IEEE Int. Conf. on Neural Networks, San Francisco, March 1993, Vol. I pp 257-262.

Annealing Networks and Fractal Landscapes

Raymond Lister

Dept. of Electrical Engineering, University of Queensland, QLD 4072, Australia

Abstract— The conventional explanation for the poor scaling of Hopfield and Tank networks is that they have di culty in balancing the trade-o between the path length and the legality components of the energy function. We first describe an experiment which suggests that the conventional explanation is either wrong, or at best incomplete. We propose an alternative explanation: these networks might scale better if their dynamics e ectively implemented a "divide-and-conquer" strategy. That is, if they recursively decomposed the problem into smaller independent sub-problems. An annealing network can do so if the energy landscape has a self-similar "quasi-fractal" structure. We believe this proposition applies to both discrete and analog networks. We support our proposition by describing our work on finding low cost solutions for traveling salesman problems. We then consider the implications for two other optimization problems: graph bisection and coloring.

I. INTRODUCTION

Since the seminal work of Hopfield and Tank [1] on the N city Traveling Salesman Problem (TSP), other workers have refined the $N\times N$ analog network to improve its performance on large TSPs [2,3]. These refinements lessen the well known problem of trading-off the length and legality components of the energy function [4, page 78-79]. In earlier work we approached this trade-off problem differently [5]. We introduced two discrete $N\times N$ networks, McRotA and McRevA. Both are initialized to legal (but random) solutions, and their respective update mechanisms preserve legality, so they both dispense with legality terms in the energy function.

McRotA moves a single city from one position in the path to another, in a single operation. This is illustrated in Fig. 1. The crossover in Fig. 1a of connections A3→C3 and B1→D1 is suboptimal. Any traveling salesman's path containing a crossover is suboptimal [6]. Fig. 1b shows that same path after city "C3" has been moved between cities "B1" and "D1". This eliminates the crossover in Fig. 1a, but introduces another crossover. Four more cities need to move to eliminate the intercluster crossover.

McRevA reverses the direction of travel along a path segment, in a single operation. This is illustrated in Fig. 1c, where the segment C3→B1 from Fig. 1a has been reversed. The suboptimal crossover in Fig. 1a has been removed in a single operation. Even if more cities were added to the four clusters, the removal of the intercluster crossover in Fig. 1a would still only require one segment reversal. Segment reversal is the simplest heuristic used in the well known non-neural algorithm by Lin and Kernighan [7].

We have since compared McRotA and McRevA to Peterson and Soderberg's [3] network, on Peterson's 200 city benchmark problem [8]. The simulation results are shown in Table I. (Reference [8] gives only the average performance of the Peterson and Soderberg network. The best/worst figures are from private communication with Peterson.) We attribute the performance improvement of McRotA over the Peterson and Soderberg network to McRotA's elimination of the length/legality problem. It could be argued that McRotA is not a "neural" network, because its units switch state in a coordinated fashion, but to argue so is to miss our intention: McRotA was designed to give an empirical guide to the performance we could expect from more traditional neural networks, if the length/legality problem could be eliminated.

TABLE I
PERFORMANCE OF SEVERAL ALGORITHMS ON PETERSON'S 200 CITY TSP.

Algorithm	No. runs	Path Length		
		Best	Average	Worst
P&S network, from [8]	5	12.137	12.660	12.930
McRotA	100	11.223	12.084	13.235
Iter. Improve. by seg. rev.	100	10.893	11.381	11.904
Elastic Net, from [8]	1	---	11.14	---
McRevA	100	10.604	10.811	11.112

This paper is not primarily concerned with the performance difference between Peterson and Soderberg's network and McRotA: it is the performance difference between McRotA and McRevA that is of primary interest. This difference cannot be attributed to the length/legality problem, as neither McRotA nor McRevA have legality constraints. The performance gap is large. The difference in their average path lengths on Peterson's 200 city problem is 1.27 units, when the square enclosing this TSP has sides of unit length, so McRotA averages approximately one extra trip across the square.

