Criticality and Parallelism in Combinatorial Optimization*

William G. Macready*
Athanassios G. Siapas†
Stuart A. Kauffman*

* Santa Fe Institute,
1399 Hyde Park Road,
Santa Fe, NM 87501

† Dept. of Electrical Engineering and Computer Science,
Massachusetts Institute of Technology,
Cambridge, MA 02139

November 15, 1995

Abstract

Local search methods constitute one of the most successful approaches to solving large-scale combinatorial optimization problems. A new result concerning the parallelization of such methods is presented. As parallelism is increased, optimization performance initially improves, but then abruptly degrades to no better than random search beyond a certain point. The existence of this transition is demonstrated for a family of generalized spin-glass models and the Traveling Salesman Problem. Finite-size scaling is used to characterize size-dependent effects near the transition and analytical insight is obtained through a mean field approximation.

*to be published in *Science*

Optimization tasks are common and very often uncommonly difficult. In most areas of science and engineering, from free energy minimization in physics to profit maximization in economics, the need to optimize is ubiquitous. In light of the importance and difficulty of optimization, much effort has gone into developing effective algorithms for finding good optima. The methods of simulated annealing [1], genetic algorithms [2], and taboo search [3] are three of the most popular techniques, inspired by ideas from statistical mechanics, evolutionary biology, and operations research respectively. All of these methods rely in part on constructing improved solutions by applying a local operator to a population of candidate solutions. Good solutions result from the accumulation of many beneficial local modifications applied one after another. An obvious speed up in the algorithm's performance can be gained if we apply many local modifications in parallel. Despite the promise that parallel algorithms hold, they have received less attention and little is known about them [4].

In this report we investigate the effects of parallelizing local search for combinatorial optimization. We demonstrate that for a wide class of search techniques increasing parallelism leads to better solutions faster. But only up to a certain point. At some degree of parallelism the quality of solutions abruptly degrades to that of random sampling. This transition is sharp and will be shown to share many of the characteristics of thermodynamic phase transitions.

We compute critical exponents that characterize this transition using finite-size scaling, a technique borrowed from statistical physics, and demonstrate our results on two important problems: energy minimization on NK energy functions [5], and tour length minimization for the traveling salesman problem (TSP) [6]. The NK model, a generalization of spin-glass models [7], was chosen as one of the few general models capable of generating tunably difficult optimization tasks. The TSP is perhaps the most famous and well studied combinatorial optimization problem and is often used as a test-bed for new ideas. Both problems, though quite different, show remarkably similar behavior.

The NK model [5] defines a family of energy functions over a discrete search space. The search space consists of all possible configurations $\{s\} \equiv \{s_1, \ldots, s_N\}$ of N variables. If each of the variables can take any of A possible values, the search space is of size A^N. We confine our investigation to the case $A = 2$, call our variables spins, and let them take the values ± 1 due to their similarity with Ising spins which arise in models of magnetism. Our

local modifications for the NK model are single spin flips where $s_i \rightarrow -s_i$. Each spin, s_i, makes a contribution to the total energy dependent on its own state and the state of K other spins. These K other spins may be selected at random or according to some specified topology. We have considered both selection mechanisms obtaining similar results in both cases. The total energy of a configuration, $\{s\}$, is defined by:

$$E\{s\} = \frac{1}{N}\sum_{i=1}^{N} E_i(s_i; s_{i_1}, \cdots, s_{i_K}) \tag{1}$$

By analogy with spin glasses [7] the local energy contributions, $E_i(s_i; s_{i_1}, \cdots, s_{i_K})$, for each of the 2^{K+1} local spin configurations, $\{s_i, s_{i_1}, \cdots, s_{i_K}\}$, are chosen at random from a uniform distribution over [0,1). However, by specializing E_i and the selection of the K neighbors we can investigate specific optimization problems, including spin-glasses, graph-coloring, number partitioning *etc.*

The number of other spins, K, that each spin interacts with, varies from 0 to $N-1$ and controls the ruggedness of the energy landscape. In the $K = 0$ limit the spins are independent and there is a single minimum. As K increases the number of conflicting constraints increases leading to multiple local minima that can trap local search algorithms. At the $K = N-1$ extreme, every spin affects every other spin, energies of adjacent configurations are uncorrelated, and there are exponentially many local minima. The $K \gg 1$ limit is analytically tractable and has been studied by many authors [8].

