On Evolution, Search, Optimization, Genetic Algorithms and Martial Arts

Towards Memetic Algorithms

Pablo Moscato

Caltech Concurrent Computation Program 158-79.
California Institute of Technology.
Pasadena, CA 91125, U.S.A.

ABSTRACT

Short abstract, isn't it?

1 Introduction

Large Numbers

"...the optimal tour displayed (see Figure 6) is the possible unique tour having one arc fixed from among 10^{655} tours that are possible among 318 points and have one arc fixed. Assuming that one could possibly enumerate 10^9 tours per second on a computer it would thus take roughly 10^{639} years of computing to establish the optimality of this tour by exhaustive enumeration."

This quote shows the real difficulty of a combinatorial optimization problem. The huge number of configurations is the primary difficulty when dealing with one of these problems. The quote belongs to M.W Padberg and M. Grotschel, Chap. 9., "Polyhedral computations", from the book *The Traveling Salesman Problem: A Guided tour of Combinatorial Optimization* [124].

It is interesting to compare the number of configurations of *real-world* problems in combinatorial optimization with those large numbers arising in Cosmology. For example, the current estimation of the number of hadrons in the Universe, which undoubtely is a huge number, is *only* 10^{80}. I use the word "only" after its comparison with the 10^{655} tours of a *solved* instance of a traveling salesman problem. The hypothetical run-time of 10^{639} years seems pretty large when we consider that the age of the Universe is approximately 5×10^{17} *seconds*. I should use the phrase *only* 5×10^{17} seconds, but I do not want to touch the susceptibility of those that consider these as "large numbers". The task of the complete enumeration of the 10^{655} tours can dissuade even the most patient Tibetan monks that Arthur C. Clarke would be able to imagine [53].

In this desperate attempt to enumerate all the possible 10^{655} tours we should put our hopes in a kind of "Planck-technology". Such a technology must control events within the Planck length ($10^{-33}cm$) and the Planck time ($10^{-43}sec$). Such an incredible technology must deal with "wormholes" or other kinds of unknown space-time "alterations", or even know how to wisely use them if indeed they exist. As suggested by H. Camblong, this is a subject "beyond science- fiction". He considers as real practical limitations the characteristic times and lengths in the atomic scale [44]. So we should move to *nanotechnologies* [188] [82] [83] [167] pionered by Shoulders and Feynman

[172] [80], or even quantum mechanical [81] or plasma computers [173] as the limit of "human" possibilities.

All these possible technologies would give an improvement of orders of magnitude over present technologies. But combinatorial optimization problems have an exponential growth of possible configurations to be evaluated, so there will be always an upper boundary that will reflect our limitations. Ironically, the development of such technologies would give arise to new and more difficult combinatorial optimization problems, the VLSI layout problem in chip design is such an example.

We should not worry too much about these limitations. This was only a description of the *desperate* approach to the problem, namely *complete enumeration*. Instead, carefully designed techniques have been developed to study some of these optimization problems. For example, Padberg and Rinaldi recently have solved a 532-city *real world* traveling salesman problem [145] and they are now working with problems that involve thousands of cities.

I do not expect to have to do a *desperate* run as complete enumeration. *Real-world* problems challenge us with peculiarities inherent to each one of the problems. So the real need is to create general purpose strategies that can deal with these peculiarities, exploiting common features of these problems. We must be interested in developing effective techniques for doing the search of the optimum in these configuration spaces.

The selection of a good representation

The fact that the number of possible configurations is enormous is only one aspect of the problem and is not the problem *per se*. The problem is that configurations that *look very different* have similar values of the objective function to optimize. There exist many configurations with values of the objective function which are very similar to the global optimum. Throughout this review I will refer to the objective function as the *cost* or *energy* function, when the problem is to minimize it, or as a fitness function when it is to maximize it. A trivial change of sign in the objective function turns a maximization problem into a minimization one, so they are equivalent.

When I said that there are configurations that look very different to the global optimum and have similar values of the objective function, I should have added that they look very different *under a given representation*. We

2

represent a configuration, which means a possible state of the system to be optimized, as a certain mathematical entity. So two of these entities would have few things in common, but similar energy (or fitness) values. So in order to do the search for the global optimum we have a mathematical representation of the states that our system can achieve, and a function associated with each one of them.

In this review, we will be interested in trying to develop iterative search procedures for combinatorial optimization problems. The purpose is to establish general procedures that can lead to search methods that would be applied to a great variety of problems. Iterative search procedures in combinatorial optimization often use the value of a function they want to optimize in order to find ways to move in the landscape that this function forms in the configuration space. So it is extremely important to decide simultaneously which are the *moves* that we will perform with the type of landscape we are dealing with. With the term "moves" I mean which transformation turns one configuration into another. Under a given representation, a move connects two of these mathematical entities.

The value of the objective function is associated to each one of the configurations, but is not associated with each one of the entities that represent them in the representation chosen. I can use this freedom in order to create another function to optimize, with the condition that it should preserve the ranking between optima in the original objective function. The selection of a good function is intrinsically related to the selection of an adequate representation of the configuration space.

The practical problem is *how* to do the search in the minimum CPU time possible. However, there are some problems in which there is no representation possible that can help us to make a search better than the *random search*. One trivial example of such a problem is the search of the *secret name of God*. Suppose that we are given a certain number of symbols, say N, and we are told that a certain permutation of them in a linear sequence, is the secret name of God (recall that this is not the problem in Ref. [53]). There is only one way to know if a given permutation is the correct one and that is by reading it: if correct God will appear. For this problem the fitness function is a flat function with only one correct configuration. Informally, it is known as the *golf-course* problem. There is only one hole in a very large and flat surface making it imposible to find. In the search of the secret name, there is no representation that can help us in this theoretical case (see also

the related discussion and the references in [21] [22]). But in many practical cases, although a better representation can exist, sometimes it is so difficult to find that it is to all intents and purposes impossible to do so.

Who has put so many local minima in *my* optimization problem ?

In combinatorial optimization problems, although discrete, there also appears the concept of *local minima* (or local maxima). Under a given representation, a local minima is a configuration, I should say one of the mathematical entities associated with it, such that all possible moves lead to entities with higher values of the objective function.

Now suppose that somebody asks the question stated above in the title of this subsection. He is working alone in his office... Elemental ! He did. He has chosen a representation and a set of moves, then he has created local minima in *his* problem. And since he is alone in the office, who else can be ?

To clarify this with an example, suppose that we are given the following optimization problem: Find the integer that minimizes the function $F(x) = (x-8)^2$. For the purpose I have in mind, let us suppose that x is constrained to have values within the sixteen integers between 0 and 15. Now, we must select the move. The move can be to add or substract a unity. So being in configuration (integer) i we can reach only configurations (integers) $i+1$ and $i-1$. We can add boundary conditions and connect 0 with 15. The search strategy would be *adaptive walks via best configuration*. So in each configuration, I will check to the *left* and to the *right* and we will move towards the one that has the smaller value of the two. What we are doing is from a configuration, checking all the neighbours connected by the move in this representation, and moving in the direction of the best neighbour found. If from a given configuration all the neighbours accessibles have a higher value of F, we will stop because we found a minimum. No matter which is the initial configuration, we will evolve towards the only minmum $x = 8$. This situation is shown in Figure 1.

Let me now choose a boolean representation. Each integer is now represented with a four-bit string. I will represent 0 with 0000, 1 with 0001, 2 with 0010 and so on until 15 which will be 1111. Let me also choose as a move the usual *one-bit flip*. So from each configuration now we can reach

4

four instead of the two we could before. This fact will not make things easier. To have a picture, we can associate each configuration with the vertex of a four-dimensional hypercube. I will draw an arrow from each configuration to another with a smaller value of $F(x)$. If we use the same search strategy described above, we will find that we now have the possibility of getting stuck in *local minima*. In Figure 2 we can see how the configuration represented as 0111 is a local minimum.

As a conclusion, the choice of representation, search strategy, and moves in the configuration space can create or avoid some *local optima*.

Mapping your problem

This freedom to choose the representation of your combinatorial optimization problem explains the proliferation of "approaches" to them. It is intended that a better representation can give better results. And regarding the previous example, there exists such a representation in some cases. However, it is not an easy task to find the best representation.

As an example, I will mention the Traveling Salesman Problem (TSP). Recently it has received the attention of many researchers who have employed different representations. I will only refer to the *non-orthodox* approaches to this problem. One of them was initiated with Hopfield's idea of using neural networks for combinatorial optimization problems [105]. Although analogous values can be taken by the network, it evolves towards a final state which is a permutation matrix and corresponds to a feasible tour. A *neural elastic-net* has also been proposed as an approach to the TSP [71] [152] although these two representations may actually be considered as only one [174]. A good optimization behaviour and the convergence property to a tour can only be achieved by a careful selection of parameters in the network [194] [30]. Another recent approach uses, as the magnitude to be updated, the probability of selecting a given intercity link as an actual part of a tour in a kind of "learning" scheme [52]. Generally speaking, all are exploiting the possibilities of different representations to avoid *local minima* which, as we have seen, would differ according to the chosen mapping [183] [109].

All these approaches are *physical computations*. They involve the creation of a physical system that makes the search. We can regard this system as *one individual* searching in a space. The original problem was discrete while the actual space being searched is not.

2 Two optimization problems

Most of the results that will be analysed in this review are related with three combinatorial optimization problems. They have attracted a great deal of attention in the past and they also have the advantage of large instances solved to optimality. In general, effective heuristics have been constructed for them and it is important to note how these heuristics can be improved. I will describe some features of two of them, the TSP and the optimization tasks in Kauffman's NK model. A third problem that is also mentioned is the quadratic assignment problem [121] [89] [125] [10] [170]. It arises in the processor mapping in parallel processing [72], data analysis [107], VLSI layout design and location of facilities [101] [192] [102] [140] [159] [103] [25] [169] [100].

The Travelling Salesman Problem

The TSP is an important combinatorial optimization problem due to its *academic* significance and its *real-world* applications. The TSP is generally defined as the task of finding the *cheapest* way of connecting N cities in a closed tour where a cost is associated with each link between cities. One version of it, which belongs to the NP-complete class, the Euclidean TSP in two dimensions, has been one of the most studied optimization problems. In the Euclidean version, the cost of linking two cities is proportional to the Euclidean distance between them. It is interesting since belonging to the NP-complete class of combinatorial optimization problems, it is conjectured that it can not be solved by any polynomial-time algorithm [88].

The TSP is a common test-bed for many optimization algorithms. During the long battle that science has waged with the TSP, only some large instances have been solved to optimality [144] [124] and they give an opportunity to check a new method using those instances to compare the quality of the final ordering of cities. In addition, for a random uniform distribution of N cities over a rectangular area of R units, an asymptotic expected length formula for the optimal tour has been derived [23]. The expected length L_e of the optimal tour is given by

$$L_e(N, R) = K\sqrt{NR} \tag{1}$$

With computational experiments [179], the value K has been bounded by

$$0.765 \leq K \leq 0.765 + \frac{4}{N} \qquad (2)$$

note also the derivation given by Bonomi and Lutton in which they give a value of $K = 0.749$ for large N [26]. Many *ad hoc* TSP algorithms have been constructed during the last 50 years [124] and thus it represents a challenging test-bed for combinatorial optimization algorithms.

The TSP, like some other optimization problems, involves an ordering problem in which N cities must be ordered in a ring-like array. Representing it with strings, the N different elements, the city names, the task consists of finding the string that has the minimum length. To a given string, there are $2N$ equivalent strings that have the same length and are constructed by shifting all the elements or by inverting the order. So an iterative search must deal with a huge configuration space, since the number of tours is $(N-1)!/2$.

The Kauffman NK model

A "tunable" family of correlated landscapes

This model has been used to study adaptive somatic evolution in the immune response [114] [117]. It can also be interpreted as a model of genetic interactions in a multigene system. It is based on an entity (chromosome, haploid genotype) with N parts where each of these parts (genes) can be of A different "types" (alleles). In its simplest version we have only two different "types" 0 and 1. Obviously, the number of such possible entities (genotype) is 2^N. There is a value of an objective function (phenotype, fitness) associated with each one of these genotypes. The association of phenotypes with the values of this objective function is correct but should be understood as a *first-order* approximation since in nature the phenotype understands a dependence with the epigenetic enviroment which is not considered now. The model requires a "reflexivity" condition; if gene i depends on gene j, then j depends on i. K genes are assigned to each i, so for each gene i there are 2^{K+1} combinations of alleles 0 or 1 and for each one of these combinations a "fitness" (objective function) contribution is assigned, a real value between 0 and 1. These interdependencies between alleles are called *epistatic interactions*. So for a gene i, a different contribution is assigned according each of the different combinations of its associated genes. This procedure is followed

for each of the N genes, so the fitness of a given genotype is found out by summing the fitness contribution of each one of the genes and normalizing the sum dividing by N. The NK model has also been described in Ref. [115] and a more complete description can be found in Ref. [116]. Regarded as an optimization problem, the task consists in finding the configuration of alleles that has the biggest fitness.

The two extreme cases $K = 0$ and $K = N - 1$

According to the above description, the case $K = 0$ is an N-locus, 2-allele *additive* fitness model. In this case, each part is independent of other parts. In each gene, one of the two possible alleles, 0 or 1 is fittest. In that case there exists a special genotype with the more fit allele at each of the N loci which is the optimum genotype. So there is a connected pathway via fittest variants to the unique optima of the landscape. An adaptive walk via 1-mutant neighbours can reach the single global optimum. The $K = 0$ additive model corresponds to a very smooth, highly correlated landscape.

The case $K = N - 1$ corresponds to a fully random fitness landscape. Each gene is epistatically affected by all the other genes in the genotype. The fitness value of one genotype has no information, about the fitness of its 1-mutant neighbors, so there is no correlation between their values. As a result, the number of local optima is extremely large.

For intermediate values of K, the landscape is correlated in some way. This correlation can be studied using the autocorrelation function of a random walk in the landscape [114]. We can understand K as a parameter that "tunes" the correlation structure of the landscape.

The complexity catastrophes

Kauffman has pointed out an important feature of his model. As the number of genes N increses while $K = N - 1$ *the expected value of the local optima falls towards the mean fitness of the space of genotypes.* This fact combined with the huge number of local minima would make the optimization task a very hard one. When $K = N - 1$ any kind of search cannot be better than random search since the landscape is not correlated at all. However, the above mentioned characteristics, part of the complexity catastrophe, seems to appear also when K is a fraction of N in the limit of N large. We should

not be worried about the complexity catastrophe behaviours if it were only a phenomena of the NK model when $K = N - 1$, but numerical simulations have shown that this complexity catastrophe can also appear when K is a fraction of N (Ref. [116], pp. 557). When K is fixed, the complexity catastrophe is avoided.

