

## Fixed-Point Attractors in Analog Neural Computation

F. R. Waugh, C. M. Marcus, and R. M. Westervelt

*Division of Applied Sciences and Department of Physics, Harvard University, Cambridge, Massachusetts 02138*

(Received 20 November 1989)

We show analytically that the expected number of fixed-point attractors in an associative memory neural network with analog neurons decreases exponentially as the neuron gain is reduced. Eliminating fixed-point attractors by using analog neurons has beneficial effects similar to stochastic annealing but can be easily implemented in a deterministic dynamical system such as an analog electronic circuit. Numerical data based on fixed-point counts in small networks support the analytical results.

PACS numbers: 87.10.+e, 06.50.Mk, 64.60.Cn

A general feature of optimization problems is the existence of a large number of possible solutions, with few good solutions (as defined by some cost measure) and many bad ones. This multiplicity of solutions is a common feature of spin glasses,<sup>1-3</sup> neural networks,<sup>4</sup> problems in artificial vision such as edge detection<sup>5</sup> and stereoscopic depth perception,<sup>6</sup> as well as the more traditional optimization problems such as the traveling salesman problem<sup>7,8</sup> and the graph matching problem.<sup>9</sup>

One approach to finding good solutions to an optimization problem is to recast the problem into a dynamical system which evolves on a complicated energy landscape. In this formulation, the difficulty of the problem becomes avoiding the many local minima in the landscape which correspond to bad solutions. Stochastic algorithms such as simulated annealing,<sup>10</sup> in which occasional uphill steps allow the system to climb out of shallow local minima, work well at finding near-optimal solutions but are notoriously slow and poorly suited for implementation in massively parallel hardware. For the case of optimization using neural networks, which we focus on in this Letter, finding a good solution quickly is crucial to making the system practical for real-time problems.

Neural-network models based on two-state neurons have received particular attention within the physics community recently because of their similarity to Ising spin glasses.<sup>11</sup> One aspect of these models which differs from biological networks is that real neurons have a continuous rather than a binary input-output transfer function.<sup>12</sup> Whether this difference is computationally significant and, in general, how the neuron transfer function affects the performance of the network has not been studied systematically. However, the benefits of analog computation have been observed numerically in a variety of applications.<sup>13-16</sup> For example, Hopfield and Tank<sup>7</sup> found that the steepness of the analog neuron's transfer function strongly affects the ability of an analog neural network to compute good solutions to the traveling salesman problem. Using analog neurons with low gain has also been shown to improve performance in associative memory neural networks,<sup>14,15</sup> as well as artificial vision systems<sup>5</sup> and in a recent "elastic net" approach to the

traveling salesman problem.<sup>16</sup>

In this Letter we show analytically that analog neural networks do indeed have a computational advantage over their binary-neuron counterparts. The advantage is that analog neural networks have exponentially fewer spurious fixed-point attractors, as indicated schematically in Fig. 1. In particular, we show that the expected number of fixed points in an analog associative memory network with Hebb-rule interconnections behaves as  $e^{NF(\alpha,\beta)}$ , where  $N$  is the size of the network and the function  $F(\alpha,\beta)$  depends on the neuron gain parameter  $\beta$  and the ratio  $\alpha$  of stored memories to neurons as shown in Fig. 2. Our results do not depend on the details of the network dynamics; for example, the network may evolve as a set of differential equations or as an iterated map with either parallel or sequential updating. We also show that the analytical results for  $F(\alpha,\beta)$  agree well with numerical counts of stable fixed points in small analog networks ( $20 < N < 90$ ) at several values of storage ratio  $\alpha$  and gain  $\beta$ .

We emphasize that analog dynamics is not the same as Ising dynamics at finite temperature. Besides the important conceptual difference—deterministic dynamics on a smooth landscape versus stochastic dynamics on a

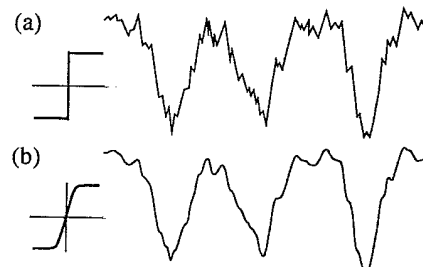


FIG. 1. Schematic landscapes showing energy or a Lyapunov function (vertically) as a function of an abstract phase-space coordinate (horizontally) for networks with (a) two-state neurons and (b) analog neurons with finite gain. Inset: neuron transfer function. Decreasing neuron gain smooths the energy landscape and greatly reduces the number of spurious states.

