1987

In: Vision, Brain and Cooperative Computation,
Arbib, M. A. & Hanson, A. R., eds.,
Cambridge: MIT Press, 1985

Separating Figure from Ground with a Boltzmann Machine

Terrence J. Sejnowski

Department of Biophysics
The Johns Hopkins University
Baltimore, Maryland 21218

Geoffrey E. Hinton

Computer Science Department
Carnegie-Mellon University
Pittsburgh, PA 15213

Address correspondence to:

Dr. Terrence Sejnowski
Biophysics Department
Johns Hopkins University
Baltimore, MD 21218
Phone: (301) 338 8687

Introduction

Many problems in visual processing can be formulated as searches: Given an image or sequence of images find the best interpretation from amongst a large set of possible internal models. That we are able to recognize three-dimensional objects in images within a few hundred milliseconds implies an effective search strategy. Mistakes, when they do occur, are usually confusions among similar objects. These fast, effortless and generally reliable searches are carried out in parallel by a large number of neurons in the visual cortex. The architecture of visual cortex in primates has inspired recent parallel models of visual computation (Arbib 1975; Marr, 1982; Feldman & Ballard, 1982; Ballard, Hinton & Sejnowski, 1983).

In this chapter we review a class of parallel visual algorithms that use relaxation to perform rapid best-fit searches and we examine some of the difficulties inherent with this search technique. In particular, we analyze the problem of separating figure from ground in an image and show how a parallel relaxation algorithm can be trapped in states that are locally optimal but globally incorrect. A general parallel search method is introduced, based on statistical mechanics, that overcomes this shortcoming and finds globally optimal solutions with a high probability (Kirkpatrick et al., 1983; Hinton & Sejnowski, 1983). This approach is effective in small-scale simulations of parallel visual algorithms; its usefulness for large problems is still uncertain.

An intriguing aspect of the stochastic search procedure is that it depends on the presence of noise, which normally is considered a nuisance and typically degrades the performance of a system. There

1970; Julesz, 1971; Dev, 1975; Nelson, 1975; Marr & Poggio, 1976). The problem is then reduced to finding the matches that best satisfy all the local constraints.

In the Marr-Poggio (1976) algorithm for random-dot stereograms, each unit stands for a binary hypothesis about the correspondence of a particular pair of dots and therefore represents the existence of a patch of surface at a particular depth. There are excitatory interactions between neighboring units with the same depth to ensure continuity of surfaces, and inhibitory interactions between units that represent different depths at the same image location to ensure that depth assignments are unique; if the sum of all the inputs to a unit from the two images and from local interactions is above threshold, the value of the unit is set to 1, and otherwise to 0. Starting from all 0's, the units are iteratively updated: During the relaxation various combinations of depth assignments are tried and the network eventually "locks" into a generally consistent solution in a way that resembles our perceptual experience when we fuse random-dot stereograms (Julesz, 1971).

In general it is not possible to prove that this algorithm always converges to the correct depth assignments, in part because small clusters of units may form coalitions that are locally optimal but are not the globally best solution (Burt, 1977; Marr, Palm & Poggio, 1978). Another drawback of this relaxation method is the large number of iterations required to reach the final solution. If there are only nearest-neighbor interactions between units, then at least as many iterations are required as there are units across the image,

that a globally optimal solution can no longer be assured. Discrete decisions must therefore be made together with the estimation of continuous variables: similar problems occur in many other computations of intrinsic surface properties in early vision (Ballard, Hinton & Sejnowski, 1983).

Figure-Ground Separation

One of the simplest problems in visual perception where a discrete choice at a boundary affects subsequent processing is the organization of figure and ground in an image (Weisstein, this volume). An illustration of the classic drawing that can be interpreted as either a vase or two faces is shown in Fig. 1. The drawing gives rise to two percepts depending on whether the figural part of the drawing is on the inside of the outside of the closed outline. We are remarkably good at performing the separation and can report within a few hundred ms whether a small spot is inside or outside a briefly-flashed closed outline (Ullman, 1983). The discrimination probably requires two steps: a segmentation of the figure and ground, and a subsequent decision about whether the spot is located in the figure.