In addition to the NK model, there is another combinatorial optimization problem with a similar behaviour in the limit of N large, the quadratic sum assignment problem [125] [41]. It consists of determining the best assignment of N interacting sources, given N sites independently distributed inside a bounded convex region of the Euclidean space. An instance of this problem is defined by a $N \times N$ distance matrix $D = (d_{ij})$ and a matrix $T = (t_{kl})$ of traffics which represents the traffic which is to be routed from source k to source l. So the task is to find the best assignment of sources with sites, having only one source per site, in order to minimize a total cost which is the sum over all traffic elements t_{kl} of the products $(d_{ij}t_{kl})$ where i and j are the location of sources k and l respectively. If S_N is the set of all permutations of $(1, 2, ..., N)$

$$H(\pi) = \sum_{i,j} d_{ij} t_{\pi(i)\pi(j)} \tag{3}$$

and the task is to find the permutation that minimizes such a cost. R.E. Burkard and U. Fincke have shown that the relative difference between the worst and the best solutions tend to zero with a probability of one when N tends to infinity [43]. Bonomi and Lutton have applied the statistical mechanics formalism and confirmed the result via computer simulations [27]. Burkard and Fincke have also addressed bottleneck problems [42]. However they differ in the fact that while in the NK model the task of finding the best genotype still makes sense, in the quadratic sum assignment problem, the task itself vanishes.

In the NK model, the complexity catastrophe has two origins, the *conflicting constraints* between genes and the normalization used for the fitness, mathematically combined with the central limit theorem. It is an open question which are the optimization problems that present this or other kind of complexity catastrophes for large N and to classify the different behaviours (see also Ref. [158]). Kauffman simulations with the NK model and the fact that the complexity catastrophe seems to still be present when K is a fraction of N should be important to consider as a warning when dealing

with an optimization problem and can also guide in the selection of a better representation.

3 Population approaches

This review is mainly concerned with the population approaches to combinatorial optimization problems. I understand them as those strategies that are based in more than one individual to do the search. The search is usually provided by an iterative improvement heuristic. I will analyse the characteristics of some new population approaches, its performances and I will show how some of them are outperforming present heuristics.

Sometimes we are faced with an optimization problem in which a near optimum output of a certain algorithm is enough to satisfy our needs. Refering to the TSP, if we want to construct a tour for a candidate on an election trip, we do not care very much if our tour is three or four percent above the optimum, but if we are faced with a mail problem in which the points to be connected are always the same and they are visited daily; in that case, we would like to have better tours to solve that constant drain of resources. For these cases, another method would be suitable.

It is my purpose to show how such a method can be constructed using heuristics and a population approach. I will show some research projects that, as far as I know, have been developed independently. Common features will be analysed and this would lead to a better comprehension of how they can outperform present techniques.

To simulate a population takes more computer time and more effort in writing a code than to use one individual. This is true in a sequential computer. In a parallel computer, things are different. We are often confronted with the *parallelization* of intrinsically sequential algorithms [2] [37] [86] [9]. So a population approach which is *intrinsically parallel* seems the natural choice. Ultimately, this dilemma is a version of the compromise between quality of the solution obtained versus the CPU time used. This compromise appears in parallel or sequential computers. It can also depend on the particular parallel computer used since we have to regard the memory requirements per node, architecture type, etc. After the initial discussion about the large numbers involved in the search, it would be wise to ask ourselves: "Is there any difference in doing a search with only one individual or with one hundred

?" We can also ask ourselves, as H. Mühlenbein did, **"How can a very small population search such a huge space in an efficient way?"** [138]. I will show some examples that will clarify these questions. We will see that there exists a difference and it favours a population approach if we are interested in the quality of the final solution. I will also show how Mühlenbein and other researchers have parts of the answers to the second question. First, I should briefly introduce one of the most popular *one-individual general-purpose* optimization techniques, Simulated Annealing; and one of the most popular *multi-individual*, Genetic Algorithms.

Simulated Annealing

Simulated Annealing (SA) is a technique for global optimization borrowed from Statistical Mechanics. It has proved to be suitable in a wide variety of problems such as the min-cut partitioning problem [119], global wiring [191], least square fitting of many unknowns [190], image analysis [88], the problem of finding an efficient manner to execute parallel computation [85] and the *alma mater* spin glass ground states localization and evaluation.

Kirkpatrick *et al.* [119] have developed Simulated Annealing (SA) which is based in a random walk on configuration space. Starting in an initial state $\mathbf{x_0}$, we pick up another state from the neighbourhood of $\mathbf{x_0}$ and we compute the quantity $\Delta C = C(\mathbf{x_1}) - C(\mathbf{x_0})$. Then if $\Delta C > 0$, it would be accepted according to the Boltzmann factor $e^{-\Delta C/T}$ where T is an external control parameter interpreted as a temperature, and C is the cost function of the problem. If $\Delta C \leq 0$ the new state is always accepted. This is an iterative stochastic procedure, and at a low temperature is expected that it will jump between the low-cost (low-energy) configurations. The temperature T is not fixed during this process. It was found that starting with a high initial value T and cooling slowly, significant improvements can be achieved. If the initial temperature is decreased abruptly, many "frozen" imperfections remain in the final state, a situation that recalls the idea of being trapped in a local minima. One dilemma appears here, because although a MC method will always find the lowest energy for a finite system, the computer time involved would be prohibitive. The cooling schedule is also an important question and theoretical criterions for fixing it are now being developed [1] [3] [164] [165] [141] [128].

Genetic Algorithms

Genetic Algorithms (GA) [90] [104] is a population based approach for searching that is based on some biological mechanisms for generating more fit individuals. This evolutionary approach have also been considered by G.E.P. Box [31], G.J. Friedman [84], W.W. Bledsoe [24], H.J. Bremermann [38] [39], L.J. Fogel [75] [76] [77] [78], D.B. Fogel [79], I. Rechenberg [156] [157], H. Schwefel [168], K.A. Dewdney [68] [69], and R.M. Brady [36].

In general, a GA is composed of three different operators: Reproduction, Crossover and Mutation. Usually, it underlies a string representation of individuals where generally this codes the parameter set, not the parameters themselves. It uses probabilistic rules to search. For example during reproduction, the population is copied according to the values of the objective function we are optimizing. It tries to mimic natural selection. Crossover is the mechanism by which two individuals interchange information. This is done generally by the creation of a new individual by taking parts of the two regarded as parents. I will discuss this mechanism later. Mutation is the alteration of one part of an individual and in GA it is regarded as the mechanism that generates the necessary amount of noise in the search and the population will evolve under the force of selection.

Parameters

It is beyond the scope of this review to introduce both techniques in more detail. For an introduction to GA refer to Refs. [90] [61] [104]. SA has been also reviewed in Ref. [61]. Applications of SA and GA can be found almost everywhere. Parallel implementations of SA can be found in Refs. [2] [86] [37].

Both techniques have to deal with the adjustment of parameters. In SA, the cooling schedule is critical. In GA, the problem is the population size and mutation rate [166]. It would be interesting to have a strategy that would not need such a tuning of parameters or that it would improve monotonically as the number of individuals increase (or use more temperature steps in SA). At present, when we have an optimization problem, if we want to use SA or GA we have to solve *two* optimization problems, the optimal value of our parameters and the problem itself [141] [128] [92].

I will postpone until the final discussion some of the characteristics of SA

and GA after the new strategies have been introduced. Now I will introduce some analogies that, like the analogy with biological evolution in GA, would be useful to describe the new techniques I want to discuss.

4 Life and Science as a result of an evolutionary process

Since the trip that Darwin made around the world and the subsequent publication of his book *"On the Origin of Species"* [59], we have been enlightened with a theory, which later would turn into the corner-stone of modern biology and set the basis to understand the diversities and similarities of all life forms. He was condensing in simple terms a complex pattern of different observations he made, and finally he succeded in the task of searching for a simple rule that would fit the data he observed on his expeditions. At the same time he was addressing how adaptation and selection would yield better individuals. The genes make use of a population based approach to search in the space of possible biological organisms, trying to create the best *coadapted set of genes* and as a result *"survive"*. Is interesting to remark, that inside Darwin's brain a similar process would have been taking place. Surely, as any theorist would, he constructed *pre-theories* during his trip, tested them against the data he collected, and questioned some parts to observe the effects. Then, common features that explain many observations are preserved and the others are deleted. So he was dealing with a *population of ideas*, mental individuals that help to perform the search of the best explanation to understand the data.

These concepts are not new. P.W. Anderson in his comments on Francis Crick's autobiography quotes him saying: **"Do not be afraid of making mistakes, he says: No single idea, no matter how brilliant, is going to solve a hard problem; persistence is all; evolution seldom chooses the elegant solution. He stresses that "professionals know that they have to produce theory after theory before they hit the jackpot.""** [7].

We know intuitively how this mental process works or, at least, we have learned it by our own experience. We know how a rigid assumption that pretends to explain everything can be lethal when confronted with data not

considered before or misunderstood. A *multiscale* approach is then generaly used, trying to cluster the data that can be explained with a similar pre-theory, a statement which is not completely developed and requires modifications to reach its final form. It should also be entirely disregarded, unified with another by creation of other theory that involves these as subcases or merged with other pre-theories. Most of these processes should take place unconciously and should lead, in the *"happy-ending"* cases, to a *Eureka* situation when all is clear and, in some cases, the final formulation seems obvious. What all this mental machinery is doing is to help us to deal with the complex data because the information given by the data is not hierarchically ordered according its relevance.

A different kind of reasoning is the one employed by the aviation accident experts when a plane crashes. The previous knowledge of the possible causes that can lead to specific kinds of accidents, make their work easier. In a certain way, they have classified causes with effects, and because they have made that before, the reasoning can be more *"sequential"*. If that would have been the case in the hypothetical construction of an explanation of a given phenomena, a reasoning strategy like that of the *Twenty Questions* game would be indicated. We would test the pre-theory, or many of them, with the data, represented by questions hierarchically organized in a tree according its relevance. The pre-theory must satisfy sequentially the tests, giving a *Yes-No* output if it explains or not. The first time a "No" is found, others options from different branches at the same level in the tree must be considered. In the general case, even in cases with complete information to describe the phenomena, this order is not known.

Only few scientists, like Newton, Maxwell, Darwin or Einstein, have the rare privilege, and the rare ability, of creating a unified theory and to give order to the chaos. In Science, the whole group of scientists, work as a large, parallel, decision machinery that try to deal with the data. Each one is not trying to find the ultimate explanation to the subject under study. Instead, they are contribuiting, in different magnitudes, solving sub-problems and making logical conections between data.

A comparison with the NP-completeness theory

A clear example of this process is the theory of NP-completeness and we mention it not to talk about the unification of the fundamental laws of physics

which has been widely regarded as an example. The computer scientists and mathematicians that study the theory of NP-completeness prove theorems and link hypotheses in a logical way. At the center of this collection of data or collection of theorems, there is a central assumption; that NP $\neq$ P. This conjecture is not proved, but it does not interfere with the evolution of the theory. Many results are proved regarding this conjecture as true and others are proved assuming it false. If at a certain moment somebody would be able to show the relation between the classes NP and P, this will make useless, at first glance, many previous results, but the knowledge derived from having made the wrong assumption and the techniques and mathematical tools developed, will be preserved in the whole theory of computational complexity. It is here where the diversity of points of view enriches the subject.

What science shares with biological evolution: A population based, competitive and cooperative task

The advance of science works with some of the same principles that mark biological evolution. A big group of individuals regularly submit their ideas to the consideration of other individuals, their results are analysed and then this triggers new works. This process is governed by the forces of Competition and Cooperation. The latter can take the form of *team-work* or usually is restricted to regular interchanges of information.

In Biology, recombination is viewed as a mechanism responsible of the flow of genetic information from one generation to the next one. Biological evolution can also be regarded as a population based approach to optimize the genetic code, based on the survival of the individual [110] [111] [189] [142]. It is made via the optimization of the *code* of living organisms, the DNA molecule. This optimization is made by rearrangements of the sequence of nucleotides which constitutes the code. It seems that here the analogy also works since a life form tries to maximize a certain, complicated and in a certain way unknown *fitness function*. This life form does not compute the value of this function but it is felt as the task of surviving against aggressive external factors, biological or not.

5 Towards Memetic Algorithms

When I said that biological evolution is based on the survival of the individual, I was using these words in a kind of *double-talk*. What I should say is that biological evolution, due to the fact that only those individuals that survive can reproduce, goes in the direction of optimizing the genetic code. I am indebted to Dr. Scott John, who after reading an early draft of this paper, pointed out to me that many of the ideas in this discussion between scientific and biological evolution are very similar with those stated by Richard Dawkins in his book *"The Selfish Gene"* [60]. Dawkins also recognizes that these analogies have been investigated before by Sir Karl Popper [153], L.L. Cavalli-Sforza [45], and more recently in Refs. [46] [48] [47] [49] [50], F.T. Cloak [54] and J.M. Cullen. This subject was also studied by R. Boyd and P.J. Richerson [32] [33]. Present work of these authors is related with the interactions between the individuals within the group [98] [34]. I will not enter into details but note the analogies between cultural and genetic evolution. There are many and the talented scientists referenced above would show them better than I. Instead, I will concentrate in the differences between them.

Life deals with the combinatorial optimization problem of survival by coding the information in the form of a linear structure and performing point mutational operations like the substitution, insertion or deletion of nucleotides in the DNA or RNA. Other rearrangements of the structure are chromosomal mutations like the deletions, inversions, duplications, transpositions, translocations, conversions or even the genetic recombination mechanism in sexual organisms.

Due to the way that nature decides to do the search in the combinatorial optimization problem of finding the best genetic codes, many approaches have been developed trying to mimic biological evolution for optimization. In particular, Genetic Algorithms have been quite succesful when applied in many different frameworks. It is a population based approach that is based in three fundamental operations; Reproduction, Crossover and Mutation.

The evolution of martial arts

While biological evolution is a good example of a self-organizing process, there are others that should also be regarded as possible candidates from

which we can learn as examples of complex adaptive systems and apply to combinatorial optimization. One of them is the evolution of martial arts. In particular we can consider the chinese Kung-Fu, that has evolved in less than four thousands years. We will only study here the combat aspects of it and the way information is preserved. Studies on the human behaviour have shown that, as other primates, humans tend to fight using a very disordered sequence of movements. On the contrary, the movements of a Kung-Fu master are an extraordinary combination of simplicity and effectiveness. Its actual degree of development and the fact that it did not suffer from variations that could perturb it is a direct consequence of the *"representation"* which was used for its evolution. To my knowledge, all martial arts have exploited the ability of the brain to remember via sequences, so the basic knowledge is transmited by learning a set of selected sequence of movements called *forms*. The form, like a chromosome, is not an indivisible entity. It is composed of a sequence of defensive and aggressive sub-units which can also be divided, a pattern that resembles the structure of chromosomes, genes and alleles.

But within the form there are some movements which can be understood as an indivisible unit, and these are the ones that are really important. The whole is a support to let the brain transform them as reflexes that can be automatically triggered in real combat. The individuals can compute their *fitness* function by the evaluation of their performance in the execution of the movements of the forms and with some tournaments where they compete. It is interesting to pursue the analogy and to see how the information improve between generations. It is very important to remark that not all the individuals can teach. Only those that have the greatest values of fitness i.e. *black-belts* can have that right. This equates to the mating processes in GA which select with bigger probabilities those individuals with the best fitness.

