Parallel Error Tolerance Scheme Based on the Hill Climbing Nature of Simulated Annealing

Bruce M. McMillin
Missouri University of Science and Technology, ff@mst.edu

Chul-Eui Hong

Recommended Citation

B. M. McMillin and C. Hong, "Parallel Error Tolerance Scheme Based on the Hill Climbing Nature of Simulated Annealing," *Proceedings of the 35th Midwest Symposium on Circuits and Systems, 1992*, Institute of Electrical and Electronics Engineers (IEEE), Jan 1992.
The definitive version is available at https://doi.org/10.1109/MWSCAS.1992.271230

PARALLEL ERROR TOLERANCE SCHEME BASED ON THE HILL CLIMBING NATURE OF SIMULATED ANNEALING†

Chul-Eui Hong and Bruce M. McMillin

Department of Computer Science
University of Missouri-Rolla, Rolla, MO 65401 - ff@cs.umr.edu

Abstract: Simulated annealing is an attractive, but expensive, heuristic for approximating the solution to combinatorial optimization problems. In parallelizing simulated annealing in a multicomputer, maintaining the global state S involves explicit message traffic and is a critical performance bottleneck. One way to mitigate this bottleneck is to amortize the overhead of these state updates over as many parallel state changes as possible. However, using this technique introduces errors in the calculated cost $C(S)$ of a particular state S used by the annealing process. This paper places analytically derived bounds on this error in order to assure convergence to the correct result. The resulting parallel simulated annealing algorithm dynamically changes the frequency of global updates as a function of the annealing control parameter, i.e. temperature. Implementation results on an Intel iPSC/2 are reported.

I. INTRODUCTION

The simulated annealing algorithm is based on the analogy between simulation of the annealing of solids and the problem of solving large combinatorial optimization problems. [7] propose a Monte Carlo method, which simulates the evolution to thermal equilibrium of a solid for a fixed value of the temperature. Simulated annealing uses a control parameter, "temperature" T in this analogy and the concept of a "move" to perturb the "state space" S of the solution to reach a new state. Let ΔC be the difference of the cost of current state and new state. The probability that a candidate move is accepted or rejected in simulated annealing is determined by *the Metropolis criterion*:

$$\text{Prob}[\Delta C \text{ is accepted}] = \min\left(1, \exp(-\frac{\Delta C}{T})\right) \qquad (1\text{-}1)$$

A move which is made, despite a positive ΔC, is called a *hill-climbing* move since it perturbs the solution out of a local minima in the cost.

† This work was supported in part by the National Science Foundation under Grant Numbers MIP-8909749 and CDA-8820714, and in part by the AMOCO Faculty Development Program.

The model problem is to allocate regular and/or irregular patterns onto a large stock sheet of finite dimensions in such a way that the resulting waste area is minimized. The cost function is made up of affinity relation (strength of attraction) between patterns, the distance from the origin, and overlap penalty between patterns. Our results are equally applicable to the problems of stock cutting [5] and VLSI placement [2].

A data parallel domain decomposition of the stock cutting problem gives each node approximately the same number of patterns. Each processor performs internal move, rotate and exchange operations as well as participating in moves between processors. While each processor can independently calculate the distance from origin component of cost function $C(S)$, global state information is needed for calculating the affinity relation between any two patterns. In the latter case, an error may occur in the calculated $C(S)$.

Under the right conditions, annealing algorithms can evaluate the cost using old state information, but still converge to a reasonable solution. Such algorithms are called *error-present* algorithms. These are attractive since partially synchronous algorithms, that function in the presence of old state information, attain better parallel speedup by reducing synchronization bottlenecks; rather than broadcasting each internal move to all other processors, a processor saves up its changes and broadcasts them all at once thus incurring synchronization overheads infrequently. Finding an upper bound on the cost error at a particular temperature allows creation of a partially synchronous algorithm which achieves the correct result.

