Journal of Statistical Physics, Vol. 52, Nos. 5/6, 1988

Nonlinear Optimization Simplified by Hypersurface Deformation

F. H. Stillinger[1] and T. A. Weber[2]

Received March 14, 1988

A general strategy is advanced for simplifying nonlinear optimization problems, the "ant-lion" method. This approach exploits shape modifications of the cost-function hypersurface which distend basins surrounding low-lying minima (including global minima). By intertwining hypersurface deformations with steepest-descent displacements, the search is concentrated on a small relevant subset of all minima. Specific calculations demonstrating the value of this method are reported for the partitioning of two classes of irregular but nonrandom graphs, the "prime-factor" graphs and the "pi" graphs. We also indicate how this approach can be applied to the traveling salesman problem and to design layout optimization, and that it may be useful in combination with simulated annealing strategies.

KEY WORDS: Nonlinear optimization; hypersurface deformation; ant-lion method; irregular graphs; graph partitioning; global minima; simulated annealing.

1. INTRODUCTION

A broad class of challenging and important problems in the natural sciences and mathematics involves nonlinear optimization. This class includes the traveling salesman problem,[1,2] the identification of ground-state configurations for biopolymers[3] and for spin glasses,[4] the optimal layout of circuit elements on computer chips,[5] and graph partitioning and coloring[6] problems. A feature common to all of these is the search for global minima of a "cost" function (equivalently "objective" or "potential" function) $\Phi(x_1 \cdots x_N)$ over a multidimensional space of points $\{x_i\}$. The

[1] AT & T Bell Laboratories, Murray Hill, New Jersey 07974.
[2] Permanent address: National Science Foundation, Washington, D.C.

0022-4715/88/0900-1429$06.00/0

$\{x_i\}$ may be either discrete or continuous, and most problems of significance are NP-complete.[7]

A few generally applicable methods have been advanced which empirically facilitate computational searches for global, or near-global, minima. This includes the "simulated annealing" method,[5] suggested by analogy between the multidimensional optimization problems of interest and the statistical mechanical problem of evaluating many-body partition functions with variable temperature. Our purpose in this paper is to explore another potentially useful strategy which can be applied either by itself, or perhaps even better in conjunction with a simulated annealing program. We illustrate the approach with some calculations for the partitioning of two families of irregular graphs. These calculations establish the dramatic simplification of the forbidding combinatorial enumeration that in principle must be carried out to effect the optimal partitions. They also suggest similar strategies for other nonlinear optimization problems, as mentioned in the Section 7.

2. HYPERSURFACE MODIFICATION STRATEGY

We shall be concerned with the case where the $\{x_i\}$ constitute the full N-dimensional Euclidean space E^N. The cost function Φ, bounded below, will be at least twice differentiable in E^N.

The shape of the Φ hypersurface in E^{N+1}, $(x_1 \cdots x_N, \Phi)$, leads to a natural division of the E^N configurational space $(x_1 \cdots x_N)$ into "basins." Each of these basins B_α contains exactly one local minimum α of Φ within its interior. B_α is defined to be the collection of all configurations $\{x_i\}$ which map onto minimum α by means of a steepest-descent connection.[8,9] Except for an insignificant zero-measure set of boundary points, all $\{x_i\}$ in E^N can thereby be assigned to basins.

At least one of the B_α will contain a global minimum, whose identification is the central objective. When the number of variables N is large, however, these global-minimum basins are typically very difficult to locate; the total number of local minima and associated basins tends to grow at least exponentially with N. Roughly speaking, then, the chance of randomly encountering a correct (i.e., global-minimum) basin is exponentially small in N.

This pessimistic situation for large N can be relieved somewhat if it is possible to increase the aperture of the global-minimum basins at the expense of those which surround higher-lying relative minima. This motivates the search for a continuous deformation of the Φ hypersurface with the following properties: (a) basins for high-lying relative minima are drastically reduced in size, or those minima and their basins are eliminated

altogether; (b) global minima before deformation continuously transform into deep minima with large basins as the deformation is "turned on." Ideally, the deformed hypersurface would exhibit only a single minimum which could be traced back to the desired global Φ minimum as the deformation is continuously switched off. More realistically, several large basins would continue to persist, but the correspondingly simplified hypersurface topography should substantially shorten the search for global minima, particularly by using random initial conditions in E^N followed by steepest-descent mappings.

Larval insects of *Myrmeleon* or related genera trap their ambulatory food (frequently ants) by digging conical pits in sandy soil (at the bottom of which they lie in wait), and for that reason they are conventionally called "ant lions." By analogy, our deformed hypersurfaces should be arranged to trap random input data and to guide it to the relevant minimum. This approach hence might well be called the "ant-lion strategy."