The concept of the meme

R. Dawkins in the last chapter of his book *"The Selfish Gene"*, has introduced the word *meme* to denote the idea of *a unit of imitation in cultural transmission* which in some aspects is analogous to the gene [60]. In the case of martial arts, those undecomposable movements in the form that I mentioned above should be considered as *memes*. A defensive movement generally is composed by the coordinated action of many of these *memes*.

We can understand the martial arts in the context of the evolution of a *coadapted set of memes*. For Dawkins, examples of *memes* are: **"tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches"**. But later he adds: **"So far I have talked of memes as though it was obvious what a single unit-meme consisted of. But of course, this is far from obvious. I have said that a tune is one meme, but what about a symphony: how many memes is that ? Is each movement one meme, each recognizable phrase of melody, each bar, each chord, or what?"** In comparison with music, I believe that the martial arts are one of the best examples of the meme concept. It also has a linear representation to code information, analogous to the genetic case, which can help somebody to understand it better although the concept of a *meme* is not limited by the representation. A way of punching with the fist can be one meme. Each finger has to be in a given, fixed position. In the context of an aggressive sequence of movements, it should be used with only some of the other memes. For example, in the Kung-Fu case only some fist positions have sense with some arm movements, giving sense to the coadaptation of movements as noted above.

But the analogies with the genetic coding and natural selection can not include the mutations. In GA they are considered to be the operator that includes the necessary amount of noise to do *hill-climbing*, while it is very improbable to find a *good* improvement in martial arts as a consequence of the introduction of some *random* movements into a form. The process seems to be different. Only the masters have the sufficient knowledge that permits them create a new movement and to incorporate it to the form. And this happens with a very low frequency. So, there is much *problem specific knowledge* that is applied to each modification. Almost all modifications give improvements rather than create a disorder. This *fast-feedback* flow of information from *high order* phenotype knowledge to genotype level, seems to have differences with the processes of biological evolution, but we must consider the latter is constrained with the physical structure of the DNA and their processes are direct consequence of the primitive replicating macromolecules that gave its origin.

I believe that the analogy of cultural and genetic evolution breaks down in the *copying-fidelity* aspects of them in addition with mutation. And that these break-down points are the reasons of the tremendous speed-up observed in cultural evolution. I will quote again Dawkins when he says: **"This**

brings me to the third general quality of successful replicators: copying-fidelity. Here I must admit that I am on shaky ground. At first sight it looks as if memes are not high-fidelity replicators at all. Every time a scientist hears an idea and passes it on to somebody else, he is likely to change it somewhat. I have made no secret of my debt in this book to the ideas of R.L. Trivers. Yet I have not repeated them in his own words. I have twisted them round for my own purposes, changing the emphasis, blending them with ideas of my own and of other people. The memes are being passed on to you in altered form. This looks quite unlike the particulate, all-or-none quality of gene transmission. It looks as though meme transmission is subject to continuous mutation and also to blending. It is possible that this appearance of non-particulateness is illusory, and that the analogy with genes does not break down". I find it difficult in talking about *blending* of memes when we have defined them as a a *unit of imitation in cultural transmission.* At least I would say that the meme is a structure with *internal consistency.* If I regard what should be a good scientific idea to explain a phenomena, at least it should have no contradictory statements. This is a point where the breaking of the chains of copying-fidelity, combined with the freedom of blending concepts gave the new meme the necessary degree of refinement to even improve the previous one. The raw combination of good ideas is not always a good idea. A scientist does not pass on an idea after blending it with his own without checking the logic of what he is saying or his reputation would be in trouble. Although there are some exceptions, science does not improve by random errors.

The Memetic algorithm

While Genetic Algorithms have been inspired in trying to emulate biological evolution, Memetic Algorithms (MA) would try to mimic cultural evolution. They are a step further in the direction pointed by Kauffman when he recognizes the importance of correlated landscapes in the success of population based approaches for optimization [113].

Memetic algorithms is a marriage between a population-based global search and the heuristic local search made by each of the individuals. The GA community would like to say that MA are only a special kind of GA with

local hill-climbing. Goldberg in his book about GA, has called some similar variations of GA more close to a MA **hybrid genetic algorithms** [90].

Given a representation of an optimization problem, a certain number of individuals are created. The state of these individuals can be randomly chosen or according to a certain initialization procedure. An heuristic can be chosen to initialize the population. After that, each individual makes local search. The mechanism to do local search can be to reach a local optima or to improve (regarding the objective cost function) up to a predetermined level. After that, when the individual has reached a certain development, it interacts with the other members of the population. The interaction can be a competitive or a cooperative one. The competition can be similar to the one which will be described in the Competitive and Cooperative method (CCA) or can be similar to the selection processes of GA. The cooperative behaviour can be understood as the mechanisms of crossover in GA or other types of breeding that result in the creation of a new individual. More generally, we must understand cooperation as an interchange of information. The local search and cooperation (mating, interchange of information) or competition (selection of better individuals) are repeated until a stopping criterion is satisfied. Usually it should involve a *measure of diversity* within the population.

The above description is somewhat general, but it must be that way. For example, I am not constraining a MA to a genetic representation. While a genetic, or a zero-one representation would be useful under certain circumstances, for some problems they are not the best representations. Sometimes they are useful for proving theorems but, if we are interested in a good optimization algorithm, we must use those that naturally belong to the problem. If I am solving a problem with an intrinsic two-dimensional structure, I do not see any reason for not using a *two-dimensional gene* if I want to use a GA. Dawkins says **"I am an enthusiastic Darwinian, but I think Darwinism is too big a theory to be confined to the narrow context of the gene"**. I have the same impression regarding GA or MA to be confined to have only genetic representations.

We can also draw a border with GA, saying that a GA understands a "genetic" (linear) representation and that the individuals do not make local search. However, my impression is that the only clear separation is the local search, which was considered the **hybrid** characteristic for the eyes of the GA community [90].

6 A Competitive-Cooperative Approach to Complex Combinatorial Search

The central idea of the cooperative-competitive approach for searching in large configuration spaces is to use collective properties of a group of distinguishable individuals, which are separately performing the search. As a result of the collective behaviour of the population, solutions are generated which are better than those which would be obtained by each individual without interactions within the group. We also expect these solutions to be better than the *brute force* approach of doing many independent runs and to pick up the best result of them.

In our approach, the individuals are arranged on a ring and each one *searches* locally, *competes* with its two immediate neighbours in the ring, and *cooperates* with individuals which are very distant within the ring. The arrangement introduces a different neighbourhood for cooperation and competition.

We can see that, for the sixteen element ring shown, an individual competes with its nearest neighbours in the ring, and cooperates with individuals that are four links away in the ring.

The local search is supplied by Monte Carlo simulated annealing [119]. The cooperative aspect is supplied by a crossover operator identical in form to that used in GA [104] as applied to the TSP by Grefenstette [94]. The competitive aspect is supplied by a procedure where individuals subsume each other's positions according to their relative fitnesses. The acceptance of the changes involved in *all three* components of the search is governed by temperature, as described below, and this value is subject to a cooling schedule.

Local Search

One step of the local Monte Carlo search process for a tour of N cities can be understood as N attempted rearrangements of the tour. The moves used for rearrangement are of three different types: the inversion of a sub-tour, the insertion of one city in a different part of the tour, or the insertion of two connected cities in another part of the tour. The first move changes two links and the latter two moves both change three links.

The cities and the places of insertion are selected with a random uniform distribution among all possible values and all processors have the possibility of doing all rearrangements, so we have a stochastic Markovian process. Each of the changes is accepted with a probability

$$p(\Delta E_{MC}, T) = \frac{1}{1 + e^{\Delta E_{MC}/T}} \qquad (4)$$

where

$$\Delta E_{MC} = \frac{\Delta L}{\sqrt{N}} \qquad (5)$$

and ΔL is the change in length produced by the rearrangement.

Competition

Competition occurs between individuals on the basis of a challenge by an individual currently residing in one location on the ring to an individual in another location. In a given competition phase all individuals both challenge and are challenged, and so are involved in two interactions with neighbours. The competition procedure can be clarified with an example. In the ring shown above, the tour in location 0 would compete with that in location 1 by an *issued* challenge, and with that in location 15 by a *received* challenge. If the challenge to location 1 is successful then the tour in location 1 is removed and it is replaced with a *clone*, an exact copy, of tour 0. Clearly tour 0 can itself be replaced by tour 15. The *battle* is decided according to the probability

$$p(\Delta E_{comp}, T) = \frac{1}{1 + e^{\Delta E_{comp}/T}} \qquad (6)$$

If each tour is of length L_i. where the sub-index i stands for the sequence number of the location of the tour around the ring, for the competition between 0 and 1 we compute ΔE_{comp} according to

$$\Delta E_{comp}(0, 1) = \frac{\Delta L_{0,1}}{N} = \frac{L_0 - L_1}{N} \qquad (7)$$

so if tour 0 challenges tour 1, we generate a random number q with uniform distribution in the interval $[0, 1]$ and if $q \geq p(\Delta E_{comp}(0, 1), T)$ nothing hap-

22

pens but if $q < p(\Delta E_{comp}(0,1), T)$ tour 1 is deleted and replaced with a copy of tour 0.

Cooperation

The cooperation procedure is based upon the crossover operator of genetic algorithms. As a result, components of configurations are exchanged, allowing the combination of subcomponents of successful individual searches into configurations that may develop to be better than either of their generating configurations.

The operator used is that defined as the *order crossover* or OX operator by Goldberg [90]. Of the two configurations to be combined, an arbitrary subtour is chosen from one tour, and inserted into a second. In order that generated tour should obey the constraint that each city is visited exactly once, the cities that are inserted are excised from their original locations in the second tour, and those cities that were connected on each side of them are re-connected to each other. The result bears a structural relationship to both parents, although the excision of cities means that achieving the subtour often makes significant changes to the tour into which it is inserted.

In contrast to the two and three link changing operations of the Monte Carlo procedure, the number of links in the second tour which change during crossover may be any value, up to the number of links it contains.

Cooperation occurs on a similar basis to competition in that a challenge, which may be considered in this case as a proposition, is issued between neighbours in the locality defined for cooperation. A proposition is assessed by the same criteria as a challenge, scaled with temperature in the same way. If the proposition is accepted, crossover is performed between the tours and the result replaces the recipient of the proposition. The length of the *result* of crossover is not used to determine its acceptability.

The Optimisation Schedule

Competition, cooperation and local search are interspersed so that a period of local search is followed by a competition phase, another period of local optimisation, a cooperation phase, and then back to local optimisation. The first period of local optimisation consists of many Monte Carlo steps. Later periods consist of a single step. As in some implementations of simulated

annealing, the temperature is initially set to a value where 40 percent of the rearrangements with $\Delta L > 0$ are accepted and reduced using the standard geometrical schedule $T_n = 0.98 T_{n-1}$ on the completion of each Monte Carlo step.

The optimisation is judged to be completed if the diversity of the group falls to a low value, usually zero. To be more specific, samples of the connections of 128 random cities are made in random pairs of individuals within the group. If, for all such pairs, the selected city is connected to the same two other cities then the group is judged to have reached a solution. The sampling is implemented in a MIMD machine by considering only pairs of cities either competing or cooperating, and by globally asynchronously monitoring the diversity of pairs of tours.

The advantage of the cyclical sequence of phases outlined above are first that the results of cooperation do not compete until they have undergone local optimisation to ameliorate the damage caused by the OX operator, and second that if both an individual and its clone are victors of competition, they are allowed to optimise along separate paths before their components are propagated in cooperation.

This said, there is no real reason, apart from simplicity of implementation, that the phases should run synchronously in all localities. Indeed, the method requires only occasional communication between individuals during cooperation and competition and so is not likely to suffer the performance penalties usually associated with message-passing in an asynchronous environment.

The moves used for the individual search

The moves we used in the Monte Carlo procedure are of three different types. We have seen how the moves in the configuration space define which configurations can be reached from a given one. A move generates a graph, each of its vertices is one configuration and there is an edge if one configuration can be transformed into the other by the application of the move. This creates a notion of distance in the graph and in connection with the definition of a energy or fitness function, it gives the concept of local minima.

If we have a local search with only one type of move, the introduction of a different move must be motivated not only by its effectiveness but due to the *innovation* that it would give. To develop this concept, we will return

to our image of a graph in which vertices represent configurations and they are connected with an edge if there exist a move that can transform one into the other. Within this picture, if we want to introduce a new move to complement another, we should also try to find one move, say move B, such that its generated graph has a minimum overlap with the graph of the other move. In that way we would guarantee the existence of *channels between the walls* of the valleys that contain local minima, so generating more possibilities for easily leaving them. So given a move of type A we should try to construct move B that makes near two configurations that are far by application of move A and we are also opening new channels to those local minima by only application of move B.

Is this the ultimate solution for the problems that the search involves ? Is it wise to use a set of many different moves, to continue adding different moves *ad infinitum* ? Certainly not. Effective moves are those that, on the average, create a new configuration with similar values of the objective value, reflecting the efficient use of the correlation between the configurations given by the representation. In the TSP, this requirement is satisfied by moves that involve the deletion and creation of few intercity links. Even with three different moves we are still faced with local minima. An example of such a situation is given in Figure 3.

It is clear from the figure that no move that involves the deletion and creation of few links can help to avoid this kind of situation. Figure 3 results from a real simulation with the Competitive and Cooperative approach using the three moves described and performing a uniform decrement of the temperature according with a geometric schedule.

The moves used have been chosen due to its reported efficiency in the TSP among the literature of iterative edge-exchange heuristics [57] [126] [127] [143] [124]. I would like to remark again that a MA does not need to start from scratch in a given optimization problem. Usually there are good iterative improvements procedures which can be used to do the local search and to reach local optima. The interactions between individuals, as in the GA case, may involve the design of a crossover or recombination operator. This can be designed with the purpose of interchanging information from parents and trying to preserve the information adquired. For example, in the TSP and other permutation problems, the relative order is an important feature of both parents and it must be preserved in the offspring. There is a need of establishing some rules to design recombination operators.

Research should concentrate on this point to move the creation of recombination operators from the present state of *art* [91] to a more rational design. The possibility of using a correlation function to study recombination will be addressed in the final discussion.

Simulation results

The performance of the CCA was tested with instances of the TSP [135]. We have studied a random distribution of 100 cities in square and the 318-cities Lin-Kerninghan problem [127]. Using the CCA as described above, good performance was observed when the number of individuals doing the search was of the same magnitude as the number of cities. The resemblance of the final solutions in the 100 cities case (see Figures 4.a-d) and the similarity of the tour found with the optimal one in the 318-cities case shows the effectivness of the approach (Figures 5 and 6). Other artificially created problems were studied. Figure 7 shows a 100-cities TSP. For each value of the horizontal component, 10 cities are created, 8 have been asigned random vertical coordinates.

However, there exists a clear need to reduce the number of individuals that perform the search. In a sequential computer, it increases the computer time and in a parallel computer it would be wise to use as many individuals as processors available. Pursing the objective of the reduction of computer time, a deterministic update was incorporated for the local search. While the usual simulated annealing accepts a new configuration with

$$P(\{S'\}) = \begin{cases} \exp(-\Delta E/T) & \text{if } \Delta E > 0 \\ 1 & \text{otherwise} \end{cases}$$

$$(8)$$

the deterministic update is governed by

$$P(\{S'\}) = \begin{cases} 0 & \text{if } \Delta E > T \\ 1 & \text{otherwise} \end{cases}$$

$$(9)$$

where T is a parameter that can be considered as playing the same role as the "temperature" of a usual simulated annealing. The parameter T is decreased

as in simulated annealing. A similar procedure was used previously by Dueck and Scheuer [70]. The acceptance probabilities for positive increments is shown in Figure 8.

