arXiv:1407.1904v1 [quant-ph] 7 Jul 2014

Adiabatic Quantum Optimization for Associative Memory Recall

H. Seddiqi and T. S. Humble‡

Quantum Computing Institute, Oak Ridge National Laboratory, Oak Ridge TN, USA

E-mail: humblets@ornl.gov

Abstract. Hopfield networks are a variant of associative memory that recall information stored in the couplings of an Ising model. Stored memories are fixed points for the network dynamics that correspond to energetic minima of the spin state. We formulate the recall of memories stored in a Hopfield network using energy minimization by adiabatic quantum optimization (AQO). Numerical simulations of the quantum dynamics allow us to quantify the AQO recall accuracy with respect to the number of stored memories and the noise in the input key. We also investigate AQO performance with respect to how memories are stored in the Ising model using different learning rules. Our results indicate that AQO performance varies strongly with learning rule due to the changes in the energy landscape. Consequently, learning rules offer indirect methods for investigating change to the computational complexity of the recall task and the computational efficiency of AQO.

Keywords: quantum computing, adiabatic quantum optimization, associative memory, content-addressable memory, Hopfield networks

1. Introduction

Content-addressable memory (CAM) is a form of associative memory that recalls information by value [1]. Given an exact or approximate input value, a CAM returns the closest matching key in memory by making comparisons with stored keys. This is in contrast to random access memory (RAM), which returns the value stored at a provided key or address. CAMs are of particular interest for applications tasked to quickly search large databases including, for example, network switching, pattern matching, and machine vision [2]. An auto-associative CAM is a memory in which the key and value are the same and partial knowledge of the input value triggers the recall of its completed value.

Auto-associative CAMs have proven of interest for modeling neural behavior and cognition [3]. This is due partly to their properties of operating in massively parallel mode and being robust to noisy input. Indeed, these connections were part of the motivation for Hopfield to propose a model for an auto-associative CAM based on a

‡ MS 6015, Oak Ridge National Laboratory, One Bethel Valley Road, Oak Ridge, TN, 37831-6015 USA

network of computational neurons [1, 4]. The Hopfield neural network stores memories in the synaptic weights describing the connectivity between computational neurons using an unsupervised learning rule. The initial network state propagates discretely by applying an update to each neuron based on the synapses and all other neural states. Hopfield showed that stored memories can become fixed point attractors under Markov dynamics, thus enabling a new type of CAM. It is now understood that the memory capacity for a Hopfield network depends strongly on the learning rule used in setting the synaptic weights [5, 6, 7].

The theoretical underpinning of the Hopfield network is a classical Ising model in which each binary neuron is mapped into a spin-1/2 system [3]. The synaptic weights define the coupling between these spins and the susceptibility for a neuron to activate is set by the applied bias. Because the energy of the Ising model represents a Lyapunov function, the network dynamics guarantees convergence to a fixed point attractor in the asymptotic limit [1]. Conventionally, Hopfield networks are formulated in terms of a stochastic update rule governed by the Ising energy. However, finding stable points of the Lyapunov function can also be viewed as minimization of the network energy [8]. In the case of the Hopfield network, spin configurations that minimize the network energy are the fixed point attractors and solutions to the stochastic dynamics.

In this work, we develop an auto-associative CAM that performs memory recall using the principle of energy minimization as implemented by adiabatic quantum optimization (AQO). AQO represents a novel approach to optimization that leverages quantum computational primitives for minimizing the energy of a system of coupled spin states [9, 10]. As part of the broader adiabatic quantum computing mode, AQO has been investigated for a number of applications, including classification [11, 12], machine learning [13], graph theory [14, 15, 16], and protein folding [17] among others [18]. In each of these uses, the respective problems have undergone reduction first to a discrete optimization problem that is subsequently mapped into the AQO paradigm. By comparison, Hopfield networks offer a direct application of AQO when the latter is implemented using an Ising model in a transverse field [19]. This is because no reduction in the original problem is required to apply the principle of energy minimization. Consequently, our theoretical model does not constitute a quantum neural network but rather a conventional neural network solved with quantum computing principles.

The use of the AQO algorithm for performing memory recall in a Hopfield network has been investigated previously by Neigovzen et al. in the context of pattern recognition [20]. Specifically, they employ AQO to minimize the energy of a Hopfield network expressed as an Ising Hamiltonian. Neigovzen et al. performed an experimental demonstration of these ideas using a 2-neuron example in the context of NMR spin-based encoding. Their results confirmed the recall accuracy using AQO for this small network and invite questions as to how the details of Hopfield networks influence performance of AQO. Our investigation complements these efforts by quantifying how different network parameters, including size, memories, and learning rules, influence recall accuracy.

In a more general sense, Hopfield networks and CAMs are tasked with finding an

unknown value within an unsorted database suggesting there is a strong connection between this type of tagged search and Grover's search algorithm. Previous work by Roland and Cerf on using the AQO algorithm to perform search tasks makes this connection clear [21]. In particular, Roland and Cerf show how Grover's search algorithm can be mapped into the AQO algorithm and how he terminal Hamiltonian serves the role of the oracle. Our Hopfield-based CAM using AQO for memory recall is almost equivalent to the search of Roland and Cerf for a specific learning rule (the Hebb rule). The primary distinction is that an auto-associative CAM must differentiate between members of a set of stored memories as well as the set of unstored memories, whereas the oracle of Roland and Cerf deal with only acceptable states or memories. Our results extend the search task to the context that the oracle must discriminate between tagged and untagged keys.