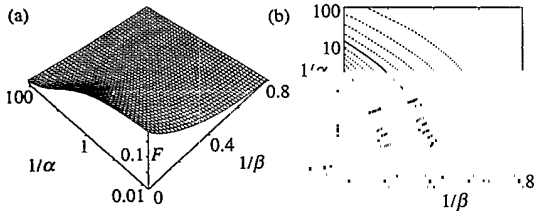


FIG. 2. Theoretical values of  $F(\alpha, \beta)$  as a function of inverse neuron gain  $1/\beta$  and inverse storage ratio  $1/\alpha$ , for the neuron transfer function  $f(x) = \tanh(x)$ . The expected number of fixed points in the network is  $e^{NF(\alpha, \beta)}$ . All quantities plotted are dimensionless. Contours lines of the surface (a) are shown in (b) for every 0.01 (dashed curves) and every 0.05 (solid curves).

rough landscape—the results of the two approaches differ quantitatively. Furthermore, deterministic analog networks can be built in very-large-scale-integrated electronic hardware<sup>17</sup> much more easily than stochastic networks requiring local noise generators.

The calculation of  $F(\alpha, \beta)$  generalizes calculations by Gardner<sup>4</sup> on the number of metastable states in an Ising-model neural network, and by Bray and Moore<sup>2</sup> on the number of solutions of the mean-field equations of an Ising spin glass given by Thouless, Anderson, and Palmer (TAP).<sup>18</sup> The connection to these results is especially apparent if one chooses the neuron transfer function to be  $f(x) = \tanh(x)$ . For this choice, the large- $\beta$  result,  $F(\alpha, \infty)$ , numerically approaches the result of Gardner.<sup>14</sup> The large- $\alpha$  result,  $F(\infty, \beta)$ , is formally equivalent<sup>19</sup> to the result obtained using an interconnection matrix with random Gaussian-distributed off-diagonal elements, but is *not equivalent* to the result of Bray and Moore<sup>2</sup> for the mean-field spin glass. The difference is due to the absence of a reaction field in the neural-network dynamics.<sup>15</sup> Finally, in the limit in which both  $\alpha$  and  $\beta$  are large, we obtain the familiar Ising spin-glass result  $F(\infty, \infty) = 0.1992$ .<sup>1,2</sup>

We consider networks of analog neurons which evolve according to either the system of coupled differential equations

$$dm_i(t)/dt = -m_i(t) + f\left[\beta \sum_{j=1}^N J_{ij}m_j(t)\right], \quad (1)$$

or the iterated map<sup>15,20</sup>

$$m_i(t+1) = f\left[\beta \sum_{j=1}^N J_{ij}m_j(t)\right], \quad (2)$$

with either parallel or sequential updating. The quantities  $m_i(t)$  describe the state of neuron  $i$  ( $i=1, \dots, N$ ) at time  $t$ ,  $\beta$  is the gain parameter, and  $f$  is a nonlinear neuron transfer function which we assume to be single valued and continuous. The  $J_{ij}$  are given by the Hebb

rule

$$J_{ij} = (N\sqrt{\alpha})^{-1} \sum_{\mu=1}^{\alpha N} \xi_i^\mu \xi_j^\mu, \quad i \neq j, \quad J_{ii} = 0, \quad (3)$$

where the memory patterns  $\xi_i^\mu$  take the values  $\pm 1$  randomly and without bias,  $\alpha$  is the ratio of the number of stored patterns to number of neurons, and the normalization  $(N\sqrt{\alpha})^{-1}$  is chosen to make the magnitude of  $J_{ij}$  independent of  $\alpha$ . In the following analysis, sums and products over italic indices are assumed to range from 1 to  $N$  and those over greek indices are assumed to range from 1 to  $\alpha N$  unless otherwise noted.