In the following sections we will show how these ideas can be applied to simplify graph partitioning, first by defining appropriate random graphs, and then embedding them in E^N and implementing a specific hypersurface distortion.

3. IRREGULAR GRAPHS

Edge graphs for n vertices are defined by specifying bond matrices whose elements $B(i, j)$ $(1 \leqslant i, j \leqslant n)$, are equal to $+1$ if i and j are connected by a bond or 0 if they are unconnected. To illustrate our method, we employ two families of irregular, but nonrandom, graphs. We call these respectively the prime-factor graphs and the π graphs. Neither has any fundamental mathematical significance (to the best of our knowledge), but both provide a convenient testing ground for the simplification strategy.

3.1. Prime-Factor Graphs

Every positive integer m has a unique set of prime factors:

$$m = 2^{q_2(m)} 3^{q_3(m)} 5^{q_5(m)} \cdots \tag{3.1}$$

Define

$$\varepsilon(m) = \sum_p q_p(m) \qquad (\text{mod } 2) \tag{3.2}$$

that is, $\varepsilon(m)$ is $+1$ if m has an odd number of prime factors or 0 if it has an even number of prime factors. Prime-factor graphs are defined by requiring

$$B(i, j) = \varepsilon(i + j) \tag{3.3}$$

Table I contains the first few values for $\varepsilon(m)$. From these values one sees that the $n=2$ and $n=3$ prime-factor graphs are simply connected, while that for $n=4$ is multiply connected. The graph for $n+1$ vertices contains that for n vertices. We believe that all prime-factor graphs for $n \geqslant 4$ are multiply connected.

Figure 1 illustrates the $n=20$ prime-factor graph. It seems to be a reasonable assumption that in the large-n limit, the fraction of bonds present converges to 1/2.

3.2. π Graphs

The decimal representation for the transcendental number π is available to high order.[10] Symbolically it can be written

$$\pi = \sum_{j=0}^{\infty} C(j)\, 10^{-j} \tag{3.4}$$

Table I. Values of $\epsilon(m)$ for Small m

m	$\varepsilon(m)$
3	1
4	0
5	1
6	0
7	1
8	1
9	0
10	0
11	1
12	1
13	1
14	0
15	0
16	0
17	1
18	1
19	1
20	1
21	0
22	0
23	1
24	0
25	0

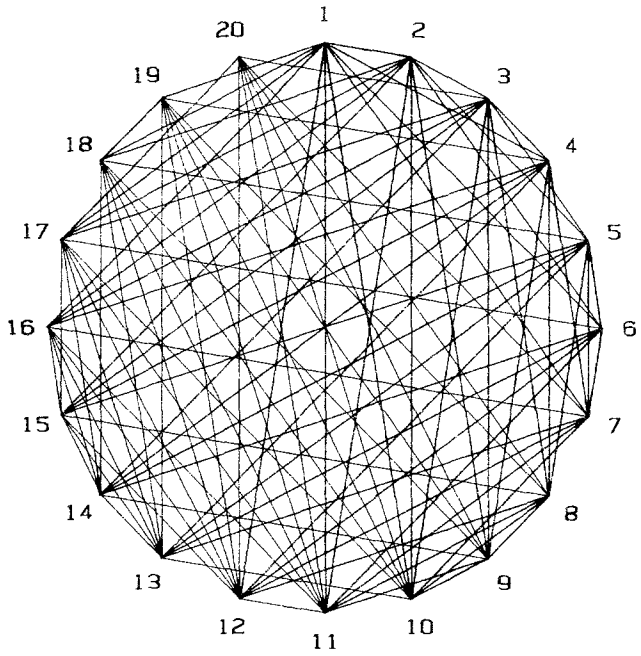

Fig. 1. Prime-factor graph for $n = 20$.

where $C(0) = 3$, $C(1) = 1$, $C(2) = 4$, etc. Next set

$$\eta(j) = C(j) \pmod 2 \tag{3.5}$$

so the η's are 0 or 1 according as the C's are even or odd. Table II lists the first few values of $\eta(j)$.