In Sec. 2, we define the task of memory recall using a conventional Hopfield network and describe the Hebb, Storkey, and projection learning rules for preparing the synpatic wieghts. In Sec. 3, we introduce adiabatic quantum optimization, its use for memory recall, and the basis for our numerical simulation studies. In Sec. 4, we present results for example instances of Hopfield networks that demonstrate the behavior of AQO for memory recall while in Sec. 5 we present calculations of the average recall success for an ensemble of different networks. We present final conclusions in Sec. 6.

2. Hopfield Networks

We define a classical Hopfield network of n neurons with each neuron described by a bipolar spin state $z_j \in \{\pm 1\}$. Neurons i and j are symmetrically coupled by the synaptic weights $w_{ij} = w_{ji}$ while self-connections are not permitted, i.e., $w_{ii} = 0$. The energy of the network in the spin state $\mathbf{z} \in (z_1, z_2, \ldots, z_n)^T$ is

$$E(\mathbf{z};\theta) = -\frac{1}{2}\sum_{i,j=1}^{n} z_i w_{ij} z_j - \sum_{i=1}^{n} \theta_i z_i, \tag{1}$$

with $\theta = (\theta_1, \theta_2, \ldots, \theta_n)^T$ and θ_i the real-valued activation threshold for the i-th neuron. This form for the energy represents a classical version of the Ising model in which the spin configuration describes the orientation of the n dimensional system. The dynamics of the Hopfield network are conventionally modeled by the discrete Markov process

$$z_i = \begin{cases} 1 & \text{if } \sum_j w_{ij} z_j > \theta_i \\ -1 & \text{otherwise} \end{cases} \tag{2}$$

where the state of the i-th neuron may be updated either in series (asynchronously) or in parallel (synchronously) with all other neurons in the network. The network is initialized in the input state $z_i = z_{0,i}$ and subsequently updated under repeated application of Eq. (2) until it reaches a steady state

$$z_i = \text{sign}\left(\sum_j w_{ij} z_j\right) \tag{3}$$

Steady states of the Hopfield network represent fixed point attractors and are local minima in the energy landscape of Eq. (1) [3]. The stable fixed points are set by the choice of the synaptic couplings w_{ij} and the network converges to the memory state closest to the initial state z_0. However, the network has a finite capacity to store memories and it is well known that when too many memories are stored the dynamics converge to a spurious mixture of memories. The emergence of spurious states places a limit on the storage capacity of the Hopfield network that depends on both the interference or overlap between the memories and the learning rule used to set the synaptic weights.

2.1. Synaptic Learning Rules

Learning rules specify how memories are stored in the synaptic weights of a Hopfield network and they play an important role in determining the memory capacity. The capacity $c_n = p/n$ is the maximum number of patterns p that can be stored in a network of n neurons and then accurately recalled [22]. Different learning rules yield different capacities and we will be interested in understanding how these differences influence performance of the AQO algorithm. Setting the synaptic weights w_{ij} for a Hopfield network is done using a specific choice of learning rule that in turn generates a different Ising model. Learning rules represent a form of unsupervised learning in which the memories are stored in the network without any corrective back-action. We make use of three learning rules that have been previously found to yield different capacities for Hopfield networks in the classical setting.

2.1.1. Hebb Rule The Hebb learning rule defines the synaptic weights

$$w_{ij} = \frac{1}{n} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu \tag{4}$$

for a set of p memories $\{\xi^1, \xi^2, \ldots, \xi^p\}$, each of length n with bipolar elements $\xi_i^\mu \in \{\pm 1\}$. Geometrically, each summand corresponds to the projection of the neuron configuration into the μ-th memory subspace. These projections are orthogonal if all p patterns are mutually orthogonal. More generally, the Hebb rule maps non-orthogonal memory states into overlapping projections. This leads to interference during memory recall as two or more correlated memories may both be close to the input state. In the asymptotic limit for the number of neurons, the capacity of the Hebb rule is $c_n = n/2 \ln n$ under conditions of perfect recall, i.e., no errors in the retrieved state. By comparison, under conditions of imperfect recall the asymptotic capacity is $c_n \approx 0.14$ [5]. It is worth noting that the Hebb rule is incremental as it is a sum over individual patterns. The rule is also local since the synaptic weights depend only on the value of the adjacent neurons.

2.1.2. Storkey Rule The Storkey learning rule defines the synaptic weights in an iterative fashion as

$$w_{ij}^{\nu} = w_{ij}^{\nu-1} + \frac{1}{n}\xi_i^{\nu}\xi_j^{\nu} - \frac{1}{n}\xi_i^{\nu}h_{ji}^{\nu} - \frac{1}{n}h_{ij}^{\nu}\xi_j^{\nu} \tag{5}$$

where ξ^{ν} is the memory to be learned in the ν-th iteration for $\nu = 1$ to p and

$$h_{ij}^{\nu} = \sum_{k=1,k\neq i,j} w_{ik}^{\nu-1}\xi_k^{\nu} \tag{6}$$

is the local field at the i-th neuron [7]. The final synaptic weight storing p memories is given by $w_{ij} = w_{ij}^{p}$. The Storkey rule is found to more evenly distribute the fixed points and increases the capacity of the network. The asymptotic Storkey capacity under prefect recall is $n/\sqrt{2\ln n}$, which represents an improvement over the Hebb rule. As with the Hebb rule, the Storkey rule is incremental and permits the addition of new memories.

