RESEARCH ARTICLE

Are binary synapses superior to graded weight representations in stochastic attractor networks?

Jason Satel · Thomas Trappenberg · Alan Fine

Abstract Synaptic plasticity is an underlying mechanism of learning and memory in neural systems, but it is controversial whether synaptic efficacy is modulated in a graded or binary manner. It has been argued that binary synaptic weights would be less susceptible to noise than graded weights, which has impelled some theoretical neuroscientists to shift from the use of graded to binary weights in their models. We compare retrieval performance of models using both binary and graded weight representations through numerical simulations of stochastic attractor networks. We also investigate stochastic attractor models using multiple discrete levels of weight states, and then investigate the optimal threshold for dilution of binary weight representations. Our results show that a binary weight representation is not less susceptible to noise than a graded weight representation in stochastic attractor models, and we find that the load capacities with an increasing number of weight states rapidly reach the load capacity with graded weights. The optimal threshold for dilution of binary weight representations under stochastic conditions occurs when approximately 50% of the smallest weights are set to zero.

Keywords Synaptic plasticity · Binary versus graded · Associative memory · Point attractor networks

J. Satel · T. Trappenberg (✉)
Faculty of Computer Science, Dalhousie University, 6050
University Avenue, Halifax, NS B3H 1W5, Canada
e-mail: tt@cs.dal.ca

J. Satel · T. Trappenberg · A. Fine
Neuroscience Institute, Dalhousie University, 5850 College
Street, Halifax, NS B3H 1X5, Canada

Introduction

Synaptic plasticity is widely believed to be the basis of learning and memory in neural systems (Martin et al. 2000). However, the exact nature of synaptic weight representation is still unresolved (Dobrunz 1998; Chklovskii et al. 2004; Peterson et al. 1998; O'Connor et al. 2005a, b; Liao et al. 1995; Isaac et al. 1995; Brunel et al. 2004; Poirazi and Mel 2001). While some early modeling work considered binary weight representations (Willshaw et al. 1969), more emphasis has been given to models with continuous-valued weights to represent synaptic efficacies (Amari 1972; Grossberg 1969). This approach has been called into question by neurophysiological studies that have been unable to show graded learning in vitro (Peterson et al. 1998; O'Connor et al. 2005a, b). These experiments with rat hippocampal slices indicated that synapses that have already been potentiated are unable to exhibit further potentiation. The authors of these studies have supported the case for binary synapses with theoretical arguments, proposing that binary synapses would be advantageous for biological systems in that they would be less susceptible to noise than graded synapses. Driven in part by these results, many modelers have begun to switch from graded to binary weight representations in their models of learning and memory (Amit and Mongillo 2003; Brody et al. 2003; Koulakov et al. 2002; Abarbanel et al. 2005; Braunstein and Zecchina 2006; Vladimirski et al. 2006; Senn and Fusi 2004; Giudice et al. 2003; Baldassi et al. 2007; Graupner and Brunel 2007; Fusi and Abbott 2007).

The aim of our study was to investigate this issue to see whether or not graded weights are more susceptible to noise than binary weights when using the weights as the basis of an associative memory system, specifically in the form of attractor networks. It is known that attractor
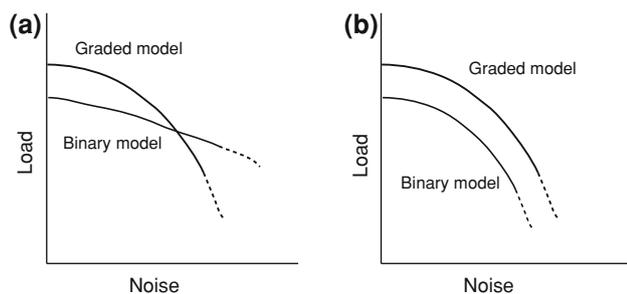
Fig. 1 Load capacities in attractor networks with graded and binary weights. **a** Scenario in which the binary weights are less susceptible to noise. **b** Scenario in which the effect of noise is similar for models with binary and graded weights