Next, order the $(1/2)\,n(n-1)$ vertex pairs (i, j) in "dictionary" fashion:

$$\begin{gathered}(1, 2)\\(1, 3)\\\vdots\\(1, n)\\(2, 3)\\\vdots\\(2, n)\\(3, 1)\\\vdots\\(n-1, n)\end{gathered} \tag{3.6}$$

Table II. Values of $\eta(j)$ for Small j

j	$\eta(j)$
1	1
2	0
3	1
4	1
5	1
6	0
7	0
8	1
9	1
10	1
11	0
12	1
13	1
14	1
15	1
16	0
17	1
18	0
19	0
20	0
21	0
22	0
23	0
24	1
25	1

It is easy to see that if $1 \leqslant i < j \leqslant n$, the pair (i, j) occurs at line

$$l(i, j) = (i-1)\, n - \tfrac{1}{2} i(i+1) + j \tag{3.7}$$

in the list. We use these line numbers as variable for the η's [Eq. (3.5)] to provide the π-graph B's:

$$B(i, j) = \eta[l(i, j)] \tag{3.8}$$

The π graphs for $n = 2$ and 3 are simply connected, while those of larger n all seem to be multiply connected. But unlike the previous case, the π graph of order $n+1$ does not generally contain that of order n as a subgraph. Due to the inherent irregularity of the decimal representation for π, $\eta(j)$ will be an irregular (and presumably unbiased) string of 0's and 1's; consequently, the fraction of bonds present in the large-n limit should again be 1/2.

Figure 2 displays the π graph for $n = 20$.

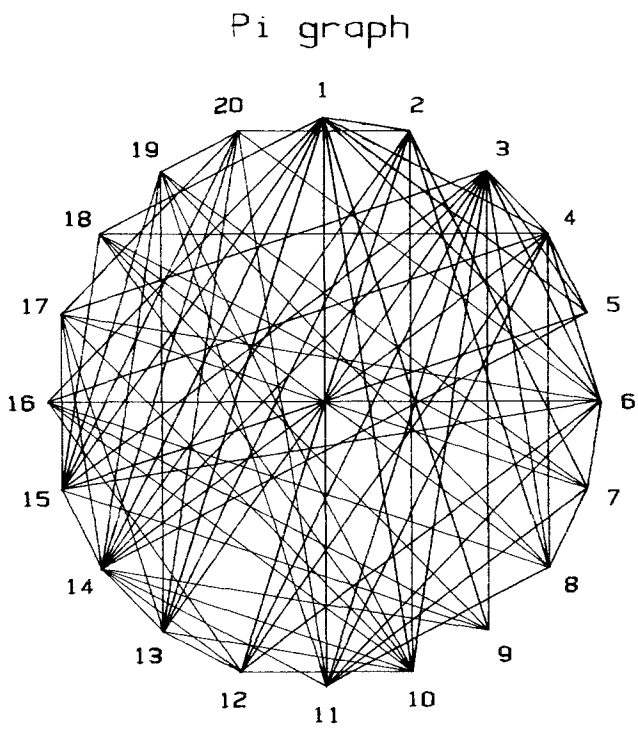

Fig. 2. π graph for $n = 20$.

4. GRAPH PARTITIONING

Given any one of the irregular graphs defined in Section 3, its n vertices can be divided in 2^n distinct ways between two sets **S** and **T**. For any such partition, call it P, let $b_+(P)$ stand for the number of bonds of the given graph that connect vertices in the same set, whether **S** or **T**. Likewise, let $b_-(p)$ stand for the number of bonds connecting vertex pairs, one member of which is in **S**, the other of which is in **T**. Our basic problem then is to find the minimum of

$$V(P) = b_+(P) - b_-(P) \tag{4.1}$$

for the given graph with respect to all partitions P. Note that the maximum of $V(P)$ is trivially achieved by placing all vertices in the same set. Notice also that switching all vertices between **S** and **T** leaves V unchanged.

The required minimum, or minima, can be found for small n by examining all partitions. But it is clear that this process bogs down quickly

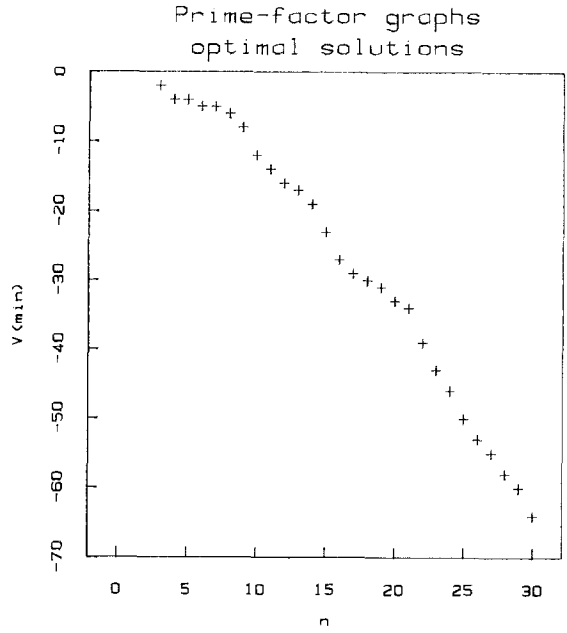

Fig. 3. Global minimum $V(P)$ values for n-vertex prime-factor graphs.