As a result of the incorporation of the deterministic update, it was found that similar results can be obtained with 16 individuals in the 100 random cities problem. Previous results reported in Ref. [135] using the sigmoid function as the accepting probability, need 128 individuals to have the same quality of the final solution. In the 318 cities problem, the use of the deterministic update gave final solutions two percent longer than the optimum using only 16 or 32 individuals. The result reported in Ref. [135] was 1.2 percent above the optimal tour, but it need more individuals than the number of cities.

These results and the knowledge of deterministic algorithms that perform better than SA annealing for rapid cooling [95] [96], suggested that the deterministic update would deserve more investigation. With J.F. Fontanari, we have studied this way of updating, checking its performance in both the TSP and quadratic assignment problem, outside the framework of the CCA [134]. In this comparison, we used only one individual, the usual procedure in SA. We have found that the deterministic update has no statistical advantage and that it is equivalent to the usual SA when the number of attempted rearrangements at each given temperature is of the same order than the number of neighbours from a given configuration. However, better observed results, when the deterministic update was used in the local search procedure of the CCA, would indicate that it should be better than the stochastic update only when the number of attempted rearrangements is small.

Reheating ?

In the CCA, the information about the structure of a good tour is shared by the population. We can say that information is distributed. So we can perform *reheating* of the system that would improve the current solutions. With the word "reheating" I mean a sudden increment of the temperature. This can be useful to correct defects or to return to the same local minima if no good improvements have been found.

We have applied this procedure of reheating to some instances of TSP. It was implemented as a jump in the value of the temperature, that is a sudden increment of the temperature, triggered by a loss in diversity. Since

27

we compute the diversity between the tours, we used it to trigger the new value of the temperature, that is when the diversity is smaller that 0.03 we reset the temperature to the value it had after 100 or 150 Monte Carlo steps. So when 3 per cent of the intercity links are different, which is a good indication that we are reaching a local minima, the temperature is suddenly increased.

In Figure 9 we can see one of the screens which can show the behaviour of the simulation. The lower-left window presents one of the tours after one of the intervals of local search, that is before a competitive or cooperative interaction takes place. The tour to be shown is selected randomly. In the upper-left window, some parameters give information about the simulation. The solid curve that presents an abrupt decrement is the average tour length. The diversity is also decreasing, then it reaches a near constant value for the interval shown. The oscilating curve is a parameter that shows how *just* are the competitive interactions. It is important the average value over some steps. In the shell to the right, the diversity, temperature are printed. Jump is equal to zero, which means that there was no reheating up to now.

An artificial instance of the TSP was created. Many researchers use perfect square grids to study the behaviour of their algorithms. The reason they invoke is that the solution of the optimal tour is known, and they consider them *simpler* than one of the solved instances like the 318-cities Lin-Kerninghan. However, the arrangement of cities in a square grid has as a consequence, that the optimal tour is degenerate. In order to create an instance that would have a certain degree of simplicity and only one optimal solution, the distribution of 210 cities shown in Figure 10 was used. Figure 10 shows one tour output of a CCA run with 16 individuals and with a deterministic update. When the reheating procedure was applied, sometimes the optimum solution (Figure 11) was found.

In many cases, after some reheatings, the individuals have left a certain configurations and avoid a local minima. However the possible application of this technique is under consideration since there are not enough simulations to make a statistical significant argument about its efficiency. This discussion has been included mainly due to its analogies with SAGA, one of the techniques described as examples of memetic algorithms.

The 532-cities Padberg and Rinaldi problem [145] was also investigated. All the simulations performed, with 16 individuals have ended with tour lengths near two percent of the optimum. For all the TSP instances studied

with this method, the final tours found have a length similar to the average of those tours found by a Lin-Kernighan procedure, which are usually the input to a 'branch and bound' algorithm. The tours found seem to have a smaller standard deviation than those used for solving to optimality the 318 and 532-cities problems. A typical result is shown in Figure 12.

7   Parallel GA towards Memetic Algorithms

Due to its intrinsic parallelism and the fact that multiprocessor architectures are each day more available, there is considerable interest in the GA community to exploit this advantage [55] [56] [99] [150] [151] [186] [187]. I will not discuss these in particular. Instead, I will concentrate in some parallel GA that have turned into examples of memetic algorithms.

SAGA

A Parallel "Genetic" Heuristic for the Quadratic Assignment Problem

SAGA is one of the two parallel heuristics that I will analyse as examples of memetic algorithms. It has been described by Brown, Huntley and Spillane [40] as a cascaded hybrid of a genetic algorithm and simulated annealing customized to solve permutation problems. When SAGA was applied to the Quadratic Assignment Problem, they found SAGA superior to CRAFT, one the most commonly employed heuristic in solution quality and for large problems also superior in solution time [108].

The algorithm can be summarized as follows:

Step 1. Initialize the parameters of the GA.

Step 2. Generate an initial population of solutions for the GA.

Step 3. Use the GA to produce k "good" solutions.

Step 4. For each of the k solutions, do the following: a) Initialize the parameters of the SA. b) Improve the "good" solution using SA, and return to the GA.

Step 5. Repeat steps 3 and 4 as needed

In a parallel computer, step 4 can be done in parallel. Each of the offspring generated is improved using SA. The input configuration given to the

SA procedure is generally quite good, so a high initial temperature would destroy the work previously done to develop that configuration. The following heuristic is used to set the annealing schedule:

1. Approximate the expected change in the cost incurred by random pairwise interchanges terms in p. 100 pairwise interchanges are tried and then the mean absolute deviation (MAD) of the cost is calculated.

2. The initial temperature is set to $\beta\ MAD$ where β is a user-defined constant. Hence, β controls the probability of accepting and "average" cost change in the early stages of the search.

3. Set the scalar constant α to the value given by $(\frac{T^*}{T_{init}})^{1/\eta}$ where η is the number of temperatures in the schedule.

η controls the expected run-time given β and T^*.

The selection of one of the two parents who will create a child is made choosing it from a list of the best s structures, where s is a user-defined constant. The second parent is chosen as usually is done in a GA, selected with a probability equal to the ratio of its fitness to the sum of all the fitness values in the population.

The performance of SAGA

SAGA was compared to CRAFT for two test instances. The first was selected from a work of Nugent *et al.* [140], and the second one is from Scriabin and Vergin [169]. CRAFT, a steepest-descent-pairwise-interchange heuristic, was the technique used to compare its results against SAGA due to the fact that CRAFT was superior to some other techniques [140] [159]. For the 20 object problem [10] [169] SAGA in ten runs found four unique permutations all with a cost of 110030. In ten runs CRAFT has been always between the values 112588 and 124246. The optimality of the four permutations found by SAGA could not be confirmed due to the fact that a parallel version of Gilmore and Lawler's branch and bound procedure [89] [125] proved intractable. For a reduced problem with 18 objects, branch and bound was tractable and the optimal solution was found in 18 hours using a 32-nodel Intel iPSC/2 hypercube. SAGA was reported to have found the same solution in 2.4 minutes [40].

ASPARAGOS

An Asynchronous Parallel "Genetic" Optimization Strategy

ASPARAGOS is defined by its creators, M. Gorges-Schleuter and H. Mühlenbein [93] [138], as an asynchronous parallel genetic algorithm. Due to its characteristics I should classify it as a asynchronous *memetic* algorithm.

Step 0. Define a genetic representation of the optimization problem

Step 1. Create N individuals

Step 2. Each individual does local hill climbing (increases its fitness)

Step 3. Each individual chooses a partner for mating (local selection)

Step 4. Creation of new offspring (crossover and mutation)

Step 5. Replace the individual

Step 6. If not finished, go to Step 2

This algorithm has found a new optimum for the largest published quadratic matching problem and it also showed strikingly good performance in two of the biggest TSPs solved to optimality, the Padberg and Rinaldi 532-cities problem [145], and the Grötschel's 442-cities problem [161].

As the CCA, it is based in a physical neighborhood where the individuals that compose the population are allocated and it also has the advantages of few interproccessor communications. They also share with the CCA the fact that there is no global knowledge of the entire system; selection is done locally, within neighbours. These groups are called deme or "tribe" and are defined as the subpopulation in the immediate locality, a set of potential partners. The neighborhood acts as the selective enviroment of an individual. The population number of a deme is determined by the mobility of the individuals and those with better fitness have a better chance of being selected for mating. Due to the fact that different neighborhoods overlap, a diffusion process, inherent to the isolation by distance, gives the opportunity for good schemata belonging to well developed individuals a higher chance to propagate. In the CCA, this diffusion process is present in the mechanisms of competition and cooperation.

One run of the algorithm is dependent on some parameters. They are: M, which is the population size, D, the neighborhood (deme or "tribe") size,

C, the size of the crossover interval, P_M, the mutation rate, W, the window size, and S, the selection strategy. A window size of $W = n$ indicates that the base value which local fitness is computed is determined by the locally least fit individual from $n - 1$ generations in the past. M. Gorges-Schleuter has studied the effect of these parameters, for a more complete description the reader can see Ref. [93] and [139].

The computer simulations show that the quality depends on the parameter settings but also in the number of generations computed. A stopping criterion was established based on the diversity of the gene pool. In the TSP the diversity can be understood as the number of different edges between the tours present in a given generation. Experiments with more than 1600 generations showed a very small probability of finding better solutions, so a certain number of generations is selected as the stopping criterion.

Simulation Results

Although the effects of mutation and migration can be considered similar, since both introduce modifications into a neighborhood, migration is more useful than mutation. It was found that it is a poor strategy to use a mutation rate $P_M = 0.02$ to prevent loss of diversity. It is interesting to remark that M. Gorges-Schleuter conlcudes that in ASPARAGOS **"chance is less important than cooperation"** as in CCA cooperation being understood as the result of application of a crossover (recombination) operator.

The algorithm was not very dependent of the crossover parameter C but it was dependent of the selection strategy and the population size. The population size seems to be problem size dependent. There are also dependences that involve the window size, the selection strategy and the mutation rate.

ASPARAGOS has found a new optimum to the largest published Quadratic Assignment Problem. It has also found the optimum tours for TSP of less than 100 cities and in the 532-cities problem it has found a solution less than one percent above the optimal tour.

8   Discussion

The purpose of this section is to discuss some of the properties of these strategies to do the search stressing their analogies and differences with previous approaches.

These techniques are examples of what I called *memetic algorithms*. First

of all, they are population approaches that need few individuals. They combine a very fast heuristic to improve a solution (and even reach a local minima) with a recombination mechanism that creates new individuals. Recombination is a mechanism to *look between* good solutions.

The blind-fold search of the tallest building in New York

I want to be a little more clear in my description of why recombination is a good strategy for optimization in some kind of landscapes. To do that I would introduce an optimization problem: the task of finding blind-folded the tallest building in New York. Suppose a certain hypothetical *tireless* man is blind-folded. His task is to find the position of the tallest building in NY. Although he is blind-folded, he can walk to a given address, enter a building and compute its height. He has to have a good strategy to do the search. Of course he can not go to every place in NY and compute the height of the building there. This would be *complete enumeration*, we have discussed that possibility before.

So he can decide as his strategy the following: start from a random initial point, walk ten miles in a random direction, measure the building found, and repeat this for a certain number of iterations. Then report the address of the tallest building found. This is *random walk*, a very poor strategy for this problem. I should say a very poor strategy *for this landscape.* The quality of the final solution is expected to be bad, and also there is a waste of *computer time*, measured as unnecessary height evaluations (our objective function) in "*unpromising areas*". As an example, we would like to avoid searching in low-height areas like Central Park.

We can add a probability to the search: start from a random initial point, walk ten miles in a random direction, measure the building, accept this as the new configuration with a given probability (we can adopt the exponential transition probability of SA), and repeat the sequence for a given number of iterations. At the end the highest building found is reported.

Now suppose we have the results of these two *theoretical simulations.* It is hard to believe that the second one would be better than random search. But if instead of going from place to place in ten-miles long steps, he made steps of few blocks, it is clear that the result will be better than the random case. The reason is that we do not expect any *correlation* between the values of two buildings that are separated by ten miles. If we are in a tall building, a

building at ten miles can be of any possible heigth. Due to the fact that cities use to have tall buildings in clusters, the *downtown growth*, a smaller step-size would be more appropiate. Our blind *"optimizer"* would be attracted towards the cluster first, then, in the cluster, the small step size would avoid him to leave such a favourable region and search within the cluster.

To be fair with an analogy with SA in a discrete space, I should say that the step-size is fixed. The reader familiar with SA would point out that a better technique would be to use a big step-size in the begining and reduce it related with the *temperature* or statistical information from the heights of the last positions searched. A procedure that is more related to SA in continuous spaces [190].

Why we should adopt such a strategy for selecting the step-size ? If we do not and we use a short (few-block lengths) step-size, we have a high chance of being attracted by the nearest cluster form our departing point. Since our hypotetical *optimizer* is a tireless individual, he can adopt this long-step-size-first strategy. This will allow him to search more widely at the begining and more locally at the end of the search (we are always supposing that there is a stopping criterion). A probabilistic criterion for determining the lenght of the step would be more efficient than the monotonically decreasing fixed step-size [184] [185]. It would add the possibility of jumping between clusters.

Parallel approaches to the blind-fold search

I recall that the above description is the analogy of the development of a strategy to search an optimum in a combinatorial optimization problem. In particular, we have been discussing an analogy with SA.

Now, forgeting the analogies for a moment, we have to consider that we plan to do the search with a certain computer available. In principle, it can be a sequential or a multi-processor computer. In this case a question arises, the parallelization of our algorithm. Parallelization is often viewed as the correct application of the *divide-and-conquer* concept. Here we should ask ourselves, what we will "divide" ? What will we distribute among processors?

Studying the TSP, Felten et al. have decomposed the *physical* problem [74]. In the TSP case, a certain number of cities would be assigned to only one processor, in a VLSI design problem, a certain number of modules is assigned, etc.. Similar decompositions can be found in Refs. [26] [122] [58]

[154] [19]. Finally, the decision would be to *speed-up* the search or to be more interested in the final quality of the solution [122].

Returning to our analogy with the blind-fold search in NY, suppose that we have a parallel computer. Our two possibilities can be analogous with having a faster individual to do the search or to have *many* individuals. Using a parallel machine to do a fast search, it would be the same as if this blind-folded optimizer would have used a *taxi* to go from point to point. If we leave invariant the number of evaluations he made, the use of a parallel computer is just a way of speeding-up the search. If instead of that we would like to use the *population approach*, a certain number of blind-folded guys, all of them walking, a natural question arises, how can I organize them to make a more efficient search ?