networks with clipped Hebbian (Sompolinsky 1987; Gutfreund and Stein 1990) and binary (Krauth and Mezard 1989) weights have a lower storage capacity than models with graded weights. Although many of the classical papers (Amit et al. 1985, 1987; Sompolinsky 1987, 1986) include some discussion of noise in the systems, comparison of the models with graded and binary weights in the above papers has only been made explicit in the case of deterministic updates. Sompolinsky also analyzed the effect of noise and non-linearities in the learning process on the storage capacity of attractor networks in a separate paper (Sompolinsky 1986) and found that static noise and models with nonlinearities are related, and both reduce the storage capacity of such attractor networks. However, a direct comparison of models with graded and binary weights under increasing levels of noise has, to our knowledge, never been done explicitly. If binary weights are less susceptible to noise, then there would be a point at which the load capacity of a model with binary weights would surpass that of a graded weight model as noise in the networks is increased (see Fig. 1a). Otherwise, the load capacity of a binary weight model would never surpass that of a graded weight model (Fig. 1b). The results reported in this paper support the scenario of Fig. 1b. Binary weights are not less susceptible to noise than graded weights in stochastic attractor network models in terms of storage capacity of these networks. The results show that the theoretical argument by Peterson et al. (1998), which is widely accepted in the neuroscience community, does not hold for associative attractor memory networks.

Model and assumptions

A standard attractor network (Amit et al. 1985, 1987) was used in order to compare the noise susceptibility of binary versus graded weight representations. A network of

$N = 1000$ nodes was used where each node i is connected to all other nodes j via a connection w_{ij} without self-couplings, so $w_{ii} = 0$. The simulations were split into a training and a testing phase.

During the training phase, the strengths of the connections are learned according to a Hebbian covariance learning rule (Sompolinsky 1987):

$$w_{ij}^{\text{GRD}} = 1/\sqrt{\rho} \sum_{\mu=1}^{\rho} \xi_i^\mu \xi_j^\mu + \delta_{ij}, \tag{1}$$

where μ numbers all ρ patterns that are to be stored in the network, and $\xi = \pm 1$ represents the state of each element in the random, uncorrelated patterns that are trained. The scaling term in this learning algorithm, $1/\sqrt{\rho}$, is necessary to compare performance between models with different weight representation by ensuring that the spread of the weight values remain of order unity with ρ patterns (Sompolinsky 1987). Without this normalization procedure, different weight models being compared would essentially be comparing different levels of noise, since a rescaling of the weights can be absorbed by a rescaling of the noise scale such as the temperature in the dynamic equations below. The term δ_{ij} is a random variable to include static noise in the weight matrix.

The function, sign(), is then used to obtain a binary weight representation where all positive weights are converted to $+1$, negative weights to -1:

$$w_{ij}^{\text{BIN}} = \text{sign}(w_{ij}^{\text{GRD}}). \tag{2}$$

In addition, we study a model with varying numbers of discrete weight states:

$$w_{ij}^k = F(w_{ij}^{\text{GRD}}), \tag{3}$$

where an approximately equal number of weights are assigned to each of the k discrete states using the function $F()$. This function sorts the graded weights, finds the appropriate values to obtain as close to an equal division of weights as possible, assigns each of the graded weights to one of the discrete states, and then normalizes the discrete weight values to ensure the weights are of order unity so that we can compare performance of the different models with binary and graded weight representations. Different choices of the function $F()$ can be considered, as discussed below, but no significant difference has been found in the results we report.

Once the networks have been trained, all patterns that have been stored are tested by presenting each of them to the network, one at a time, and updating the states of all nodes, s_i, in the network. The state of a node at any timestep is a function of the sum of the activities of all other connected nodes, s_j, multiplied by the weight of the connection between them, w_{ij}. The local fields:

$$h_i = \frac{\sqrt{\rho}}{N} \sum_{j=1}^{N} w_{ij} s_j, \tag{4}$$

represent the sum of the activity of all other connected nodes in the network multiplied by the synaptic weight that connects them to node i. The scaling term $\sqrt{\rho}/N$ is introduced in order to maintain the standard formulation of attractor networks by ensuring that the local fields are of order unity (Sompolinsky 1987). For deterministic networks (no noise), the node activities are then updated according to the sign activation function:

$$s_i = \text{sign}(h_i). \tag{5}$$

For finite temperature models, which include varying degrees of stochasticity, the activation function used is a logistic function, so that nodes are updated according to:

$$\text{Prob}(s_i = +1) = (1 + e^{-\beta h_i})^{-1}, \tag{6}$$

where the parameter $\beta = \frac{1}{T}$ reflects the amount of noise in the update. The limit of $\beta \to \infty$ corresponds to the deterministic attractor model using the sign() activation function.