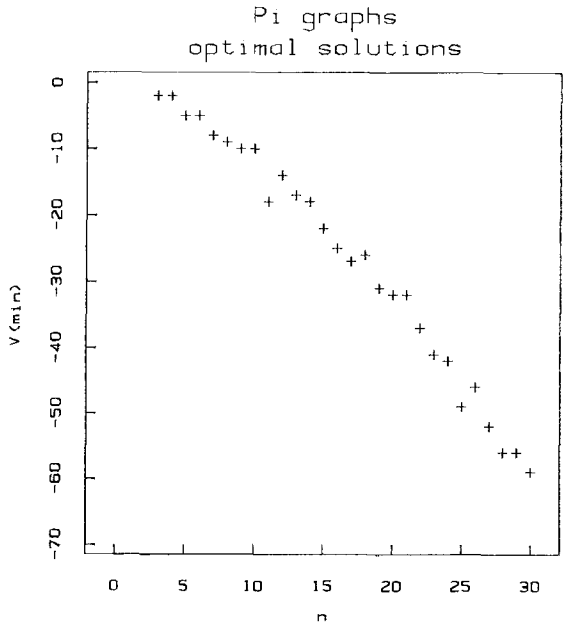

Fig. 4. Global minimum $V(P)$ values for n-vertex π graphs.

as n increases for irregular graphs of the type we are considering: this combinatorial optimization problem is evidently NP-complete.[7]

For the purpose of numerical orientation, we have calculated $V(P)$ directly for all partitions of all prime-factor and π graphs with orders $n \leqslant 30$. Figures 3 and 4 show plots versus n of the global-minimum V values for the two graph families. The impressions coveyed by these plots are similar (though not identical) to each other, namely that $V_{\min}$ is negative and asymptotically decreasing in a manner roughly proportional to n. Even when n is of the order of 20–30, the trends exhibited by these figures are sufficiently regular to justify extrapolations to $n \cong 10^2$ to give a rough idea about the optimal V's for those much larger problems.

Figures 5 and 6 provide distributions of $V(P)$ values, in histogram form, for the $n = 30$ irregular graphs. Not surprisingly, the most probable $V(P)$ values cluster around zero, as any naive probabilistic argument would indicate for random division of vertices between **S** and **T**. The key feature of these latter two figures is that the global minima are indeed far removed from the most probable values, in a region where the distributions

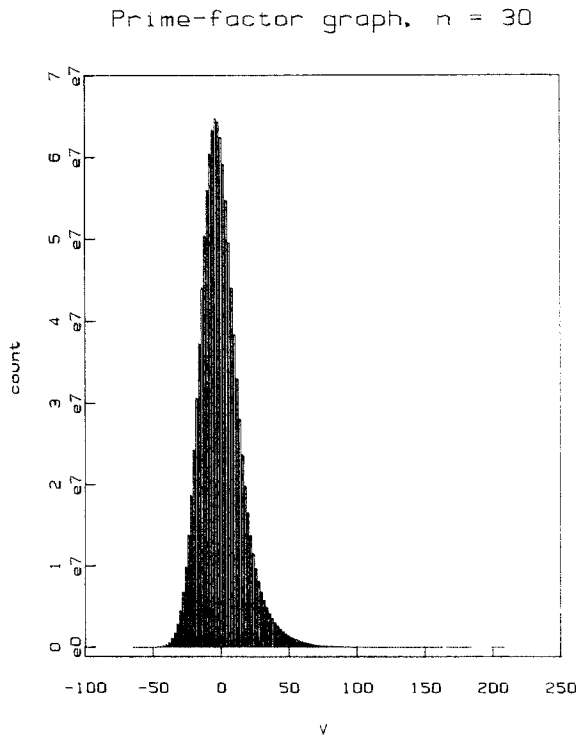

Fig. 5. Distribution of the 2^{30} values of $V(P)$ for the $n = 30$ prime-factor graph.