2.1.3. Projection Rule The projection rule defines the synaptic weights for p memories as

$$w_{ij} = \frac{1}{n}\sum_{\mu,\mu'=1}^{p} \xi_i^{\mu} C_{\mu\mu'}^{-1} \xi_j^{\mu'} \tag{7}$$

where $C_{\mu\mu'} = \frac{1}{n}\sum_{k=1}^{n}\xi_k^{\mu}\xi_k^{\mu'}$ is the covariance matrix and C^{-1} is the inverse of C. This rule has a theoretical capacity of n for linearly independent patterns and approximately $n/2$ for interfering memories [6, 23]. The projection rule is neither local nor incremental as adding memories to the network requires resetting each element using knowledge of all other memories. In the limit of orthogonal memories, all three learning rules reduce to the Hebb rule.

3. Memory Recall by Adiabatic Quantum Optimization

The learning rules defined in Sec. 2.1 offer different methods for preparing the synaptic weights and the fixed points of a Hopfield network. Conventionally, the network finds those states that satisfy the equilibrium condition of Eq. (3) by evolving under the discrete Markov process of Eq. (2). However, the fixed points of a Hopfield network are also minima of the energy function and hence known as stable fixed points. The stability of these solutions is due to the quadratic form of the energy function, Eq. (1), which is a Lyapunov function that monotonically decreases under updates of network state [3].

As an alternative to fixed point convergence, we apply the principle of optimization for finding the global minima of the energy function and for recalling a stored memory. Our formulation uses the same synaptic weight matrix and underlying Ising model of a conventional Hopfield network. However, we use this information to set the activation thresholds θ_i in place of initializing the network into a state z_0. This feature casts

recovery of an unknown memory in terms of minimizing the energy of the network. We formally define the energy minimization condition as

$$z = \arg\min_{z'} E(z'; \theta). \tag{8}$$

in which the vector θ represents the activation thresholds $\theta_i = \Gamma z_{0,i}$ and Γ is an energy scale for the applied bias. The activation threshold θ now serves as an energetic bias towards network states that best match the partial memory input. The classical dynamics of the stochastic Hopfield network are recovered by initializing the state of all neurons to an indeterminate value, i.e., $z_i = 0$, whereupon the first update will prepare the state z_0.

In the absence of any bias, finding the global minima of $E(z, \theta)$ is equivalent to computing the lowest energy eigenstates of the synaptic weight matrix w_{ij} with the constraint $z_i \in \{\pm 1\}$ (indeterminate values are not valid output states). Due to the symmetry of the unbiased energy, the complement of each memory is also an eigenstate. However, the presence of a non-zero bias breaks this symmetry and leads to a lower energy for one memory state relative to the other stored memories. Global minimization then returns the spin configuration that encodes the memory recalled.

Whether the expected memory is recalled depends on several factors. First, if the applied bias is too large then the input state itself becomes a fixed point and the global minimum is the input state $z_{0,i}$. This behavior is unwanted since it does not confirm whether the input or its closest match were part of the memory. This effect can be detected by decreasing Γ and monitoring changes in the recall. An upper bound on Γ can be calculated for the projection rule by comparing network energies of a memory state ξ^k with a non-memory state z_0, i.e.,

$$\Gamma < \frac{\sum_{i,j} \xi_i^k w_{ij} x_j^k - z_{0,i} w_{ij} z_{0,j}}{2(n - \sum_i z_{0,i} \xi_i^k)} \tag{9}$$

In the limit that the memories are orthogonal to each other as well as the input key, this reduces to the result $\Gamma < 1/(2n)$ previously noted by Neigovzen et al.[20].

Interference between memories prevents their discrimination when insufficient knowledge about the aought-after memory is provided by the activation threshold θ and input state z_0. The number of memories stored in the network may also exceed the network capacity and lead to erroneous recall results. As an example, perfect recall is observed when using the Hebb rule in a classical network storing p orthogonal memories provided $p \leq n$, since there is no interference in these non-overlapping states. However, the capacity for non-orthogonal memories is much lower and varies with learning rule, as described above. In our optimization paradigm, interference manifests as degeneracy in the ground state manifold. These degeneracies are formed from superpositions of stored memory states. They are valid energetic minima that corresponding to the aforementioned spurious states. Differences between learning rules seek to remove the presence of spurious states while also increasing the network capacity.

3.1. Adiabatic Quantum Optimization Algorithm

Adiabatic Quantum Optimization is based on the principle of adiabatically evolving the ground state of an initial well-known Hamiltonian to the unknown ground state of a final Hamiltonian. By defining the final Hamiltonian in terms of the Ising model representing a Hopfield network, we use AQO to recover the ground state expressing a stored memory. The Ising model for AQO will use the same synaptic weights and activation thresholds discussed in Sec. 2 for the Hopfield network. The recall operation begins by preparing a register of n spin-1/2 quantum systems (qubits) in a superposition of all possible network states and adiabatically evolving the register state towards the final Ising Hamiltonian. Assuming the adiabatic condition has remained satisfied, the qubit register is prepared in the ground state of the Ising Hamiltonian. Upon completion of the evolution, each qubit in the register is then measured and the resulting string of bits is interpreted as the network state, z_i.