The general model as introduced so far has two different sources of noise, one from the probabilistic update and one from the noise in the weight matrix. The method of introducing noise into the system through stochastic updating without noise in the weight matrix ($\delta_{ij} = 0$) is the method most often used in attractor models (Amit et al. 1985, 1987). The noise in this model can be interpreted in various ways. While a common interpretation is that of probabilistic synapses, the statistical mechanics formulation of the system compares a signal in form of h_i to a noise term following a Boltzmann distribution with parameter β. Thus, this formulation can also be interpreted as fluctuating weight values with a deterministic update rule.

An alternative method of modeling noise is to introduce static synaptic noise directly into the weights by adding a random variable, δ_{ij}, to the weights (Sompolinsky 1987). This noise term is taken to be Gaussian distributed and is uncorrelated with the patterns. Although this model of stochasticity is not exactly equivalent to the former, it has been shown analytically that they result in similar network behavior (Amit et al. 1985, 1987; Sompolinsky 1987). Simulations were performed using both models of adding noise to the networks, with results showing similar behaviour, as described below.

In the simulations below, each node is updated based on the states of all other nodes at the previous timestep, for ten timesteps. Once the systems have been updated for ten timesteps, the final state of the network is compared to the original trained pattern to obtain a measure of retrieval error. Since a binary representation of patterns is used with quiescent elements represented by -1 and active elements represented by $+1$, an appropriate measure of similarity is to take the average number of node states that differ between the original and retrieved patterns:

$$\text{err}_{\text{avg}} = \langle \langle 1/N \sum_{i=1}^{N} |(\xi_i^{\mu} - s_i^{\text{final}})/2| \rangle_{\mu} \rangle_{\text{trial}}. \tag{7}$$

For any given load, this procedure is performed for every trained pattern and the average percentage of incorrectly retrieved node states over all patterns is used as the measure of error for that model. When multiple trials are performed, these average error rates are then averaged over the trials.

Using random patterns with binary, equally distributed and uncorrelated components is a convenient choice commonly made in studies of attractor memory that allows for comparison with detailed analytic results (Amit et al. 1985, 1987; Hopfield 1982, 1984; Sompolinsky 1987; Morgenstern 1986). In contrast, patterns in the brain are likely more correlated, yet sparse. Correlations between patterns influence the storage capacity of the networks, although sparse representations, typically associated with hippocampal processing, can reduce correlations. Experimental results in the temporal cortex of monkeys have shown that less than 10% of nodes in a given pattern are active (Rolls and Tovee 1995), although recent imaging studies indicate that around 25% of CA1 neurons are active after fear conditioning (Matsuo et al. 2008). The standard formulation is satisfactory for the arguments derived in this study since we are investigating the influence of noise with different weight representations in attractor networks and are less concerned with absolute values for storage capacities.

Finally, discretizing the weights *after* correlation learning is not a truly binary learning rule, as the graded weights must be still be computed before converting them to a simpler, binary representation. Post-hoc discretization is an artificial and unrealistic necessity imposed by current uncertainty about precisely which (if any) discretization mechanisms are used by the synapse, and this method has been the basis for many discussions in the literature (Morgenstern 1986; Sompolinsky 1987). Even the learning algorithm of Baldassi et al. (2007) for perceptrons with binary weights uses a hidden graded representation of the pattern history. Also, it is well known that online learning rules result in different storage capacities (Fusi and Abbott 2007), and that resulting weight distributions depend on the details of the learning rule, such as the nature of boundary conditions (Gutig et al. 2003; Kepecs et al. 2002) and the number of states. Again, however, we are mainly concerned with investigating the influence of noise with different representations of final weight distributions in attractor networks rather than absolute storage capacities or equilibrium properties of different learning rules.