We focus on simulated annealing [1] as our representative local search algorithm. In simulated annealing, local modifications are accepted according to the *Metropolis criterion* [9], a method for simulating the evolution of a physical system in a heat bath to thermal equilibrium. A modification is applied and the resulting change in energy, ΔE, is computed. The modification is always accepted if $\Delta E \leq 0$,and accepted with probability $p(\Delta E) = \exp(-\Delta E/T)$ if $\Delta E > 0$. T is a *temperature* parameter that controls the fraction of uphill moves that are accepted; at zero temperature, $T = 0$, only downhill moves are accepted. It is important to occasionally accept uphill moves (*i.e.* $T > 0$) to prevent trapping on poor local minima where *all* local modifications raise the energy. Simulated annealing derives its name from the *annealing*, or gradual lowering, of the temperature parameter, so that uphill moves are accepted with decreasing frequency.

Beyond the obvious computational speedups of applying many local modifications simultaneously, parallel local search also makes it harder to get

trapped in poor local optima. Even when only greedy moves are accepted ($T = 0$), allowing for simultaneous spin updates introduces outdated information and the possibility of "mistakes": if the energy of s_i depends on s_j and both spins simultaneously attempt to flip, s_i may conclude that $\Delta E < 0$ under the assumption that s_j did not flip. However, ΔE may actually be positive if s_j did flip. This similarity between parallelism and temperature is only superficial: we will demonstrate that parallelism has qualitatively different effects on the behavior of optimization algorithms.

We parameterize the degree of parallelism by $0 < \tau \leq 1$ which denotes the probability that a spin attempts to flip. If there are N spins, then on average $N\tau$ of them are updating under the local operator at the same time. The $\tau = 1/N \to 0$ limit corresponds to a sequential algorithm, which at $T = 0$ will result in the system trapping into (usually poor) local optima. Maximal parallelism is obtained for $\tau = 1$ when all spins update simultaneously. At this extreme, the system typically cannot converge since too many of the spins use outdated information.

The transition between these two extremes is surprisingly sharp. Figure 1 shows the asymptotic energy obtained under a search at fixed $T = 0$ [10]. Identical results are obtained under annealing when the temperature is cooled exponentially. An interesting feature of Figure 1 is the initial improvement in optimizing performance with increasing τ. This improvement is most likely caused by the fact that parallel updating prevents premature trapping in poor local optima and guides the search towards low energy regions of configuration space. Premature convergence is prevented due to the fact that initially many spin flips are energy decreasing and are therefore accepted in parallel resulting in large-scale configuration changes. Since each spin flip is individually energy decreasing these large-scale changes move the system in a direction which, on average, appears to lead to low energy regions of configuration space. Numerical experiments support this conjecture: local energy minima around the configuration at which the parallel dynamics converges have statistically significant lower energies than randomly sampled local minima. As τ increases, more spins flip in parallel, more averaging is done and coarser trends of the energy landscape are discerned. Therefore the energy decrease with τ should be more significant for smoother landscapes (smaller K) where there are coarse features that can be exploited. This prediction is supported by the data in Figure 1

These observations point out two key differences between parallelism induced noise and temperature induced noise: (1) no phase transition occurs