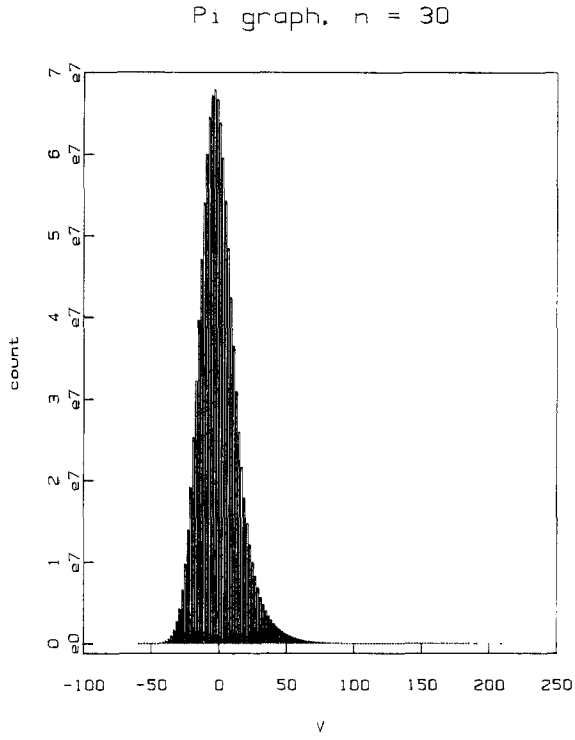

Fig. 6. Distribution of the 2^{30} values of $V(P)$ for the $n=30\,\pi$ graph.

have become very sparse. This is a visual indication of the general difficulty of finding the global minima, which becomes ever more severe as the graph order n tends to infinity.

5. EMBEDDING AND DEFORMATION

The quartic polynomial

$$v(x)=x^4-2x^2 \tag{5.1}$$

has minima at $x=\pm 1$, at which $v=-1$. Consequently, setting

$$\Phi_0(x_1\cdots x_n)=\sum_{i=1}^{n} v(x_i) \tag{5.2}$$

we know that Φ_0 will have 2^n equivalent minima (at which this function is $-n$) at the points $(\pm 1, \pm 1,..., \pm 1)$ in E^n. It is easy to see that the steepest-

descent basins for Φ are all equivalent, all stretch to infinity, and all meet at the multidimensional origin in E^n.

Next consider the n-dimensional potential function:

$$\Phi_\lambda(x_1 \cdots x_n) = \Phi_0(x_1 \cdots x_n) + \lambda \sum_{i<j=1}^{n} B(i, j)\, x_i x_j \tag{5.3}$$

where $\lambda \geqslant 0$, and $B(i, j)$ is the bond matrix for an irregular graph of order n. If λ is very small, the shape of the Φ_λ hypersurface will be close to that Φ_0, and in particular it will exhibit 2^n minima differing in positions and depths only by $O(\lambda)$ from those of Φ_0. At any one of these minima the x_i are partitioned into two subsets, namely those clustered around -1 and those clustered around $+1$. The obvious analogy to the vertex partitioning of the preceding section leads to the conclusion that to leading order in λ the depths of the Φ_λ minima will be given by

$$-n + V(P)\,\lambda + O(\lambda^2) \tag{5.4}$$

where the partition P now refers to the separation of x_i into negative and positive subsets. To this order in λ the optimal graph partitions correspond exactly to absolute Φ_λ minima and vice versa.

While λ remains small, the sizes and shapes of basins for Φ_λ obviously will differ little from those of Φ_0. A uniform random distribution of points inside a large hypersphere or hypercube centered at the E^n origin to leading order would uniformly sample the basins. The corresponding sampling of minima would then substantially reproduce the $V(P)$ distributions as exemplified by Figs. 5 and 6.

The simple expedient of increasing λ causes the Φ_λ hypersurface to deform in the manner required. This can be easily illustrated by examining the elementary $n = 2$ graph (identical for both the prime-factor and the π graph families), which consists of a single bond connecting two vertices. For this case we have

$$\Phi_\lambda(x_1, x_2) = v(x_1) + v(x_2) + \lambda x_1 x_2 \tag{5.5}$$

It is trivial to show that for small, positive λ the expected four minima occur in pairs, the deeper of which are located at

$$x_1 = -x_2 = \pm\tfrac{1}{2}(4 + \lambda)^{1/2} \tag{5.6}$$

with corresponding potential

$$\Phi_\lambda = -\tfrac{1}{8}(4 + \lambda)^2 \tag{5.7}$$

The shallower pair is located at

$$x_1 = x_2 = \pm\tfrac{1}{2}(4-\lambda)^{1/2} \tag{5.8}$$

and at these relative minima Φ_λ has the higher value

$$\Phi_\lambda = -\tfrac{1}{8}(4-\lambda)^2 \tag{5.9}$$

As λ increases toward 4, the latter pair, Eq. (5.8), move toward one another while displaying ever shallower basins. Just at $\lambda = 4$ they become coincident with each other and with the origin, and at that point cease to be minima. When $\lambda > 4$ only the pair in Eq. (5.9) survive, moving farther from the origin and deepening as λ continues to increase. Two of the ant-lions have expanded their trapping craters so much that the other two have been totally squeezed out!