Results

Binary versus continuous weight representations

The following simulations investigate differences in the retrieval performance of stochastic attractor networks when using models of binary and graded weight representation. Two different noise models were investigated, with noise added either to the weights themselves during training in Eq. 1, or incorporated into updates using probabilistic updating in Eq. 6. For a given noise level, each model was trained on the same patterns using Eqs. 1 and 2. Network performance was tested by applying each original stored pattern to the networks and allowing the models to update themselves according to Eqs. 5 or 6 for ten timesteps, finally comparing the retrieved patterns to the original stored patterns using Eq. 7. This procedure was repeated for 20 trials and the average retrieval error over all trials was recorded for each load and each noise level. Figure 2a–d illustrate how the retrieval error increases with increasing load for both the graded and binary weight representations as the level of noise is increased. As more patterns are loaded into the models, the retrieval error continues to increase. The figures show that retrieval error increases more quickly with a binary weight representation in both noise models, especially under higher levels of noise.
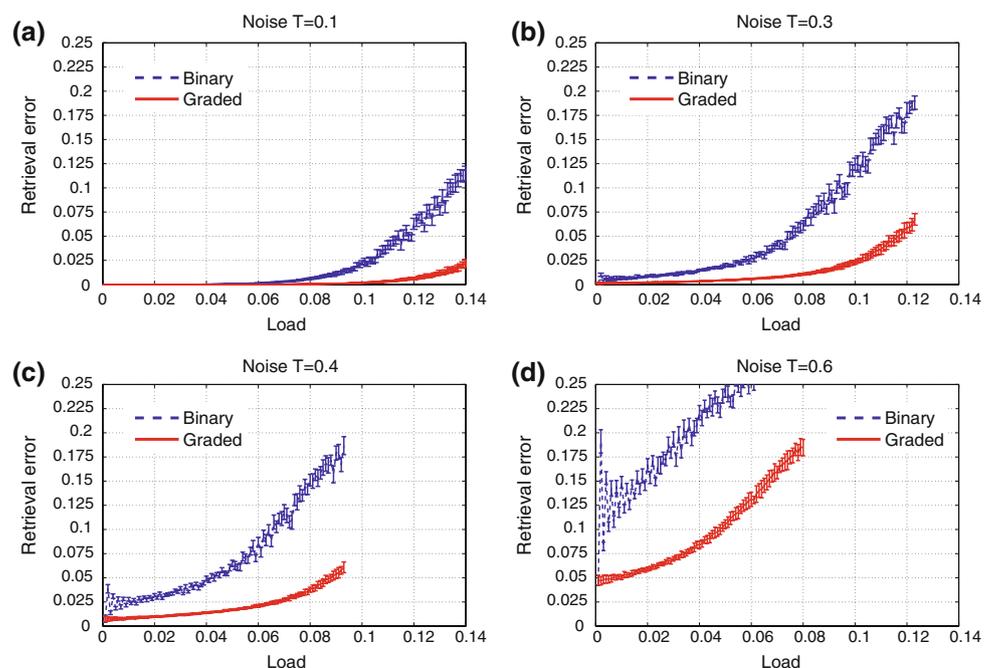
Load capacity

Analytic calculations have shown that there is a critical load, beyond which attractor networks will no longer function as memory systems (Amit et al. 1985, 1987). That is, as more and more patterns are loaded into a network, eventually the system fails to retrieve the original patterns with any degree of accuracy. This critical load, α_c, corresponds to the point at which a first-order phase transition occurs.

The phase transition in attractor networks has been shown to occur when an error threshold of approximately 0.0165 is reached in the deterministic case (no noise at all), in the limit of an infinite number of nodes, with continuous-valued weights (Amit et al. 1985, 1987; Sompolinsky 1987). That is, when the load capacity is reached, the network is retrieving patterns with an average error of 1.65%. As stochasticity is increased, this critical error threshold increases, as shown in Table 1 (Amit et al. 1985, 1987) and Table 2 (Sompolinsky 1987), which list the error thresholds for varying amounts of noise in each noise model. This shows that the amount of retrieval error that is present when the networks reach load capacity will increase with the amount of noise in the systems. These analytically determined thresholds correspond to models using continuous weight values, in the limit of an infinite number of nodes, and so may not be exact when applied to a finite number of nodes or when using binary weights, as in these numerical simulations. To address these issues, analogous results were computed using ±10% of the analytic error threshold results, as denoted by the error bars in the phase diagrams of Fig. 3 which graphically illustrate the numerically determined load capacities of both models under varying noise levels as well as the original analytic results for graded weights.