Fig. 2 compares a run each of McRotA and McRevA on Peterson's 200 city problem. The final solutions of each run (i.e. the rightmost two solutions of Fig. 2) illustrate the large performance difference. Fig. 2. also shows three intermediate solutions for both runs. The two networks

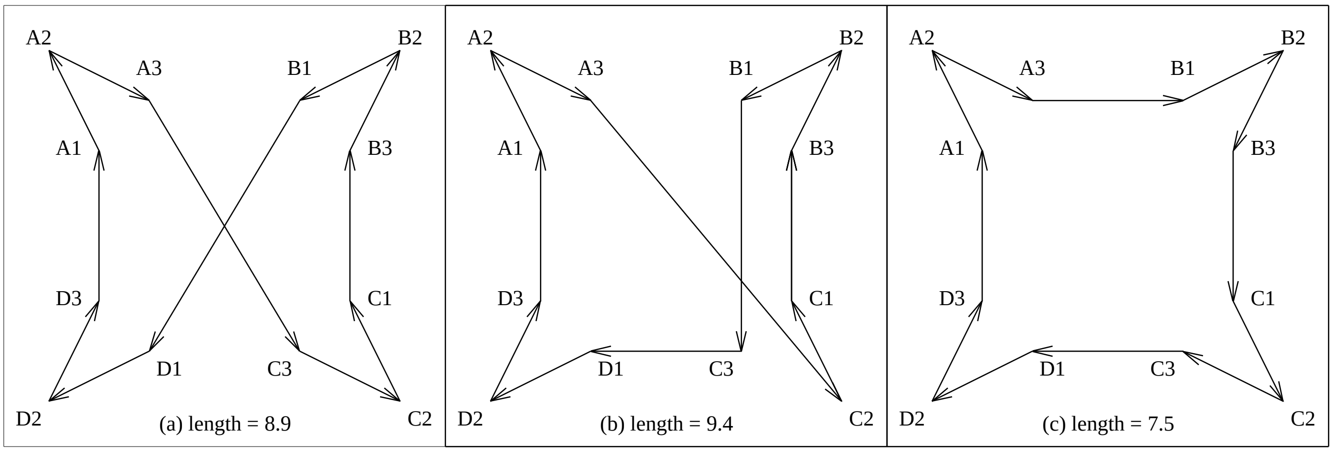

Fig. 1 1a) A salesman's path. 1b) The path from Fig. 1a after city "C3" has been moved between cities "B1" and "D1". 1c) The path from Fig. 1a after segment C3→B1 has been reversed.