Previous work of the cost-error-tolerant schemes cannot measure the cost error correctly, and cost error has been tolerated empirically. In this paper, we define maximum bound of tolerable cost error as a function of the broadcast (global update) frequency.

II. CONCEPTS OF COST ERROR

By allowing processors to work with old state information, errors in the calculated cost will result. Cost error is a measurement of how much error, with respect to the actual cost, is present. Potentially, the cost error can be defined as [6] the difference between the actual (real) cost change and the estimated (measured) cost change.

$$\Delta C_a - \Delta C_e = (C_{af} - C_{ai}) - \sum_{i=1}^{P} \Delta C_i$$

where ΔC_a is the actual cost change, ΔC_e is the estimated cost change. C_{af} is the actual final cost and C_{ai} is the actual initial cost. ΔC_i is the estimated cost change in processor i, and P is the total number of processors.

There are three shortcomings in the this error measurement scheme [3] for an error-present annealing algorithm. First, this method counts only the accepted moves. Second, when both the actual and the estimated cost changes are negative, the difference in cost, $\Delta C_a - \Delta C_e$, is added to the cost error, even though the acceptance of the move is correct. Finally, an optimistic error ($\Delta C_a > \Delta C_e$) and a pessimistic error ($\Delta C_e > \Delta C_a$) compensate for each other. [1] suggests the adaptive stream length control scheme based on the traditional error measurement scheme and the empirical optimal acceptance ratio.

When a cost error affects the annealing process, there are some interesting phenomena. First, [6, 2] report that the cost error is mostly negative and the absolute value of the cost error is very large at high temperatures and goes to zero as the temperature decreases. Second, the acceptance ratio of the error-present algorithm is smaller than that of the sequential simulated annealing algorithm in the high temperature region. However, the acceptance ratio of the error-present algorithm is increased slightly in the low temperature region [8].

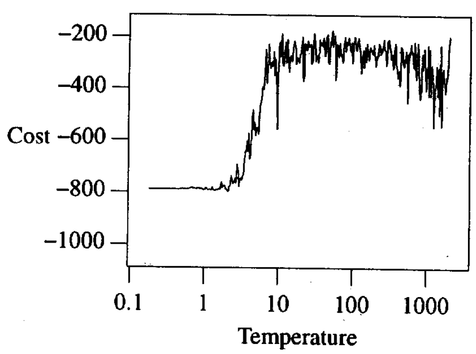

Figure 2-1: Compare Fluctuations of Cost
(··· : sequential annealing, — : error-present simulated annealing

The third phenomenon in the presence of cost error is the reduced fluctuation of the average cost change at high and intermediate temperature regions. Pessimistic error-based moves occur more frequently so that the fluctuations in cost are reduced in the high temperature region. Thus, the system is likely to be kept in the high local minimum (Figure 2-1).

III. A NEW ERROR TOLERANCE METHOD

In this section, we define a new cost error measurement scheme. Bounds on the cost error are proved analytically to be a function of global update frequency, or stream length s. Using the measured amount of cost error, an optimal stream length is derived based on the hill climbing nature of simulated annealing.

Since the estimated and actual cost changes are different, erroneous moves can result. Consider two possible cost changes, ΔC_1 and ΔC_2 where $\Delta C_1 < \Delta C_2$ where it is not known which is the actual and which is the estimated cost change. If a move is accepted with a smaller cost change, ΔC_1, while the move is rejected with the larger cost change, ΔC_2 then an erroneous move of error $\Delta E = \Delta C_2 - \Delta C_1$ has occurred.

Theorem 3-1: [4] The erroneous move decision is exponentially distributed with respect to the parameter $T > 0$, given that the candidate move is accepted with smaller cost change, ΔC_1, between the actual and the estimated cost changes.

$$\begin{aligned} &Prob[\text{The erroneous move decision with cost error } [0, \Delta E]\,] \\ &= Prob[\text{Move rejected with cost change } \Delta C_2 \\ &\qquad \mid \text{Move accepted with cost change } \Delta C_1] \\ &= 1 - e^{-\frac{\Delta E}{T}} \square \end{aligned}$$

Associating move decisions on a case-by-case basis [4] with the above cost changes allows us to calculate the probability of optimistic and pessimistic errors.

