Optimization by Simulated Annealing: A Review

Aly El Gamal
ECE Department and Coordinated Science Lab
University of Illinois at Urbana-Champaign

Abstract

Prior to the work in [1], heuristic algorithms used to solve complex combinatorial optimization problems, were based on iterative improvements, where in each step, a further decrease in cost is required. This results in locally optimal solutions, that are, for many settings, far from optimal. In this work, we review the *Optimization by Simulated Annealing* algorithm, that permits *uphill* moves with a variable probability. A connection to a statistical mechanical study of the evolution of condensed systems at low temperatures is of particular interest. Concrete examples are used to highlight the advantage of the considered approach to the greedy iterative improvement based counterpart.

I. Introduction

Finding an optimal value (maximum or minimum) for a combinatorial optimization problem, where the objective is a function of many independent parameters, is a problem of significant engineering interest, due to its relative high frequency in practice. One well known example is the traveling salesman problem (TSP), where, given locations of N distinct cities, and a defined distance metric, the objective is to traverse the set of cities with the minimum possible route length. It is natural to first consider seeking a systematic framework to find such an optimum. However, this characterization is possible only for special cases. Moreover, some optimization problems like TSP belong to the class of NP-hard problems, which means that the computational complexity needed to find a solution is non-polynomial in the arguments (number of cities). Instead, heuristics are commonly employed to reach a solution by iteratively moving in a descent direction. These solutions can guarantee to only stop at a point that is optimal with respect to at least one of its neighborhoods (local optimum).

In [1], the authors proposed a heuristic method for the general optimization problem, that is based on an observation of the behavior of matters with a large number of atoms, in thermal equilibrium at the limit of low temperature. In particular, noting that melting a crystal, then rapidly cooling the substance does not lead it back to its original state, rather a metastable configuration (glass), or a defected crystal. On the other hand, slowly cooling the substance, such that it does not get out of equilibrium at any given temperature, can rearrange the atoms in their original structure. This process is known as annealing.

In this review, we are interested in highlighting the analogy between methods of simulating the annealing process, to obtain statistical mechanical models for the low-temperature configuration of atoms for a given substance, and

the proposed approach to find an optimal solution for a general optimization problem. To this end, some physical definitions have to carry over to the model of interest. Assuming our optimization problem is concerned with finding a minimum for the objective function, this cost function will be the analogous to the notion of energy of a configuration of atoms. while the *configuration*, will be a given assignment to the function parameters. A challenging task, would be to define a temperature for the combinatorial optimization problem. In other words, one needs to define a distance between different arrangements or assignments to the function parameters. Consequently, at high temperatures, drastic rearrangements are allowed, while as we cool down, only those changes within a local neighborhood of the current configuration, can take place.

We start by an overview of the annealing process, that highlights useful observations for the proposed combinatorial optimization solution in Section II. We then present the proposed approach in Section III, and two case studies, that illustrate, through numerical results the advantage over greedy iterative improvement based approaches. Finally, in Section V, we end the paper with a brief summary.

II. Statistical Mechanics: Annealing [2]

In this Section, we are interested in shedding the light on studies in the discipline of statistical mechanics, that are concerned with finding the low-temperature equilibrium state of a system with a large number of atoms. This state is known as the ground or zero-energy state of the substance. Algorithms that simulate the ground state of substances are based on iterative improvements, where in each step a new rearrangement of atoms is produced. Alike the optimization counterpart, it is natural to first consider rapidly cooling the melt to obtain the original configuration of atoms in the crystal, at the freezing point. That corresponds to accepting only rearrangement with lower energy, at each step. This observation, along with the two facts, that rapid cooling does not lead the substance to its original crystal, rather a defected crystal or a metastable configuration, and that iterative improvement based algorithms can only result in locally optimal solutions, suggests the similarity between the two models.