The expectation of the number of fixed points of (2) or (3) is given by

$$\langle N_{\text{fp}}(N, \alpha, \beta) \rangle = \left\langle \int_{\Gamma} \prod_i (dm_i) \prod_i [\delta(G_i)] |\det A_{ij}| \right\rangle. \quad (4)$$

In Eq. (4), the brackets  $\langle \dots \rangle$  denote an average over the variables  $\xi_i^\mu$ ,  $\Gamma$  is the  $N$ -dimensional state space of the system as defined by the range of  $f$ , and the  $\delta$  functions enforce the fixed-point conditions  $G_i = 0$  for all  $i$ , where

$$G_i \equiv f^{-1}(m_i) - \beta \sum_j J_{ij}m_j. \quad (5)$$

The determinant of the matrix  $A_{ij} \equiv \partial G_i / \partial m_j$  normalizes the  $\delta$  functions for integration over  $m_i$ . The matrix  $A_{ij}$  is also the Hessian of a Lyapunov or energy function<sup>20,21</sup> for Eqs. (1) and (2); thus  $A_{ij}$  characterizes the local curvature of the energy landscape. We are interested in counting stable fixed points, where  $A_{ij}$  is positive definite, and will restrict the integration in Eq. (4) to an appropriate region of state space in an approximate way described below.

We note that, because the extensive quantity in this problem is  $\ln(N_{\text{fp}})$  rather than  $N_{\text{fp}}$  itself, it is actually  $\ln(N_{\text{fp}})$  which should be averaged to give the expected number of fixed points for a typical realization. Instead, we calculate  $\langle N_{\text{fp}}(N, \alpha, \beta) \rangle$ , which gives an upper bound for the expected number of fixed points.<sup>4</sup> At large  $N$ , nearly all fixed points are uncorrelated and the two averages should be the same.

To simplify the calculation of (4), we average  $\prod_i [\delta(G_i)]$  and  $\det A_{ij}$  separately, which assumes that as  $N \rightarrow \infty$  the distribution of  $\det A_{ij}$  at local minima becomes narrowly peaked about a single value. This assumption is reasonable, as such behavior is observed in other properties of the fixed points, such as their energies and overlaps with memory patterns.<sup>4</sup> The average of  $\prod_i [\delta(G_i)]$  is computed using an integral representation of the  $\delta$  function:

$$\delta(G_i) = \int_{-\infty}^{\infty} (dx_i/2\pi) \exp(ix_i G_i).$$

The average of  $\det A_{ij}$  is computed<sup>2</sup> by introducing replicas, indexed by  $\gamma$ ,

$$\det A_{ij} = \int_{-\infty}^{\infty} \prod_{i\gamma} \frac{d\lambda_{i\gamma}}{(2\pi)^{1/2}} \exp\left[-\frac{1}{2} \sum_{i\gamma} \lambda_{i\gamma} A_{ij} \lambda_{j\gamma}\right] \quad (6)$$

and setting the number of replicas equal to  $-2$  after averaging. The replica symmetric solution is

$$\langle \det A_{ij} \rangle = \exp \left\{ N \left[ -\frac{B\sqrt{\alpha}}{\beta} - \alpha \ln \left( 1 - \frac{B}{\beta\sqrt{\alpha}} \right) \right] \right\} \times \prod_i [a(m_i) + B], \quad (7)$$

where  $a(m_i) \equiv (d/dm)[f^{-1}(m)]_{m_i}$  and the value of  $B$  is determined variationally.<sup>22</sup> We point out that Eq. (6) is valid only if  $A_{ij}$  is positive definite, which implies  $\det A_{ij} > 0$ . We may therefore drop the absolute value brackets around  $\det A_{ij}$  but must take care that the solution of Eq. (7) makes sense, as  $A_{ij}$  is not positive definite throughout

the state space. We find that for some configurations  $\{m_i\}$  the solution of Eq. (7) yields a negative result.<sup>23</sup> We interpret a negative result in Eq. (7) as indicating that  $A_{ij}$  is not positive definite in that region of state space. As we are interested in counting stable fixed points, we limit integration over the range of  $f$  to a subregion  $\Gamma'$  for which Eq. (7) yields a positive result. This provides an approximate way of counting stable fixed points and in fact is necessary to obtain meaningful results from the saddle-point equations below.