Formally, we consider a time-dependent Hamiltonian

$$H(t) = A(t)H_0 + B(t)H_1 \tag{10}$$

with piece-wise continuous annealing schedules $A(t)$ and $B(t)$ that satisfy $A(0) = 1, B(0) = 0$ and $A(T) = 1, B(T) = 1$. Together, the initial Hamiltonian

$$H_0 = -\sum_{i}^{n} X_i \tag{11}$$

and the final Hamiltonian

$$H_1 = -\sum_{i,j}^{n} w_{ij} Z_i Z_j - \sum_{i}^{n} \theta_i Z_i \tag{12}$$

represent an Ising model in a transverse field. In the latter equations, the Pauli Z_i and X_i operators act on the i-th qubit while the constants θ_i and w_{ij} denote the qubit bias and coupling, respectively. Of course, the latter quantities are exactly the activation threshold and synaptic weights of the Hopfield network. We choose the computational basis in terms of tensor product states of the $+1$ and -1 eigenstates of operators Z_i denoted as $|0\rangle$ and $|1\rangle$, respectively. In this basis, the correspondence between the binary spin label $z \in \{0, 1\}$ and the bipolar spin configuration label $s \in \{\pm 1\}$ is $s = 2z - 1$.

The quantum state of an n-qubit register is prepared at time $t = 0$ in the ground state of H_1,

$$|\psi(t = 0)\rangle = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle, \tag{13}$$

with $|x\rangle = |x_1\rangle \otimes |x_2\rangle \ldots \otimes |x_n\rangle$ and

$$x = \sum_{i=1}^{n} x_i 2^{i-1}, \qquad x_i \in \{0, 1\} \tag{14}$$

the binary expansion of the state label x. The register state $\psi(t)$ evolves under the Schrodinger equation

$$i\frac{d\,|\psi(t)\rangle}{dt} = H(t)\,|\psi(t)\rangle \tag{15}$$

from the initial time 0 to a final time T. We set $\hbar = 1$. The time scale T is chosen so that changes in the register state $\psi(t)$ are slow (adiabatic) relative to the inverse of the minimum energy gap of $H(t)$. The minimum energy gap Δ_{min} is defined as the smallest energy difference between the instantaneous ground state manifold and those excited states that do not terminate as a ground state. Provided the time scale $T \gg \Delta_{min}^{-1}$, then the register remains a ground state of the instantaneous Hamiltonian and evolution to the time T prepares the ground state of $H(T) = H_1$. However, the exact scaling for the minimal T with respect to Ising model size and parameterization is an open question.

After preparation of the final register state $\psi(T)$, each qubit is measured in the computational basis. Because the final Hamiltonian H_1 is diagonal in the computational basis, the prepared ground state may be directly related to a valid state of the Hopfield network. The state of the i-th qubit is measured in the Z_i basis and the result $z_i \in \{\pm 1\}$ is the corresponding solution for the i-th neuron in the network.

3.2. AQO Recall Accuracy

The accuracy with which a memory is recalled using the AQO algorithm can be measured in terms of the probability that the correct (expected) network state is recovered. We define a measure of the probabilistic success for recall as

$$f_x = \begin{cases} 1, \; P_{ans} \geq x \\ \\ 0, \; P_{ans} < x \end{cases} \tag{16}$$

where P_{ans} is the probability to recover the correct memory and $x \in [0, 1]$ is the threshold probability. Denoting the correct memory state as ϕ_{ans}, the probability to recover the correct memory can be computed from the simulated register state as

$$P_{ans} = |\,\langle \phi_{ans}|\psi(T)\rangle\,|^2 \tag{17}$$

We assume in this analysis that the register state is a pure state and therefore neglect sources of noise including finite temperature and external couplings.

From this definition for probabilistic success, we consider average success for an ensemble of N instances as

$$\langle f_x \rangle = \frac{1}{N}\sum_{i=1}^{N} f_x^i, \tag{18}$$

where f_x^i represents the probability for success of the i-th instance of n neurons storing p memories. This is a binomial distribution with variance $\langle \Delta f_x \rangle = \langle f_x \rangle (1 - \langle f_x \rangle)$. This statistic will be used for characterizing the behavior of an ensemble of simulated recall operations.

3.3. Numerical Simulations of the AQO Algorithm

We use numerical simulations of the time-dependent Schrodinger equation in Eq. (15) to compute the register state $\psi(T)$ prepared by the AQO algorithm. These simulations provide the information needed to calculate the probabilistic success f_x as well as the average success with respect to network size and learning rule. Our methods are restricted to pure-state simulations, which provide an idealized environment for the AQO algorithm and permit our analysis to emphasize how learning rules influence success via changes to the Ising model.

Our numerical methods make use of a first-order Magnus expansion of the time-evolution operator

$$U(t_{j+1}, t_j) = \exp\left[-i \int_{t_j}^{t_{j+1}} H(\tau) d\tau\right] \tag{19}$$

over the interval $[t_j, t_{j+1}]$ for $j = 0$ to $J - 1$. We use a uniform time step $\Delta t = t_{j+1} - t_j$ such that $T = J\Delta t$. Starting from the initial state Eq. (13), an intermediate state is generated from the series of time evolution operators

$$|\psi(t_{j'})\rangle = \prod_{j=0}^{j'-1} U(t_{j+1}, t_j) \, |\psi(0)\rangle \tag{20}$$

In these calculations, the action of the jth time-evolution operator onto the appropriate state vector is calculated directly [24, 25]. The simulation code is available for download [26]. In our simulations, we use annealing schedules $A(t) = 1 - t/T$ and $B(t) = t/T$, and we do not place any constraints on the qubit connectivity or the coupling precision.

4. Recall Instances

We present some example instances that demonstrate the behavior of the AQO algorithm for memory recall. We begin by considering the case of p orthogonal memories. A convenient source of orthogonal bipolar states is the n-dimensional Hadamard matrix for $n = 2^k$, whose columns are orthogonal with respect to the usual inner product. We use these memories to prepare the synaptic weights and corresponding Ising Hamiltonians. For orthogonal memories, all the learning rules prepare the same weights.