Fig. 2 Retrieval error increases with increasing load more rapidly for binary than graded synaptic weights. Load refers to the number of patterns stored in the network, relative to the number of nodes in the system. Each *inset* shows the retrieval error versus load for a set amount of noise in the system using probabilistic updating. The *top curves* in each graph were produced by a model using binary weights, while the *bottom curves* were produced by a model using graded weights

In these simulations, the load capacity is defined as the load just before the retrieval error surpasses the error thresholds specified for the given level of stochasticity. The figures illustrate that a binary weight representation is not less susceptible to degradation than a graded weight representation.

Multiple discrete weight states

Simulations were also performed with differing numbers of discrete weight states to investigate how performance changes when adding additional possible weight states. Discretization of the graded weights was performed after training according to Eq. 3.

One issue with the weight discretization process is whether to use an even or odd number of states, since this attractor network model uses both positive and negative weight values. The question here is whether or not a weight value of 0 should be considered as a separate state. After a number of simulations, it was determined that performance is enhanced by using an odd number of states, which results in diluted models (i.e., models with more zero-valued weights). Figure 4a and b illustrate the differences in retrieval performance of models which do and do not have a diluted representation, under both deterministic and stochastic conditions. Models that include zeros as a state, and therefore have an odd number of discrete states, have sometimes been referred to as ternary models (Sompolinsky 1987; Morgenstern 1986); we refer to them as diluted models in this work. Although the zero state is considered as a separate state here, there is no active synaptic connection when the weights are set to zero.

Each of the different weight representation models was trained on the exact same patterns, increasing the number of patterns loaded into the systems one by one. For each of these loads and each noise level, the models were tested on all the trained patterns, for 20 trials. The error was recorded for each load and noise level, averaged over all trained patterns and all trials. As before, the load capacities of the different models were determined by finding the load at which the error crosses a particular threshold, as described in Sect. 2 and seen in Tables 1 and 2.

Figure 4c illustrates the increase in load capacity under different levels of stochasticity as more and more possible

Table 2 Table of error thresholds used to define the point at which network load reaches load capacity, α_c, for different levels of stochasticity using a noise model with static noise added directly to the weights

Δ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
err_c	0.0165	0.0170	0.0225	0.0355	0.0555	0.0865	0.1380	0.2395

The critical error thresholds, err_c, are calculated through a simple scaling procedure from the overlap values, m_c, that are taken from analytic work (Sompolinsky 1987)

discrete weight states are used. When the number of weights states is increased from two, as in the binary weight model, to an infinite number of possible states, as in the graded weight model, the load capacity continues to increase, regardless of the noise level. Interestingly, the increase in load capacity is relatively slight once the number of weight states has been increased beyond a small number.

Diluted weight representations

In the previous section, diluted binary networks were created by placing an approximately equal number of weights into each of the high, low, and zero states. Alternatively, we explored the use of a threshold, $|z|$, where all weights from the graded weight model in between $-z$ and $+z$ are changed to zero, while the rest of the weights are set to a high or low value based on their sign:

$$w_{ij}^{\text{diluted}} = \begin{cases} -1 & \text{if } w_{ij} < -z \\ +1 & \text{if } w_{ij} > +z \\ 0 & \text{otherwise} \end{cases}$$

The original, undiluted, binary networks are thus equivalent to diluted binary networks with a threshold of $z = 0$. Simulations were run as before, except that performance was also measured for different values of $|z|$ between 0 and 1.2.

Figure 4d illustrates the retrieval error at load capacity for different values of $|z|$ in networks with varying levels of stochasticity. Performance is maximal, and relatively stable, when using threshold values between $z = |0.4|$ and $z = |0.8|$. This reflects the number of zero valued weights in the model. Since the update equation reflects an average of all connected node activities multiplied by their weight values, it is only the largest weights that really count. In

Table 1 Table of error thresholds used to define the point at which network load reaches load capacity, α_c, for different levels of stochasticity using a probabilistic updating model for noise

T	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
err_c	0.0165	0.0175	0.0220	0.0295	0.0440	0.0645	0.0965	0.1405	0.2025	0.3000

The critical error thresholds, err_c, are calculated through a simple scaling procedure from the overlap values, m_c, that are taken from analytic work (Amit et al. 1985, 1987)