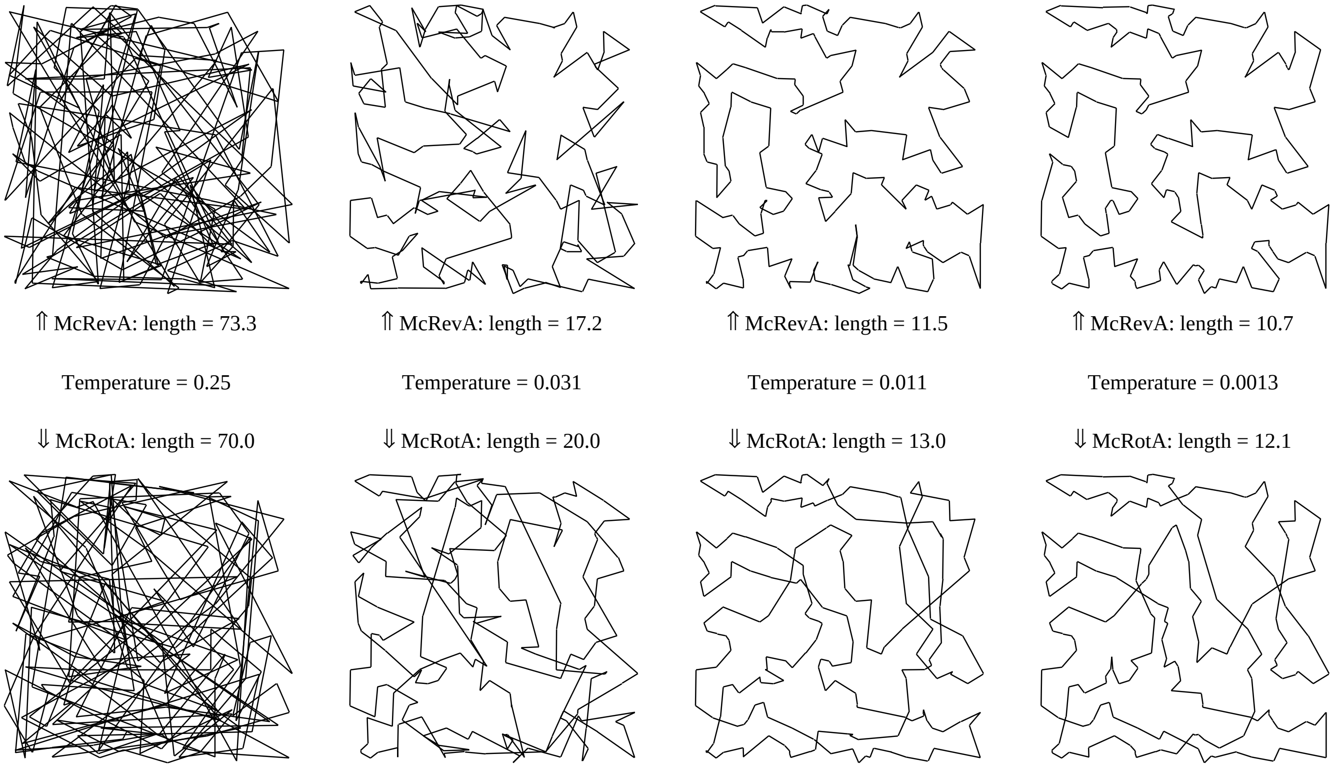

Fig. 2. Three intermediate solutions, and the final solutions, from a run of McRevA (top) and McRotA (bottom), on Peterson's 200 city TSP [8]. The two solutions in each column are from the same temperature stage of each run.