We briefly summarize here a simple parallel relaxation model of one type of process that occurs during figure-ground separation (Sejnowski, Hinton, Kienker & Schumacher, 1985; for previous work on scene segmentation using relaxation algorithms see Prager, 1980; Zucker & Hummel, 1979; Danker & Rosenfeld, 1981). The model is designed to mark the inside or the outside of connected figure when given some lines that represent its edges and an "attentional spotlight" that provides a bias to either the inside or outside. Examples of these

the two edge units inhibit each other.

To implement the constraint that lines in the input require interpretation, each line segment provides equal excitatory input to the two relevant edge units. To implement the constraint that edges are implausible in places where there are no lines in the input, edge units have high thresholds that normally require excitatory input to overcome them. To implement the constraint that edges tend to be continuous, a figure unit supports the colinear neighbors of its bounding edge units. This was found to work better than direct support between the colinear edge units themselves, because it allows edge completion to occur around the figure region, but not elsewhere.

The complete set of interactions of a figure unit and an edge unit are shown in Figure 2. The precise strengths of the interactions were chosen by trial and error using a variety of outlines and were guided by the following two considerations:

1. The region within the attentional spotlight should tend to be figure and the region outside should tend to be background.

2. The discontinuity between figure and background should normally appear as a line in the image, and so there should be a penalty for "open frontier" where the figure region ends without there being a line in the image.

Whenever the spotlight of attention does not precisely align with the lines in the image, these two considerations are antagonistic and it is the frustration between them that makes it necessary to perform a best-fit search.

One of he simplest updating algorithms consists of choosing a

separation previously discussed have the property that the connections, considered a matrix, are symmetric. A large class of constraint-satisfaction problems can be implemented with symmetric weights, including ones that require asymmetric constraints between hypotheses. For example, two hypotheses related by implication can be implemented by two units connected by symmetric weights and having different thresholds (Hinton & Sejnowski, 1983). Symmetric connectivity has the significant advantage that optimization techniques and variational principles can be used to analyze the performance of the network (Hummel & Zucker, 1983). In particular, Hopfield (1982) has shown that one can define an "energy" for a symmetric network of binary hypotheses that can be used to analyze its convergence. Each state is assigned an energy according to

$$E = -\frac{1}{2}\sum_{i\neq j} w_{ij}s_i s_j - \sum_i (\eta_i - \theta_i)s_i \qquad (1)$$

where s_i is the state of unit i, w_{ij} is the strength of connection between the units i and j, η_i are the inputs to unit i, and θ_i the threshold of unit i. A simple asynchronous algorithm for finding the combination of hypotheses that has a local energy minimum is to choose asynchronously a unit at random and set its state to the one with the lowest energy. Because of the symmetric weights, this updating rule requires that the unit be set to 1 if the "energy gap"

$$\Delta E_i = \sum_j w_{ij}s_j + \eta_i - \theta_i \qquad (2)$$

dynamic systems (Binder, 1978) and has recently been applied to problems of constraint satisfaction (Kirkpatrick et al., 1983; Hinton & Sejnowski, 1983; Smolensky, 1983; Geman & Geman, 1985; Bienenstock, 1985). Boltzmann Machines (Fahlman, Hinton & Sejnowski, 1983) are networks of binary processors that use as their update rule a form of the Metropolis algorithm that is suitable for parallel computation: If the energy gap between the 1 and 0 states of a unit is ΔE_i then regardless of the previous state set the unit to 1 with probability

$$p_i = \left(1 + e^{-\Delta E_i / T} \right)^{-1} \qquad (3)$$

where T is a parameter that acts like temperature (see Fig. 5). Observe that as T approaches zero, Eq. (3) approaches a step function, the deterministic update rule for binary threshold units already introduced.