The CCA was a step in this direction. Suppose that I start with a certain number of these individuals in random points scattered in NY. I leave them for a certain time searching. After that, each one radioes a message to a neighbour (they form a ring as described in the CCA). They compare the height of the buildings from where they are actually sending the messages. The individual that is on the smallest building (there is a probabilistic rule in the CCA) abandons it and goes to the address of the neighbour who has sent the message. Now both would start from this new position and since we expect that there is a correlation between the heights of buildings we have now *two* individuals searching that area. A new period of individual search takes place and then again this interaction occurs. In this case we can say that we have a *competitive search*, a *competitive annealing* if we use simulated annealing to do the local search.

But a competitive search like that looses some of the information created by the individual that must abandon its present position. So there is a need for the creation of a different interaction, a cooperative one that would interchange information between individuals. Recombination, the crossover operator in the CCA, is such an interaction. As a result of these interactions we expect a group of individuals, instead of a very fast individual, the two possible implementations in a parallel computer, to be a much more efficient search strategy.

Correlation of local optima

In the previous example, we are exploiting the fact that tall buildings are organized in clusters to develop a search strategy. We have discussed how our search can fail if it is not adapted to the correlation of the objective function we are optimizing (eg. the search with a very big step-size). I can affirm that any success in the application of SA or a GA to an optimization problem has as a key feature the correlation of local optima. However, SA and GA exploit this correlation in different ways.

Kauffman has found evidence of correlation of local optima in his NK model [113] [116]. Figure 13 shows a sketch of one of his numerical experiments. First a search via 1-mutant (one-flip) fitter-variants is made until no further local optima are uncovered or until 10,000 have been discovered. Having all these optima, for some problems it was founded that there is a global structure to the fitness landscape (the objective function). In Figure 13 the global structure is evident when we display the fitness of the local optima as a function of the Hamming distances of all the 1-mutant local optima from the fittest local optima found in the whole set. The Hamming distance is the natural definition of distance between configurations since we are using the 1-mutant as the move in the space. Figure 13 shows that the position of the local optima *is not random*. For K small in comparison with N, the highest local optima have very small Hamming distances between them. Local optima with succesively greater Hamming distance from the highest optimum are succesively less fit. Figure 14 shows how the fittest local optima seems to have the biggest basins of attraction. This property disappears when K is increased. These two facts suggest a *hilly* landscape. The familiar *multi-valley* structure of spin-glasses [8] [63] [64] [147] and the existence of a region where all good optima are located leads Kauffman to suggest the metaphor of the "Massiff Central" in the Alps.

Kauffman's numerical experiments are biased by the distribution of sizes of basins, but they can be considered as overall properties. Optima with very small basins of attraction might not be found with the 1-mutant fitter-variant technique used to find the local optima. However, I have previously discussed the problem of the golf-course landscape and the impossibility of finding a good strategy in that case. The interest here is to analyze a general property of the landscape and how to exploit it when we are moved by an optimization purpose.

In my opinion, the study of correlation between minima and the structure of the landscape is one of the most important results that came from research in disordered systems for the analysis of optimization problems. Computer scientists should look at those results in order to analyze and construct heuristics for combinatorial optimization problems. Failure of heuristics in a given problem can be understood regarding features of disordered systems. For example, the task of finding the ground state of a three-dimensional spin-glass has been proved to belong to the NP-complete class [18] [12]. The task of finding ground states of spin-glasses, regarded as an optimization problem, has to deal with the asymptotic behaviour of the height of energy barriers as one of its central features. This height grows with N, the number of spins [8].

The physics of disordered systems is addressing these kind of questions through the works of P.W. Anderson [73] [87] [14], J.R Banavar [14] [15] [16] [17], B. Derrida [63] [64] [65] [66] [67], R.G. Palmer [146] [181], G. Parisi [148] [149], N. Sourlas [175] [14] [176]. In particular the works in ultrametricity tried to understand some of the properties of the configuration space [129] [155] [11] [13] [97]. A significant amount of effort was directed to analyze the connections between the physics of disordered systems and other fields, [180] [182] and some tools of Statistical Mechanics have been applied to understand optimization problems [87] [112] [123] [130] [131] [14] [171] [175] [177]. The message to Computer Science is explicitly addressed by P.W. Anderson [6]. All these works are trying to understand the properties of the landscape, to use Kauffman's words, and as an example of similar work he did with the NK model. I should remark the work of Kirkpatrick and Toulouse in the TSP [120].

The memetic algorithm at work: Exploiting correlations in the landscape

Suppose we want to apply a memetic algorithm to a completely correlated landscape. An example of such a landscape can be Figure 15, a bowl-like cost function. It can be viewed as an analogous picture of the $K = 0$ Kauffman model, for example the N-locus, 2-allele *additive* fitness model previously discussed. First, create a certain number of individuals in random locations. Then evolve them to a local minima. Here we should compute the diversity of

them, and we will find that all have reached the only optima in the landscape. Then, the search stops and the location of the local (which in this case is global) optima is reported.

It is interesting to see how SA and GA would behave in such a landscape. To make the analogy with SA, an individual is placed in a random location. As a result, at each temperature many *up-hill* moves are accepted, which leads in a waste of computer time, slowing the search. In some implementations of SA, where the number of iterations at a given temperature and the way of decreasing the temperature are fixed, SA would take the same amount of computer time in this landscape than in a *"very complex"* one.

A GA will also be time-consuming. It will start with a group of individuals in random locations as the memetic algorithm. Then the operators of crossover and mutation would act, selection of the best individuals, and this procedure would be repeted many times. In contrast the MA does not make any recombination for this problem, and if implemented in a distributed architecture, there would not be any communication during the search.

Unfortunately, most combinatorial optimization problems do not present such a correlated landscape, but in many cases where SA or GA is being applied there exists a certain correlation. A picture of somewhat correlated landscapes can be found in Figure 16. It is clear the advantage of a MA in this case as a strategy that exploit the correlation and avoids the problem given by the ruggedness of the landscape. Because each one of the individuals recombines to create a new individual after they have reached a local optima, they are using information about the location the local optima to find potential good regions where to search. GA would work in a similar way, but because they are based in random mutations, the individuals to be recombined are not necessarily *good*. So a positive effect of a recombination operator is masked by the lack of correlation between two configurations that are not local minima. They are not completely uncorrelated, but we expect that the correlation would be smaller than in the case that both are local minima. SA seems to be based in a kind of *multiscale* process by which the landscape *"looks"* more smooth at high *temperature* and increasing its roughness when the temperature is decreasing. It seems that the stochastic hopping over barriers does not play a fundamental role in the efficiency of SA [134].

Measuring correlations of the landscape

Returning to the primary stage of the election of a representation for a combinatorial optimization problem and how to select the set of moves associated, we need to develop a technique to choose them. The landscape is a consequence of this election, and we have pointed out the correlations between entities in that representation as one of the advantages that SA, GA and MA are exploiting to do the search. Faced with a set of possible representations, $r_1, r_2, \ldots r_m$ and a set of moves $s_1, s_2, \ldots s_m$ a good criterion for choosing a certain pair (r_i, s_j) as the optimal would be based in the correlation of its associated landscape. A measure of correlation would be of primary interest.

E.D. Weinberger has suggested the use of autocorrelation functions to study the landscape of the NK model [114] [193]. Suppose we choose a given pair (r_i, s_j), we will study the correlation of its associated landscape with the following procedure. We start with a randomly generated configuration (an entity in the representation). A random walk is generated by the successive application of the move s_j), so a set of entities is generated $e_1, e_2, \ldots e_h$. We compute the fitness (objective value) of each of these entities. The autocorrelation function relates the fitness of two entities that there are p steps apart:

$$C(r_i, s_j, m) = \frac{E(F_n F_{n+m}) - E(F_n)E(F_{n+m})}{variance(F)} \tag{10}$$

where E is the expected mean value. In the NK model it was found the expected correlation for K small and the deterioration when K is increased. It was also found an exponential fall of the autocorrelation as a function of m, the number of steps apart of two entities.

Of particular interest would be the autocorrelation function $C(r_i, s_j, 1)$. However, it has to be proved that the selection of a pair (r_i, s_j) which is optimal in the sense of the autocorrelation $C(r_i, s_j, 1)$ would also be the best pair for a technique like SA. My impression is that successful implementations of SA have a good value of $C(r_i, s_j, 1)$ but there are other correlations involved. To be more precise, we have talked about the correlation of local minima, and this measure of correlation does not give such information. The random walk is started in a randomly chosen entity, and since the walk is random, we are computing the correlation of entities that can only appear in the very high, extremely high indeed, phase of a SA algorithm. Although I

must introduce another measures of correlation, this measure has proved in the NK model *the existence of a natural correlation length* for each one of the landscapes generated varying the value of K.

Towards a rational design of moves and recombination operators

It is undoubtely true that the success of a GA implementation is a direct consequence of the utilization of appropiate recombination operators. A clear result of that is the extraordinary improvement in performance in the TSP when new and more efficient crossover operators and representations have been used. However, there is no clear understanding of the reasons that lead a certain crossover operator to be better than others.

One possible way to try to understand these better performances is to use correlation functions to analyze the behaviour of crossover operators. And it would also be a tool to design more effective recombination operators. We have discussed the bowl-like landscape and the GA implementation. A certain crossover operator that takes two configurations, here two points in the bowl, and creates a child that is half-way between the parents, would be an effective recombination operator for the problem. While another operator that most of the time creates a child farther from the parents than the inter-parents distance, would be a poor way to introduce recombination.

This leads to the introduction of a measure of correlation to somewhat measure this behaviour. Let p_1, p_2 be two parent configurations randomly selected. Let $d(p_1, p_2)$ be the distance between two configurations. Let rec be a certain recombination operator and ch the child created by application of rec to p_1, p_2. The values of the objective functions $F(p_1), F(p_2), F(ch)$ are computed. Let F^* be the best value of the pair p_1, p_2. Then we can create a correlation function of the type.

$$C1_{GA}(r_i, rec, d) = \frac{E(F^* F(ch)) - E(F^*)E(F(ch))}{variance(F^*)} \qquad (11)$$

where E is the expected mean value. Although this would be more interesting for GA, it has a similar problem. Since the two parents have been selected randomly, their values of fitness are not near-optimal. As a consequence, this measure of correlation, is reflecting properties of the recombination operator

40

at the begining of the run of a GA.

Another measure of correlation can be used, supposing that the two parents selected are *local minima*. In that case they are two local minima under a given set of moves s_j. It is obvious that this measure would be natural for a MA. So we can compute

$$C1_{MA}(r_i, s_j, rec, d) = \frac{E(F^*F(ch)) - E(F^*)E(F(ch))}{variance(F^*)} \qquad (12)$$

where the use of a "1" means that is a child *of the next generation*, the equivalent of *one step* after the application of the recombination operator. I should write $C1_{MA}(r_i, s_j, rec, d) = C_{MA}(r_i, s_j, rec, d, 1)$ and open the posibility of analyzing $C_{MA}(r_i, s_j, rec, d, m)$.

We can think of $C1_{MA}(r_i, s_j, rec, d)$ as a possible tool to analyze the performance of a recombination operator in the landscape generated by the pair (r_i, s_j). For example, in Figure 17, we can see that there exists a correlation between local optima. They are located over a ring. So the recombination operator that creates a child over a straight line over two local optima is not the more adequate for this problem. It will create offspring in the central region which is flat. In conclusion, for a certain optimization problem, a memetic algorithm can be developed if a certain hidden correlation can be exploited by the use of the most effective recombination operator. Figure 18 shows how we can have local optima correlated although the landscape would be not very smooth. H. Mühlenbein says in Ref [138]: **"There is lots of evidence that for many applications the crossover operator is the key to the success of the genetic algorithm, but there are only some qualitative arguments explaining the above observation"**. Perhaps the study of these kind of correlation functions would be able to explain this success and be useful in the rational design of recombination operators.

9   Future Directions of Memetic Algorithms

In this section I want to discuss some possible directions of research in Memetic Algorithms. One of the possibilities which have not been much explored both in SA and in GA is a kind of *annealing in the complexity* of the task. I am using the word "complexity" as Kauffman uses it to relate it with the *difficulty* of the task. This differs with the use given by computer

scientists and complexity theory. For them, the real complexity catastrophe is the existence of NP problems so I believe they would like me saying *annealing in the difficulty* of the task. In the TSP, it is natural to associate the complexity (and the difficulty) with the number of cities that compose a tour. Although it was not well considered in SA and GA, *constructive heuristics* are common-place in the design of algorithms for combinatorial optimization problems.

In the beginning was simplicity

The N-var approach

Usually, SA and GA when applied to the TSP have used a tour composed of N cities during all the optimization process. In that way, we are always trying to find the low-distance tours between the (N-1)!/2 possible tours. In other words, a 100-city TSP have a configuration space 99 times bigger than a 99-city TSP and we are always working out the most difficult task. However, usually a small number of cities can give an approximation of the overall shape of the tour. Obviously we have more possibilities of finding an optimum tour if we have less cities, and then we can continuously adding cities, perturbating the present solution until we complete a tour with the N cities. Reviewing the foundations of SA we address the following question: why can't we leave the number of cities as an external control parameter in the same way as we do with temperature? Pursuing the analogy that if T is decreased from an initial value T_0, the number of cities of the tour (n) can be increased *ad hoc* from an initial number n_0 to N.

Pursuing an analogy with biological systems, which is much more clear after the introduction of GA, life forms have two possibilities, at the genetic level, to improve their fitnesses. One is the rearrangement of the sequence in a gene that can lead to better "genetic codes", and they can also improve them by the addition of new nucleotides in the sequence, so increasing the dimensionality of the configuration space of that gene. These mechanisms, in addition with natural selection permit to develop only those individuals that make the most effective use of the genetic material they can use to build a code.

Playing with the words, we can say that using the number of cities as an external control parameter, we would also be able to do an *annealing at zero*

temperature. This statement seems strange, so we need to clarify in which sense it must be understood . SA, it is said, used the temperature, an artificial parameter to escape to *"local minima"* situations. Now suppose that we have set the temperature to zero, while increasing the number of cities; so the acceptance probability functions of SA go to the step function. In that case, as a result of the addition of a new city, we can leave many situations in which without this new degree of freedom, it would have been impossible to improve the tour, leaving an *artificial local minima* if we consider that the real problem has N cities.

Due to the fact that after five years, the optimal annealing schedule is still a practical difficulty, the analogous question of how many attempted rearrangments we have to make for a given number of cities, can not be answered, even at zero temperature. Empirically, we found that a number of attempted rearrangements of the order of $O(n_t \; ln(n_t))$ per MC step seems to be adequate, where n_t stands for the actual number of cities in the tour. So starting with n_0 cities we can generate a nearest-neighbour tour or, for research purposes we can start with a random tour. Then we make $n_0 \; ln(n_0)$ attempted rearrangements and then we add a new city, and so on. This procedure can be complemented with the use of the temperature, so we can have the two control parameters at the same time.

Another question that must be addressed is the order of insertion of cities. That is, which cities are the initial n_0 and which ones are inserted further on. One lesson from the experience in SA is that the large structures of the tour anneal first and then, at low temperatures, the small defects are corrected. If we perform the insertion of the cities while doing the decreasing of the temperature, it would be wise to insert first those cities that will give the *gross features* of the tour.