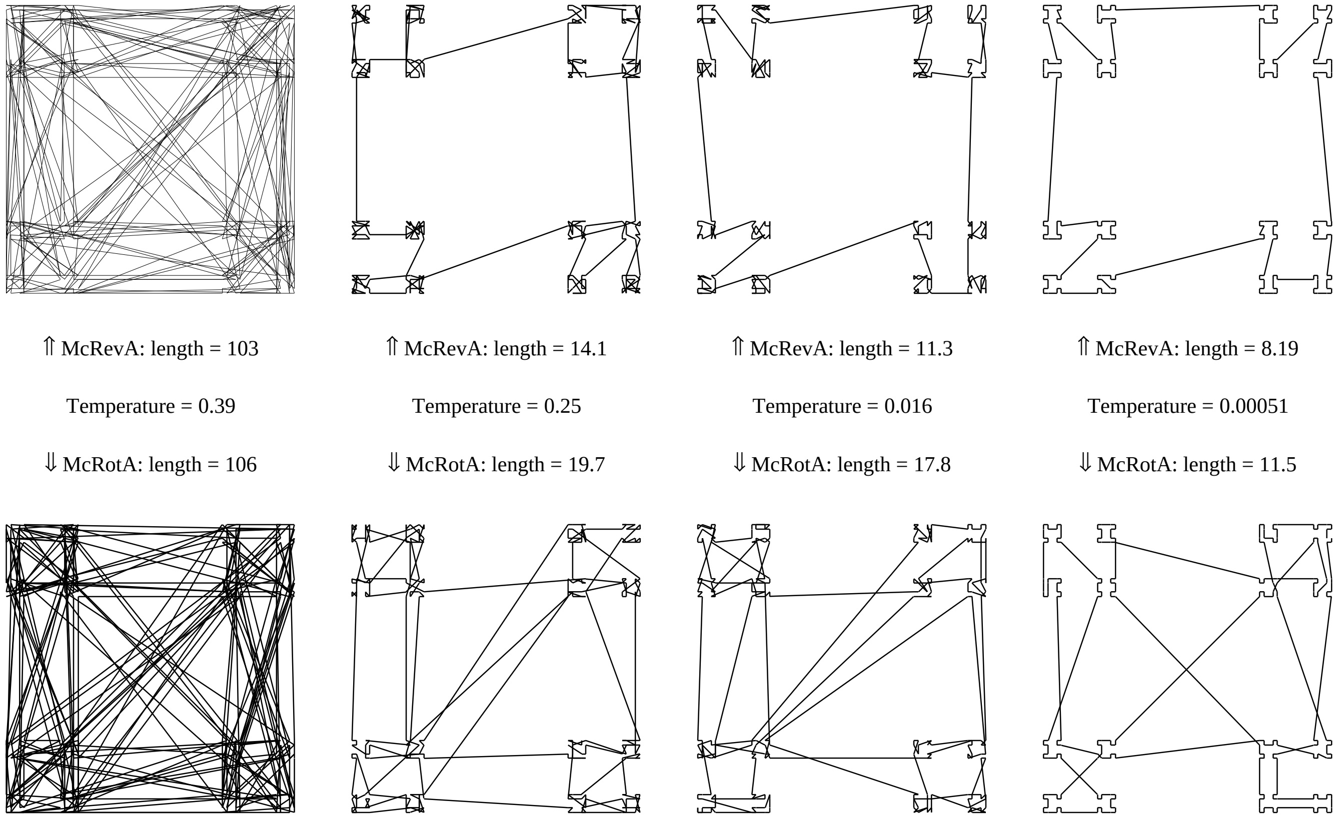

Fig. 3. Three intermediate solutions, and the final solutions, from a run of McRevA (top) and McRotA (bottom), on a recursively clustered 256 city TSP. The two solutions in each column are from the same temperature stage of each run.

evolved toward their respective final solutions quite differently. When Temperature = 0.25, both McRevA and McRota maintained random solutions. However, by the time Temperature = 0.031, the McRevA solution exhibits good coarse structure, whereas the McRotA solution contains many long range connections. At Temperature = 0.011, McRevA had almost settled upon its final solution, apart from some localized problems, whilst the McRotA solution still contains many long range connections. Both runs were terminated at Temperature = 0.0013, by which time McRevA had cleaned up its localized problems, but McRotA was still locked into a poor overall solution. McRotA's final solution exhibits the same type of structure that Wilson and Pawley observed in their study of Hopfield and Tank's analog network: the McRotA final solution is composed of "*several segments, each of which is locally a good tour, but which are joined together in such a way as to make a bad tour overall*" [9, page 67].

The McRotA final solution in Fig. 2. has approximately the same length as Peterson and Soderberg's best solution for this problem (see Table I). Even iterative improvement (a simple and widely used technique) when combined with segment reversal can produce better solutions than McRotA (see Table I). The poor relative performance of McRotA raises the following three questions:

- Why does the segment reversal heuristic endow McRevA with better scaling characteristics than McRotA?
- Is the cause of McRotA's scaling problem relevant to analog networks?
- Is the cause peculiar to the TSP, or is it relevant to the application of neural networks to other combinatorial optimization problems?

The answer to the first question is well known within the simulated annealing community, but it is an issue not widely recognized within the neural network community. We explain the answer in the remainder of the introduction. The rest of the paper considers the other two questions.

A. McRevA Scales Better Than McRotA Because McRevA Uses Divide-And-Conquer

A classical (i.e. pre-neural) approach to general problem solving is "divide-and-conquer", in which a large problem is decomposed recursively into smaller independent sub-problems. The quality of the final solution depends crucially upon the means of selecting each decomposition.

Fig. 3 contrasts the decomposition capabilities of McRevA and McRotA, on a 256 city TSP. The cities are enclosed by a square of unit side. The clustering is recursive: there are four *major clusters*; each major cluster contains four *medium clusters*; and each medium cluster contains four *miniclusters*.When Temperature = 0.39, both McRevA and McRota maintained random solutions. However, by the time Temperature = 0.25, McRevA had settled upon a relatively good route between the major clusters, whereas McRotA had not. At Temperature = 0.016, McRevA had further subdivided the problem by finding good routes between the four medium clusters within each major cluster, whilst McRotA had still not even determined a good route between the major clusters. When the runs were terminated at Temperature = 0.00051, McRevA had found good routes within the miniclusters. In contrast, McRotA's final solution still contains many repeated "visits" to the same major and medium clusters; the same type of poor structure that Wilson and Pawley observed in their study of Hopfield and Tank's analog network [9].