In [2], the authors introduced a novel algorithm to simulate the ground state of atoms in thermal equilibrium, that can accept rearrangements with higher energy, at any given step. The key idea is to prevent the algorithm from being stuck at local optima, by allowing *uphill* moves. In each step, the algorithm executes as follows,

- An atom is given a small random displacement from its position in the current configuration.
- The energy of the new configuration $\left(\bar{E}\right)$ is computed, and the change in energy

$$\Delta E = \bar{E} - E$$

  where E is the energy of the current configuration.
- if $\Delta E \leq 0$, then the displacement is accepted.
- if $\Delta E > 0$, then the displacement is accepted with probability,

$$P\left(\Delta E\right) = e^{\frac{-\Delta E}{k_B T}} \tag{1}$$

  where k_B is Boltzmann's constant, and T is the temperature of the system.

We note from (1), that at high temperatures, rearrangements with higher positive energy change, are likely to take place. While as the system cools down, more restrictions, in a stochastic sense, are imposed on the set of acceptable displacements. One can use the above algorithm to find the ground state of a substance, by first melting it to an effectively high temperature, i.e. starting with an initial random arrangement, then slowly cooling down, while running the algorithm at each temperature, possibly many times. As in carefully annealing physical materials. Spending long enough time near the freezing point, assures that the system reaches equilibrium. Thus, large enough number of rearrangements should be attempted at each temperature.

In combinatorial optimization problems, it is often the case that the objective is a function of conflicting requirements, thereby the final solution achieves the optimal trade off between these parameters, with respect to a defined cost. In that regard, one would identify a major difference between annealing a system where the ground state is a regular crystal and the optimization analog problem. Fortunately, nature provides us with systems where competing orderings of atoms are simultaneously present. A well studied example, where this feature, known as *frustration* is present, is that of magnetic alloys known as spin glasses, which exhibit competition between ferromagnetic and anti ferromagnetic spin orderings. This is particularly useful to find a faithful physical analog for complex optimization problems with conflicting requirements.

As will be clear throughout the examples in the next Section, a fundamental challenge to iterative improvement based approaches, is that of identifying global features of the objective function, from local views. In other words, how to extract macroscopic properties from microscopic average. For a physical annealing process, a quantity called the *specific heat* in statistical mechanics is important for this purpose. Let $E(T)$ denote the average energy of all possible configurations of the atoms, at a given temperature T. The specific heat is given by,

$$C(T) = \frac{dE(T)}{dT} \tag{2}$$

The larger the value of $C(T)$ is, the slower the cooling process should be, as the energy of the system becomes more sensitive to changes in temperature.

III. Optimization by Simulated Annealing

For complex optimization problems, where a large number of independent parameters determine each instance of the problem, finding an algorithm that reaches the exact optimal solution in a time that is polynomial in the number of arguments may not be feasible. However, as a direct consequence of the law of large numbers, one may not need to assess such an algorithm by its worst case performance, the average performance may be more appropriate in that case. In this Section, we study important aspects of the *Optimization by Simulated Annealing* algorithm, first introduced in [1]. The key observation is that of the average behavior of atoms in condensed matter systems at ground states.

One useful result of the analogy between some combinatorial optimization problems and the study of system's behavior under annealing in statistical mechanics, is the possibly gained insight into the structure of the problem. In Section IV-A, this point will be clarified through an example, where finding the physical analog to the problem,

helps identifying the relative value of the objective function near the minimum. Accordingly, the design of the heuristic can be simplified by relaxing the requirements.

The following natural question arises at this point. Given an optimization problem, that is a typical candidate for using the simulated annealing heuristic (e.g. N-P hard, many independent parameters,...etc), how would one model it such that the corresponding physical model becomes obvious?. In particular, we will be concerned with defining the analogs to the following physical quantities,

- **Configuration:** Each configuration of the parameters of the objective function, correspond to a configuration of atoms in the condensed matter.
- **Energy:** An appropriate cost function should be set, such that the corresponding configuration of parameters to the ground state, incurs a minimum cost.
- **Temperature:** A real valued mapping $f : R \rightarrow R$ should be defined between the set of temperatures, and a set of values, that has the same units as the cost function. Only rearrangements with cost below $f(T)$ are allowed at a given temperature T.