The best way seems to use a *farthest selection procedure*. The *farthest-selection cheapest-insertion* procedure has proved to be a good heuristic for the TSP. One of the standard versions would be [124]:

Step 1. Start with a subgraph consisting of city choosen at random. We will call it as city i .

Step 2. Find a city k such that c_{ik} is maximal and form the subtour (i, k). The value of c_{ik} is the cost of connecting these two cities and it is equal to the value of the distance between them if we are considering an euclidean problem.

Step 3. (*Selection*) Given a subtour, find a city k not in the sub-

tour and city l in the current subtour such that $c_{lk} = \max_j(\min_i(c_{ij}))$, where j denotes a city *not in* the current subtour and i denotes a city *in* the current subtour.

Step 4. (*Insertion*) **Find the edge (i, j) in the subtour which minimizes $c_{ik} + c_{kj} - c_{ij}$. Insert k between i and j.**

Step 5. Go to Step 3 unless we have a Hamiltonian cycle.

or in Step 3 can be replaced with

Step 3'. (*Selection*) **Given a subtour, find a city k not in the subtour farthest from any city in the subtour.**

Both versions, being deterministic, only depend on the initial city i. Using the farthest-insertion solution as the starting tour in the r-opt Lin-Kernighan procedure [127], the optimal solution was found in eight of twelve non-Euclidean problems studied in a computational study performed by Adrabinski and Syslo [5] and it still performs well on big problems [160].

Another obvious possibility is to add the cities in a random order, that would leave the complexity of the memetic algorithm without changes.

I would like to remark before ending this subsection, that the concept here is to add to the evolutionary strategy that is the core of the memetic algorithm, the component of an *increasingly* complex task. However, this idea should not only be associated with a gradual increase in the dimensionality of the configurations. In other systems, like the NK Kauffman model, it would be wise to develop a K-var strategy. At the beggining, a small value of K is selected, the landscape is more correlated than with the final value. A slow increment of the epistatic interactions may also have an interesting biological analogy.

Exploiting asynchronism and heuristics

ASPARAGOS was an example of how a MA does not need to have a synchronous implementation. I should add that it does not need that each of the processors to be of the same type. Different computers can be connected with suitable protocols. Taking the CCA ring structure as an example, it is interesting to remark that a certain network of computers can be doing a certain optimization task using idle time. When a certain computer is idle, it can send a message to the ring structure and position itself between two other computers presently working in the ring. It can start with one of the configurations actually considered by one of the neighbours in the ring.

When a certain computer is needed for another task, it would leave the ring in a similar way.

This possibility, is given only due to the advantages of the memetic approach. For it, the network *really is* the computational device. As a newcomer to Computer Science, I can not avoid wondering about the coincidence that, a memetic algorithm which is inspired in emulating cultural evolution, has as its natural computational framework what computer scientists have called "Social Systems". Social Systems are asynchronous distributed processors characterized by a large and variable population of small individuals and a random and changing communication architecture [178].

Another advantage that can be exploited is that the most powerful computers in the network can be doing the most time-consuming heuristics, while others are using a different heuristics. The program to do local search in each individual can be different. This enriches the whole, since what is a *local minima* for one of the computers is not a local minima for another in the network. Different heuristics may be working fine due to different reasons. The collective use of them would improve the final output. In a distributed implementation we can think in a division of jobs, dividing the kind of moves performed in each computing individual. It leads to an interesting concept, where instead of dividing the physical problem (assignment of cities/cells to processors) we divide the set of possible moves. This set is selected among the most efficient moves for the problem.

What are the general rules ?

One of the most important questions that research in memetic algorithms should address is the search of general principles [118] [106] [51]. For example, one of them is related with the topology of the interconnection network. The CCA and ASPARAGOS have used similar configurations while SAGA has no such a structure. So, is the isolation by distance a better strategy ? It has a certain advantage in the sense that the computation is distributed, but it would be interesting to prove that it also benefits the quality of the final solutions or the speed of the algorithm or both. The CCA and ASPARAGOS differ since in the CCA the neighbourhoods of cooperation and competition are different. Is this ingredient important ? If it is important, how can we exploit it ? Is the ring the best topology ? Some simulations with the CCA have shown that the use of more than 16 processors does not give

better solutions (at least with the OX operator and the determistic update). ASPARAGOS, which was using a better crossover operator, has found that **"...with a 100-city problem a population size of only 16 is sufficient to always converge to the global optimum."** [93]. M. Gorges-Schleuter, working with ASPARAGOS, supports the advantage of the isolation between individuals. She says in Ref. [93] **"To get high quality solutions it is important that most demes are separated, and only locally near demes overlap. This means local information should only propagate to other demes through a diffusion process"**.

She also remarks that **"In comparing the effects of mutation and migration we conclude that chance is less important than cooperation"**. This seems to me to be a result of the fact that ASPARAGOS, regarded as a memetic algorithm, is exploiting the correlation of local minima. The use of a mutation rate, which is necessary in a GA, is not so important here. In a GA it is needed to improve the quality of solutions after the application of the recombination operator. Being a technique that mimics biological evolution it is sure that it needs to be incorporated. On the contrary, the individuals in a memetic algorithm are improved, slightly in the CCA, more significantly in SAGA and reaching the maximum in ASPARAGOS. In a conversation with authors of SAGA and ASPARAGOS, I found myself using the same words to describe the improvement of the individuals doing the phase of local search. In Ref. [40], Brown et al. describe it as: **"...each of the offspring generated by the GA in a given generation is improved using SA. In other words, each offspring is required to "mature" before being allowed to have offspring, much as it would be in a natural system"**. If all these methods are using the correlation of local minima as the reason of sucess, the introduction of mutations would not benefit the search process since it will only generate "noise". However, for those that would like to equate noise with a benefical contribution, I must remark that the necessary random effects are provided by the recombination operators and not by the mutation.

Other questions are related to the breeding procedures. In the CCA, the introduction of an *acceptance mate factor* was benefical. It improved the quality of solutions in aproximately four percent of the length of the optimum, for different size of instances. The existence of an *acceptance mate factor*, helps an individual to avoid to mate if the other individual is worse than itself. SAGA **"uses a rank ordering of the costs when selecting**

a pair of parents for the CrossOver operator" and in ASPARAGOS a selection of the parents was also a good strategy. Kauffman as a result of one of his numerical experiments note: " **Preferential mating and recombination of the highest local optima is a selective force which tends to pull the entire population toward the highest actual local optimum discovered. Indeed, in the present case, the entire population climbs to the actually fittest opimum uncovered in the entire adaptive procedure"**. He also found that "...a different bias in **recombination such that "marriage" occurs preferentially between nearby peaks, regardless of their fitness, aids recombination. To test this, we required peaks to be less than half the current mean Hamming distance among all peaks encountered by 100 walkers in order that recombination might ocurr between them. Somewhat to our surprise, this non-random mating rule helps adaptive hill climbing compared to random mating and recombination"**. In my opinion this is the result of a correlation of local minima combined with the use of a recombination operator that generates offspring which are *near* the parents. There is not more suprise than the fact that the recombination operator is wisely exploiting the correlation of local optima.

In the search of general rules, other similar techniques would be taken into account (see for example Refs. [29] [29] [162] [163]) and careful performance mesures should be derived, perhaps using some instances of this combinatorial problems as a benchmark of the strategies.

Kauffman is using memetic techniques

The last quotes of Kauffman are a result of numerical experiments that I should also classify as memetic optimization. It is interesting to remark that although his work is inspired in trying to develop some the mysteries of biological evolution, to show the advantages given by the use of recombination operators he had to use a memetic approach. Perhaps a genetic algorithm would have been considered more adequate regarding the biological assumptions and scope of his work. However, I believe that he is right in the application of a memetic technique and to prove by his experiments the advantage of using recombination operators. He describes his experiments in this way: **"As hinted above, some fitness landscapes may be self-similar... It is intuitively plausible that in such a land-**

47

scape, which tends to have "Massifs Central", recombination will be helpful. But will it ? In order to test this numerically, my colleague Lloyd Clark and I have studied the NK model. We have schematized the effects of recombination in a simple way. We release a fixed number (100) of randomly chosen genotypes upon the same NK fitness landscape, and allow each to walk via randomly chosen, 1-mutant, fitter variants to a local optimum. In general, 100 or fewer independent local optima are found. Thereafter, we mated and recombined randomly chosen pairs of local optima at randomly chosen positions within each genotype, to form 100 new recombined genotpes. These 100 recombinants were then allowed to walk via randomly chosen 1-mutant fitter variants to local optima. Thereafter, the cycle of recombination followed by hill climbing to optima was repeated. This numerical procedure clearly asks whether the regions between local optima help direct the adaptive process to yet higher local optima". I have no doubt he has chosen this strategy because of the similar constraint that we faced in combinatorial optimization, i.e. the lack of computer time available. I have enjoyed seeing his Figures where a complete convergence was found after some generations. The recombination operators, which can be the more expensive computational operation in some problems, was applied no more than 50 times. This fact constrasts with the large number of generations we would have to wait to see if we mimic biological evolution instead of cultural evolution. It can easily be the answer of one of Mühlenbein questions. He asks in Ref. [138]: "Why should a complex crossover operator lead to a faster evolution than mutation? Or in more algorithmetic terms: Should we use a large population which evolves by small mutations or a small population evolving by sexual reproduction and crossover? Which algorithm is faster (in number of computer instructions) ? " It can be the case that a crossover operator exploits the correlation of local minima (and a fast heuristic to reach configurations near the local optima), the analgous to cultural evolution, is much faster than the big population evolving with small mutations.

Kauffman has wisely remarked that an important feature of adaptation that can be related to one of the central subjects of study in genetic algorithms. He made an analogy with what he calls a *weak Maxwell's Demon.* He remarks that "...if selection is too weak to hold an adapting pop-

ulation in very small volumes of the ensemble, then even in the presence of continuing selection the adapting population will almost certainly exhibit the "typical" ordered properties of most ensemble members. Hence I use to tend to use the phrase that such adapting systems would exhibit order not because of selection, but despite it".

In genetic algorithms, since we are interested in finding the best optimization technique, a natural question arises in trying to find the best relation between the population size and the mutation rate. This seems an endless question similar to the best annealing schedule in SA. It can be possible that the optimum value of this parameters can not be found, that there is no general rule for them. They may depend on the problem, or even on the instance of the problem being under consideration. Returning to the initial discussions, it may depend on the landscape, and we know that few things can be said *a priori* about the landscape of an optimization problem. The mutation-rate population-size problem of GA may be closely related with the *weak Maxwell's Demon* that Kauffman analyzes.

On the contrary, the memetic algorithms I described seem not to deal with that problem. It has been remarked that present versions need few individuals. In the implementation of the CCA, the use of a decreasing temperature that controls the process of competition and cooperation, makes the population concentrate in good solutions in a gradual manner, avoiding the system to spend time in potentially *bad* regions of the *tour-space*. It seems that this mechanism is controling a problem described by Ceccatto and Huberman in Ref. [51]. Similar mechanisms arise in SAGA and in ASPARAGOS. Mühlenbein has noted this fact. He says in Ref. [138] that **"...the evolution is driven totally by the system itself. There is no need for artificial control parameters. Especially there is no need for a sharing function to mantain variability"**. The use of a sharing function in some implementation of genetic algorithms is discussed in Ref. [90]

10   Conclusions

In this review, I have presented a unified view of a certain kind of distributed algorithms which have been introduced very recently and have shown a ex-

traordinary performace dealing with some of the biggest instances of certain combinatorial optimization problems. The analogies of these kind of algorithms with some features of cultural evolution have been remarked and explored. They suggest a framework for them, playing a similar role to the biological inspiration of genetic algorithms. Due to some of these analogies and the fact that they clearly diverge of some other approaches, I found that they can be labeled as *memetic algorithms*. I have also shared the hypothesis that the correlation of local minima is responsible of this amazing success.

The scope of these techniques goes beyond the range of combinatorial optimization problems. They would be applied in other optimization problems where the representation and the search strategy are suitable selected. Memetic algorithms are not a new heuristic that can be chosen to be applied in an optimization problem. They are not motivated to replace present heuristic. Instead they are a framework to exploit all previous knowledge about the problem, combining methods to improve their perfomance.

Aknowledgments

References

[1] E.H.L. Aarts and P.J.M. van Laarhoven, "Statistical Cooling: A General Approach to Combinatorial Optimization Problems", Philips Journ. Res. **40**, pp. 193 (1985).

[2] E.H.L. Aarts et al., "Parallel Implementations of the Statistical Cooling Algorithm", INTEGRATION, the VLSI Journal **4**, pp. 209 (1986).

[3] E.H.L. Aarts and P.J.M. Laarhoven, "Simulated Annealing: Theory and Applications", D. Reidel Publishing Company (1987).

[4] D.H. Ackley, "An Empirical Study of Bit Vector Function Optimization", in *Genetic Algorithms and Simulated Annealing*, Ed. by L. Davis, Morgan Kaufmann, Los Altos CA (1987).

[5] A. Adrabinski and M.M. Syslo, "Computational Experiments with some Approximation Algorithms for the Travelling Salesman Problem", Zastos Mat. **18**, 91 (1983).

[6] P.W. Anderson, "What Statistical Mechanics has to say to Computer Scientists", Physica A, **140**, pp. 405 (1986).

[7] P.W. Anderson's comment of Francis Crick's autobiography *"What Mad Pursuit: A Personal View of Scientific Discovery"*, Basic Books, New York, (1988). Anderson's comment appeared in Physics Today, July (1989).

[8] P.W. Anderson, "Spin Glass V: Real Power Brought to Bear", fifth article on a series appeared in Physics Today, pp. 9, July (1989).

[9] I. G. Angus et al. *Solving Problems on Concurrent Processors. Vol. II. Software for Concurrent Processors* (to appear) (1989).

[10] G. Armour and E. Buffa, "A Heuristic Algorithm and Simulation Approach to the Relative Allocation of Facilities", Management Science, **9**, pp. 294 (1963).

[11] M. Aschbacher et al. "Embbedings of Ultrametric Spaces in Finite Dimensional Structures", SIAM J. Alg. **8** (4), pp. 564 (1987).

[12] C.P. Bachas, J. Phys. A **17**, L709 (1984).

[13] C.P. Bachas, "Invariant Geometry of Spin-Glass States", Phys. Rev. B **35**(4) pp. 1965 (1987).

[14] J.R. Banavar, D. Sherrington and N. Sourlas, "Graph Bipartitioning and Statistical Mechanics", Jour. of Phys. A: Math. Gen. **20** (1), (1987).

[15] J.R. Banavar and A.J. Bray, "Chaos in Spin Glasses: A Renormalization Group Study", Phys. Rev. B **35**(16), pp. 8888 (1987).

[16] J.R. Banavar, "Heisenberg and Potts Spin-Glasses: A Renormalization Group Study", Phys. Rev. B **38**(4), pp. 2564 (1988).

[17] J.R. Banavar, "Trasport Properties of Disoredered Continuum Systems", Phys. Rev. B **39**(16) pp. 11965 (1989).

[18] F. Barahona et al., J. Phys. A **15**, 673 (1982).

[19] V.C. Barbosa and E. Gafni, "A Distributed Implementation of Simulated Annealing", Journal of Parallel and Distributed Computing, **6**, pp. 441 (1989).

