In the *Proceedings of the Fourth International Conference on Neural Networks and Genetic Algorithms (ICANNGA' 99)*. Springer-Verlag.

Transposition: A Biologically Inspired Mechanism to Use with Genetic Algorithms

A. Simões[1,2], E. Costa[2]
[1] Escola Superior de Educação, Instituto Politécnico de Leiria, Rua Dr. João Soares, 2400 Leiria, Portugal
[2] Centro de Informática e Sistemas, Universidade de Coimbra, Polo II, 3030 Coimbra, Portugal
E-mail: {abs, ernesto}@dei.uc.pt

Abstract

Genetic algorithms are biological inspired search procedures that have been used to solve different hard problems. They are based on the neo-Darwinian ideas of natural selection and reproduction. Since Holland proposals back in 1975, two main genetic operators, crossover and mutation, have been explored with success. Nevertheless, in nature there exist much more mechanisms for genetic recombination based in phenomena like gene insertion, duplication or movement. The goal of this paper is to study one of these mechanism, called transposition. Transposition is a context-sensitive operator that promotes the movement intra or inter chromosomes. In this preliminary work we empirically study the performance of the genetic algorithm where the traditional crossover operator was substituted by transposition. The results are very promising but must be confirmed by a more extensive empirical study and the correspondent theoretical justification.

1 Introduction

Genetic Algorithms (GA's) are a search paradigm that applies ideas from evolutionary biology (crossover, mutation, natural selection) in order to deal with intractable search spaces. The power and success of GA's is mostly achieved by the diversity of the individuals of a population which evolve, in parallel, following the principle of "the survival of the fittest". In the standard GA the diversity of the individuals is obtained and maintained using the genetic operators crossover and mutation which allow the GA to find more promising solutions and avoid premature convergence to a local maximum [7].

In order to find the most efficient ways of using GA's, many researchers have carried out extensive studies to understand several aspects such as the role of types of selection, space representation and how to apply the genetic operators. Several studies were made concerning the genetic operators crossover and mutation. For instance, Schaffer and Eshelman empirically compared mutation and crossover and concluded that mutation alone is not always sufficient. Spears and De Jong analyse the role of crossover and mutation in terms of disruption theory, trying to understand the power of the two operators [14]. Later, De Jong and Spears in [3] present a formal study of the role of multipoint crossover in GA, in order to analyse their recombination potential and exploratory power. Their work provides a better understanding of when and how to use n-point and uniform crossover. Spears in [15] proposes an adaptive GA which decides between 2-point and uniform crossover as it runs. He concludes that this adaptive mechanism works well especially with larger populations.

Although the classical GA uses these two main genetic operators to achieve population diversity, in nature the diversity of the species genetic material is obtained by several mechanisms which involve gene insertion, duplication or movement. With this respect, Mitchell and Forrest point out the importance of study other "mechanisms for rearranging genetic material (e.g., jumping genes, gene deletion and duplication, introns and exons)" to know if any of these is significant algorithmically [10].

Some authors proposed other biological inspired genetic operators besides crossover and mutation. Harvey and Smith suggest alternative genetic operators inspired in a bacterial form of recombination called conjugation [6], [12], [13]. This process involves the uni-directional transfer of genetic material by direct cellular contact between a donor bacterial cell and a recipient cell. Harvey suggests a type of conjugation based on tournament selection [6]. Parents are selected on a random basis, the two parents "fight" in a tournament. The winner of the tournament becomes the donor and the loser the recipient of the genetic material. Smith uses conjugation as a method of genetic recombination to solve hard satisfiability problems [12]. He constructed a simple model using a GA operating directly the phenotype (the satisfiability expressions) and using mutation operator. A random population is created and placed in a 15x15 matrix. The individuals are allowed to move in the matrix and to

conjugate genetic material if placed in adjacent positions. Later, Smith proposes a simple conjugation operator involving two individuals randomly chosen [13]. Both authors achieved good results using those substitute genetic operators.

In this paper we will introduce a genetic operator alternative to crossover, inspired in real biology. This mechanism is known as transposition and consists in the presence of genetic mobile units called transposons, that are capable of relocating themselves, or transposing, onto the chromosome and subsequently jumping into new zones of the same or other chromosome.

We will compare the performance of the GA in finding an optimal solution to a given function using either crossover and transposition followed by mutation.