B. Simulated Annealing can perform divide-and-conquer in a quasi-fractal energy landscape.

The above heading paraphrases a conjecture made by Solla, Sorkin and White [10], and which is illustrated in the energy landscape of Fig. 4. At very high temperatures, simulated annealing may wander across the entire landscape represented in the figure. At a lower temperature, the transition A→B→C is substantially more probable than the reverse transition. As the temperature drops, this transition asymmetry increases until the solution is effectively confined to the region between C and C'. This *ergodicity breaking* locks in the broad structure of the final solution: the problem has been decomposed into smaller independent sub-problems. Confinement to a sub-region allows for that portion of the landscape to be explored more fully. As the temperature continues to fall, similar *phase changes* lock in progressively finer structure. Sorkin [11] proved that simulated annealing converges on a [0,1] normalized fractal energy landscape to an expected energy within $0<1/f<1$ energy units of the global optimum, in time polynomial in *f*.

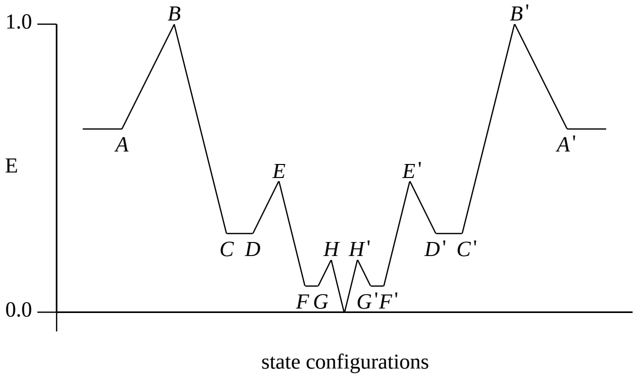

Fig. 4. A fractal energy landscape.

C. Segment reversal defines a quasi-fractal landscape.

Figs. 1a and 1c illustrate the proposition in this section's heading. Only one reversal is required to remove the sub-optimal cross-over. The number of cities in each cluster is irrelevant. When simulated annealing and segment reversal are combined, there exist temperature ranges where each cluster of cities is effectively a single "super city". The ordering of these "super cities" can be altered by single applications of segment reversal. The McRevA run in Fig. 3 illustrates how McRevA can recursively decompose a TSP into "super cities".

The energy landscape defined by segment reversal is only quasi-fractal, because up to three segment reversals are required to move a cluster to an arbitrary new position. A fully fractal landscape would require the more sophisticated k=3 path arrangement technique from Lin and Kernighan's algorithm [7]. However, the k=3 technique requires a network with N^3 units, and the simpler segment reversal heuristic is sufficient for most "real world" applications. Kirkpatrick and Toulouse studied the distribution of minima in TSP landscapes defined by segment reversal, and found they exhibited a high degree of *ultrametricity* [12], which is a measure of self-similarity [10].

Moving a single city at a time, as in McRotA, does not define a quasi-fractal energy landscape. The number of moves required to reposition an entire cluster of cities is proportional to the size of the cluster. Furthermore, there is no substantial shortening of the path until all cities in the cluster have been moved - McRotA's energy landscape contains "flat spots". These flat spots become larger with problem size. The McRotA runs in Figs. 2 and 3 illustrate how McRotA struggles to cross these flat spots.

II. RELEVANCE TO ANALOG NETWORKS

Hopfield and Tank conjectured that their analog $N\times N$ network has a fundamental computational advantage over similar discrete $N\times N$ networks, in that discrete versions are confined to moving between vertices of the solution hypercube, whereas the analog net moves through the volume enclosed by the hypercube. The argument has genuine intuitive appeal; never-the-less it as an hypothesis to be tested by experiment. The poor performance of Hopfield and Tank networks, even with the refinements introduced by later workers, suggests that Hopfield and Tank's intuitive argument is flawed.