Definition 3-1: The probability of an optimistic cost error

$$P_{opt} = Prob[\Delta C_a \ge \Delta C_e > 0] \cdot \left| e^{-\frac{\Delta C_e}{T}} \cdot \left(e^{-\frac{\Delta E}{T}} - 1\right) \right|$$

Definition 3-2: The probability of a pessimistic cost error

$$P_{pes} = Prob[\Delta C_e \ge \Delta C_a > 0] \cdot \left| e^{-\frac{\Delta C_e}{T}} \cdot \left(e^{\frac{\Delta E}{T}} - 1\right) \right|$$

The total amount of cost error (E) can be calculated from the *Definition 3-1* and *3-2*.

Theorem 3-2: The cost error is

$$E \le \Delta C_e \cdot Prob[\Delta C_e > 0] \cdot e^{-\frac{\Delta C_e}{T}} \cdot \left(e^{\frac{\Delta E}{T}} - 1\right)$$

Proof: The proof is obvious.

$$\begin{aligned} E &= \Delta C_e \cdot (P_{opt} + P_{pes}) \\ &\le \Delta C_e \cdot Prob[\Delta C_e > 0] \cdot e^{-\frac{\Delta C_e}{T}} \cdot \left(e^{\frac{\Delta E}{T}} - 1\right) \qquad \square \end{aligned}$$

This error measurement scheme overcomes the three shortcomings of the work in Section II. First, the new cost error measurement method includes the cost error of the rejected moves (*Definition 3-2* and *Theorem 3-2*). Second, this method does not include the cost error of the negative cost change moves. Finally, there are no compensating pessimistic and optimistic cost errors (*Theorem 3-2*).

Since a cost error is tolerated by hill climbing moves, we now turn our attention to deriving a maximum bound on the cost error using a maximum bound on the hill climbing move. Let $d(s)$ be the maximum amount (or depth) of cost which can be hill-climbed at a given temperature T and stream length s. [9] gives bounds on choosing $d(s)$ hill climbing moves in s moves.

$$e^{-\frac{d(s)}{T}} \geq \frac{1}{s} \quad \Rightarrow \quad d(s) \leq T \ln s \qquad (3\text{-}1)$$

If an error-present algorithm produces a high local minimum, as in Figure 2-1, it is due to a decreased hill-climbing power. Since the decreased hill climb power is due to the error, we have the following theorem.

Lemma: In the error-present algorithm, the hill climbing depth (power) for one move is

$$d_a \leq d_e + E$$

where d_a is the hill climbing depth of sequential simulated annealing and d_e is the hill climbing depth of the error-present algorithm.

Proof: Since a loss of hill-climbing power is introduced only for pessimistic errors, $\Delta C_e > \Delta C_a > 0$. Hill climbing, probabilistically, is $d_a = \Delta C_a \cdot e^{-\frac{\Delta C_a}{T}}$ and $d_e = \Delta C_e \cdot e^{-\frac{\Delta C_e}{T}}$. Using E (from *Theorem 3-2*), and ($1 \leq \frac{\Delta C_e}{\Delta C_a}$), we have

$$\Delta C_a \cdot e^{-\frac{\Delta C_a}{T}} \leq \Delta C_e \cdot e^{-\frac{\Delta C_e - \Delta E}{T}} + \Delta C_e \cdot \left(e^{-\frac{\Delta C_e}{T}} - e^{-\frac{\Delta C_e}{T}}\right)$$

$$\Delta C_a \cdot e^{-\frac{\Delta C_a}{T}} \leq \Delta C_e \cdot e^{-\frac{\Delta C_e}{T}} + \Delta C_e \cdot \left(e^{-\frac{\Delta C_e - \Delta E}{T}} - e^{-\frac{\Delta C_e}{T}}\right)$$

$$d_a \leq d_e + E \quad \square$$

The optimal stream length can now be calculated on the total amount of cost error ($E(s)$).