This paper is organised in the following manner. First, in section 2, we introduce the classical way to use the traditional GA. In section 3, we describe how transposition works in nature and discuss how we implement it. In section 4, we present our case study and we make an exhaustive comparison of the results obtained with transposition, 1-point, 2-point and uniform crossover. Finally we conclude with a discussion and direction for future research in this area.

2 The Classical Genetic Algorithm

The mechanics of a simple GA is very simple, involving nothing more complex than copying strings and swapping partial strings [4]. It starts with an initial population of individuals created at random. Then, this population is evolved through time by a string manipulation process based in three genetic operators: reproduction, crossover and mutation (see Figure 1).

1. Generate population
2. Do
   - 2.1. Evaluate population
   - 2.2. Reproduction (Select parents)
   - 2.3. Crossover
   - 2.4. Mutation
   - 2.5. Substitute old population

Until (DONE)

Fig. 1. - The classical GA

Reproduction is a process in which individuals are copied to a mating pool according to their fitness. Individuals with higher fitness have higher probability of generate offspring in next generation. The "goodness" of a solution is measured by the fitness function and is typically defined with the respect to the current population. This operator mimics the natural selection process in which the fittest individuals are determined by their ability to survive predators, sickness and other obstacles.

After reproduction, crossover may proceed in two steps. First, members of the individuals in the mating pool are selected and mated at random. Second, each pair of strings is crossed-over, exchanging genetic material between them. There are several types of crossover operator, but the general idea of all of them is to swap genetic material between two strings. For instance, in 1-point crossover, a cut point is chosen at random and the genes are swapped according the cut point (see Figure 2):

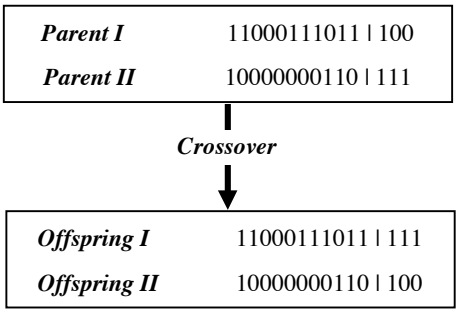

Fig. 2. Crossover operator

The power of the GA is mostly due to crossover. It is the most important operator to the GA. Diversity is indispensable to evolution. The populations diversity is obtained and maintained by crossover. which allows the GA to find better solutions in the search space [9].

The offspring generated by reproduction and crossover can be affected by mutation. The effect of this operator is to change the value of a single gene (see Figure 3). Although mutation plays a secondary role in the operation of the GA, it is needed to avoid premature convergence of the GA to a local optima. Mutation is applied with a low rate and has the ability to "shake" the GA enabling it to continue evolving.

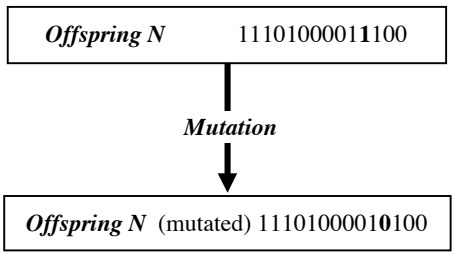

Fig. 3. Mutation operator

3 Transposition

In nature the genetic diversity of the individuals is preserved by several mechanisms that involve operations like gene insertion, duplication or movement. In each one of these categories there are several processes that in one way or another produce changes in the genome of the species enabling the genetic diversity , so important to the evolutionary process. For instance we can find phenomenon like transformation, transduction, conjugation, retroinsertion, etc. (involving gene insertion); break and fusion, unequal recombination, transposition, etc (involving either gene duplication or gene movement). Table 1 shows the categorisation of these mechanisms [5], [11].
In this paper we explore one of these mechanisms, the transposition.