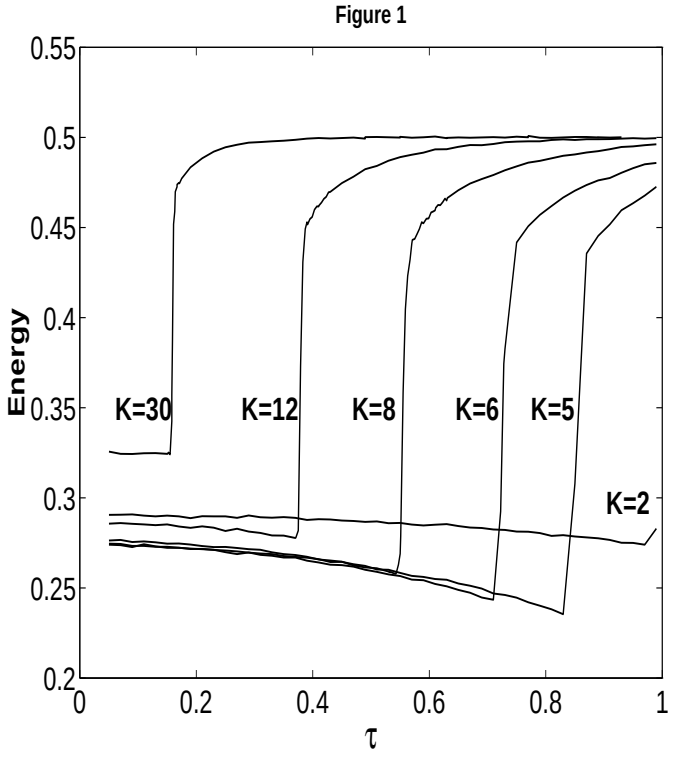

Figure 1: Energy reached at the end of 500 generations versus τ for $N = 5000$, temperature $T = 0$ and several different K values. The search space is vast, of size $2^{5000} \approx 10^{1500}$. A *generation* is defined to be N update attempts, an average $N\tau$ of which are simultaneous. The average over 30 randomly generated NK landscapes is plotted for each τ value.

as temperature is varied, and (2) in contrast to temperature induced noise which only depends on local energy differences, parallelism induced noise also depends on long range correlations in the energy landscape.

The second interesting feature of Figure 1, the abrupt degradation in optimizing performance, is associated with an order-disorder transition in the update dynamics. Above the transition where the energy remains high, spins continue to flip indefinitely. Below the transition, reaching low energy configurations is accompanied by a freezing of the spins (at $T = 0$) to particular values. This motivates the definition of the following *order parameter* which distinguishes the behavior on either side of the transition: Let p_i be the probability that at equilibrium the i^{th} spin will flip if asked. We define the order parameter for the i^{th} spin to be its entropy, $S_i = -p_i \log_2 p_i - (1 - p_i) \log_2 (1 - p_i)$. This is 0 if the spin is frozen and 1 if the spin is flipping randomly. Images of the spatial distribution of S_i for a lattice of spins can be found in Figure 2. The order parameter, S, for the entire system is defined as the average entropy per spin, $S = (1/N) \sum_{i=1}^{N} S_i$. Plots of the order parameter for $K = 6$ near the transition can be found in Figure 3(a). The order parameter is close to 0 below the critical point τ_c and rises sharply towards 1 above τ_c.

The sharpness of the transition, as well as the location of τ_c, depend upon the size N of the system [11]. We use finite-size scaling [12], a method from statistical physics in which the observation of how the critical point $\tau_c(N)$ changes with the size of the system gives direct evidence for critical behavior at the transition. As the size of the system N increases, the transition sharpens and the transition point shifts according to:

$$\tau_c(N) - \tau_c(\infty) \sim N^{-1/\nu} \tag{2}$$

Figure 3(b) shows the results of the finite-size scaling analysis for $K = 6$. The empirical observation behind this analysis is that sufficiently close to the critical point, systems of all sizes are indistinguishable except for an overall change of scale. By defining a rescaled parameter as

$$y = N^{1/\nu} (\tau - \tau_c(\infty)) / \tau_c(\infty) \tag{3}$$

the rescaled curves fall on a universal (N-independent) curve (Figure 3(c)).

Analytical insight into the transition for the NK model can be obtained by making two approximations: we assume that all spins have the same energy E_t at time t (mean-field approximation) and that energies induced by