The quantity  $\langle N_{fp}(N, \alpha, \beta) \rangle$  is found from saddle-point integration to be

$$\langle N_{fp}(N, \alpha, \beta) \rangle = \max_{q, \lambda, \Delta} \min_B [e^{NF(\alpha, \beta)}], \quad (8)$$

where

$$F(\alpha, \beta) = \ln \int_{\Gamma'} dm I(m) - (\sqrt{\alpha}/\beta)(\Delta + B) + (\alpha/2) \ln \{ [(\Delta + \beta\sqrt{\alpha})^2 - 2\beta^2\lambda q] / (B - \beta\sqrt{\alpha})^2 \} \quad (9)$$

and

$$I(m) = [\beta(2\pi q)^{1/2}]^{-1} [a(m) + B] \exp \{ -[f^{-1}(m) - \Delta m]^2 / 2\beta^2 q + \lambda m^2 \}. \quad (10)$$

In Eq. (8), the extremum over the variables  $q, \lambda, \Delta,$  and  $B$  is evaluated by setting partial derivatives of  $F(\alpha, \beta)$  with respect to these variables equal to zero. We emphasize that the integral in (9) is over a region  $\Gamma'$  which covers the range of  $f$  where  $a(m) + B \geq 0$ . The saddle-point equations for  $q, \lambda, \Delta,$  and  $B$  are given by

$$q = \{ [(\Delta + \beta\sqrt{\alpha})^2 - 2\beta^2\lambda q] / \beta^2 \alpha \} \langle m^2 \rangle, \quad (11)$$

$$\lambda = - \{ [(\Delta + \beta\sqrt{\alpha})^2 - 2\beta^2\lambda q] / 2\beta^2 \alpha q \} [1 - \langle [f^{-1}(m) - \Delta m]^2 \rangle / \beta^2 q], \quad (12)$$

$$0 = [(\Delta + \beta\sqrt{\alpha})^2 - 2\beta^2\lambda q] [(\beta\sqrt{\alpha})^{-1} - \langle m f^{-1}(m) \rangle / \beta^2 \alpha q] - \beta\sqrt{\alpha}, \quad (13)$$

$$B = (\beta B / \sqrt{\alpha} - \beta^2) \langle [a(m) + B]^{-1} \rangle, \quad (14)$$

where the double brackets  $\langle \langle \dots \rangle \rangle$  are defined by

$$\langle \langle g(m) \rangle \rangle \equiv \left[ \int_{\Gamma'} dm I(m) \right]^{-1} \int_{\Gamma'} dm g(m) I(m) \quad (15)$$

and, again,  $\Gamma'$  denotes the range of  $f$  where  $a(m) + B \geq 0$ .<sup>24</sup> The resulting function  $F(\alpha, \beta)$  is plotted in Fig. 2 as a function of  $1/\alpha$  and  $1/\beta$ . The data were obtained by solving Eqs. (11)–(14) for  $q, \lambda, \Delta,$  and  $B$  and computing  $F(\alpha, \beta)$  from Eqs. (9) and (10) for each of 1600 points in the  $(\alpha, \beta)$  plane.

The values obtained for  $F(\alpha, \beta)$  for the case  $f(x) = \tanh(x)$  were tested numerically at six points in the  $(\alpha, \beta)$  plane by the following procedure. At each of the points, twenty networks with random-pattern Hebb matrices were generated for six different values of  $N$ . The values of  $N$  were chosen so that the number of fixed points was roughly in the range of 20 and 400. The number of fixed points in each network was counted by choosing random initial conditions  $m_i(0) \in \{-1, +1\}$  and iterating the map (2) with sequential updating until convergence to a fixed point was reached. The search was terminated after  $10^5$  initial conditions or when no new fixed points had been found for  $10^4$  consecutive initial conditions and for every fixed point found, the inverse point ( $m_i \rightarrow -m_i$  for all  $i$ ) had also been found. Figure 3 shows the resulting values for  $\langle N_{fp} \rangle$  vs  $N$  at the

six values of  $\alpha$  and  $\beta$ , along with weighted least-squares fits to the data. The slopes generated by these least-squares fits are compared with the theoretical values of  $F(\alpha, \beta)$  computed from Eqs. (9)–(14) in Table I. Agreement between numerical and theoretical results is good, especially at large  $\alpha$  and  $\beta$ . We note that each of the