We note the increasing difficulty in defining the three quantities. While determining the parameters contributing to the objective function might look like an easy task, defining the appropriate weights of those parameters is, in many cases, not as obvious. At the same time, the latter task is crucial to the structure of the problem and the efficiency of the used heuristic. It is worth noting at this point, that a major advantage of the simulated annealing algorithm, is that it might facilitate the choice of the objective function, as it identifies the expected performance of the solution through the physical analog. This will be clear in the example in Section IV-A. Finally, defining the *temperature map* and annealing schedule, usually require a *trial and error* approach, as it determines the behavior of the algorithm. One should be careful to let the algorithm be tolerant enough to random arrangements at effectively high temperatures, and the band of allowed rearrangements be narrow enough at low temperatures, such that the optimal point is reached, with a sufficiently high probability, at the freezing point.

IV. Case Studies

Since the algorithm under consideration is a heuristic, it is necessary to study concrete examples, where some of its frequently arising features are present. To this end, we review the examples discussed in [1], and connect some of the results to the ideas discussed above. We first define a systematic framework, within which, we study each of the below examples.

For each of the discussed problems,

- We start by a brief introduction of the problem, and its use in practice.
- We then provide an intuitive explanation to why this example is appropriate for the simulated annealing algorithm, and its advantage over greedy iterative improvements.
- A model is then formulated for the problem, by first determining the function parameters, then the objective function, and finally the temperature map, as discussed above.

- A possibly available physical analog is used to provide expectations of the algorithm performance, and the structure of the optimization problem solution(s).
- Numerical results are provided, that show the advantage of using the simulated annealing algorithm, while highlighting some of its important features, e.g. identifying the analog of *specific heat* in the context of the optimization problem.

A. Physical Design of Computers

The design of electronic circuits is commonly divided into two major tasks. The first is the logical design, where the basic logical components, and connections between them are determined. Then follows the physical design, which is concerned with the placement of the listed circuits and the detailed map of their wirings. We are interested in studying ways for using the simulated annealing to handle the second task, and analyzing their performance. First, the placement of circuits is determined, then the physical routes of the wires. Due to the possibly large size of the problem, and the different requirements at each design stage, it is typically the case where further subdivsion is required. However, one major drawback of having a large number of sub-problems is the possible loss of accuracy of the final solution. The reason behind that, shall be clear in the following discussion of the second step in our framework.

We note that the solution to each design stage, defines the structure of the problem of a subsequent stage. While a partitioning algorithm would search for a configuration that groups circuits into logically meaningful clusters, it shall inherit the following requirements from the wiring stage,

- Overlaps and areas of high wire density (congestion) should be avoided, to reduce the inter-signal thermal noise.
- Wires should not be too long, to meet a speed requirement of the overall system.

Hence, further subdivision of the original problem, increases the difficulty of determining the appropriate objective function at each stage. Here, we can see an advantage of the considered heuristic's ability to handle complex combinatorial optimization problems. Also, it is clear that, the inaccurate ouput of a greedy algorithm causes an error that propagates to subsequent stages. This suggests the utility of simulated annealing in this context.

1) Coarse Placement, Partitioning: Consider the following problem, where we have N circuits and two chips. A placement algorithm that determines which chip, each of the circuits should belong, is needed. We start the model of the objective function parameters by N binary variables $\mu_i = \pm 1, i \in \{1, 2, \ldots, N\}$, where $\mu_i = 1$ if the i^{th} circuit is placed on the first chip, and $\mu_i = -1$ otherwise. An obvious requirement, is to minimize the number of wires passing through the boundary between the two chips. For that, we need to define a formula to compute that quantity, and insert it into the objective function. Let a_{ij} be a binary variable that takes the value 1 if circuits i and j are connected, and 0 otherwise. The number of cross boundary connections can then be found as,

$$c_1 = \frac{1}{2} \sum_{i \neq j, i, j \in \{1, \ldots, N\}} a_{ij} \frac{(\mu_i - \mu_j)^2}{4} \tag{3}$$

Alternatively, one could write the above formula as,

$$c_1 = \frac{1}{2} \sum_{i \neq j, i,j \in \{1,\ldots,N\}} a_{ij} \mathbf{1}_{\mu_i \neq \mu_j} \tag{4}$$