The advantage resulting from increasing λ can be expected to extend to cases with arbitrary n. Unfavorable graph partitions will have their small-λ minima moved upward and inward, and eventually annihilated, as λ increases. At the same time the minima for the most favorable partitions will be moved downward, outward from the origin, and be surrounded by expanding basins. The next section tests this scenario quantitatively.

6. NUMERICAL RESULTS

The hypersurface deformation method has been tested on the partitioning of both $n = 30$ irregular graphs by several sets of specific calculations. For one of these sets the following procedure was used.

1. The parameter λ was initially set equal to 10.
2. A random collection of 1000 starting configurations $(x_1 \cdots x_{30})$, uniformly distributed over the hypercube $-1 \leqslant x_i \leqslant +1$, was created.
3. Steepest-descent trajectories were calculated from each of these random starting configurations to the basin minima for $\lambda = 10$.
4. The parameter λ was reset to 0.01.
5. Starting with each of the $\lambda = 10$ minimum configurations generated in step 3, above, steepest-descent trajectories were created on the $\lambda = 0.01$ hypersurface to locate the relevant collection of small-λ minima.
6. The x_i were rounded to the nearest integer (inevitably ± 1 at this stage), and then used to evaluate the partitioned graph energy,

$$V = \sum_{i<j} B(i,j)\, x_i x_j \tag{6.1}$$

The intention of this two-stage deformation was to use the large λ value as a means for selecting appropriate sign patterns among the x_i, and then to use the small λ value to move each of the x_i into the close neighborhood of ± 1.

Figure 7 shows the raw distribution of the 1000 V values obtained for the prime-factor graphs. This distribution should be compared with that in Figure 5 for all graph partitions. The strong bias toward the lowest possible V values is obvious, and prominently includes the global minimum -64 that was identified by complete enumeration.

The raw distribution presented in Fig. 7 contains multiple occurrences of specific partitions. Consequently, the data were analyzed to produce a derived distribution for the number of distinct partitions for each V. Figure 8 shows the result. The count for the global minimum energy $V = -64$ has been strongly depressed, indicating that it was highly redundant in the raw distribution. This is expected for strongly distended basins when $\lambda = 10$. However, only three of the six globally optimal minima

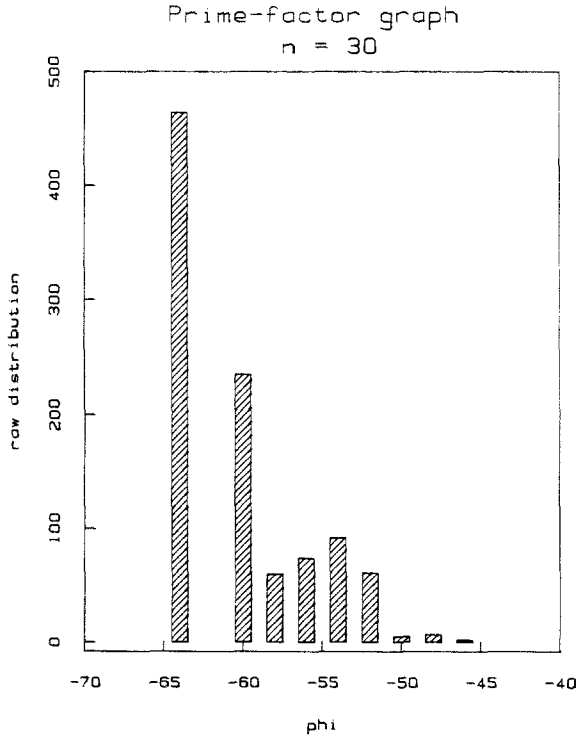

Fig. 7. Distribution of $n = 30$ prime-factor graph partition energies for the $\lambda = 10, 0.01$ deformation sequence.