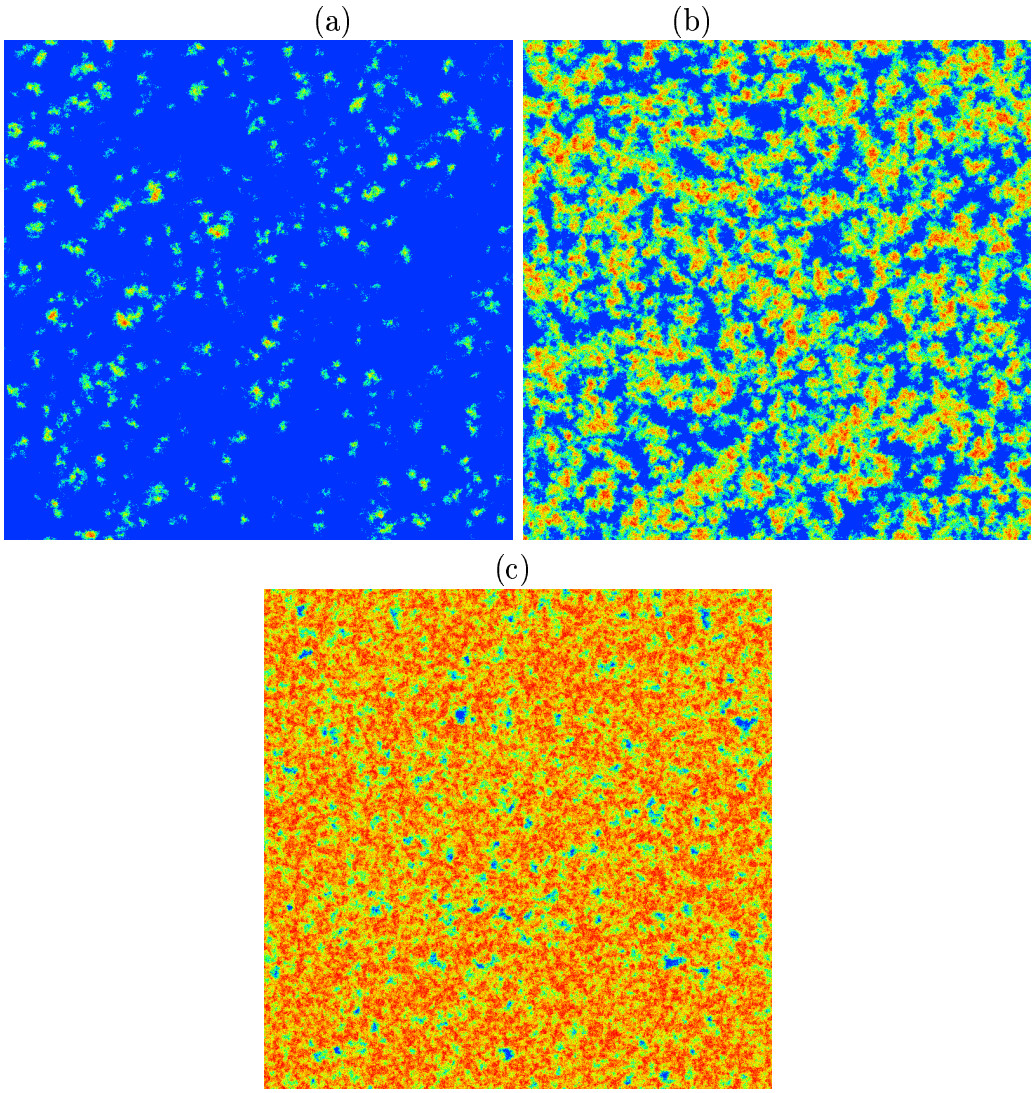

Figure 2: Rendering of the entropy fields for a 800×800 lattice of spins at $T = 0$. The energy of each spin depends on the states of the nearest $K = 12$ spins in the lattice. The color of each spin represents the entropy computed over the last 40 of a total of 80 generations. A rainbow color map is used with the violet end of the spectrum corresponding to low entropy. The images correspond to (a) $\tau = 0.43 < \tau_c$, (b) $\tau = 0.47 \approx \tau_c$, (c) $\tau = 0.51 > \tau_c$.