Our analysis of Boltzmann Machines is based on the statistical mechanics of physical systems (Schroedinger, 1946). The probabilistic decision rule in Eq. (3) is the same as the equilibrium probability distribution for a system with two energy states. A system of particles in contact with a heat bath at a given temperature will eventually reach thermal equilibrium and the probabilities of finding the system in any global state will then obey a Boltzmann distribution. Similarly, a network of units obeying this decision rule will eventually reach "thermal equilibrium" in which the relative probability of two global states of the network follows the Boltzmann distribution:

$$\frac{P_\alpha}{P_\beta} = e^{-(E_\alpha - E_\beta)/T} \qquad (4)$$

at golf (Andrew Witkin, personal communication). It is therefore not at all clear whether simulated annealing would be useful when trying to satisfy multiple weak constraints such as those found in visual algorithms.

As a test case we have applied the Metropolis algorithm and simulated annealing to the parallel algorithm for separating figure from ground introduced previously (A more detailed account can be found in Sejnowski, Hinton, Kienker & Schumacher, 1984). At high temperatures the figure and edge units make a structureless pattern, as shown in Fig. 6a. As the temperature is exponentially reduced the figure units around the center of attention tend to remain on, and these on average support those edge units whose orientation is consistent with them (Fig. 6b). As the temperature is further reduced local inconsistencies are resolved and the entire network "crystalizes" to the correct solution. In a series of 1000 annealings from random starting configurations every trial reached the correct solution, as shown by the histogram in Fig. 7. A single iteration consisted of 2,000 updates in which one of the 2,000 units in the problem was chosen at random. Similar results have been obtained for a variety of simple figures, including ones where the outline is incomplete. In contrast, the performance of the algorithm on spirals using the same annealing schedule is very poor and humans also have great difficulty with the same figures; however, with a much slower annealing schedule the algorithm reliably finds the correct solution.

The model of figure-ground separation presented here is clearly much too simple to explain how the problem is solved in our visual

Relationship Between Boltzmann Machines and Neural Models

The energy gap for a binary unit has a role similar to that played by the membrane potential for a neuron: both are the sum of the excitatory and inhibitory inputs and both are used to determine the output state through a nonlinear transformation. However, a neuron produces action potentials, which are brief spikes that propagate down its axon, rather than a binary output. When the action potential reaches a synapse, the signal it produces in the postsynaptic neuron rises to a maximum and then decays with the time constant of the membrane (typically around 5 msec for neurons in cerebral cortex). The effect of a single spike on the postsynaptic cell body may be further broadened by electrotonic transmission down the dendrite to the spike-initiating zone near the cell body.

This suggests a neural interpretation for the binary pulses in a Boltzmann Machine: If the average time between updates is identified with the average duration of a postsynaptic potential then the binary pulse between updates can be considered an approximation to the postsynaptic potential. Although the shape of a single binary pulse differs significantly from a postsynaptic potential, the sum of a large number of pulses stochastically impinging on a processing unit is independent of the shape of the individual pulses. Thus for networks having the large fan-ins typical of cerebral cortex (several thousand) the energy gap of a binary unit should behave like the membrane potential of a spike-producing neuron.

In addition to the nonlinear membrane currents in axons that

inputs.

Intracellular recordings in the central nervous system reveal stochastic variability in the membrane potential of most neurons, in part due to fluctuations in the transmitter released by presynaptic terminals. Other sources of noise may also be present and could be controlled by cellular mechanisms (Verveen & Derksen, 1968; Holden, 1976). If some sources of noise in the central nervous system are gated or modulated, it should be possible to experimentally identify them. For example, the noise could be regularly cycled and this would be apparent in the massed activity. Alternatively, noise may always be present at a low level and be increased irregularly whenever there is an identified need.

In the visual cortex of primates single neurons respond to the same visual stimulus with different sequences of action potentials on each trial (Sejnowski, 1981). In order to measure a repeatable response, spike trains are typically averaged over 10 trials. The result, called the post-stimulus time histogram, gives the probability for a spike to occur as a function of the time after the onset of the stimulus. However, this averaging procedure removes out all information about the variance of the noise, so that there is no way to determine whether the noise varies systematically during the response to the stimulus or perhaps on a longer time scale, while the stimulus is being attended. Such measurements of the noise variance over a range of time scales could provide evidence that this parameter has an active role in neural processing.