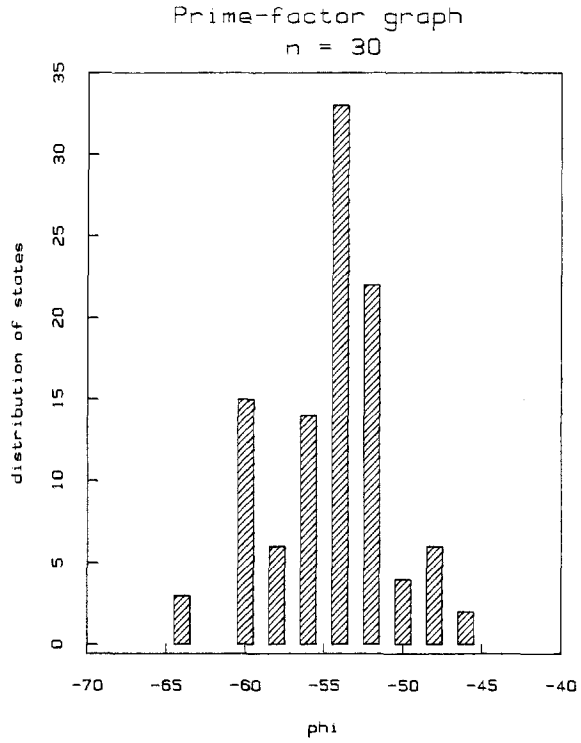

Fig. 8. Distribution of distinct partitions, by energy, derived from the data of Fig. 7.

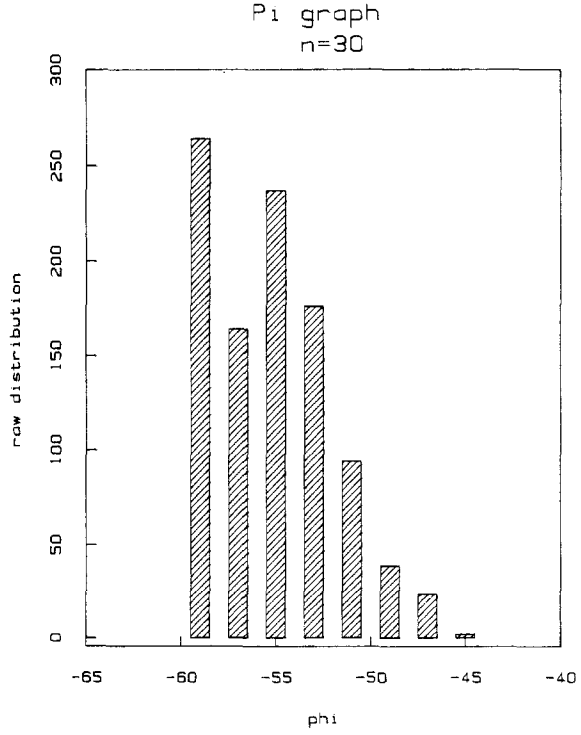

Fig. 9. Distribution of $n = 30$ π-graph partition energies for the $\lambda = 10$, 0.01 deformation sequence.

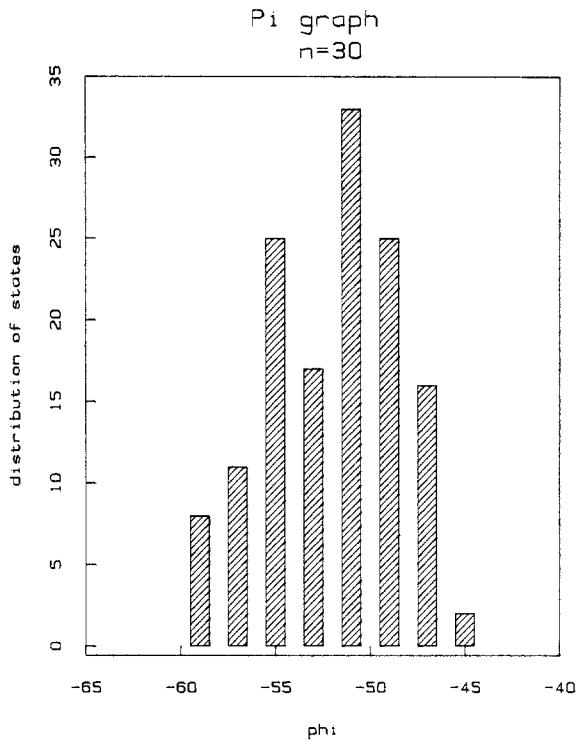

Fig. 10. Distribution of distinct partitions, by energy, derived from the data of Fig. 9.

uncovered by enumeration were produced by the hypersurface distortion search.

Figures 9 and 10 display the analogous distributions for the π graph of order 30. The pattern is quite similar to that conveyed by Figs. 7 and 8 for the prime-factor case, though the global-minimum V is now -59. Once again the deformation and steepest-descent sequence 1–6 strongly biases the outcome toward optimal or near-optimal partitions. Eight out of ten global minima (identified by enumeration) were encountered.

Several variants of the specific sequence 1–6 have also been examined. These include, first, use of $\lambda = 100$, or alternatively $\lambda = 1000$, in step 1 above (with subsequent steps (2–6) unaltered). Second, a three-stage λ reduction scheme was employed, using successively the values 100, 10, and 0.01. The final distributions were very little changed by these alternatives, while the execution times were increased somewhat. In addition, all of these procedures were carried out for the two $n = 20$ graphs as well, with similar (but more quickly attained) results.