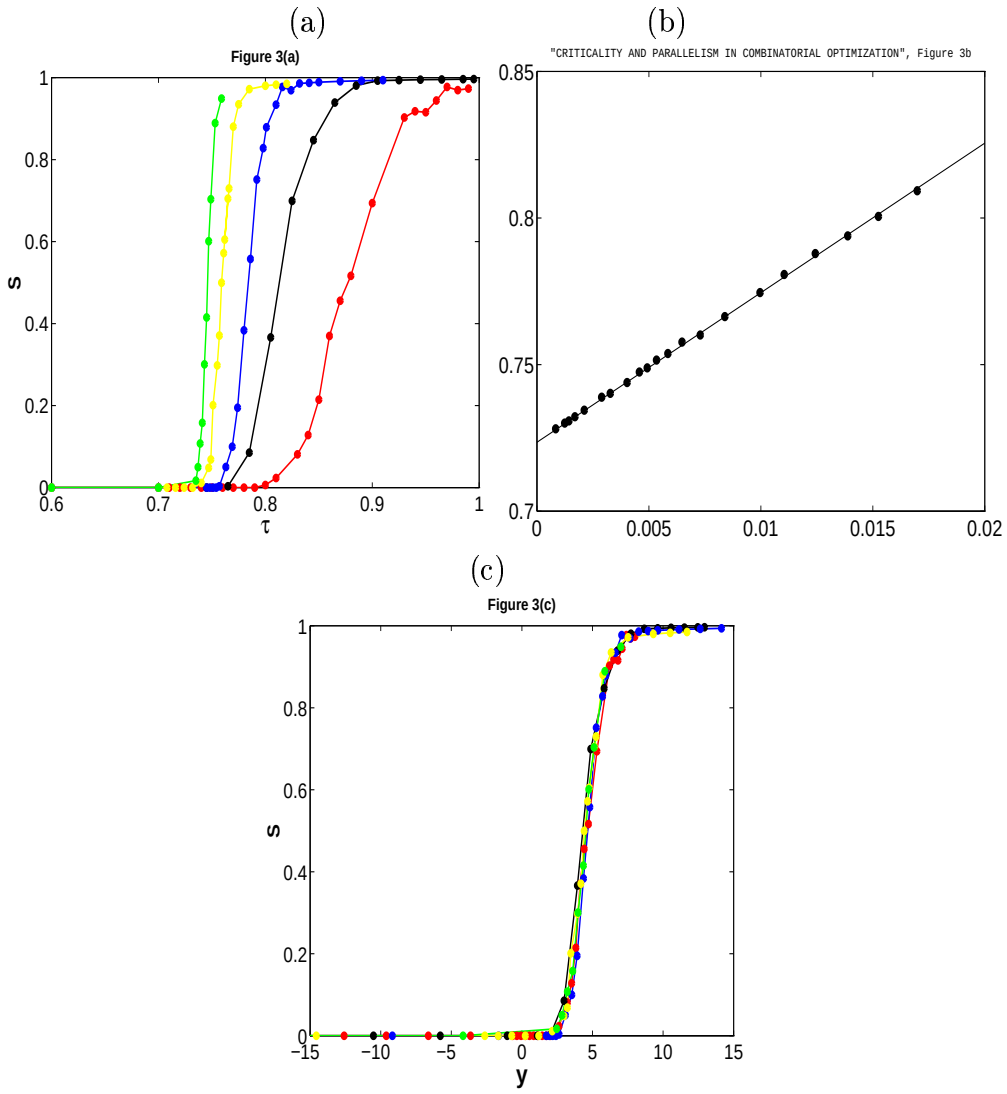

Figure 3: (a) Order parameter of $K = 6$ energy functions with different N values ($N = 100$ red, $N = 200$ black, $N = 400$ blue, $N = 800$ yellow, $N = 1600$ green). (b) $\tau_c(N)$ vs. $N^{-1/\nu}$, where $\nu \approx 1.30 \pm 0.04$, $\tau_c(\infty) \approx 0.723 \pm 0.001$, computed by a nonlinear least squares fit from a collection of $(N, \tau_c(N))$ values. Here we define $\tau_c(N)$ to be the τ value at which the entropy is 0.5. (c) the order parameter curves of figure 3(a) plotted with respect to $y = N^{1/\nu}(\tau - \tau_c)/\tau_c$ instead of τ. Within the error bars (not shown), all curves collapse onto a universal (N-independent) curve.

spin flips are chosen uniformly from the unit interval (annealed approximation). Then $p_n(E) = \binom{K+1}{n}(\tau E)^n(1-\tau E)^{K+1-n}$ is the probability that n spins flip given that the system is at energy E, and the time evolution of E is described by:

$$E_{t+1} = p_0(E_t)E_t + p_1(E_t)\beta E_t + (1 - p_0(E_t) - p_1(E_t))\Phi \qquad (4)$$

where $\beta E_t = (1/2)E_t$ is the expected energy that results from single flips which lower the energy, and Φ represents the "reset" energy that results when two or more interdependent spins flip simultaneously, and is modeled as $\Phi = (1 + \tau/5)/5$ [13]. The fixed point of Equation 4 is the asymptotic energy. Under the above simplifying assumptions, $E = 0$ is the only fixed point for low τ, and τ_c is the smallest τ for which a non-zero fixed point exists. By setting $E_{t+1} = E_t = E$ in Equation 4 and expanding to second order in E we find that $\tau_c(K) = 5(\sqrt{1+4/K}-1)/2$ [13]. A comparison of this theoretical prediction and the results from simulation is shown in Figure 4. This result is consistent with the fact that the transition occurs only for $K > 4$ since $\tau_c \geq 1$ for $K \leq 4$. At $K = 0$ no phase transition occurs because all spins are independent of each other and it requires sufficient interaction between spins (sufficient K) in order to create enough "resetting" of energies causing the phase transition.