While the second equation is intuitively easier to understand, the former structure make it easier to extend the objective function to include other requirements, and to find a matching energy equation for a physical system. Having only this requirement, a trivial optimal solution would set all the μ variables to have the same sign. However, it is typically a requirement to balance the distribution of circuits on different chips, as it facilitates the selection of wire paths in a subsequent design stage. A cost of a possibly imbalanced distribution can be,

$$c_2 = \frac{1}{2} \sum_{i \neq j} \mu_i \mu_j \tag{5}$$

We note that $c_2 = 0$ when the chips are equally distributed between the two chips. A simple choice of the objective function that is linear in both c_1 and c_2 is,

$$f = w_1 c_1 + w_2 c_2 \tag{6}$$

Letting $\lambda = \frac{w_2}{w_1}$, and noting that μ_1^2 and μ_2^2 are fixed, the above function can be simplified to,

$$f = \sum_{i>j} (\lambda - \frac{a_{ij}}{2}) \mu_i \mu_j \tag{7}$$

Under the assumption that the connectivity variables a_{ij} are uncorrelated, the function f has the same form as an energy function of a frustrated magnet. where the physical term *frustration*, is the analogous to conflicting requirements of the combinatorial optimization problem. The physical analog to this problem has the following properties,

- There are many different *degenerate* ground states with almost equal energies.
- The average difference in energy between a random high temperature state and one of the ground states is fairly high.
- Reaching one ground state configuration from another, requires a relatively large number of rearrangements.

As mentioned above, a major advantage of the simulated annealing algorithm, is the possible conclusions that can be made prior to running the algorithm, through the physical analog. The above properties of the analog frustrated magnet leads to the following conclusions about the partitioning problem (with respect to order),

- There are many distributions of the circuits to the two chips, that lead to *nearly optimal* solutions to the objective function.
- Even with conflicting requirements, simulated annealing results in solutions that are, on average, significantly better than possible random distributions.
- Reaching an approximate solution by simulated annealing, may be fairly cheaper in terms of runtime, than an alternative that guarantees reaching the exact minimum, yet the first property concludes that the error of the former solution is small.

Next, we show numerical results for a partitioning problem of the logical design of an IBM microprocessor. Throughout the rest of the paper, we assume the temperature map to be the identity map. For this example, we also define the schedule of temperatures and the *time* spent at each temperature (*annealing schedule*) as,

$$T(k) = \gamma T(k-1) \tag{8}$$

where γ is a temperature attenuation factor ($0 < \gamma < 1$). Also, the index k changes whenever the number of different rearrangements reached at temperature $T(k)$, is N_k, or the number of attempted rearrangements is MN_k. As in the statistical mechanical analog, it is usually preferable to spend a longer time near the freezing point. However, maybe due to the availability of numerous ground states in that example, in [1], the authors selected $N_k = n, \forall k$, for this simulation.

Figure 1 plots the results of using the simulated annealing algorithm, at different temperatures, for a similarly formalized partioning problem, where the objective is to minimize the total number of pins required for both chips. The $x-$axis shows the sum number of external connections (pins) over the two chips, while the $y-$coordinate represents the corresponding probability at the stated temperature value. The width of each band represents the number of possible configurations at each temperature. We note the following observations from the figure,

- The $T = \infty$ configuration is an initial random arrangement.
- The system reaches equilibrium as we cool down (sharpening intervals).
- Simulated annealing achieves considerable improvement over a greedy iterative improvement approach (result pointed at by the arrow pointer), for this example.
- The fact that the *band width* decreases with lower temperature values is the statistical mechanical analog of the decrease of entropy of the system as we cool down.

2) Fine Placement: After the partitioning of the circuits involved in the logical design, comes the task of selecting their exact locations. We consider a simplified problem of placing 98 chips, where the space available onboard is discretized to a 10×10 grid. Each square in the grid can be occupied by only one chip.

The requirements of the objective function in this design stage, are similar to, but different from, the ones above defining (6). While the notion of *cross* connection, was natural above, due to the presence of different partitions, we define *regions* for the considered problem and identify accordingly the number of cross connections. The definition of regions is also useful for defining the cost function of the other requirement, which aims at avoiding congestion of wires to reduce thermal noise. The latter requirement is imposed by defining a capacity for each region. The number of wires passing through a given region should not exceed its capacity for the final solution. By taking into consideration the structure of the problem, the exact characterization of the objective function follows in a similar fashion to that of Section IV-A.1, and is omitted for brevity.