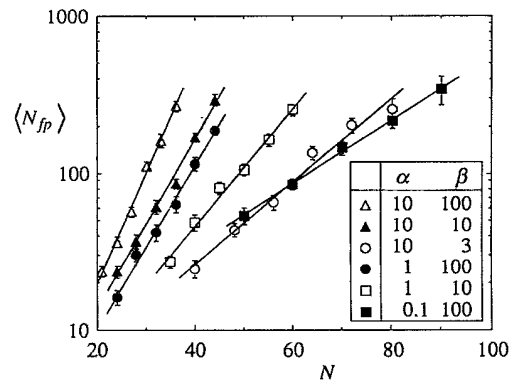


FIG. 3. Results of numerical counts of stable fixed points for six values in the  $(\alpha, \beta)$  plane. Data are averages over twenty randomly generated interconnection matrices. Error bars indicate standard error. Lines are weighted least-squares fits to the data.

TABLE I. Comparison of theoretical and numerical values of  $F(\alpha, \beta)$ . Standard errors for the numerical values are given in parentheses.

$\alpha$	$\beta$	$F(\alpha, \beta)$	
		Theoretical	Numerical
10	100	0.166	0.165(0.006)
10	10	0.121	0.126(0.006)
10	3	0.054	0.064(0.004)
1	100	0.124	0.121(0.006)
1	10	0.088	0.086(0.005)
0.1	100	0.059	0.047(0.004)

data points in Fig. 3 required roughly ten CPU hours on a Sun 4 workstation.

To illustrate the dramatic effect that lowering the neuron gain  $\beta$  has on the number of spurious states, consider a change from  $\beta=100$  to  $\beta=10$  in a network with a storage ratio of  $\alpha=0.1$ . Using the theoretical values  $F(0.1, 100)=0.059$  and  $F(0.1, 10)=0.040$ , we find that the expected number of spurious attractors is reduced by over 97% for  $N=200$  and by 8 orders of magnitude for  $N=1000$ . Finding an *optimal* value of gain for a particular neural-network application depends on the learning rule used, and lower gain does not always mean improved recall, as might be inferred from the fact that  $F(\alpha, \beta)$  decreases monotonically as the gain is lowered. The recall states, which are the desired attractors, are also affected by the gain: For the Hebb rule, for example, too low a gain destabilizes all recall states, resulting in a spin-glass phase.<sup>15</sup> Learning-rule-dependent constraints on the gain are discussed in Ref. 15.

In conclusion, we have shown that using analog neurons to smooth the energy landscape in an associative memory neural network greatly reduces the number of spurious fixed points. The analytical results are found to be in good agreement with computer experiments on relatively small networks. The analysis presented may also be applied to other optimization problems in which smoothing the energy landscape with analog dynamics has been found numerically to be effective, including the traveling salesman problem and problems in artificial vision.

We thank L. Abbott and T. Kepler for valuable discussions. We also thank the Robotics Laboratory, the Tinkham Group, and the Condensed Matter Theory Group at Harvard for use of their computers. One of us (F.R.W.) acknowledges support from the Army Research Office as a Joint Services Electronics Program (JSEP) Graduate Fellow, and one of us (C.M.M.) acknowledges support as an AT&T Bell Laboratories Scholar. This work was supported in part by ONR Contract No. N00014-89-J-1592 and by JSEP Contract No. N00014-89-J-1023.

<sup>1</sup>F. Tanaka and S. F. Edwards, *J. Phys. F* **10**, 2769 (1980); C. De Dominicis, M. Gabay, T. Garel, and H. Orland, *J. Phys. (Paris)* **41**, 923 (1980).

<sup>2</sup>A. J. Bray and M. A. Moore, *J. Phys. C* **13**, L469 (1980).