Furthermore, we offer the following intuitive counter-argument. Hopfield and Tank noted that if the final solution of the analog network is to be physically interpretable, then "*the interior of the continuous decision space is smoothly related to the corners*" [1, page 150]. Whereas discrete-state McRotA moves a single city from one position to another in a path, analog networks can be thought of as moving "portions of cities" about in the path. Consequently, analog networks are also slow to move entire clusters of cities around in a salesman's path. That is, the analog energy landscape inside the solution hypercube contains flat spots that are "smoothly related" to the flat spots defined on the vertices of the solution hypercube. So, if flat spots present difficulties for an algorithm that moves between vertices of the hypercube, then flat spots could also present difficulties for an algorithm that moves within the hypercube. (The algorithm moving on the vertices might experience difficulty at a smaller problem size, but the analog algorithm will also experience difficulty at some problem size that is not substantially larger.)

The behavior of the Elastic Net Algorithm [13] suggests that it implements a divide-and-conquer strategy. When the net is distant from a cluster of cities, each of these cities has a similar effect on the net, and so the cluster acts as a "super city". The elastic net does not appear to perform as well as McRevA on the 200 city problem, but both of these divide-and-conquer algorithms appear to outperform the other algorithms (see Table I).

III. GENERALITY

The above observations about annealing and fractal landscapes are relevant to other combinatorial optimization problems. In this section, we consider two graph problems: bisection and coloring [14]. In both cases, the illustrative examples are small, and only serve a pedagogical role. As with the TSP, the need for recursive decomposition only arises in large problems.

A. Graph Bisection.

The object of bisection is to divide the nodes of a graph into two *sections* of equal size, so that the number of edges between the sections is minimized. Consider the graph bisection problem in Fig. 5. (Note that the figure is not an energy minimization network, but directly represents a graph, where the vertical triple-line indicates the sectioning. Any change to the sectioning moves at least one vertex from one side of the triple-line to the other side.) There are two links between the two sections. Exchanging any two vertices between the sections may remove one of the existing inter-section links, but it will also introduce new inter-section links. However, exchanging two cliques of three nodes can eliminate the inter-section links in a single operation. Thus, if a neural network approach is to scale to large graphs, it is not sufficient to exchange single vertices between sections, as Peterson and Soderberg's graph partitioning network does [3]. A more promising approach is to represent the problem hierarchically, so that it can treat highly intra-connected subgraphs as "super nodes". The resultant energy landscape is quasi-fractal.

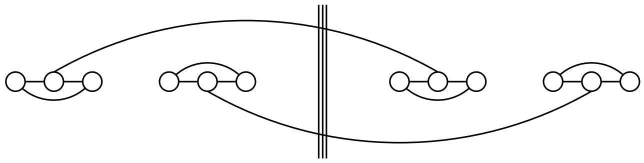

Fig. 5. A small graph bisection problem.

Subgraph exchange is commonly used in conventional heuristic [15]. However, whereas these conventional algorithms employ complicated move selection mechanisms, the stochastic selection mechanism of simulated annealing is particularly easy to implement. Annealing networks will probably not give better solutions than sophisticated conventional heuristic algorithms; the advantage of annealing networks lies in their ease of implementation.

B. Graph Coloring.

For our purposes in this paper, it is sufficient to consider two-coloring, where the object is to assign states to the units of an energy minimization network so that the number of connected nodes in the same state is minimized. (In the case of analog networks, the activation is rounded to zero/one.)

Consider the coloring problem in Fig. 6. The assignment of states within each four-node clique is optimal, but the state assignment between the two cliques is suboptimal. The problem could be fixed by a series of state changes to individual nodes within one clique, but as problem size increases, so will the size of the flat spots in the energy landscape. Alternately, the suboptimal state assignment could be remedied in a single operation, if the cliques were considered to be "super nodes", allowing the states of all nodes in a clique to be toggled simultaneously. Such a coordinated toggle does not alter the energy within a super node, but only between super nodes. The resultant energy landscape is quasi-fractal.

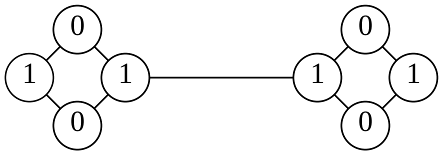

Fig. 6. A small graph coloring problem.

(We note in passing that graph coloring is a similar problem to finding low energy state configurations in the Boltzmann Machine [16]. We suspect that the very long run times required by the Boltzmann Machine are due, at least in part, to the way its units switch individually, instead of in the recursively decomposed way we advocate here.)