Theorem 3-3: When the total amount of cost error (E) occurs during stream length s, $s \cdot u$ stream length is needed to tolerate the cost error.

Proof:

$$e^{-\frac{d_a}{T}} \geq e^{-\frac{d_e + E}{T}} \qquad \text{(Lemma)}$$

$$e^{-\frac{d_a(s)}{T}} \geq e^{-\frac{d_e(s)+E}{T}} = e^{-\frac{d_e(s)}{T}} \cdot e^{-\frac{E}{T}} \geq \frac{1}{s} \cdot \frac{1}{u} \qquad \square \text{ (ergodicity)}$$

Thus, in the presence of cost error, $s_r = s \cdot u$ stream length is required in order to have the same hill climbing depth as the regular algorithm has in stream length s. Conceptually, the additional Markov chain length, $s \cdot u - s$ provides additional move generations to compensate for the decreased hill climbing power.

The maximum error bound E can be calculated using equation (3-1). By setting this error bound, *a priori*, an adaptive algorithm was developed which decreases or increases the stream length, s, if the measured amount of cost error exceeds the maximum bound E. Thus, by adjusting the global update frequency, improved computation speed is attained while maintaining the convergence properties of the sequential simulated annealing algorithm.

IV. EXPERIMENTAL RESULTS

The *adaptive method* from Section III was implemented on a 16-node Intel iPSC/2 multicomputer. The target problem was the stock cutting problem. Decomposition was performed spatially on the dimensions of the stock sheet with a random, initial, assignment of patterns.

We ran experiments using the adaptive method varying the number of patterns, composition of patterns, and the number of processors. As a basis for comparison, the fixed stream length method, the *static method*, was implemented on each stream length.

Comparing the stream length at each temperature (Figure 4-1) with the annealing curve (Figure 4-2) for the adaptive method, the stream length reduces to 2 in the *critical region*, or region of rapid descent. However, the stream length increases to 125 far from the critical region. The stream length varies dynamically according to the annealing curve. This means the cost error has little affect on the annealing process away from the critical region, but affects it greatly in the critical region. This corresponds to the fact that the annealing process proceeds rapidly away from the critical region, but much more slowly in the critical region.

We compared, for a small problem of 4 nodes over 16 irregular patterns with a Markov chain length of 500 and allowing for a 10% increase in the Markov chain length ($u = 1.1.$), the adaptive and static method (over a fixed stream length of 10). The standard deviation, over 12 runs, of the adaptive method was smaller (8.29) than that of the static method (11.7), as expected. The average number of global updates was reduced 6.3 times in the adaptive method over the static method. On a larger problem, we varied number of processors for 128 patterns to measure the speedup of running time comparing with a sequential processor in Figure 4-3. The speedup of static stream length (5 stream length) is 7.6 and the speedup of adaptive error control scheme is 11.9 in 16 processors.