In the absence of any bias, $\theta_i = 0$ and we expect recall to recover each of the p encoded memories with uniform probability. The quadratic symmetry of the energy in Eq. (1) also makes the complement of each memory state a valid fixed point. This implies a total ground state degeneracy of $2p$ in the absence of bias. An example of the time-dependent spectral behavior is shown in Fig. 1 for the case $p = n = 4$, and all the eigenstates converge to a single ground state energy. The same case but with θ set to the first memory and $\Gamma = 1$ is shown in Fig. 2. The presence of the bias removes the ground state degeneracy and, not apparent from the figure, the prepared ground state matches the biased input state.

We next consider an instance of non-orthogonal memories, that is, for which the inner product between memory pairs is non-zero. Interference is expected to cause failure during recall when the applied bias is insufficient to distinguish between similar states. With $p = n = 4$, we use the memory set

$$\Sigma = \begin{pmatrix} +1 & -1 & -1 & +1 \\ -1 & +1 & -1 & -1 \\ +1 & +1 & +1 & +1 \\ -1 & +1 & +1 & +1 \end{pmatrix} \tag{21}$$

where columns 1, 2, and 3 overlap while columns 2 and 4 are orthogonal. We use an input state $z_0 = (1, -1, 1, -1)$ that most closely matches the first memory $\Sigma_{i,1}$. For these simulations, we use an annealing time $T = 1000$ that was found to yield numerical convergence in the ground state amplitudes. Both time and energy are expressed in arbitrary units since all calculated quantities are independent of the absolute energy scale of the Ising model.

Figures 3 through 6 plot the probability to recall the answer state as a function of $\Gamma \in [0, 1]$. The recall probability varies with input bias, number of memories, and learning rule. For $p = 1$, there is only one memory stored in the network and any non-zero bias distinguishes between the memory and its complement. Similarly, all three rules behave the same for the case of $p = 2$ in Fig. 4. This is because there are not significant energetic differences between the rules using the first two memories above. The Hebb and Storkey rules coincide exactly, while the projection rule is identical for the lowest energy eigenstates. However, for the case of $p = 3$ in Fig. 5, there is a distinction between all three rules. The answer state probability using the projection rule is nearly the same as was observed for fewer memories while the Hebb rule shows a shift to larger bias. This is a result of the added memory having a lower unbiased energy than the answer state for this learning rule. As a result, larger bias must be applied to lower the answer state below that of the new memory. In contrast, the Storkey shifts to smaller bias as a result of memory addition. This is because the rule mitigates interference using the local field calculation. However, with the addition of another memory, $p = 4$, the Hebb rule becomes more evenly distributed in energy across the degenerate memory states while the Storkey rule shows a slight shift to larger bias and the projection rule remains unchanged.

5. Statistical Recall Behavior

Our results for recall success of individual Hopfield networks indicate there is a large degree of variability in performance with respect to the stored memory states. We have found it useful to average performance across a range of problem instances. Under these circumstances, we use the average success probability defined by Eq. (18) to quantify the relative performance of each learning rule in terms of neurons n, memories p, and bias Γ. As noted earlier, these statistics correspond to a binomial distribution with

parameter $\langle f_x \rangle$.

We first investigate the average AQO recall behavior with respect to the bias Γ. An ensemble of problem instances is constructed for $n = 5$ neurons in which each instance consist of p memories with elements sampled uniform random from $\{\pm 1\}$. Among the p memories, one is selected as the answer state while the other memories are selected to be distinct from the answer state. The answer state is then chosen as the input state, which defines the activation threshold θ. The simulation computes the full quantum state using an annealing time $T = 1000$. The probability to occupy the expected answer state is then computed using Eq. (16) with a threshold $x = 2/3$. The exact value is not expected to be significant provided it is above the probability for a uniform superposition.

Figure 7 shows the average success probability for recovering the answer state as the bias increases from 0 to 1. Relative to the number of p memories, we find the learning rules exhibit distinct behaviors. For the Hebb rule there is a step-wise decrease in success as the number of memories increases and none of the Γ values read unit success except for the trivial $p = 1$ case. A similar but weakened version of this behavior is seen for the Storkey rule at values of Γ below 0.25. However, above this value the Storkey rule recovers unit success. The projection rule demonstrate a very different behavior; unit success is seen for all non-zero values of Γ for all value of p except the trivial case which succeeds with $\Gamma = 0$ as expected.

The plots in Fig. 7 indicate when the prepared ground state has greater than 2/3 probability to be in the answer state given an input that matches a memory. The better performance of both the Storkey and projection rules is a result of how they modify the energy landscape. Both rules effectively raise the energy barrier between fixed stable points, while the Hebb rule preserves interference between memories. As the number of memories increases, so does the interference within the the typical problem instance. This behavior is underscored by the strong dependence of the Hebb rule on the number of stored memories p.

A related behavior is investigated by providing an input state to the network that has a Hamming distance 1 from a certain memory state, and all other memory states are chosen to be at least Hamming distance 2 away from the input state. This can be interpreted as a case where the input state is noisy or incomplete. In Fig. 8, the success probability for this case is plotted with respect to Γ, number of memories, and learning rule. Note the similarity between the behavior of the Hebb and projection rules.