Table 1. Biological mechanisms of changing genetic material

Mechanism	Possible Consequences
Transformation	New genes from a dead cell imported from surrounding medium incorporated into chromosome of bacterium.
Transduction	New genes accidentally picked up from previous host and imported into cell by a virus.
Conjugation	Genetic material from a donor bacteria is transferred to a recipient cell.
Lysogenic Insertion	Novel genes of temperate phage inserted into host genome.
Retroviral Insertion	cDNA copy of novel genes of retrovirus inserted into host genome.
Intron Insertion	Excised introns inserted into genome, mainly at exon-exon junctions in cDNA insertions.
Retroinsertion	cDNA copies of transcribed host DNA incorporated into genome providing duplicate copies of genes.
Breakage and Fusion	Part of one chromosome breaks off and fuses to the end of another during gamete formation; some gametes may obtain duplicate copies of genes on the broken fragment.
Unequal crossing over	Chromosomes may be misaligned during the crossover process; some gametes may obtain duplicate copies of some genes.
Transposition	Chromosomal DNA moved with genome, or both duplicated and moved.

3.1 Biological Transposition

Transposition characterises itself by the presence of mobile genetic units that move about in the genome, either removing themselves to new locations or by duplicating themselves for insertion elsewhere. These mobile units are called transposons.
Transposons (also known as jumping genes) can be formed by one or several genes or just a control unit, and can move in several ways, none of which is fully understand: some transposons move from one site on the chromosome to a new point of the same or to the other chromosome; others leave a copy behind, still others remain fixed but dispatch copies to other sites. In some cases, the transposon, before inserting in the target position, duplicates itself and the seek for another insertion point continues in the same way (see Figure 4).

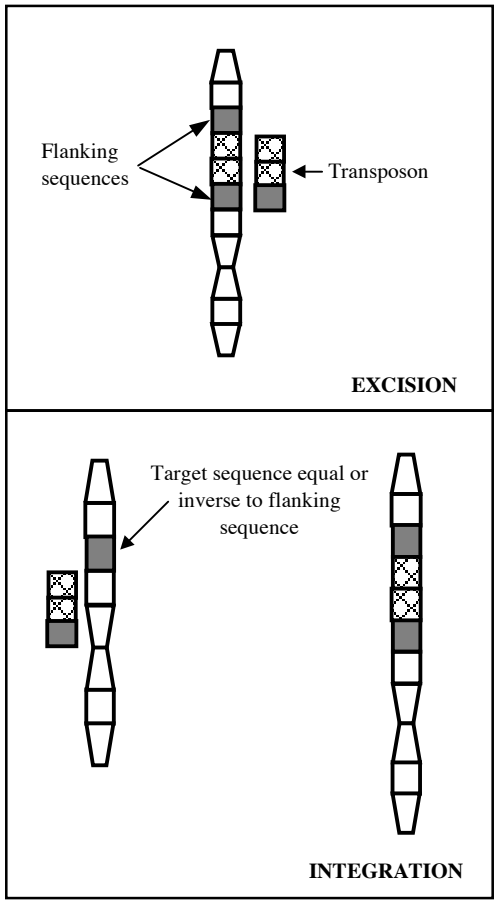

Fig. 4. Transposition mechanism

Transposition was first discovered by Barbara McClintock in the 50's (when the DNA structure was not completely understood). She proved that certain

phenomena present in living beings exposed to UV radiation could not be the result of the normal recombination and mutation processes. She found that in corn certain genetic elements occasionally move producing kernels with unusual colours that could not have resulted from crossover or mutation. Transposons were for a long time considered as some sort of abnormality, but in 1983 when she was awarded with the Nobel Prize, many such transposons had been discovered and their possible role in evolution was beginning to be recognised. For instance, the genetic alterations caused by transposons are responsible for the growth of cancers in human or the resistance to antibiotics in bacteria [5], [11].

In order for a transposable element to transpose as a discrete entity it is necessary for its ends to be recognised. So, transposons within a chromosome are flanked by identical or inverse repeated sequences, some of which are actually part of transposon (see Figure 5).

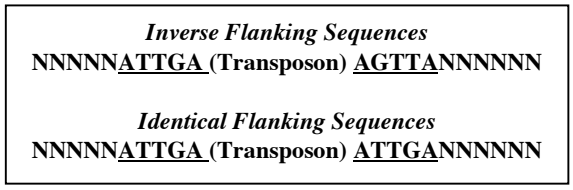

Fig. 5. Inverse or equal flanking sequences

When the transposon moves to another zone of the genome one of the sequences goes with it. The insertion point can be chosen at random, but there are transposons that show a regional preference when inserting into the same gene. Other method can be a correspondence (identical or inverse) in the new position with the flanking sequences. The last method is described in Figure 6.