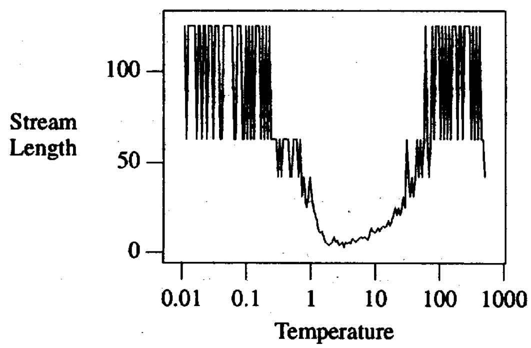

Figure 4-1: Stream Length vs. Temperature

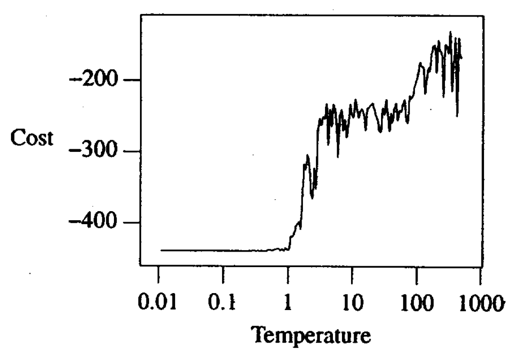

Figure 4-2: Annealing Curve

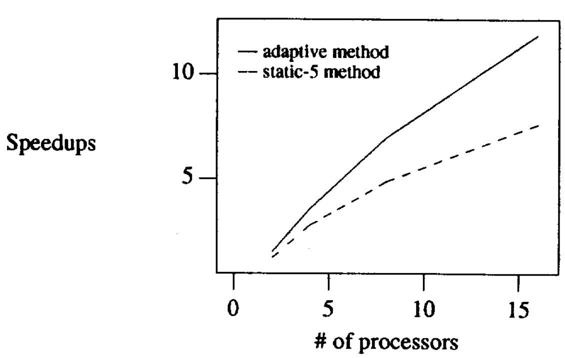

Figure 4-3: Speedup Of Adaptive And Static Method

V. CONCLUSION

In this paper we developed a cost error measurement scheme and an adaptive error control method in terms of stream length based on a cost error analysis of hill climbing power. An adaptive error control algorithm was developed that varies the stream length as a function of the annealing schedule. Experimental results show that this method reduces the frequency of global state updates and, thus, improves the parallel speedup, while still reaching the optimal configuration of the system. Additionally, the adaptive error control scheme chooses the optimal stream length dynamically rather than through the extensive experimentation required by a static stream length method.

VI. BIBLIOGRAPHY

[1] Banerjee, P., Jones, M., and Sargent, J., "Parallel Simulated Annealing Algorithm for Cell Placement on Hypercube Multiprocessors," *IEEE Trans. on Parallel and Distributed System*, Vol. 1, No. 1, January 1990.

[2] Casotto, A., Romeo, F. and Sangiovanni-Vincentelli, A., "A Parallel Simulated Annealing Algorithm for the Placement of Macro-Cells," *IEEE Transactions on Computer-Aided Design*, September 1987, pp 838-847.

[3] Durand, M., "Parallel Simulated Annealing: Accuracy versus speed in placement," *IEEE Design and Test*, June, 1989, pp 8-34.

[4] Hong, C-E., and McMillin, B. "Relaxing Synchronization in Distributed Simulated Annealing," *Department of Computer Science Technical Report, CSC 92-06*, University of Missouri-Rolla, March, 1992.

[5] Lutfiyya, H., McMillin, B., Poshyanonda, P., and Dagli, C., "Composite Stock Cutting Through Simulated Annealing," *Mathl. Comput. Modeling*, Vol. 16, No. 1, 1992, pp 57-74. Pergamon Press, New York.

[6] Jayaraman, R. and Darma, F., "Error Tolerance in Parallel Simulated Annealing Technique," *Proc. the International Conference on Computer Design*, 1988.

[7] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E., "Equation of State Calculations by Fast Computing Machines," *J. of Chem. Physics*, vol. 21, 1953, pp 1087-1092.

[8] Rose, J., Blythe, D. , Snelgrove, W., and Vranesic, Z. "Fast, High Quality VLSI Placement on an MIMD Multiprocessor," *Proc. International conference on Computer-Aided Design*, 1986, pp 42-45.

[9] White, S., "Concepts of Scales in Simulated Annealing," *Proc. IEEE Int. Conference on Computer Design*, Port Chester, November 1984, pp 646-651.