There are two ways to view the sigmoidal probability rule used

few ms. The time required for transmission of a spike down the axon to the nerve terminal, for the release of neurotransmitter, and for postsynaptic integration can delay the signal from reaching the spike-initiating zone of the target neuron by several ms. In some simulations both simultaneous updates and transmission delays were included and these appear to increase the noise in the system, effectively increasing the temperature (Sejnowski, Hinton, Kienker & Schumacher, 1985; Venkatasubramanian & Hinton, 1985). At low temperatures these effects are less pronounced because the rate of flipping is lower; thus, simultaneous decisions and time delays contribute noise that could mimic annealing even without an explicit temperature control (Francis Crick, private communication). Time delays are especially effective at introducing noise, and a delay of one iteration (2,000 updates in these simulations) starting from a random state and running at T = 1 was almost as effective as the standard exponential annealing.

Learning in Cerebral Cortex and Boltzmann Machines

The values of weights between units for the two examples of networks discussed in this chapter were chosen as much by trial and error as by plan, and it would be desireable to have an automated procedure for incorporating the constraints from a given task domain into the weights. The evolution of cerebral cortex is closely linked to the ability of mammals to learn from experience and adapt to their environment; this adaptability may be the consequence of rules for modifying the strengths of cortical synapses.

A single weight between two units can be considered a "microscopic"

areas. Many problems in speech recognition, associative retrieval of information, and motor control can be formulated as searches. However, there is a serious obstacle that appears to prevent symmetric modules from modeling sequential information processing: At thermal equilibrium there can be no consistent sequences of states. It is tempting to use asymmetrical weights to produce sequences, but this would be incompatible with the central idea of performing searches by settling to equilibrium.

An alternative that we are exploring is sequential settlings in a hierarchy of asymmetrically connected modules. The result of each search could be considered a single step in a strictly serial process, with each search setting up boundary conditions for the next. An attractive possibility for speeding up sequential settlings is to cascade partial settlings so that an approximate solution for one module could be used to start the search for the next one up the line (McClelland, 1979). Although there are some similarities between the organization of cerebral cortex and parallel stochastic search in Boltzmann machines, we need more experience with larger problems and a wider range of applications before the general usefulness of this approach can be properly assessed (Fahlman, Hinton & Sejnowski, 1983).

Acknowledgment

This research was supported by grants from the System Development Foundation and a grant from the National Science Foundation (BNS-8351331) to T.J.S. We especially thank Paul Kienker, Lee Schumacher, and Tony Yang for their assistance with the simulations.

Dosher, Sperling, & Wurst, in press.

Fahlman, S. E., Hinton, G. E. & Sejnowski, T. J., 1983. Massively-parallel architectures for AI: NETL, THISTLE and Boltzmann Machines, Proceedings of the National Conference on Artificial Intelligence, Washington, D. C., 109-113.

Feldman, J. A. & Ballard, D. H., 1982. Connectionist models and their properties, Cog. Sci. 6: 205-254.

Geman, S. & Geman, D., 1985. Stochastic relaxation, Gibbs distributions, and the Baysian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence 6:

Grimson, W. E. L., 1981. From image to surfaces, Cambridge: MIT Press.

Hinton, G. E. & Sejnowski, T. J., 1983. Optimal perceptual inference, Proceedings of the IEEE Computer Society Conference on Computer Vision & Pattern Recognition, Washington, D. C., 448-453.

Hinton, G. E., Sejnowski, T. J. & Ackley, D., 1984. Boltzmann machines: Constraint-satisfaction networks that learn, Carnegie-Mellon Computer Science Technical Report.

Holden, A. V., 1976. Models of the stochastic activity of neurones, Lecture Notes in Biomathematics 12, Berlin: Springer-Verlag.

Hopfield, J. J., 1982. Neural networks and physical systems with emergent collective computational abilities, Proceedings of the National Academy of Sciences USA 79: 2554-2558.