7. DISCUSSION

Although the partitioning of two specific families of irregular graphs has been used for illustration, it should be clear that the hypersurface deformation method should have wider applicability. First, it is obvious that a more general class of partitioning problems could be studied where the bond matrix elements $B(i, j)$ were not restricted to values 0 and 1, but could take arbitrary real values. The resulting optimization problems are equivalent to finding ground-state spin configurations for Ising spin glasses.[4]

The "traveling salesman problem" can also be cast into an N-dimensional continuum format and optimal solutions sought by deforming the resulting hypersurface. Let $\mathbf{R}_1 \cdots \mathbf{R}_n$ be the locations of n "cities" or fixed sites in D-dimensional Euclidean space ($D = 2$ for the conventional traveling salesman problem). Next introduce n movable particles whose variable positions are denoted by $\mathbf{r}_1 \cdots \mathbf{r}_n$. The basic potential hypersurface can be defined as follows:

$$\Psi(\mathbf{r}_1 \cdots \mathbf{r}_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} V_b(|\mathbf{r}_i - \mathbf{R}_j|) + \sum_{i<j=1}^{n} V_r(|\mathbf{r}_i - \mathbf{r}_j|) + \sum_{i=1}^{n} |\mathbf{r}_i - \mathbf{r}_{i-1}| \quad (7.1)$$

where by convention $\mathbf{r}_0 \equiv \mathbf{r}_n$. The V_b terms represent short-range binding potentials for particles to cities, the V_r terms produce interparticle repulsions to prevent multiple occupancy of cities by particles, and the last sum in Eq. (7.1) adds up the length of the cyclic tour. This approach views different tours of the cities as distinct ways of binding n vertices of a kind of cyclic "polymer" onto the fixed binding sites (cities). The optimal tour will be the one (aside from cyclic permutations) that minimizes the last sum in Eq. (7.1).

Specific forms for V_b and V_r might be the following:

$$\begin{aligned} V_b(r) &= -l^{-1} \exp(-r^2/l^2) \\ V_r(r) &= (2r)^{-1} \exp(-r^2/4l^2) \end{aligned} \quad (7.2)$$

containing a positive parameter l whose variation deforms the Ψ hypersurface in a useful way. In particular, the search calculation would be started with a relatively large l value, so binding attractions to cities are weak but long-ranged and indeed even overlapping, and the repulsions likewise are relatively long-ranged. Starting from virtually any initial set of particle positions, with this large l value steepest descent on the Ψ hypersurface would at best only produce a rough approximation to a legitimate tour. Nevertheless, the long range of V_b and V_r should substantially simplify the topography of Ψ. Subsequent reduction in l toward zero will shorten the

range of both V_r and V_b while tightening the binding of the latter, so that in the limit of vanishing l, single particles become coincident with their own cities. The tour length L then formally follows from

$$\lim_{l \to 0} [\Psi(\{\mathbf{r}_i\}_{\min}, l) - nV_b(0, l)] = L \tag{7.3}$$

Initial studies of this approach to solving the traveling salesman problem are favorable, but suggest that it may be desirable to combine the hypersurface deformation strategy with simulated annealing using the molecular dynamics approach for the cyclic polymer binding process.

We remark finally that integrated circuit layout problems evidently can also be stated in analogous form, using component interaction potentials whose smoothness and range are varied in a beneficial way to achieve the "ant-lion" effect.

REFERENCES

1. E. L. Lawler, J. K. Leustra, A. H. G. Rinnooy Kan, and K. B. Shmoys, eds., *The Travelling Salesman Problem* (North-Holland, Amsterdam, 1979).
2. G. Baskaran, Yaotian Fu, and P. W. Anderson, *J. Stat. Phys.* **45**:1 (1986).
3. H. A. Scheraga, *Chem. Rev.* **71**:195 (1971).
4. M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
5. S. Kirkpatrick, C. D. Gelatte, Jr., and M. P. Vecchi, *Science* **220**:671 (1983).
6. B. Bollabás, *Ggaph Theory, An Introductory Course* (Springer-Verlag, New York, 1979).
7. M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (Freeman, San Francisco, 1979).
8. F. H. Stillinger and T. A. Weber, *Science* **225**:983 (1984).
9. R. A. LaViolette and F. H. Stillinger, *J. Chem. Phys.* **83**:4079 (1985).
10. Cover design, *Math. Intelligencer* **7**(3) (1985).