<sup>3</sup>H. Gutfreund, J. D. Reger, and A. P. Young, *J. Phys. A* **21**, 2775 (1988).

<sup>4</sup>E. J. Gardner, *J. Phys. A* **19**, L1047 (1986); A. D. Bruce, E. J. Gardner, and D. J. Wallace, *J. Phys. A* **20**, 2909 (1987).

<sup>5</sup>C. Koch, J. Marroquin, and A. Yuille, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4263 (1986); A. Blake and A. Zisserman, *Visual Reconstruction* (MIT Press, Cambridge, 1987).

<sup>6</sup>A. Yuille, *Biol. Cybern.* **61**, 115 (1989); H. G. E. Hentschel and A. Fine, *Phys. Rev. A* **40**, 3983 (1989).

<sup>7</sup>J. J. Hopfield and D. W. Tank, *Biol. Cybern.* **52**, 141 (1985); *Science* **233**, 625 (1986).

<sup>8</sup>G. Basharan, Y. Fu, and P. W. Anderson, *J. Stat. Phys.* **45**, 1 (1986); N. Burgess and M. A. Moore (to be published).

<sup>9</sup>Y. Fu and P. W. Anderson, *J. Phys. A* **19**, 1605 (1986).

<sup>10</sup>S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).

<sup>11</sup>For a recent collection of papers, see *J. Phys. A* **22**, 1953-2273 (1989), special issue in memory of Elizabeth Gardner.

<sup>12</sup>G. Toulouse, *J. Phys. A* **22**, 1959 (1989).

<sup>13</sup>C. M. Soukoulis, K. Levin, and G. S. Grest, *Phys. Rev. Lett.* **48**, 1756 (1982); *Phys. Rev. B* **28**, 1495 (1983).

<sup>14</sup>G. Bilbro, R. Mann, T. Miller, W. Snyder, D. Van den Bout, and M. White, in *Advances in Neural Information Processing Systems*, edited by D. S. Touretzky (Kaufmann, San Mateo, 1989), p. 94; G. Bilbro and W. Snyder, *ibid.*, p. 594.

<sup>15</sup>C. M. Marcus, F. R. Waugh, and R. M. Westervelt, *Phys. Rev. A* **41**, 3355 (1990).

<sup>16</sup>R. Durbin and D. Willshaw, *Nature (London)* **326**, 689 (1987).

<sup>17</sup>C. A. Mead, *Analog VLSI and Neural Systems* (Addison-Wesley, Reading, 1989).

<sup>18</sup>D. J. Thouless, P. W. Anderson, and R. G. Palmer, *Philos. Mag.* **35**, 593 (1977).

<sup>19</sup>F. R. Waugh (unpublished).

<sup>20</sup>C. M. Marcus and R. M. Westervelt, *Phys. Rev. A* **40**, 501 (1989).

<sup>21</sup>J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3008 (1984).

<sup>22</sup>The value of  $B$  used gives a local minimum of  $\det A_{ij}$ , and corresponds to the solution  $B=0$  of Ref. 2, which is also at a local minimum. Having to use a local minimum results from continuing the number of replicas to  $-2$  before solving the saddle-point equations.

<sup>23</sup>This situation does not occur for the TAP spin glass, where the corresponding solution (Ref. 2) has  $B=0$  and therefore is always positive.

<sup>24</sup>In the limit  $\alpha \rightarrow \infty$ , Eqs. (11)-(15) with  $f(x) = \tanh(x)$  reduce to the saddle-point equations of Ref. 2 for the TAP equations excluding the reaction field term. We have chosen our notation for the variables  $q$ ,  $\lambda$ ,  $\Delta$ , and  $B$  to make the correspondence. The quantities  $\beta^2 q/2$ ,  $-\lambda$ ,  $\Delta + \beta\sqrt{\alpha}$ , and  $(\beta\sqrt{\alpha} - B)/2$  are the fields conjugate to the order parameters  $N^{-1} \sum_i x_i^2$ ,  $N^{-1} \sum_i m_i^2$ ,  $iN^{-1} \sum_i x_i m_i$ , and  $N^{-1} \sum_i \lambda_i^2$ , respectively.