IV. CONCLUSION

We have extended our earlier study [5] of the performance and behavior of McRotA and McRevA to see how well they scale to a 200 city Traveling Salesman Problem. As in the earlier study, McRotA and McRevA outperform the conventional analog $N \times N$ network that we benchmarked.

The primary interest in this study, however, is the substantial performance improvement of McRevA over McRotA. This disparity cannot be attributed to the conventional legality/length trade-off problem, as neither McRevA over McRotA use a legality term in their energy function. We attribute the disparity to McRevA's use of the segment reversal heuristic, which gives it a quasi-fractal energy landscape. Such landscapes are desirable because they allow an annealing network to find a solution by recursive decomposition of the problem: divide-and-conquer.

While it can be argued that McRotA and McRevA do not qualify as "neural" networks, to do so is to miss our intention: McRotA and McRevA are experimental vehicles for illustrating the utility of the divide-and-conquer approach in any type of annealing network, whether analog or digital, neural or non-neural.

The experimental work described in this paper does not establish that an analog neural network can only scale well if it recursively decomposes the problem. However, we are sympathetic to that conjecture.

ACKNOWLEDGMENTS

Thanks to Peter Bartlett and Jim Stone for their comments on drafts of this paper; and to Carsten Peterson, for his help with the comparative study.

REFERENCES

[1] J. Hopfield and D. Tank, "Neural Computation of Decisions in Optimization Problems", *Biol. Cybern.*, vol. 52, pp. 141-152, 1985.

[2] D. Van den Bout and T. Miller, "Improving the performance of the Hopfield-Tank neural network through normalization and annealing", *Biol. Cybern.*, vol. 62, pp. 129-139 1989.

[3] C. Peterson and B. Soderberg, "A new method for mapping optimization problems onto neural networks", *International Journal of Neural Systems*, vol. 1, pp. 3-22, 1989.

[4] J. Hertz, A. Krogh and R. Palmer *Introduction to the Theory of Neural Computation*, Addison-Wesley, 1991.

[5] R. Lister, "Segment Reversal and the Traveling Salesman Problem", *International Joint Conference on Neural Networks*, Washington, pp. I-424 to I-427, 1990.

[6] E. Lawler, J. Lenstra, A. Rinnooy Kan and D. Shmoys, Eds, *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, Wiley, 1985.

[7] S. Lin and W. Kernighan, "An Effective Heuristic Algorithm for the Traveling-Salesman Problem", *Oper. Res.* vol. 21, pp. 498-516, 1973.

[8] C. Peterson, "Parallel Distributed Approaches to Combinatorial Optimization: Benchmark Studies on Traveling Salesman Problem", *Neural Computation*, vol. 2, pp. 261-269, 1990.

[9] G. Wilson and G. Pawley, "On the Stability of the Travelling Salesman Problem Algorithm of Hopfield and Tank", *Biol. Cybern.*, vol. 58, pp. 63-70, 1988.

[10] S. Solla, G. Sorkin and S. White, "Configuration space analysis for optimization problems", in *Disordered systems and biological organization*, NATO ASI Series, vol. F20. E. Bienenstock et al., Eds, Berlin: Springer Verlag, pp. 283-293, 1986.

[11] G. Sorkin, "Simulated Annealing on Fractals: Theoretical Analysis and Relevance for Combinatorial Optimization", in *Advanced Research in VLSI*, W. Dally, Ed. MIT Press, pp. 331-351, 1990.

[12] S. Kirkpatrick and G. Toulouse, "Configuration space analysis of travelling salesman problems", *Journal de Physique*, vol. 46, pp. 1277-1292, 1985.

[13] R. Durbin and D. Willshaw, "An analogue approach to the travelling salesman problem using an elastic net method", *Nature*, vol. 326, pp. 689-691, 16 April 1987.

[14] M. Garey and D. Johnson, *Computers and Intractability*, W. H. Freeman, 1979.

[15] B. Kernighan and S. Lin, "An Efficient Heuristic Procedure for Partitioning Graphs", *Bell Systems Technical Journal*, vol. 49, pp. 291-307, 1970.

[16] D. Ackley, G. Hinton, and T. Sejnowski, "A Learning Algorithm for Boltzmann Machines", *Cognitive Science*, vol. 9, pp. 147-169, 1985.