Snapshots of the numerical results of this simulation, at selected temperatures, is given in Figure 2. In part a, a manual configuration is given, where we note that functionally dependent blocks (having the same marking), are placed together. Also, the density of wires increases around the center, most likely as the chips carrying the most time sensitive operations are placed in the middle to be easily reachable from anywhere else. In part b, the system

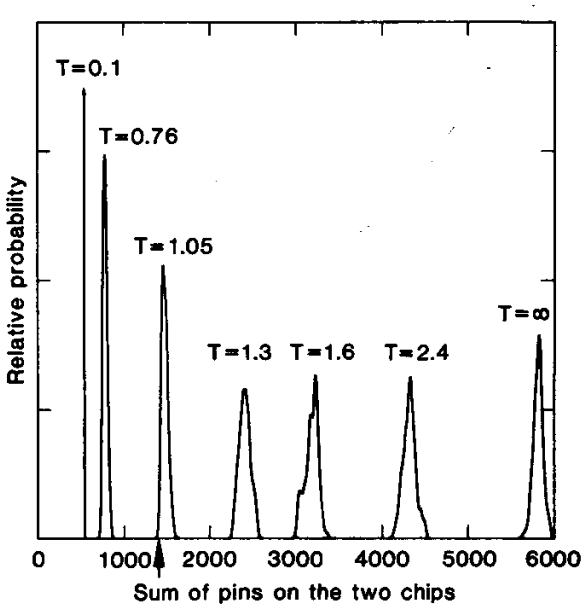

Fig. 1. Distribution of sum number of external pins required in two-way partition of the logical design of an IBM 370 microprocessor, with 5000 circuits and 200 external signals. The following values are used, $\frac{w_2}{w_1} = 0.01, \gamma = 0.9, n = 10, M = 100$. The arrow points out to the best result of multiple runs of greedy iterative improvement algorithms.

is melted, so a random arrangement is selected. As the system cools down (and a sufficiently large number of rearrangements is attempted at each temperature), we note the following from parts c and d,

- Functionally dependent chips get closer, as a result of the cost of the number of cross connections.
- One major difference from the initial manual design, is that groups of chips are now set apart, affected by the cost of congestion.
- At the freezing point, the wire distribution is flattened out, and the peaks are significantly lower than the manual design.

Let $\bar{C}(T)$ be the derivative, with respect to temperature, of the average value of the objective function observed at a temperature T, where the average is taken over all accepted rearrangements at T. As shown in Figure 3, we note that value of $\bar{C}(T)$ have two peaks, the first when orderings start to arise (functionally dependent blocks group together). The second peak happens when clusters start to form. The reason behind this relation, is that $\bar{C}(T)$ can be considered as the analog of the thermodynamic specific heat, which, as mentioned above, carries information about the *big picture*. In other words, another advantage of using simulated annealing, is that of extracting macroscopic properties, by observing the rate of change of local averages at each temperature.

B. Traveling Salesman Problem

One well known combinatorial optimization problem belonging to the set of NP-complete problems, is the traveling salesman problem. The problem definition of TSP involves N cities, and distances between cities are