Hummel, R. A. & Zucker, S. W., 1983. On the foundations of relaxation labeling processes, IEEE Transactions on Pattern Analysis & Machine Intelligence, 5: 267-287.

& Teller, E., 1953. Equation of state calculations by fast computing machines, J. Chem. Phys. 21: 1987-1092.

Miller, J., Rall, W. & Rinzel, J., 1985. Synaptic amplification by active membrane in dendritic spines (in press)

Nelson, J. I., 1975. Globality and stereoscopic fusion in binocular vision, J. Theor. Biol. 49: 1-88.

Nishihara, H. K., 1984. PRISM: A practical real-time imaging stereo matcher. MIT Aritifical Intelligence Laboratory Memo #780.

Perkel, D. H. & Perkel, D. J., 1985. Dendritic spines: role of active membrane in modulating synaptic efficacy, Brain Research (in press).

Poggio, G. F. & Poggio, T., 1984. The analysis of stereopsis, Ann. Revs. Neurosci. 7: 379-412.

Prager, J. M., 1980. Extracting and labeling boundary segments in natural scenes, IEEE Transactions on Pattern Analysis and Machine Intelligence 2, 16-27.

Prazdny, K., 1985, Detection of binocular disparities, in press.

Rosenfeld, A. & Vanderbrug, G. J., 1977. IEEE Transactions on Systems Man & Cybernetics 7: 104-107.

Rubin, E., 1915. Synoplevede Figurer, Copenhagen.

Schroedinger, E., 1946. Statistical thermodynamics, London: Cambridge University Press.

Sejnowski, T. J., 1981. Skeleton filters in the brain. In Parallel models of associative memory, Hinton, G. E. & Anderson, J. A., eds. Hillsdale, New Jersey: Erlbaum Publishers.

Sejnowski, T. J., Hinton, G. E., Kienker, P., & Schumacher, L.,

von Neumann, J., 1966. Theory of self-reproducing automata, A. W. Burks, ed., Urbana: University of Illinois.

Waltz, D., 1975. Understanding line drawings of scenes with shadows. In The psychology of computer vision, Winston, P. H., ed., New York: McGraw-Hill, pp. 19-91.

Wolfram, S., 1983. Statistical mechanics of cellular automata, Revs. Mod. Phys. 55: 601-644.

Zucker, S., 1983. Computational and psychological experiments in grouping: early orientation selection, In Human and machine vision, Beck, J., Hope, B. & Rosenfeld, A., eds., New York: Academic Press.

Zucker, S. W. & Hummel, R. A., 1979. Toward a low-level description of dot clusters: labeling edge, interior and noise points, Comp. Graph. Image Proc. 9: 213-233.

between edge and figure units.

3. Two types of inputs to the figure-ground module: Bottom-up inputs from the image to some of the edge units (arrowheads) which in this case form a 9x6 rectangle, and top-down attentional inputs to the figure units (cross-hatched squares). The strengths of the inputs to the figures units have a Gaussian distribution centered on the unit just to the right of the rectangle's center given by $15e^{-(d/2)^2}$ where d is the Euclidean distance of the unit from the center of attention. The figure units that are shown cross-hatched are those whose attentional input exceeds 1. Each figure unit has a threshold 41, so the top-down input is not enough by itself to turn the figure units on. The edges composing the outline of the 9x6 rectangle have external inputs of 60 and all edge units have thresholds of 45. Thus, there was a strong bias for edge units composing the outline to be on; however, both types edge units at each position of the outline received equal input.

4. Final state of the figure-ground module using the gradient-descent update rule (T = 0). The simulation was started from a random starting state with approximately 1 out of 10 units on. Each iteration consisted of 2,000 updates and for each update one of the 2,000 units was chosen at random, the weighted inputs from other active units were summed, and the binary threshold rule applied to determine its new state. The system reached the steady-state configuration shown here after 28 iterations. Notice that the bottom line of figure units has been incorrectly stabilized outside the rectangle.