Fig. 3 Phase diagrams in the model with probabilistic updates (**a**) and static noise in the weight matrix (**b**) for graded and binary weight models illustrating load capacities. The *vertical axis* represents load and the *horizontal axis* represents degree of stochastisity in the systems. *Error bars* represent load capacities when using ±10% variations in the values of the critical overlap used to determine the transition points
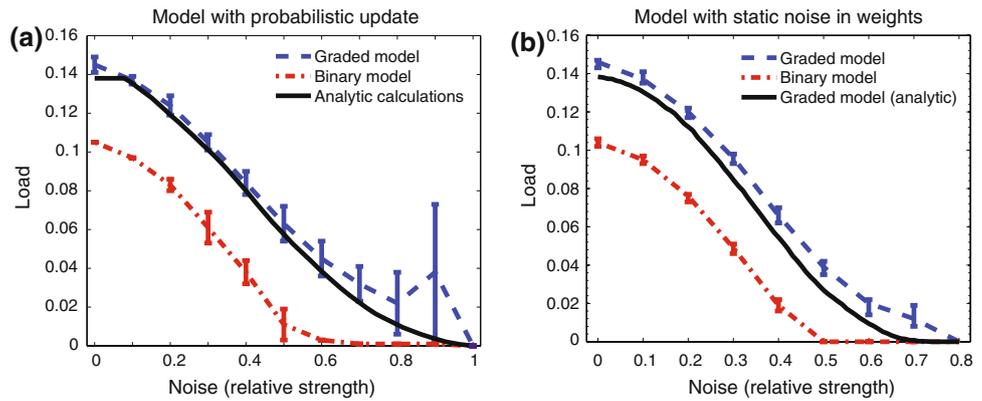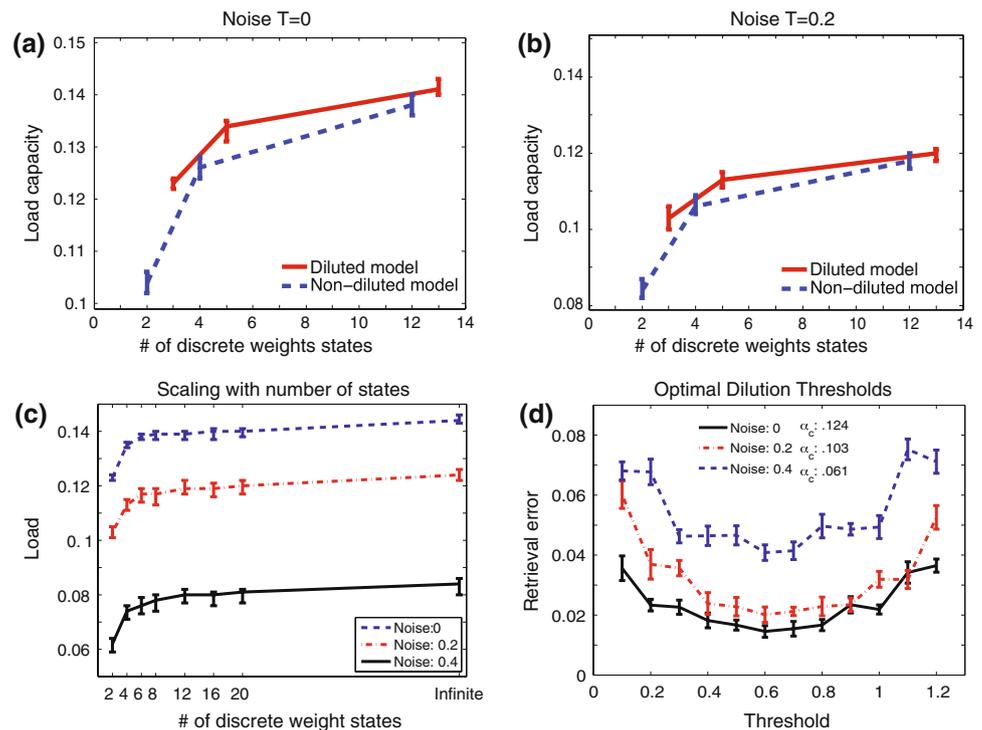

Fig. 4 a, **b** Diluted versus non-diluted models of binary weight representation. **c** Load capacity as a function of the number of discrete weight states for different levels of stochasticity. **d** Optimal dilution thresholds at load capacity for varying degrees of stochasticity. The load capacities, α_c, when using the specified thresholds are shown

fact, the increased performance of diluted binary networks over binary networks illustrates the fact that many small weights can actually have a negative effect on performance when they are discretized to values much larger than they originally were. Previous work has shown that deterministic attractor networks will continue to function as memory systems with a dilution as high as 90%, although with somewhat reduced performance (Morgenstern 1986). The optimal level of dilution found here corresponds to approximately 50% of the smallest weights being set to zero. It has been calculated analytically that the optimal value for the diluted binary threshold is $z = |0.62|$ in a deterministic model, resulting in a load capacity of $\alpha_c = 0.12$ (Sompolinsky 1987). Simulations validated this result by finding an optimal value of $z = |0.6|$ at all levels of stochasticity.