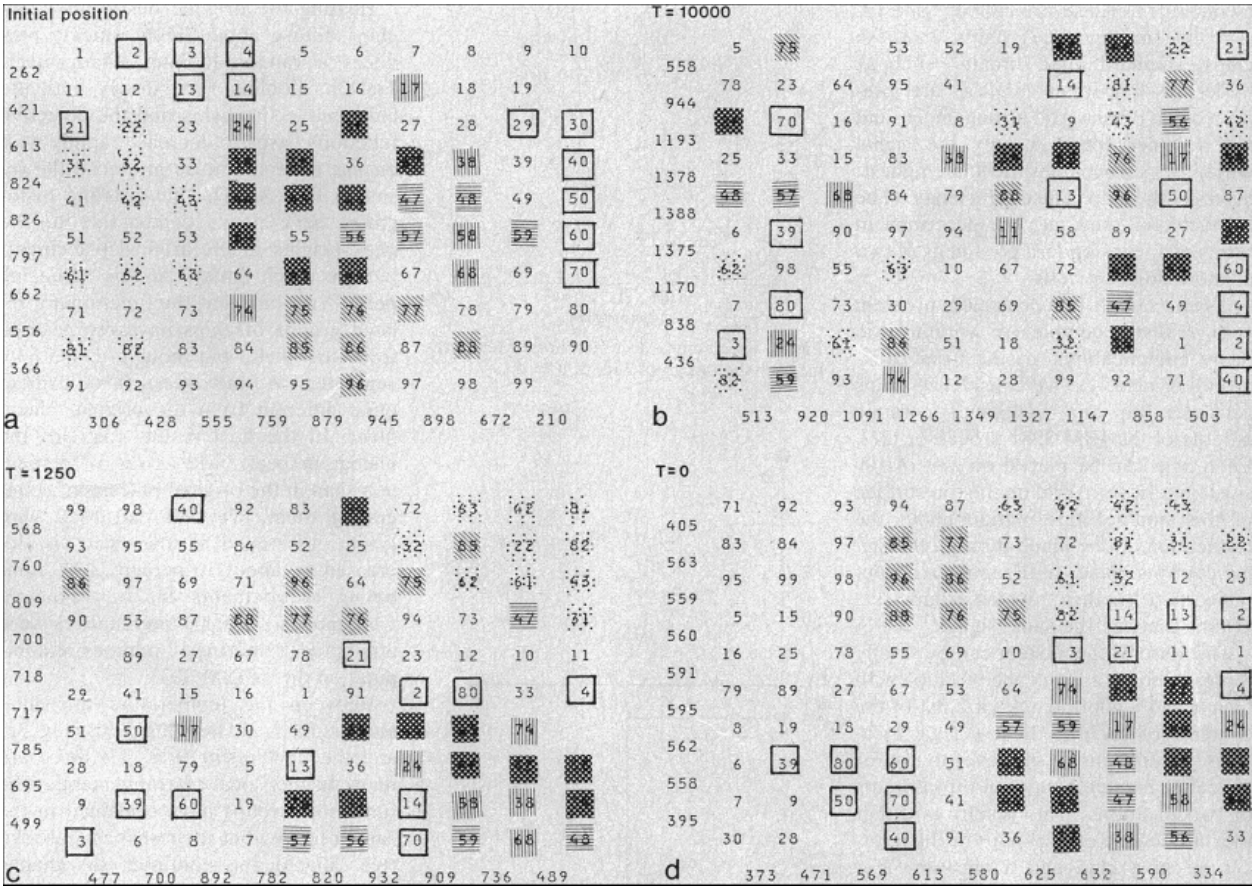

Fig. 2. Ninety-eight chips on a ceramic module from the IBM 3081. Different markings of chips, denote different functionalities. The bottom and left numbers placed between vertical and horizontal lines, respectively, denote the number of connections across the corresponding lines. (a)manual design; (b) a configuration at T=10,000; (c) T=1250; (d) final solution at freezing point.

specified according to a defined metric, and the goal is to find a traversing route of all the cities from an initial point, with minimal length.

The objective of TSP is a function of only one requirement, that is, the length of the traversing path. Hence, a distance metric is needed. The Manhattan distance is used throughout this example for this purpose. In order to make the performance analysis generic, we require the average distance between a pair of cities to be independent of N. To this end, we let the area of the space scale with N^2. In other words, let the cities be represented by randomly selected points in a square with side length $N^{\frac{1}{2}}$. In a realistic setting, cities are typically distributed in dense groups. To model this phenomenon, the space used in the simulations below, has nine distinct regions that can be occupied by the randomly located cities, and unpopulated gaps are present in between. Finally, it is worth mentioning that the results shown below for the simulated annealing algorithm, are found using an empirically generated annealing schedule.