We have also investigated the influence of overbiasing the network toward the input state. As noted previously, there are loose upper bounds on Γ for the Hebb rule based on energetic analysis [20]. Figure 9 plots the average failure, which is measured as the likelihood that AQO recalls the input state. In these plots, the input state is not among the stored memories and, in fact, is at least Hamming distance 1 away from all of them. We see that the failure rate increases as Γ increases. This indicates that the system is overbiased. It is notable that the learning rules exhibit very different behaviors for failure. Whereas the Hebb rule terminates with lower failure probability as the memories are increased, both the Storkey and projection rules reach unit failure

with sufficiently large Γ. A similar plot is shown in Fig. 10 for the case that the input is at least Hamming distance 2 from all the stored memories. The sensitivity to failure increases with the increase in Hamming distance due to the decreased interference with true memory states.

The annealing time T is expected to also play a role in the recall success. Because the state dynamics must be adiabatic relative to the minimum energy gap, the diversity of instances used for $\langle f_x \rangle$ are also likely to support a diversity of Δ_{min}. This implies that there may be some maximum T for the ensemble which ensures every instance is quasi-adiabatic. In Fig. 11, we show a series of recall averages for different annealing times. For small values of T, the average success is low, especially as p approaches n. This suggests that many instances do not meet the $x = 2/3$ threshold for success. As T increases, the average success also increases but only up to a limit that depends on each learning rule. For the Storkey and projection rule, this limit is after $T = 100$, while for the Hebb rule the limits seems to occur after $T = 20$. For annealing times larger than these limits, the average recall success for each learning rule does not make a significant change. This indicates the annealing time is not the limiting factor in recall success and, therefore, it is likely the adiabatic condition has been met.

6. Conclusions

We have presented a theoretical formulation of auto-associative memory recall in terms of adiabatic quantum optimization. We have used numerical simulations to quantify the recall performance with respect to three different learning rules (Hebb, Storkey, and projection) and we have accumulated statistics on recall accuracy and failure across an ensemble of different network instances. We have found that the probability to populate the expected ground state using AQO is sensitive to learning rule, number of memories, and size of the network. Our simulation studies have been limited in size, but for these small networks there are notable differences in both the success and failure rates across learning rules.

As noted in Sec. 1, the use of AQO for memory recall is closely related to its use for searching an unsorted database [21]. Roland and Cerf constructed the search problem around an oracle that matches the Hebb rule for the Hopfield network and they considered the task of recovering any valid memory. By contrast, we have shown how the activation threshold is interpreted as bias in the Ising model. Thus, the activation threshold serves as classical input, i.e., a tag, into the oracle construction for unsorted search. We have not attempted to optimize the annealing schedule associated with memory recall. We think it is unlikely that the optimized annealing schedule recovered by Roland and Cerf for untagged search would extend to memory recall due to the influence of the activation threshold on the energy spectrum. However, we have found that the Hebb rule is sub-optimal with respect to recall accuracy when the stored memories are non-orthogonal. This indicates that the optimized annealing schedule for memory recall also likely depends on learning rule.

We have observed a significant dependence on recall accuracy and minimal annealing time with respect to learning rule. These results support recent work by Katzgraber et al. that argued the relative height of energy barriers play a fundamental role in determining which Ising Hamiltonian are challenging for AQO [27]. Historically, learning rules that provide well separated but broad energy basins have been the goal of classical Hopfield networks. For the projection rule, which performs best of the three studied here, its known that the classical energy basins are better separated than either Hebb or Storkey for non-orthogonal memory sets. The dependency of average success rate with respect to learning rule found here bear out the expectations of Katzgraber et al. It is therefore interesting to consider as to whether learning rules may provide implicit control over the complexity of solving specific search tasks.

Acknowledgments

Acknowledgments H. S. thanks the Department of Energy Science Undergraduate Laboratory Internship (SULI) program and Prof. Mark Edwards, Georgia Southern University, for use of computing resources. This manuscript has been authored by a contractor of the U.S. Government under Contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

References

[1] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences*, vol. 79, no. 8, pp. 2554–2558, 1982.

[2] K. Pagiamtzis and A. Sheikholeslami, "Content-addressable memory (cam) circuits and architectures: A tutorial and survey," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 3, pp. 712–727, 2006.

[3] R. Rojas, *Neural Networks.* Berlin: Springer, 1996.

[4] J. J. Hopfield and D. W. Tank, "Computing with neural circuits-a model," *Science*, vol. 233, no. 4764, pp. 625–633, 1986.

[5] D. J. Amit, H. Gutfreund, and H. Sompolinsky, "Spin-glass models of neural networks," *Phys. Rev. A*, vol. 32, pp. 1007–1018, Aug 1985.

[6] L. Personnaz, I. Guyon, and G. Dreyfus, "Collective computational properties of neural networks: New learning mechanisms," *Phys. Rev. A*, vol. 34, pp. 4217–4228, Nov 1986.

[7] A. J. Storkey and R. Valabrgue, "The basins of attraction of a new hopfield learning rule," *Neural Networks*, vol. 12, no. 6, pp. 869 – 876, 1999.

[8] J. Schneider and S. Kirkpatrick, *Stochastic Optimization.* Scientific Computation, Springer, 2007.

[9] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, "A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem," *Science*, vol. 292, pp. 472–476, 2001.

[10] G. E. Santoro, R. Martonak, E. Tosatti, and R. Car, "Theory of quantum annealing of an ising spin glass," *Science*, vol. 295, no. 5564, pp. 2427–2430, 2002.

[11] H. Neven, G. Rose, and W. G. Macready, "Image recognition with an adiabatic quantum computer I. Mapping to quadratic unconstrained binary optimization." 2008.

[12] H. Neven, V. S. Denchev, G. Rose, and W. G. Macready, "Training a binary classifier with the quantum adiabatic algorithm." 2008.