Discussion

The results presented here indicate that the primary theoretical argument for the existence of binary synapses does not hold in a traditional attractor network regime: it appears that graded weight representations are, if anything, less, not more, susceptible to noise than binary representations.

Binary weight representations may have practical advantages in coding engineering applications of attractor-type memory systems, due to the relatively slight increase in load capacity obtained by including more weight states and the extra information required to store the additional states. These advantages have yet to be realized, however, inasmuch as true binary learning algorithms have not yet been developed since traditional Hebbian-type learning requires continuous valued weights before discretization.

Our analysis is strictly based on standard attractor networks, on which some of the arguments have been based. Our results do not preclude advantages of binary weights in different network architectures, such as feedforward mapping networks, for different reasons. Indeed, interesting arguments have been made for such networks with binary weights, which include an efficient learning algorithm (Braunstein and Zecchina 2006; Baldassi et al. 2007). Also, our results are derived for the classical, fully connected network, and variations in the network structure could change these conclusions. However, as argued in this paper, the difference between binary and graded synapses seems to be moderated by the large number of inputs which drive the attractor dynamics. Thus, we expect similar results as long as the networks dynamics are determined by a large number of inputs.

Some engineering advantages may, in any event, be irrelevant to biological brains. The appropriate biological equivalent of nodes in attractor network models is controversial. Our recent experimental evidence that individual synapses can display multiple stable transmitter release probabilities in association with multiple inductions of long term potentiation and depression (Enoki et al. 2009) indicates that real brains do not use binary weight representations. Furthermore, one possible interpretation of nodes in neural network models is that each node represents a population of neurons with similar response properties (Wilson and Cowan 1973; Gerstner 2000). In this formulation, the weights of neural network models represent the average connectivity strength between populations of neurons. In this case, even if the individual synaptic efficacies were binary, the representation in a neural network model would be graded since it represents an average over many binary weights.

Increasing the number of possible discrete weight states increases the load capacity of attractor networks. This increase, in our models, begins to plateau beyond a relatively small number of states, on the order of a dozen. Our recent experimental work (Enoki et al. 2009) provides evidence of at least five weight states at a single synapse, but the maximum number of states is not yet known. Thus, it appears appropriate to use graded weight representations in formulating neural network models.

References

Abarbanel H, Talathi S, Gibb L, Rabinovich M (2005) Synaptic plasticity with discrete state synapses. Phys Rev E 72:031914

Amari S (1972) Learning patterns and pattern sequences by self-organizing nets of threshold elements. IEEE Trans Comput C-21:1197

Amit D, Mongillo G (2003) Spike-driven synaptic dynamics generating working memory states. Neural Comput 15:565

Amit D, Gutfreund H, Sompolinsky H (1985) Storing infinite numbers of patterns in a spin-glass model of neural networks. Phys Rev Lett 55:1530

Amit D, Gutfreund H., Sompolinsky H (1987) Statistical mechanics of neural networks near saturation. Ann Phys 173:30

Baldassi C, Braunstein A, Brunel N, Zecchina R (2007) Efficient supervised learning in networks with binary synapses. Proc Natl Acad Sci 104:11079–11084

Braunstein A, Zecchina R (2006) Learning by message passing in networks of discrete synapses. Phys Rev Lett 96:30201

Brody C, Romo R, Kepecs A (2003) Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors and dynamic representations. Curr Opin Neurobiol 13:204

Brunel N, Hakim V, Isope P, Nadal J, Barbour B (2004) Optimal information storage and the distribution of synaptic weights: perceptron versus purkinje cell. Neuron 43:745

Chklovskii D, Mel B, Svoboda K (2004) Cortical rewiring and information storage. Nature 431:782