[20] G. Baskaran, Y. Fu and P.W. Anderson, "On the Statistical Mechanics of the Traveling Salesman Problem", Journal of Statistical Physics **45**, pp. 1 (1986).

[21] G. Baskaran and D.L. Stein, "Intractable Computations without Local Minima", Comment, Phys. Rev. Lett. **59** (3), pp. 373 (1987).

[22] E. Baum, "Intractable Computations without Local Minima", Reply, Phys. Rev. Lett. **59** (3), pp. 374 (1987).

[23] J. Beardwood, J.H. Halton and J.M. Hammersley, "The Shortest Path through many Points", Proc. Cambridge Philos. Soc. **55**, 299 (1959).

[24] W.W. Bledsoe, "The Use of Biological Concepts in the Analytical Study of Systems", Paper presented at the ORSA-TIMS National Meeting, San Francisco CA (1961).

[25] T. Block, "On the Complexity of Facility Layout Problems", Management Science **25**, 280 (1979).

[26] E. Bonomi and J.L. Lutton, "The N-city Travelling Salesman Problem and the Metropolis algorithm", SIAM Review **26** 551 (1984).

[27] E. Bonomi and J.L. Lutton, "The Asymptotic Behaviour or Quadratic Sum Assignment Problems: A Statistical Mechanics Approach", European Journal of Operations Research, **26**, pp. 285 (1986).

[28] T. Boseniuk, W. Ebeling, and A. Engel, "Boltzmann and Darwin Strategies in Complex Optimization", Physics Letters A **125** (6-7), pp. 307 (1987).

[29] T. Boseniuk and W. Ebeling, "Optimization of NP-Complete Problems by Boltzmann-Darwin Strategies Including Life Cycles", Europhysics Letters **6** (2), pp. 107 (1988).

[30] D.E. Van den Bout and T.K. Miller, "Improving the Performance of the Hopfield-Tank Neural Network Through Normalization and Annealing", North Carolina State University, preprint (April 1989).

[31] G.E.P. Box, "Evolutionary Operation: A Method for Increasing Industrial Productivity", Journal of the Royal Statistical Society, C, **6** (2), pp. 81 (1957).

[32] R. Boyd and P.J. Richerson, "The Cultural Transmission of Adquired Variation: Effects on Genetic Fitness", Journal of Theoretical Biology, **100**, pp. 567 (1983).

[33] R. Boyd and P.J. Richerson, *"Culture and the Evolutionary Process"*, University of Chicago Press, Chicago (1985).

[34] R. Boyd and P.J. Richerson, "The Evolution of Reciprocity in Sizable Groups", Journal of Theoretical Biology, **132** (3) pp. 337 (1988).

[35] R. Boyd, "Mistakes Allow Evolutionary Stability in the Repeated Prisoners Dilemma Game", Journal of Theoretical Biology, **136** (1), pp. 47 (1989).

[36] R.M. Brady, "Optimization Strategies Gleaned from Biological Evolution", Nature **317**, pp. 804 (1985).

[37] B. Braschi, "Solving the Traveling Salesman Problem with Simulated Annealing Techniques on a Concurrent Supercomputer", TIM3-INPG Grenoble, Report RR 752-I Nov. (1988).

[38] H.J. Bremermann, "Optimization through Evolution and Recombination", in M.C. Yovits, G.T. Jacobi and G.D. Goldstein (Eds.). *"Self-Organizing Systems"*, pp. 93 (Washington D.D.: Spartan Books) (1962).

[39] H.J. Bremermann, "Quantitative Aspects of Goal-Seeking Self-Organizing Systems", Progress in Theoretical Biology, **1**, pp. 59 (1967).

[40] D. Brown, C.L. Huntley and A. Spillane, "A Parallel Genetic Heuristic for the Quadratic Assignment Problem", Proceedings of the Third International Conference of Genetic Algorithms, Fairfax, VA, ed. by J.D. Schaffer, (Morgan Kaufmann, San Mateo CA) pp. 406 (1989).

[41] R.E. Burkard, "Quadratic Assignment Problems", European Journal of Operations Research, **15**, pp. 283 (1984).

[42] R.E. Burkard and U. Fincke, "On Random Quadratic Bottleneck Assignment Problems", Mathematical Programming, **23**, pp. 227 (1982).

[43] R.E. Burkard and U. Fincke, "The Asymptotic Probabilistic Behaviour of Quadratic Sum Assignment Problems", Zeitschrift für Operations Research **27**, pp. 73 (1983).

[44] H. Camblong, "The Large Numbers", private communication.

[45] L.L Cavalli-Sforza, "Similarities and Dissimilarities of Sociocultural and Biological Evolution", in *Mathematics in the archaeological and historical sciences*, (eds. F.R. Hodson et al.), University Press, Edinburgh (1971).

[46] L.L. Cavalli-Sforza and M.W. Feldman, *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton: Princeton University Press (1981).

[47] L.L. Cavalli-Sforza and M.W. Feldman, K.H. Chen and S.M. Dornbusch, "Theory and Observation in Cultural Transmission", Science, **218**, pp. 19 (1982).

[48] L.L. Cavalli-Sforza and M.W. Feldman, "Cultural versus Genetic Adaptation", Proc. Natl. Acad. Sci. USA **80**, pp. 4993 (1983).

[49] L.L. Cavalli-Sforza, A. Piazza and P. Menozzi, "Reconstruction of Human Evolution - Bringing together Genetic, Archaeological and Linguistic Data", Proc. Natl. Acad. Sci. USA **85**, (16), pp. 6002 (1988).

[50] L.L. Cavalli-Sforza, A. Piazza, P. Menozzi and J. Mountain, "Genetic and Linguistic Evolution", Science, **224** (4909) pp. 1128 (1989).

[51] H.A. Ceccatto and B.A. Huberman, "Persistence of Nonoptimal Strategies", Proceedings of the Natural Academy of Sciences USA, **86**, pp. 3443 (1989).

[52] K. Chen, "A Simple Learning Algorithm for the Traveling Salesman Problem", Brookhaven National Laboratory Preprint (1989).

[53] A.C. Clarke, *"The Nine Billon Names of God"*, Ballantine Books (1953).

[54] F.T. Jr. Cloak, "Is a Cultural Ethology Possible ?", Hum. Ecol. **3**, pp. 161 (1975).

[55] J.P. Cohoon et al., "Punctuated Equilibria: A Parallel Genetic Algorithm", *Genetic Algorithms and their Applications*, J.J. Grefenstette, editor, Lawrence Erlbaum Associated (1987).

[56] J.P. Cohoon et al., "Floorplan Desing Using Distributed Genetic Algorithms", Proceedings of the IEEE International Conference on Computer Aided Design, (Sep. 1988).

[57] G.A. Croes, "A Method for Solving Traveling Salesman Problems", Oper. Res. **6** pp. 791 (1958).

[58] F. Darema, S. Kirkpatrick and V.A. Norton, "Parallel Algorithms for Chip Placement by Simulated Annealing", IBM Journ. Res., **31** (3), pp. 391 (1987).

[59] C. Darwin, *"The Origin of Species"*, John Murray, London (1859).

[60] R. Dawkins, *"The selfish gene"*, Oxford University Press, Oxford (1976).

[61] L. Davis, ed., *"Genetic Algorithms and Simulated Annealing"*, London: Pitman (1987).

[62] K.A. De Jong, "Adaptive System design: A Genetic Approach", IEEE Transactions on Systems, Man, and Cybernetics, **SMC-10** (9), pp. 566 (1980).

[63] B. Derrida and H. Flyvbjerg, "Multivalley Structure in Kauffman's Model: Analogy with Spin Glasses", J. Phys. A: Math. Gen. **19**, L1003 (1986).

[64] B. Derrida, "Valleys and Overlaps in Kauffman Model", Philosophical Magazine B, **56** (6), pp. 917 (1987).

[65] B. Derrida and H. Flyvbjerg, "Statistical Properties of Randomly Broken Objects and of Multivalley Structures in Disordered Systems", J. Phys. A.: Math. Gen. **20** (15), pp. 5273 (1987).

[66] B. Derrida, "Statistical Properties of Valleys in the Annealed Random Map Model", Jour. of Phys. A: Math. Gen. **21** (9), pp. 509 (1988).

[67] B. Derrida and O. Golinelli, "Barrier Heights in the Kauffman Model", Journal de Physique, **50** (13), pp. 1587 (1989).

[68] A.K. Dewdney, "Exploring the Field of Genetic Algorithms in a Primordial Computer Sea Full of Flibs", Scientific American, **253**(5), pp. 21 (1985).

[69] A.K. Dewdney, "Simulated Evolution: Wherein Bugs Learn to Hunt Bacteria", Scientific American, **260**(5), pp. 138 (1989).

[70] G. Dueck and T. Scheuer, "Threshold Accepting", IBM HDSC Technical Report 88.10.011 (1988).

[71] R. Durbin and D. Willshaw, "An Analogous Approach to the Traveling Salesman Problem using an Elastic Net Method", Nature **326** (6114), pp. 689 (1987).

[72] A. Dutta, G. Koehler, and A. Whinston, "On Optimal Allocation in a Distributed Processing Enviroment", Management Science **28**, pp. 839 (1982).

[73] S.F. Edwards and P.W. Anderson, "Theory of Spin Glasses", J. Phys. **F5**, pp. 965 (1975).

[74] E. Felten, S. Karlin, and S.W. Otto, "The Traveling Salesman Problem on a Hypercubic MIMD computer", Proc. International Conference on Parallel Processing, Chicago IL, pp. 6 (1985).

[75] L.J. Fogel, "Autonomous Automata", Ind. Res. **4** pp. 14 (1962).

[76] L.J. Fogel, *"On the Organization of Intellect"*, Ph. D. Dissertation, UCLA (1964).

[77] L.J. Fogel, A.J. Owens and M.J. Walsh, "Artificial Intelligence through Simulated Evolution", Wiley, New York (1966).

[78] L.J. Fogel, "Evolutionary Computers", Technology Review **92** (4), pp. 8 (1989).

[79] D.B. Fogel, "An Evolutionary Approach to the Traveling Salesman Problem", Biol. Cybern. **60**, pp. 139 (1988).

[80] R.P. Feynman, "There's Plenty of Room at the Bottom", Engrg. and Sci. (California Institute of Technology) pp. 22, Feb. (1960).

[81] R.P. Feynman, "Quantum Mechanical Computers", Foundations of Physics **16** (6), pp. 507 (1986).

[82] A. Franks, "Nanotechnology Opportunities", J. Phys. E: Sci. Instrum. 20 Dec., pp. 237 (1987).

[83] A. Franks, "Nanotechnology", J. Phys. E: Sci. Instrum. 20 Dec., pp. 1442 (1987).

[84] G.J. Friedman, "Digital Simulation of an Evolutionary Process", General Systems Yearbook, **4**, pp. 171 (1959).

[85] G.C. Fox, S.W. Otto and E.A. Umland, "Monte Carlo Physics on a Concurrent Processor", Jour. of Stat. Phys. **43**, 1209 (1986).

[86] G.C. Fox et al, *Solving Problems on Concurrent Processors. Vol 1* (Prentice Hall, Englewood Cliffs, New Jersey, 1988).

[87] Y. Fu and P.W. Anderson, "Application of Statistical Mechanics to NP-Complete Problems in Combinatorial Optimization", Journal of Physics A: Math. Gen. **19**, pp. 1605 (1986).

[88] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, 1979.

[89] P. Gilmore, "Optimal and Suboptimal Algorithms for the Quadratic Assingnment Problem", SIAM **10**, pp. 305 (1962).

[90] D.E. Goldberg, *Genetic Algorithms in Search, Optimisation, and Machine Learning*, Addison Wesley (1989).

[91] D.E. Goldberg, "Zen and the Art of Genetic Algorithms", Proceedings of the Third International Conference of Genetic Algorithms, Fairfax, VA, ed. by J.D. Schaffer, (Morgan Kaufmann, San Mateo CA) pp. 80 (1989).

[92] D.E. Goldberg, "Sizing Populations for Serial and Parallel Genetic Algorithms", Proceedings of the Third International Conference of Genetic Algorithms, Fairfax, VA, ed. by J.D. Schaffer, (Morgan Kaufmann, San Mateo CA) pp. 70 (1989).

[93] M. Gorges-Schleuter, "ASPARAGOS An asynchronous Parallel Genetic Optimization Strategy", Proceedings of the Third International Conference of Genetic Algorithms, Fairfax, VA, ed. by J.D. Schaffer, (Morgan Kaufmann, San Mateo CA) pp. 422 (1989).

[94] J.J. Grefenstette, "Incorporating Problem Specific Knowledge into Genetic Algorithms", in *Genetic Algorithms and Simulated Annealing*, L. Davis (ed). London: Pitman (1987).

[95] G.S. Grest, C.M. Soukoulis and K.Levin, "Cooling-Rate Dependence for the Spin-Glass Ground-State Energy: Implications for Optimization by Simulated Annealing", Phys. Rev. Lett. **56**, (11) pp. 1148 (1986).

[96] G.S. Grest et al., "Monte Carlo and Mean Field Slow Cooling Simulations for Spin Glasses: Relation to NP-Completeness", Proceedings of a Colloquium on Spin Glasses, Optimization and Neural Networks. Heidelberg 1986. *Lecture Notes in Physics. Vol 275.* Springer-Verlag (1986).

[97] B. Grossman, "The Origin of the Ultrametric Topology of Spin-Glasses", Jour. of Phys. A: Math. Gen. **22**, pp. 33 (1989).

[98] A.H. Harcourt et al., "Cooperation: The Use of Alliances in Contests", Int. J. Prim. **8** (5) pp. 412 (1987).

[99] S.U. Hegde, *"Efficacy of Parallel Genetic Algorithms"*, Masters Thesis, University of Virginia, Charlottesville VA (August 1988).

[100] W. Herroellen and A. Van Gils, "On the Use of Flow Dominance in Complexity Measures for Facility Layout Problems", International Journal of Production Research **23** pp. 97 (1985).

[101] F. Hiller, "Quantitative Tools for Plant Layout Analysis", Journal of Industrial Engineering **14**, pp. 33 (1963).

[102] F. Hiller and M Connors, "Quadratic Assignment Problem Algorithms and the Location of Indivisible Facilities", Management Science **13**, pp. 42 (1966).

[103] G. Hitchings and M. Cottam, "An Efficient Heuristic for Solving the Layout Design Problem", Omega **14**, pp. 205 (1976).

[104] J.H. Holland, *Adaptation in Natural and Artificial Systems.* Ann Arbor: University of Michigan Press. (1975).

[105] J.J. Hopfield and D.W. Tank, "Neural Computation of Decisions in Optimization Problems", Biol. Cybernetics, **52**, pp. 141 (1985).

[106] B.A. Huberman ed. *The Ecology of Computation*, (North-Holland, Amsterdam) (1988).

[107] L. Hubert and J. Schultz, "Quadratic Assignment as a General Data Analysis Strategy", British Journal of Mathematical Psychology **29**, pp. 190 (1976).

[108] C.L. Huntley, "An Adaptive Aid for Facility Layout", Master Thesis, University of Virginia (1988).