Thus far we have described an abrupt transition in a family of optimization problems as the degree of parallelism is varied. Qualitatively, at least, the forces driving the transition are not unique to the NK model. We provide further support for the generality of the observed phenomena by investigating a very different optimization task, the Traveling Salesman Problem (TSP). Though part of its importance lies in its role as benchmark for new theory and optimization techniques, the TSP and its variants have many practical applications ranging from printed circuit board design to X-ray crystallography to scheduling. The task is simple: find the shortest tour passing through a set of cities, visiting each city only once.

One of the most effective solution techniques for the TSP is due to Lin and Kernighan [14]. Their method relies on a local operator, k-opt, to improve solutions. The k-opt operator removes k edges between cities and replaces them with k new edges such that the new edges still form a valid tour. At each improvement, the Lin-Kernighan heuristic intelligently looks for an improvement with any value k. Here we investigate the quality of solutions as a particular k-opt-like move is applied with increasing parallelism. The results

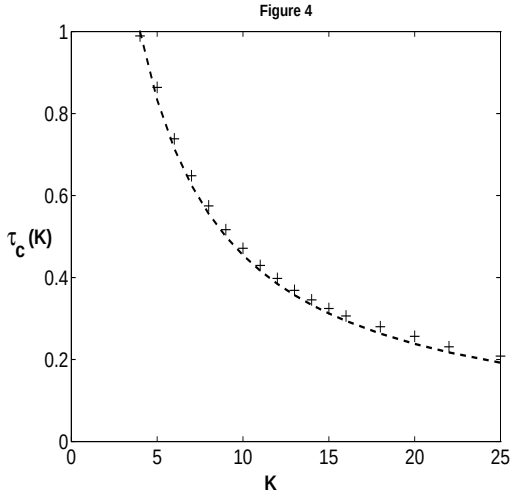

Figure 4: Comparison of the theoretical prediction $\tau_c(K) = 5(\sqrt{1+4/K} - 1)/2$ (solid curve) with estimates of τ_c at which the entropy is 0.5 (data points represented by +s). Here $N = 1000$ and $K = 1, \ldots, 25$ and parallelized greedy search is used ($T = 0$). Since N is large, the transition curves are sharp and 0.5 entropy τ points overestimate the true transition point very little.

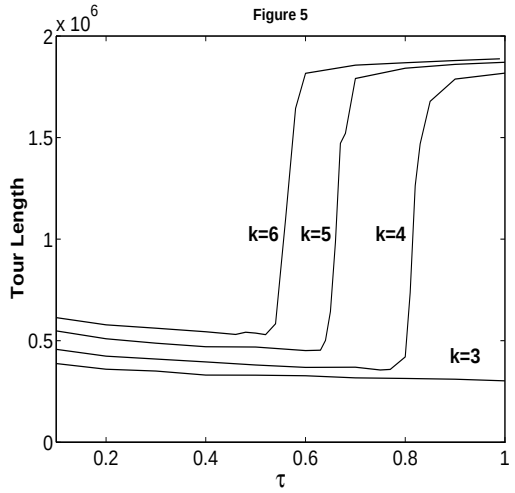

Figure 5: Expected tour lengths under $k = 3, 4, 5$ and 6 local moves as a function of τ for a set of $N = 439$ cities called `pr439.tsp`. This tour is supplied in the TSPLIB package, which is available by anonymous `ftp` through `elib.zib-berlin.de`. The results are averaged over 30 initial starting points after having run 100 generations.

we present do not depend on the exact form of the operator, but only on the fact that it has a local interaction range. The operator we use is defined as follows: for city i we select $k-1$ other cities at random yielding a set $\{c_i, c_{i_1}, \ldots, c_{i_{k-1}}\}$ of k cities. This subset of cities are then cyclicly permuted within the tour. For example, if the tour starting at city 1 and visiting cities $2, 3, \ldots, 10, 1$ in sequence is represented as $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and the subset of cities considered for a 3-opt-like move is $\{3, 5, 7\}$, then the resulting tour after the move is $\{1, 2, 5, 4, 7, 6, 3, 8, 9, 10\}$. The results of applying this operator with increasing parallelism for a set of $N = 439$ cities and various interaction ranges, k, are presented in Figure 5. A study of the order parameter for a wide variety of TSP instances reveals behavior analogous to that described for the NK model with k playing the role of K.