[13] K. Pudenz and D. Lidar, "Quantum adiabatic machine learning," *Quantum Information Processing*, vol. 12, no. 5, pp. 2027–2070, 2013.

[14] F. Gaitan and L. Clark, "Ramsey numbers and adiabatic quantum computing," *Phys. Rev. Lett.*, vol. 108, p. 010501, Jan 2012.

[15] Z. Bian, F. Chudak, W. G. Macready, L. Clark, and F. Gaitan, "Experimental determination of ramsey numbers," *Phys. Rev. Lett.*, vol. 111, p. 130505, Sep 2013.

[16] F. Gaitan and L. Clark, "Graph isomorphism and adiabatic quantum computing," *Phys. Rev. A*, vol. 89, p. 022342, Feb 2014.

[17] A. Perdomo-Ortiz, N. Dickson, M. Drew-Brook, G. Rose, and A. Aspuru-Guzik, "Finding low-energy conformations of lattice protein models by quantum annealing," *Nature Scientific Reports*, vol. 2, Aug. 2012.

[18] V. N. Smelyanskiy et al., "A Near-Term Quantum Computing Approach for Hard Computational Problems in Space Exploration." 2012.

[19] M. W. Johnson et al., "Quantum annealing with manufactured spins," *Nature*, vol. 473, no. 7346, p. 194198, 2011.

[20] R. Neigovzen, J. L. Neves, R. Sollacher, and S. J. Glaser, "Quantum pattern recognition with liquid-state nuclear magnetic resonance," *Phys. Rev. A*, vol. 79, p. 042321, 2009.

[21] J. Roland and N. J. Cerf, "Quantum search by local adiabatic evolution," *Phys. Rev. A*, vol. 65, p. 042308, 2002.

[22] R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. S. Venkatesh, "The capacity of the hopfield associative memory," *IEEE Trans. Inf. Theor.*, vol. 33, pp. 461–482, July 1987.

[23] I. Kanter and H. Sompolinsky, "Associative recall of memory without errors," *Phys. Rev. A*, vol. 35, pp. 380–392, Jan 1987.

[24] A. H. Al-Mohy and N. J. Higham, "Computing the action of the matrix exponential, with an application to exponential integrators," *SIAM J. Sci. Comput.*, vol. 33, pp. 488–511, Mar. 2011.

[25] N. J. Higham and A. H. Al-Mohy, "Computing matrix functions," *Acta Numerica*, vol. 19, pp. 159–208, 5 2010.

[26] "AdiaQC." https://github.com/hadsed/AdiaQC, 2014. Accessed 2014-06-16.

[27] H. G. Katzgraber, F. Hamze, and R. S. Andrist, "Glassy chimeras could be blind to quantum speedup: Designing better benchmarks for quantum annealing machines," *Phys. Rev. X*, vol. 4, p. 021008, Apr 2014.

Figures

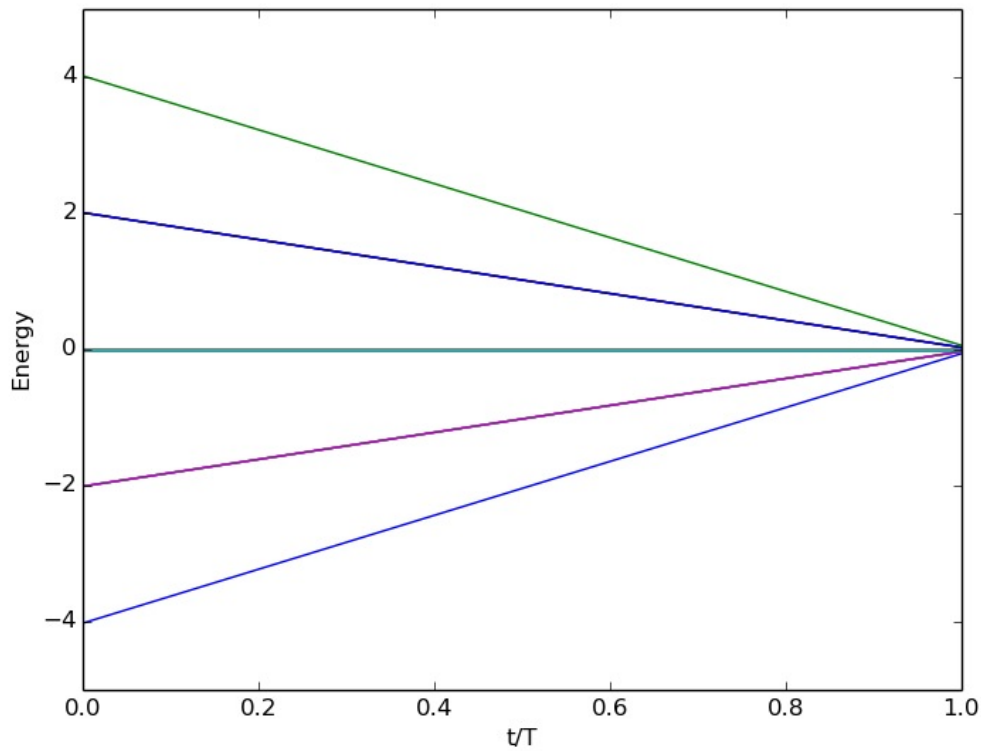

Figure 1. Time-dependent eigenspectrum for $p = 4$ orthogonal memories stored in a network of $n = 4$ neurons in the absence of bias, $\theta_i = 0$. For orthogonal memories, the spectrum is the same for the Hebb, Storkey, and projection learning rules.