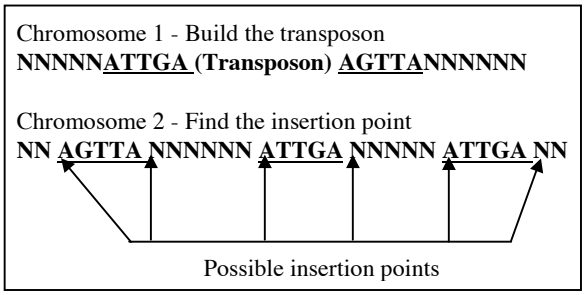

Fig. 6. Building the transposon and finding the insertion point

The sequence in into which the transposon is inserted requires no homology with the transposon. This is in marked contrast to classical recombination, where relatively long sequences of DNA must share homology to permit a recombination event to occur (same cut point(s)). As a consequence, transposition is sometimes referred to as illegitimate recombination.

3.2 Computational Transposition

We implemented the transposition mechanism following the inspiration from biology. After selecting two parents for mating we look for the transposon in one of them. The insertion point will be found in the second parent. The same amount of genetic material is exchanged between the two chromosomes according to the found insertion point.

Now, we are going to describe how the transposon will be formed, how it will move in the genome, how to define the insertion point, how to define the flanking sequences length and how the integration in the new position will take place.

Our case study uses chromosomes of fixed size. Suppose this size is CL (Chromosome Length). The transposition method will work as follow (see Figure 7):

- **FSL** is the length for the flanking sequences;
- Choose at random a gene (gene **T**) between **0** and **CL**, from which we will build the transposon;
- The **FSL** genes immediately before gene **T** will form the first flanking sequence;
- The second flanking sequence can be identical or inverse to the first one;
- Look in the chromosome, from bit **T**, for a possible second flanking sequence;
- The transposon will be formed with all the genes between gene **T** and the last gene of the second flanking sequence;
- The second flanking sequence always moves with the transposon.

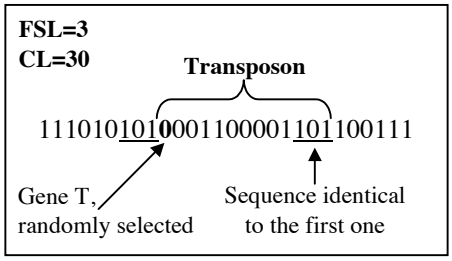

Fig. 7. Computational transposition: building the transposon

We will look in the second parent for a sequence of bits equal or inverse to the flanking sequences. The insertion point will be the first gene after that sequence. After finding the insertion point the same number of genes, equal to the transposon length, will be exchanged between the two parents.

All the process is exemplified in Figure 8.

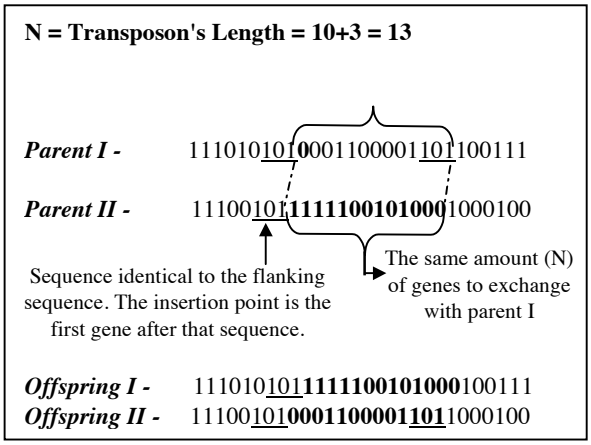

Fig. 8. Computational transposition

Some particular cases can occur:

- The search of the second flanking sequence is made between gene T+1 through the end of the chromosome. If no sequence is found the search starts at the beginning of the chromosome. In a limit situation the sequence found will be the first one. In this case there will be no transposition.
- If there is no equal or inverse sequence in the target chromosome, the insertion point is defined randomly.

4 A Case Study: Transposition versus Crossover

To study the performance of transposition we will compare it with the standard mechanism of crossover. We will use the function [8]:

$$f(x1, x2) = 21.5 + x1.\sin(4\Pi x1) + x2.\sin(20\Pi x2) \quad (1)$$

where $-3.0 \le x1 \le 12.1$ and $4.1 \le x2 \le 5.8$

The optimal solution is 38.87
This function has some interesting aspects as can be seen in Figure 9.