5. Probability for a unit to be on as a function of the energy

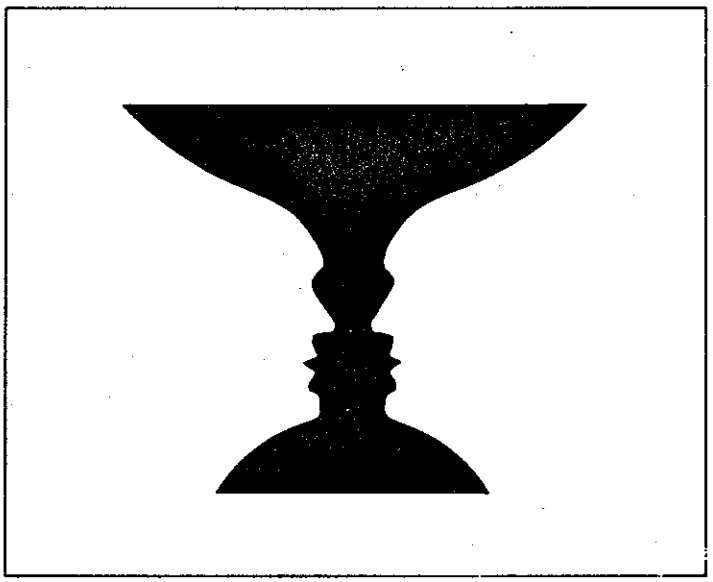

Figure 1

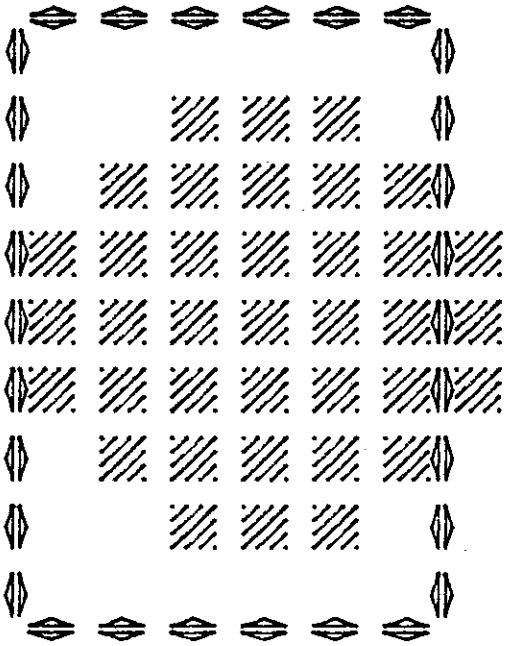

