Chemical Physics*, 110(14):6896–6906, apr 1999.