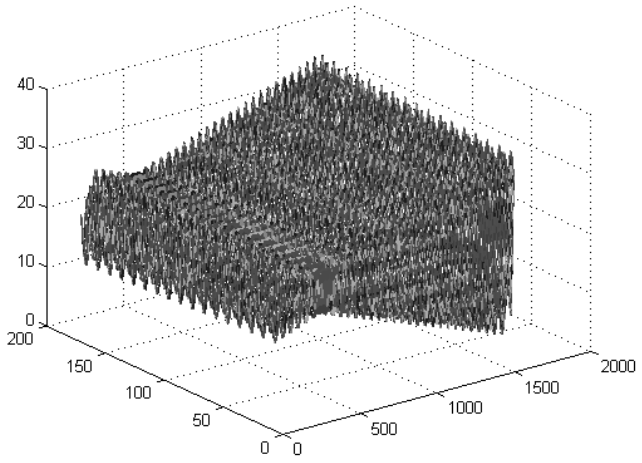

Fig. 9. Michalewicz's test function

x_1 and x_2 domains length are, respectively, 15.1 and 1.7. We assume a four decimal cases precision. So, we need 18 bits to represent x_1 and 15 bits to represent x_2. The chromosome length is 33.
In all experiments we used roulette wheel with elitism as the selection method. The elite size is 20% population size. Mutation rate used is 0.01 and crossover/transposition rate is 0.7. We made experiments involving 1-point, 2-point and uniform crossover. In each crossover type we used population size of 50, 100 and 200 individuals. Transposition was tested with flanking sequences from 1 to 10 bits and, in each case, we used populations of 50, 100 and 200 specimens. We run each experiment 10 times. All the tests were run over 1000 generations. The results analysed are the average of best individual fitness obtained in the ten trials made with each experiment.
First we will analyse transposition results alone, showing how the flanking sequences length can influence the performance of the GA.
Then, we will show the results obtained with transposition, 1-point crossover, 2-point crossover and uniform crossover. We will compare results obtained with transposition and with each one of the crossover methods to get a clear idea of the performance of the genetic operators used.

4.1 Transposition Performance

We analysed the results obtained with the mechanism of transposition using flanking sequences with lengths

from 1 to 10 bits. In each case, we use populations with 50, 100 and 200 individuals.

Observing the average of the results got in the 10 simulations we conclude two main aspects:

1. With larger populations the results are better.
2. In general, with larger flanking sequences the performance of the transposition becomes worst. (see Figure 10).

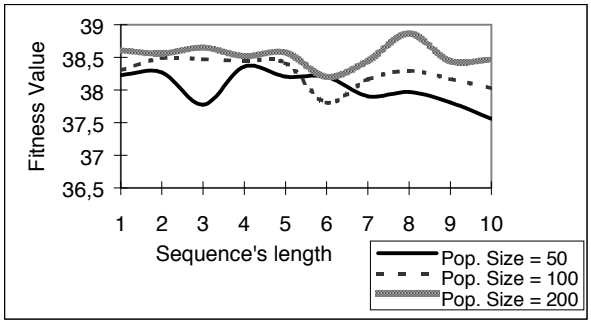

Fig. 10. The GA's performance changing sequence length

The first conclusion seems obvious.

The second one must be justified by the fact that if the flanking sequence length is greater, then the transposon length will be bigger. Thus, in most cases the transposition mechanism won't work because the second flanking sequence is never found. In practice, with bigger sequences the rate of transposition will decrease. Table 2 shows the average results of the transposon size achieved in all simulations executed with the flanking sequences from 1 to 10 bits. We also show the percentage of times that the transposon had the same size of the chromosome, i.e., no transposition occurred.

Table 2. Variation of the transposon length, increasing the size of flanking sequences.

Sequence length	Average of transposon length	Transposon length = 33 (%)
1	3	0%
2	6	6%
3	9	13%
4	15	19%
5	19	27%
6	22	33%
7	25	38%
8	27	49%
9	28	68%
10	29	73%

As we can see that, with larger sequences, the amount of genetic material exchanged is bigger. Besides this, it is harder to find the second flanking sequence so, the percentage of no occurring transposition is very high. This could lead to a loss of the population diversity and, subsequently, to the worst results achieved.