One can easily draw similar conclusions to the above examples, by carefully considering Figure 4. As the

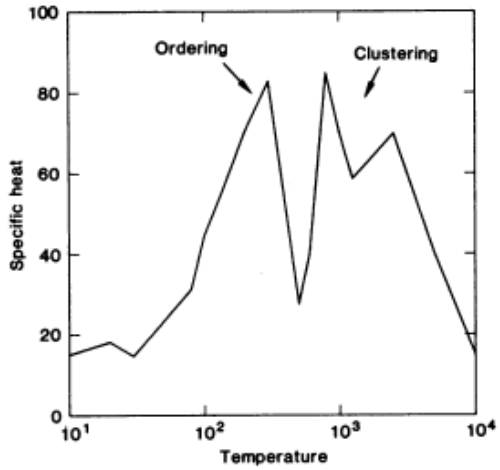

Fig. 3. $\bar{C}(T)$ for the simulation of Figure 2. We note the relation between the specific heat and coarse features of the system.

temperature is lowered, the result configurations have lower cost (on average). A locally optimal solution is reached at the freezing point. Macroscopic properties of the final solution start to appear at intermediate temperatures, this would correspond to the absence of single links crossing more than one gap. However, one should note that this desired performance is guaranteed only in a stochastic sense (on average). There may exist certain arrangement of cities, where the simulated annealing algorithm can have worse performance than deterministic or greedy iterative improvement based approaches.

V. Summary

Based on the work in [1], we provided a review of the simulated annealing algorithm for complex combinatorial optimization problems, where finding an exact solution is computationally expensive. Connection to the statistical mechanical study of finding the ground state(s) of condensed systems was drawn. Concrete examples that justify the utitlity of the considered algorithm, were given. In particular, analogs to theormodynamic specific heat and entropy, were found useful. Finally, numerical results, aided with graphical illustration, were given, and behavior of the considered systems near the freezing point was highlighted. It is worth mentioning, that the considered algorithm was independently described in [3] by V. Cerny.

References

[1] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by Simulated Annealing", *Science*, vol. 220, pp. 671–680, 1983.

[2] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, "Equation of state calculations by Fast Computing Machines", *The Journal of Chemical Physics*, vol. 21, 1953.

[3] V. Cerny, "A thermodynamical approach to the Travelling Salesman Problem:An efficient simulation algorithm.", *Journal of Optimization Theory and Applications*, vol. 45, pp. 41–51, 1985.

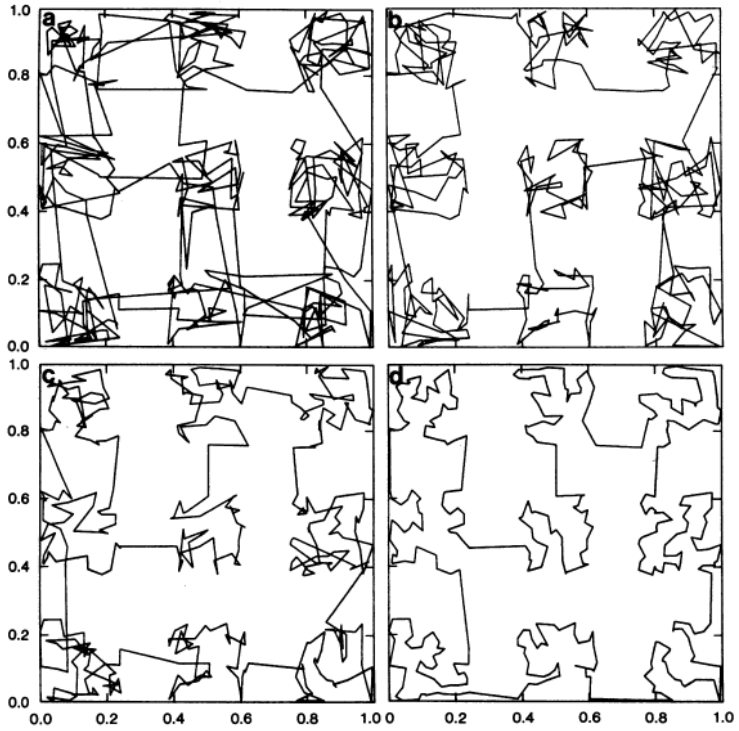

Fig. 4. Simulation Results for a $400-$city traveling salesman problem (a) T=1.2; (b)T=0.8; (c)T=0.4; (d)T=0.