[109] W. Jeffrey and R. Rosner, "Neural Network Processing as a Tool for Function Optimization", Proceeding of the International Conference on Neural Networks, pp. 241 (1986).

[110] O.C. Ivanov, P.B. Milanov and P.S. Kenderov, "Genetic Code Optimality from Mathematical and Evolutionary Points of View", Doklady Bolgarskoi Akademii Nauk **40** (1), pp. 25 (1987).

[111] S. Ji, "Watson-Crick and Prigoginian Forms of Genetic Information", Journal of Theoretical Biology **130** (2), pp. 239 (1988).

[112] I. Kanter and H. Sompolinsky, "Graph Optimization Problems and the Potts Glass", Jour. of Phys. A: Math. Gen. **20** (11), pp. 673 (1987).

[113] S.A. Kauffman and S. Levin, "Towards a General Theory of Adaptive Walks on Rugged Landscapes", J. Theor. Biol. **128**, pp. 11, (1987).

[114] S.A. Kauffman, E.D. Weinberger and A.S. Perelson, "Maturation of the Immune Response via Adaptive Walks on Affinity Landscapes", *Theoretical Immunology, Part One*, Santa Fe Institute in the Sciences of Complexity, Ed. A.S. Perelson, (Reading, MA: Addison-Wesley), vol. III, pp. 349 (1988).

[115] S.A. Kauffman, "The Evolution of Economic Webs", *The Economy as an Evolving Complex System*, Santa Fe Institute in the Sciences of Complexity, Ed. P.W. Anderson, K.J. Arrow and D. Pines, (Reading, MA: Addison-Wesley), vol. V, pp. 125 (1988).

[116] S.A. Kauffman, "Adaptation on Rugged Fitness Landscapes", *Lectures in the Sciences of Complexity*, Ed. D. Stein, Addison-Wesley Longman, pp. 527 (1989).

[117] S.A. Kauffman, *Origins of Order: Self-Organization and Selection in Evolution*, (London: Oxford Univ. Press) in press.

[118] J.O. Kephart, T. Hogg and B.A. Huberman, "Dynamics of computational ecosystems", Phys. Rev. A **40** (1), pp. 404 (1989).

[119] S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi, "Optimization by Simulated Annealing", Science **220**, 671 (1983).

[120] S. Kirkpatrick and G. Toulouse, "Configuration Space Analysis of Traveling Salesman Problems", Journal de Physique **46**, pp. 1277 (1985).

[121] T. Koopmans and M. Beckmann, "Assignment Problems and the Location of Economic Activities", Econometrica **25**, pp. 53-76 (1957).

[122] S.A. Kravitz and R.A. Rutenbar, "Placement by Simulated Annealing on a Multiprocessor", IEEE Transactions on Computer-Aided Design, **6** (4), pp. 534 (1987).

[123] W. Krauth and M. Mezard, "The Cavity Method and the Traveling Salesman Problem", Europhysics Letters, **8** (3), pp. 213 (1989).

[124] E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys, *The Travelling Salesman Problem: A Guided Tour of Combinatorial Optimization*, (Wiley-Interscience, Chichester, 1985).

[125] E. Lawler, "The Quadratic Assignment Problem", Management Science **9**, pp. 586 (1963).

[126] S. Lin, "Computer Solutions of the Traveling Salesman Problem", Bell System Tech. J. **44**, pp. 2245 (1965).

[127] S. Lin and B.W. Kernighan, "An Effective Heuristic Algorithm for the Traveling Salesman Problem", Oper. Res. **21**, 498 (1973).

[128] I. Matsuba, "Optimal Simulated-Annealing Method Based on Stochastic-Dynamic Programming", Phys. Rev. A **39** (5), pp. 2635 (1989).

[129] M. Mezard et al., Journal de Physique **45** pp. 843 (1984).

[130] M. Mezard and G. Parisi, "On the Solution of the Random Link Matching Problems", Journal de Physique, **48** (9), pp. 1451 (1987).

[131] M. Mezard and G. Parisi, "The Euclidean Matching Problem", Journal de Physique, **49** (12), pp. 2019 (1988).

[132] E.N. Miranda and N. Parga, "Ultrametricity in the Kauffman Model: A Numerical Test", Jour. of Phys. A: Math. Gen. **21** (6) pp. 357 (1988).

[133] P. Moscato, "On Genetic Crossover Operators for Relative Order Preservation", Caltech Concurrent Computation Program, Report C3P-778, (1989).

[134] P. Moscato and F. Fontanari, "Stocastic vs. Deterministic Update in Simulated Annealing", Caltech Concurrent Computation Program, Report C3P-789, (1989).

[135] P. Moscato and M. Norman, "A Competitive-Cooperative Approach to Complex Combinatorial Search", Caltech Concurrent Computation Program, Report C3P-790, (1989).

[136] H. Mühlenbein, "New Solutions to the Mapping Problem of Parallel Systems: The Evolution Approach", Parallel Computing, **4**, pp. 269 (1987).

[137] H. Mühlenbein, "Evolution Algorithms in Combinatorial Optimization", Parallel Computing, **7**, pp. 65, (1988).

[138] H. Mühlenbein, "Parallel Genetic Algorithms, Population Genetics and Combinatorial Optimization", Proceedings of the Third International Conference of Genetic Algorithms, Fairfax, VA, ed. by J.D. Schaffer, (Morgan Kaufmann, San Mateo CA) pp. 416 (1989).

[139] G. Napierala, "Ein paralleler Evolutionsalgorithmus zur Lösung des TSP", Master Thesis, University of Bonn (1989).

[140] C.E. Nugent, T.E. Vollmann and J. Ruml, "An Experimental Comparison of Techniques for the Assignment of Facilities to Locations" Operations Research Society of America, Journal, **16**, 150, (1968).

[141] J.D. Nulton and P. Salamon, "Statistical Mechanics of Combinatorial Optimization", Phys. Rev. A **37**(4), 1351 (1988).

[142] T. Ohta, "A Model of Evolution for Accumulating Genetic Information", Journal of Theoretical Biology **124** (2), pp. 199 (1987).

[143] I. Or, *"Traveling Salesman-Type Combinatorial Problems and their Relation to the Logistics of Regional Blood Banking"*, Ph.D. Thesis, Northwestern University, Evanston (1976).

[144] M. Padberg and S. Hong, "On the Symmetric Traveling Salesman Problem: A Computational Study", Math. Programming Stud. **12**, 78 (1980).

[145] M. Padberg and G. Rinaldi, "Optimization of 532-City Symetric TSP", Operations Research Letters, **6**, pp. 1, (1987).

[146] R.G. Palmer, "Statistical Mechanics Approaches to Complex Optimization Problems", *The Economy as an Evolving Complex System, SFI Studies in the Sciences of Complexity* (Reading, MA: Addison-Wesley) Vol V. (1988).

[147] N. Parga, "Overlap Distributions and Taxonomy Analysis of Spin-Glass States with Equal Weights", Journal de Physique **48** (4), pp. 449 (1987).

[148] G. Parisi, "Facing Complexity", Physica Scripta, **35** (2), pp. 123 (1987).

[149] G. Parisi, "On Spin-Glass Theory", Physica Scripta, **T19A**, pp. 27 (1988).

[150] C.B. Pettey et al. "A Parallel Genetic Algorithm", *Genetic Algorithms and their Applications*, J.J. Grefenstette, editor, Lawrence Erlbaum Associated (1987).

[151] C.C. Pettey and M.R. Leuze, "A Theoretical Investigation of a Parallel Genetic Algorithm", Proceedings of the Third International Conference of Genetic Algorithms, Fairfax, VA, ed. by J.D. Schaffer, (Morgan Kaufmann, San Mateo CA) pp. 398 (1989).

[152] J. Platt, *Constraint Methods for Neural Networks and Computer Graphics*, Ph.D. Thesis, California Institute of Technology, Pasadena CA.

[153] K. Popper, "The Rationality of Scientific Revolutions", in *Problems of Scientific Revolution*, R. Harró ed., Clarendon Press, Oxford, pp. 72 (1974).

[154] S. Rajasekaran and J.H. Reif, "Nested Annealing - A Provable Improvement over Simulated Annealing", Lecture Notes in Computer Science, **317**, pp. 455 (1988).

[155] R. Rammal et al., "Ultrametricity for Physicists", Rev. Mod. Phys. **58**, pp. 765 (1986).

[156] I. Rechenberg, "Cybernetic Solution Path of an Experimental Problem", (Royal Aircraft Establishment Translation No. 1122, B.F. Toms, Trans.). Farnsborough Hants: Ministery of Aviation, Royal Aircraft Establishment (1965).

[157] I. Rechenberg, *"Evolutionstrategie"*, Sttutgart: Frommann-Holzboog (1973).

[158] W.T. Rhee, "A Note on Asymptotic Properties of the Quadratic Assignment Problem", Operations Research Letters **7**(4), pp. 197 (1988).

[159] L. Ritzman, "The Efficiency of Computer Algorithms for Plant Layout", Management Science, **18**, pp. 240, (1972).

[160] D.J. Rosenkrantz, R.E. Stearns and P.M. Lewis, "An analysis of several heuristics for the traveling salesman problem", SIAM J. Comput **6**, 563 (1977).

[161] Y. Rossier, M. Troyon, T.M. Liebling, "Probabilistic Exchange Algorithms and Euclidean Traveling Salesman Problems", OR Spektrum, **8**, pp. 151, (1986).

[162] P. Rujan, "A Laplacian Walk for the Traveling Salesman", Europhysics Letters **7** (3) pp. 191 (1988).

[163] P. Rujan, "Searching for Optimal Configurations by Simulated Tunneling", Zeitschrift fur Physik B- Condensed Matter **73**(3), pp. 391 (1988).

[164] R.A. Rutenbar, "Simulated Annealing Algorithms: An Overview", IEEE Circuits and Devices Magazine **5** (1), pp. 19 (1989).

[165] P. Salamon et al., "Simulated Annealing with Constant Thermodynamic Speed", Computer Physics Communications **49**(3), pp. 423 (1988).

[166] J.D. Schaffer et al. "A Study of Control Parameters Affecting Online Performance of Genetic Algorithms for Function Optimization", Proceedings of the Third International Conference on Genetic Algorithms, Fairfax VA, ed. by J.D. Schaffer, (Morgan Kaufmann, San Mateo CA) pp. 81 (1989).

[167] C. Schneiker, "Nanotechnology with Feynman Machines: Scanning Tunneling Engineering and Artificial Life", *Artificial Life*, Santa Fe Institute Studies in the Sciences of Complexity, Vol VI, Ed. C.G. Langton, Addison-Wesley, pp. 443 (1989).

[168] H. Schwefel, *Numerical Optimization of Computer Models*, (M. Finnis, Trans.). Chichester: John Wiley (1981).

[169] M. Scriabin and R. Vergin, "Comparison of Computer Algorithms and Visual Based Methods for Plant Layout", Management Science **22**, pp. 172, (1975).

[170] S.I. Sergeyev, "A New Lower Bound for the Quadratic Assignment Problem", USSR Computational Mathematics and Mathematical Physics **27**, pp. 130 (1987).

[171] D. Sherrington, "Graph Partitioning and Dilute Spin-Glasses: The Minimum Cost Solution", Jour. of Phys. A: Math. Gen. **21** (2), pp. 99 (1988).

[172] K.R. Shoulders, "On Microelectronic Components, Interconnections and System Fabrication", Proc. Western Joint Computer Conference (Palo Alto, CA: National Press) pp. 251 (1960).

[173] K.R. Shoulders, "Toward Complex Systems", Microelectronics and Large Systems (Washington, DC: Spartan Books), pp. 97 (1965).

[174] P. Simic, "Statistical Mechanics as the Undelying Theory of "Elastic" and "Neural" Optimizers", Caltech Concurrent Computation Program Report 787 (1989).

[175] N. Sourlas, "Statistical Mechanics and the Traveling Salesman Problem", Europhysics Letters **12**, pp. 919 (1986).

[176] N. Sourlas, "3-Dimensional Ising Spin-Glass and Ergodicity Breaking", Europhysics Letters, **6** (6), pp. 561 (1988).

[177] N. Sourlas, "Spin-Glass Models as Error-Correcting Codes", Nature, **339** (6227), pp. 693 (1989).

[178] W.R. Stark and L. Kotin, "The Social Metaphor for Distributed Processing", Journal of Parallel and Distributed Computing **7**, pp. 125 (1989).

[179] D. Stein, *Scheduling Dial-a-Ride Transportation Systems: An Asympotic Approach* Ph.D. Thesis, Harvard University, Cambridge, MA (1977).

[180] D.L. Stein, "All Systems Slow: How Everything from Glass to the Economy Takes Shape", The Science, **28** (5), pp. 22 (1988).

[181] D.L. Stein et al., "Mean Exit Times over Fluctuating Barriers", Physics Letters A, **136** (7-8), pp. 353 (1989).

[182] D.L. Stein, "Spin Glasses", Scientific American, **261** (1), pp. 52 (1989).

[183] F.H. Stillinger and T.A. Weber, "Nonlinear Optimization Simplified by Hypersurface Deformation", Jour. of Stat. Phys. **52** (5/6), pp. 1429 (1988).

[184] H. Szu, "Nonconvex Optimization by Fast Simulated Annealing", Proceedings of the IEEE, **75**(11), pp. 1538 (1987).

[185] H. Szu, "Fast Simulated Annealing", Physics Letters A, **122**(3-4), pp. 157 (1987).

[186] R. Tanese, "Parallel Genetic Algorithms for the Hypercube", *Genetic Algorithms and their Applications*, J.J. Grefenstette, editor, Lawrence Erlbaum Associated (1987).

[187] R. Tanese, "Distributed Genetic Algorithms", Proceedings of the Third International Conference of Genetic Algorithms, Fairfax, VA, ed. by J.D. Schaffer, (Morgan Kaufmann, San Mateo CA) pp. 434 (1989).

[188] N. Taniguchi, "On the Basic Concept of Nanotechnology", Proc. Int. Conf. Prod. Eng. Tokyo, Part 2 (Tokoyo: JSPE), pp. 18 (1974).

[189] A. Tramontano and M.F. Macchiato, "Information Value and Information-Content in the Evolutionary Strategy of the Genetic Code", Nuovo Cimento D **10** (3), pp. 293 (1988).

[190] D. Vanderbilt and S.G. Louie, "A Monte Carlo Simulated Annealing Approach to Optimization over Continuous Variables", J. Comput. Phys. **56**, 259 (1984).

[191] M.P. Vecchi and S. Kirkpatrick, IEEE Trans. Comput. Aided Design: Integrated Circuits **CAD-2**, 215 (1983).

[192] T. Vollmann and E. Buffa, "The Facilities Layout Problem in Perspective", Management Science **12**, pp. 450 (1966).

[193] E.D. Weinberger, "A More Rigorous Derivation of Some Properties of Uncorrelated Fitness Landscapes", Journal of Theoretical Biology **134**, pp. 125 (1988).

[194] G.V. Wilson and G.S. Pawley, "On the Stability of the Traveling Salesman Algorithm of Hopfield and Tank", Biol. Cybernetics **58**, pp. 63 (1988).