4.2 Transposition versus 1-point Crossover

In all the simulations made with transposition (with populations size of 50, 100, 200 individuals and with flanking sequence length from 1 to 10) the results outperformed the results obtained with 1-point crossover with the same populations size.

An interesting result is that, using transposition, in most cases, with a population of 50 individuals the results were much better than 1-point crossover using 50, 100 or 200 individuals.

Only in the worst results of transposition (sequences length of 9 and 10 bits) this is not true. But in these cases, transposition with a population of 50 individual outperforms 1-point crossover with 50 and 100 individuals. In all the other experiments a population of 50 individual got better results than 1-point crossover with 50, 100 or 200 individuals. Transposition using 100 or 200 individuals, indifferently of the flanking sequence length, is always better than 1-point crossover with 50, 100 and 200 individuals in population.

For instance, in Figure 11, we show the results obtained with transposition (flanking sequence length of 4 bits, 50 individuals) and with 1-point crossover (using 50, 100 and 200 individuals). We can see that with a smaller population, transposition gets much better results. In Figure 12, we present the worst results obtained with transposition using 50 individuals (flanking sequence length = 10). As we can see the results are better than 1-point crossover with 50 and 100 population size.

4.3 Transposition versus 2-point Crossover

The results obtained with transposition and 2-point crossover were very close. We still observe the same characteristics observed with 1-point crossover, i.e. , better results with transposition using smaller populations, but the results are not so obvious. We observed that transposition with a population size of 50 individuals rarely gets the same results of 2-point crossover with 200 individuals, but frequently gets better results than 2-point crossover with 50 or 100 individuals (except when the performance of transposition is worst). With populations size of 100

and 200 specimens, transposition is often better than 2-point crossover with the same population size.

To illustrate these results we show, in Figure 13, the best values obtained with transposition using 50 individuals (sequence length = 4) and, in Figure 14, the worst ones (sequence length =10).

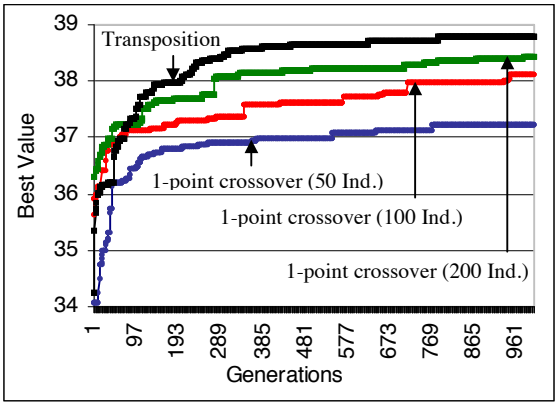

Fig. 11. - Transposition, 50 individuals, sequence length = 4; comparing results with 1-point crossover, 50, 100 and 200 individuals.

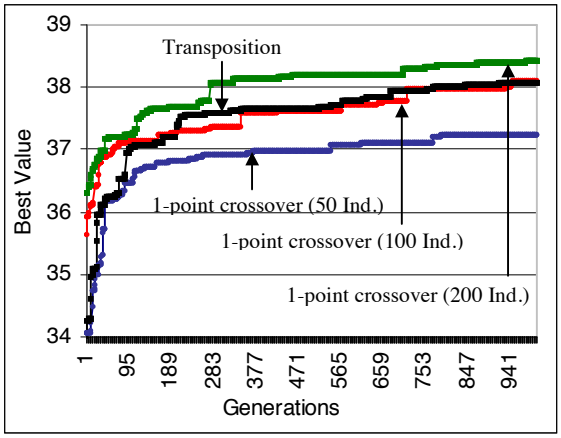

Fig. 12. - Transposition, 50 individuals, sequence length = 10; comparing results with 1-point crossover, 50, 100 and 200 individuals.

4.4 Transposition versus Uniform Crossover

Comparing the results achieved with transposition and uniform crossover we can get the same conclusions. In the best results transposition with smaller populations (50 individuals) exceed uniform crossover using 50, 100 or 200 individuals. If we analyse the worst results for transposition we conclude that with a population size of 50 individuals, the results are better than the ones achieved by uniform crossover with population of 50 and 100 individuals, but not enough to the results got with 200 individuals.

We show these results in Figures 15 and 16.