Figure 3

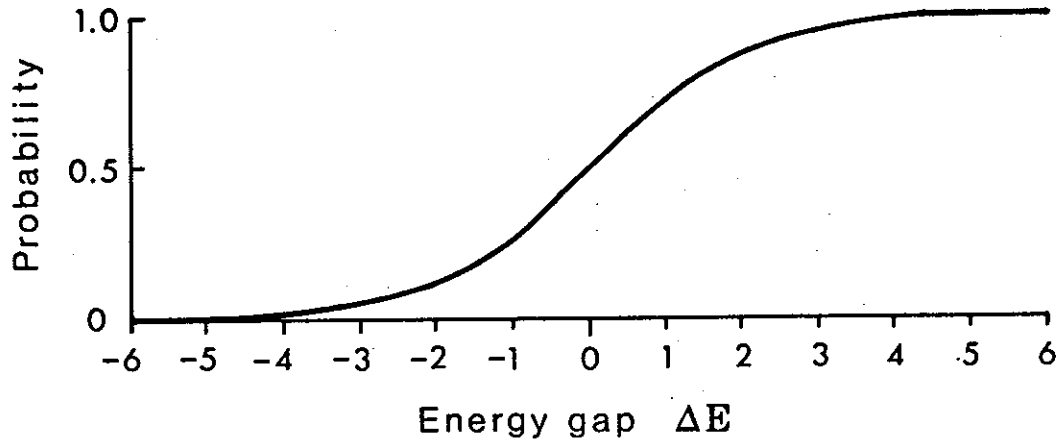

Figure 5

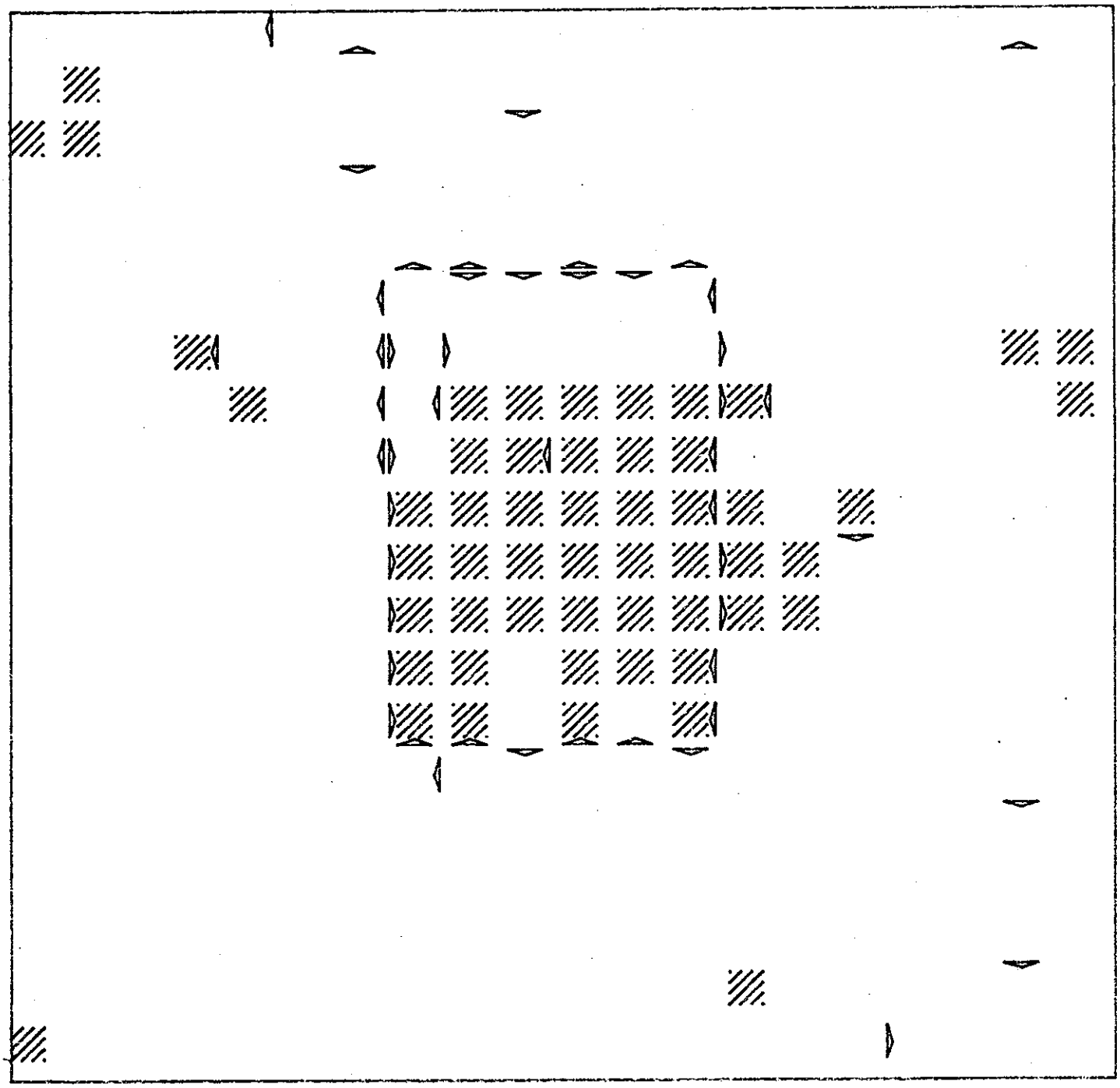

Figure 6b