CONCLUSIONS

We have discovered and characterized a phase transition that arises as local search algorithms are parallelized. This transition is driven by the syn-

ergistic/antagonistic effects of overlapping applications of local operators and its existence places a sharp upper bound on the amount of useful parallelism in local search algorithms for combinatorial optimization.

Techniques from statistical mechanics have been used to make the correspondence between the physics of phase transitions and the performance of parallelized optimizing systems more precise, demonstrating strong connections between these seemingly unrelated fields.

Phase transitions seen in the satisfiability of constraint satisfaction problems [15] have previously shown the importance of critical phenomena in artificial intelligence. The transition described here is quite distinct from this satisfiability transition but further demonstrates the importance of critical phenomena in optimization and artificial intelligence.

It is interesting to consider whether the critical behavior presented here arises in more general distributed systems of interacting agents. Any time the decision of one agent relies on information contained in the state of another agent, the possibility exists for an abrupt degradation in performance as more agents act in parallel and the amount of "stale" information that exists due to nonzero information propagation delays increases. A similar transition seen in the transmission of packets in computer networks as the traffic density rises [16] provides support for the generality of these phenomena. Such transitions may also be lurking in some human organizations.

References

[1] S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi, *Science*, **220**, 671, (1983).

[2] S. Forrest, *Science*, **261**, 872, (1993).

[3] F. Glover, *ORSA J. Comput.*, **1**, 190, (1989), F. Glover, *ibid.* **2**, 4, (1990), D. Cvijović, J. Klinowski, *Science*, **267**, 664, (1995).

[4] R. Azencott, ed., *Simulated Annealing Parallelization Techniques* (John Wiley & Sons, New York, 1992).

[5] S. A. Kauffman, *The Origins of Order*, (Oxford University Press, 1993).

[6] Gerhard Reinelt, *The Traveling Salesman, computational solutions for TSP applications*, (Springer-Verlag, Berlin, Heidelberg, 1994).

[7] K. H. Fischer, J. A. Hertz, *Spin Glasses*, (Cambridge University Press, 1991).

[8] B. Derrida, *Phys. Rev. B*, **24**, 2613, (1981), C. A. Macken, A. S. Perelson, *Proc. Natl. Acad. Sci. USA*, **86**, 6191, (1989), E. D. Weinberger, *Phys. Rev. A*, **44**, 6399-6413, (1991).

[9] N. A. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. J. Teller, *J. Chem. Phys.* **21**, 1087, (1953).

[10] This transition persists for small non-zero fixed temperatures, and broadens at larger temperatures.

[11] The position of the transition also depends on the length of time that we allow the algorithm to run. In all our numerical experiments we have waited sufficiently long that further increases in running time do not noticeably affect the transition.

[12] R. J. Creswick *et al*, *Introduction to Renormalization Group Methods in Physics*, (Wiley & Sons, 1992). J.J. Binnet, *et. al*, *The Theory of Critical Phenomena* (Oxford University Press, Oxford, 1992).

[13] A.G. Siapas, W.G. Macready in preparation.

[14] S. Lin, B.W. Kernighan, *Operations Research*, **21**, 498-516, (1973).

[15] S. Kirkpatrick, B. Selman, *Science*, **264**, 1297, (1994), P. Cheeseman, B. Kanefsky, W.M Taylor, *Proc. of the 12th IJCAI*, (1991).

[16] T. Hogg, B. Huberman, *IEEE Trans. on Systems, Man, and Cybernetics*, **21**, 1325, (1991).

[17] We would like to thank Dan Stein for helpful discussions, and H. Abelson, G. J. Sussman, J. Tsitsiklis, E. Hung, N. Berker, E. Todorov, J. Tenenbaum, A. Khwaja, P. Beebee and the reviewers for useful suggestions. This work was partially supported by D.O.E DE-FG03-93-ER61639, N.S.F. BIR-9218746, and N.I.H. grant 5-R01 GM49619.