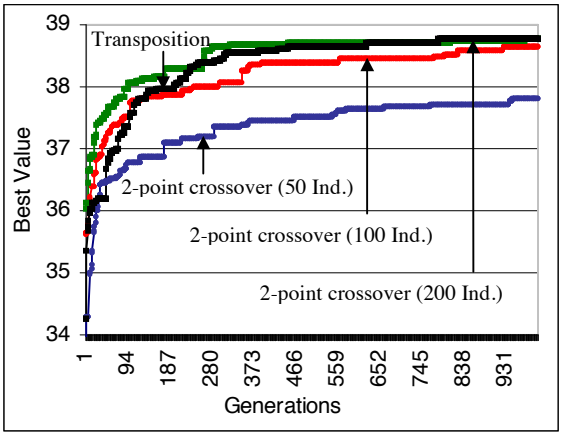

Fig. 13. - Transposition, 50 individuals, sequence length = 4; comparing results with 2-point crossover, 50, 100 and 200 individuals.

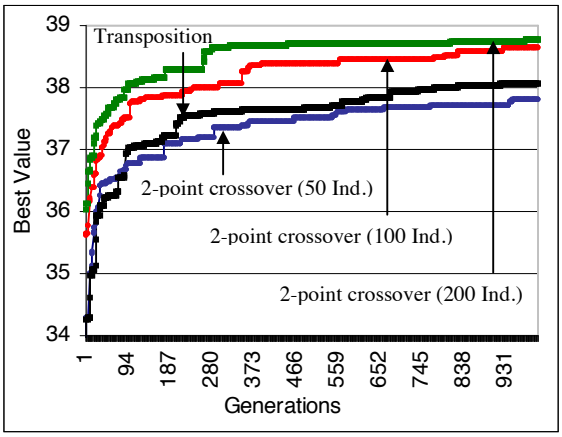

Fig. 14. - Transposition, 50 individuals, sequence length = 10; comparing results with 2-point crossover, 50, 100 and 200 individuals.

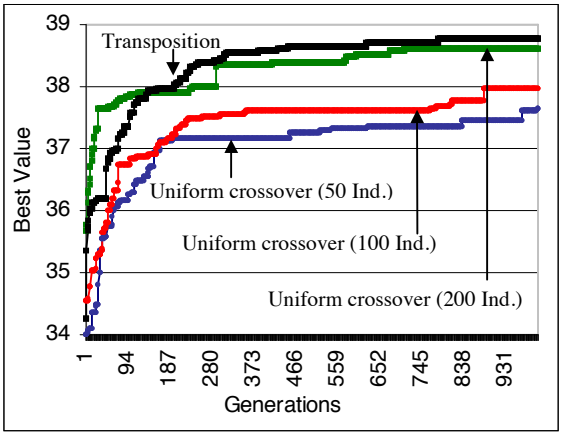

Fig. 15. - Transposition, 50 individuals, sequence length = 4; comparing results with uniform crossover, 50, 100 and 200 individuals.

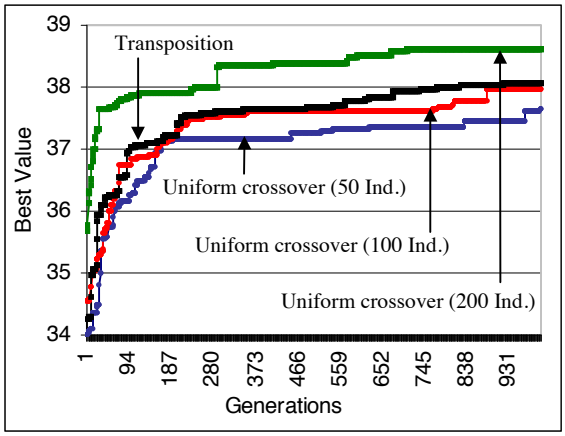

Fig. 16. - Transposition, 50 individuals, sequence length = 10; comparing results with uniform crossover, 50, 100 and 200 individuals.

Transposition using populations of 100 and 200 specimens most of the time overtake uniform crossover results using the same populations size.

If we compare transposition and 1-point crossover performances we can conclude that transposition, with a population size of 50 individuals, exceed 1-point crossover results for populations size of either 50, 100 or 200 individuals in all situations except the one presented in Figure 11.