Dobrunz L (1998) Long-term potentiation and the computational synapse. Proc Natl Acad Sci USA 95:4086

Enoki R, Hu YL, Hamilton D, Fine A (2009) Expression of long-term plasticity at individual synapses in the hippocampus is graded, bidirectional, and mainly presynaptic: optical quantal analysis. Neuron 62:242

Fusi S, Abbott L (2007) Limits on the memory storage capacity of bounded synapses. Nat Neurosci 10:485–493

Gerstner W (2000) Population dynamics of spiking neurons: fast transients, asynchronous states, and locking. Neural Comput 12:43

Giudice PD, Fusi S, Mattia M (2003) Modelling the formation of working memory with networks of integrate-and-fire neurons connected by plastic synapses. J Physiol Paris 97:659

Graupner M, Brunel N (2007) Stdp in a bistable synapse model based on camkii and associated signaling pathways. PLoS Comput Biol 3:221

Grossberg S (1969) On the serial learning of lists. Math Biosci 4:201

Gutfreund H, Stein Y (1990) Capacity of neural networks with discrete synaptic couplings. J Phys A 23:2613–2630

Gutig R, Aharonov R, Rotter S, Sompolinsky H (2003) Learning input correlations through nonlinear temporally asymmetric hebbian plasticity. J Neurosci 23:3697–3714

Hopfield J (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 79:2554

Hopfield J (1984) Neurons with graded response have collective computational properties like those of two-state neurons. Proc Natl Acad Sci USA 81:3088

Isaac J, Nicoll R, Malenka R (1995) Evidence for silent synapses: implications for the expression of ltp. Neuron 15:427

Kepecs A, Rossum MC, van Song S, Tegner J (2002) Spike-timing-dependent plasticity: common themes and divergent vistas. Biol Cybern 87:446–458

Koulakov A, Raghavachari S, Kepecs A, Lisman J (2002) Model for a robust neural integrator. Nat Neurosci 5:775

Krauth W, Mezard M (1989) Storage capacity of memory networks with binary couplings. J Phys France 50:3057–3066

Liao D, Hessler N, Mallnow R (1995) Activation of postsynaptically silent synapses during pairing-induced ltp in ca1 region of hippocampal slices. Nature 375:400

Martin SJ, Grimwood PD, Morris RGM (2000) Synaptic plasticity and memory: an evaluation of the hypothesis. Ann Rev Neurosci 23:649

Matsuo N, Reijmers L, Mayford M (2008) Spine-type specific recruitment of newly synthesized ampa receptors with learning. Science 319:1104

Morgenstern I (1986) Spin-glasses, optimization and neural networks. In: Hemmen JV, Morgenstern I (eds) The Heidelberg colloquium on glassy dynamics and optimization, Springer-Verlag, Berlin, p 399

O'Connor DH, Wittenberg GM, Wang S (2005a) Dissection of bidirectional synaptic plasticity into saturable unidirectional processes. J Neurophysiol 94:1565

O'Connor DH, Wittenberg GM, Wang S (2005b) Graded bidirectional synaptic plasticity is composed of switch-like unitary events. Proc Natl Acad Sci USA 102:9679

Peterson CCH, Malenka RC, Nicoll RA, Hopfield JJ (1998) All-or-none potentiation at ca3-ca1 synapses. Proc Natl Acad Sci USA 95:4732

Poirazi P, Mel B (2001) Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. Neuron 29:779

Rolls E, Tovee M (1995) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. J Neurophysiol 73:713

Senn W, Fusi S (2004) Slow stochastic learning with global inhibition: a biological solution to the binary perceptron problem. Neurocomputing 58:321

Sompolinsky H (1986) Neural networks with nonlinear synapses and static noise. Phys Rev A 34:2571–2574

Sompolinsky H (1987) The theory of neural networks: the hebb rule and beyond. In: Hemmen JV, Morgenstern I (eds) The Heidelberg colloquium on glassy dynamics and optimization. Springer-Verlag, Berlin, p 485

Vladimirski B, Vasilaki E, Fusi S, Senn W (2006) Hebbian reinforcement learning with stochastic binary synapses: theory and application to the xor problem. Preprint from Elsevier Science

Willshaw D, Buneman O, Longuet-Higgins H (1969) Nonholographic associative memory. Nature 222:960

Wilson H, Cowan J (1973) A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. Kybernetik 13:55