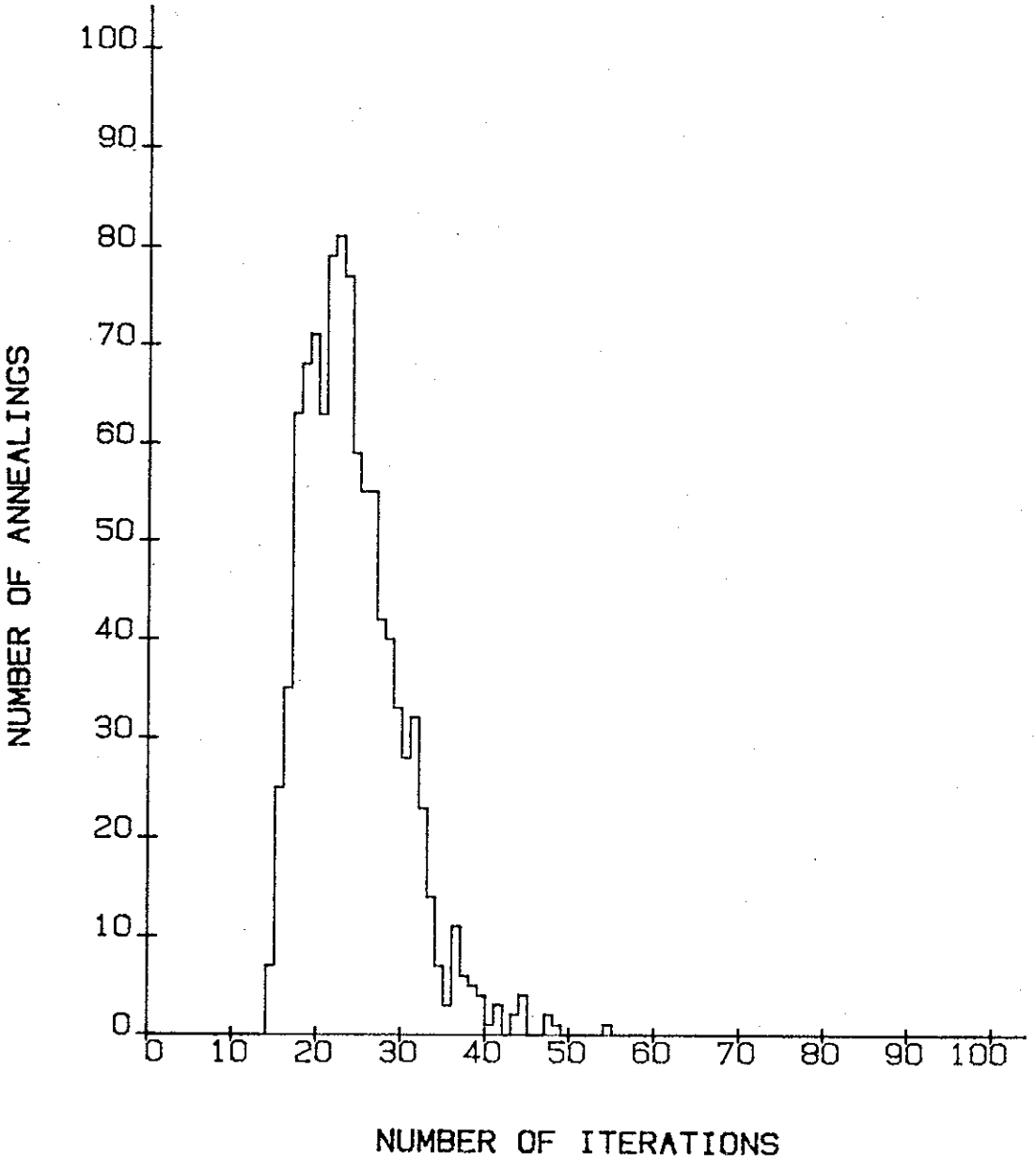
NUMBER OF ANNEALINGS
100
90
80
70
60
50
40
30
20
10
0
0 10 20 30 40 50 60 70 80 90 100
NUMBER OF ITERATIONS

Figure 7