Comparing transposition with 2-point and uniform crossover performances we observe that, for the same population size, then transposition results always exceed 2-point and uniform crossover results. However, in 2-point crossover the results were much closer to transposition results than in uniform crossover, we can say that, for both cases, if we choose with some care the flanking sequences length we can obtain better results using transposition with smaller populations than 2-point crossover and uniform crossover.

5 Conclusions and Future Work

In this paper we presented a new biological-inspired genetic operator, alternative to the traditional crossover. This genetic operator is called transposition. We used both transposition and crossover (1-point, 2-point and uniform) to solve the same function optimisation problem. We analysed the average results of the best individuals in all the experiments made, changing parameters such as the population size and the flanking sequences length.

We conclude that transposition performance is related with the flanking sequences size: bigger sequences implies worst results due to a loss of diversity. Comparing the results with crossover we saw that transposition is always better than traditional crossover when using the same populations size, and in most cases with smaller populations we can get better results with transposition.

Besides the good results obtained with this new genetic operator we intend to make more empirical work, namely, study the implications of changing other parameters besides population size and sequence length, for instance the crossover and mutation rates. An exhaustive study will be made using other selection methods such as tournament selection and roulette wheel without elitism. We will extend our work to other functions. For instance, some preliminary work already done with the five De Jong Test Bed functions [2], was very promising but need to be completed in order to make well supported conclusions.

Also, we will extend our work to other domains than function optimisation, namely, applications which use based-order genetic operators. It will be important to make the correspondent theoretical justification.

We will also analyse another version of transposition for problems using variable length chromosomes and follow the suggestion made by [1] and see how this operator can be used in genetic programming.

6 Acknowledgements

This work was partially financed by the Portuguese Ministry of Science and Technology under the program Praxis XXI.

7 References

[1] Banzhaf, W., Nordin, P., Keller, R. E., Francone, F. D.: Genetic Programming - An Introduction - On the Automatic Evolution of Computer Programs and its Applications. Morgan Kaufmann 1998.

[2] De Jong, K. A.: Analysis of the Behavior of a Class of Genetic Adaptive Systems. Ph.D. Dissertation, Department of Computer and Communication Science, University of Michigan 1975.

[3] De Jong, K. A., Spears, W. M.: A Formal Analysis of the Role of Multi-Point Crossover in Genetic Algorithms. Annals of Mathematics and Artificial Intelligence, *5 (1)*, 1-26 (1992).

[4] Goldberg, D. E.: Genetic Algorithms in search, optimization and machine learning. Addison-Wesley Publishing Company, Inc 1989.

[5] Gould, J. L., Keeton, W. T.: Biological Science. W. W. Norton & Company 1996.

[6] Harvey, I.: The Microbial Genetic Algorithm. Submitted as a Letter to Evolutionary Computation. MIT Press 1996.

[7] Holland, J. H.: Adaptation in Natural and Artificial Systems: an introductory analysis with applications to biology, control and artificial intelligence. MIT Press 1992.

[8] Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. Springer-Verlag 1994.

[9] Mitchell, M.: An Introduction to Genetic Algorithms, MIT Press 1996.

[10] Mitchell, M., Forrest, S.: Genetic Algorithms and Artificial Life. Artificial Life, *1 (3)*, 267-289 (1994).

[11] Russell, P. J.: Genetics. 5th edition, Addison-Wesley 1998.

[12] Smith, P.: Finding Hard Satisfiability Problems Using Bacterial Conjugation. Presented at the Workshop on Evolutionary Computing, University of Sussex, April 1996.

[13] Smith, P.: Conjugation - A Bacterially Inspired Form of Genetic Recombination. Late Breaking Papers of the First International Conference on Genetic Programming. Stanford University, California, July 1996.

[14] Spears, W. M., De Jong, K. A: An Analysis of Multi-Point Crossover. Proceedings of the First Workshop on Foundations of Genetic Algorithms, pp. 301-315, Morgan Kaufmann 1991.

[15] Spears, W. M.: Adapting Crossover in a Genetic Algorithm. Technical Report AIC-92-025, Washington, DC, Naval Research Laboratory, Navy Centre for Applied Research on Artificial Intelligence, 1992.

[16] Spears, W. M.: Crossover or Mutation?. Proceedings of the Second Workshop on Foundations of Genetic Algorithms, pp. 221-238, Morgan Kaufmann 1993.