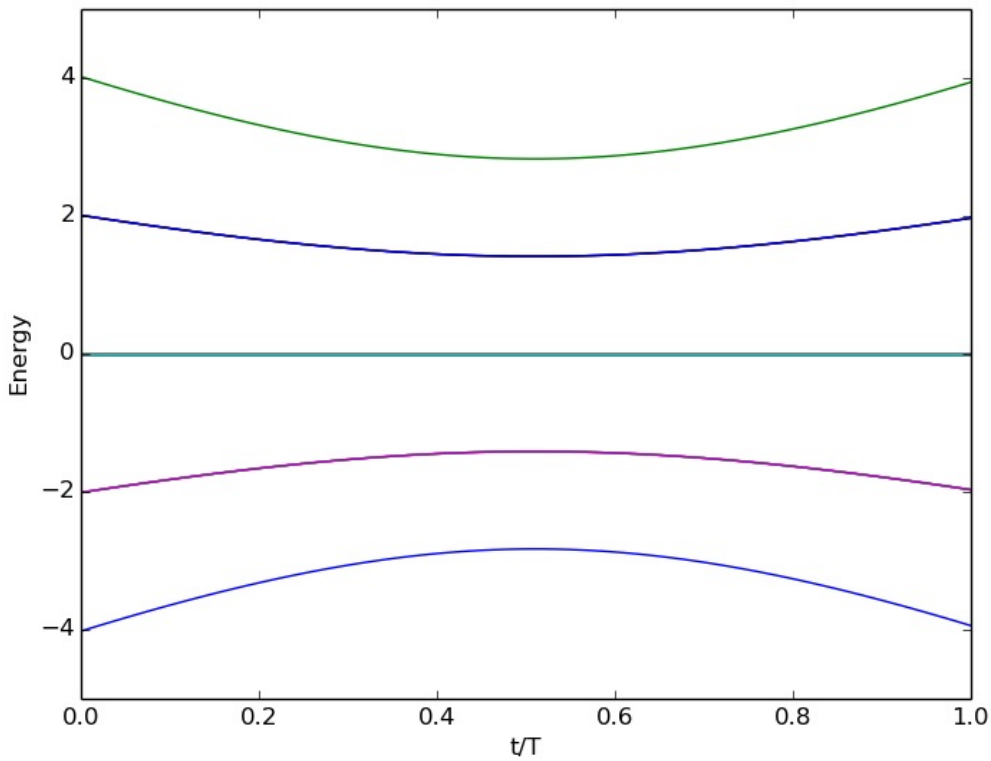

Figure 2. Time-dependent eigenspectrum for $p = 4$ orthogonal memories stored in a network of $n = 4$ neurons in the presence of bias. We define θ in terms of the first input memory and $\Gamma = 1$. For orthogonal memories, the spectrum is the same for the Hebb, Storkey, and projection learning rules.

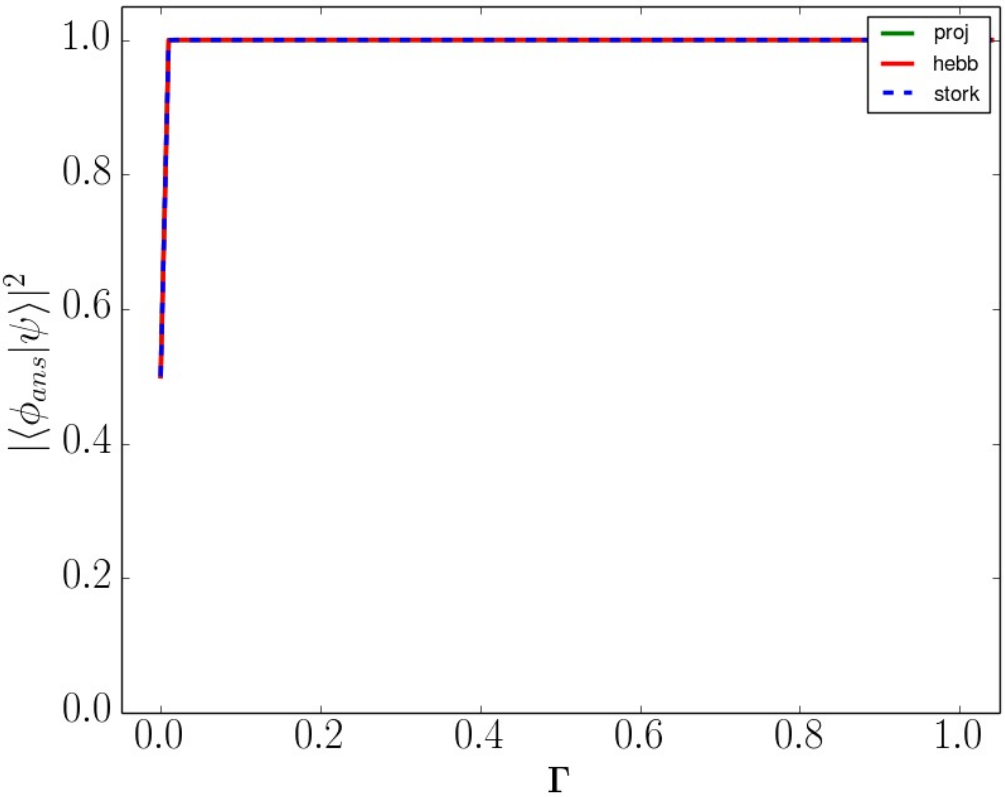

Figure 3. Recall probability with respect to applied bias Γ for $n = 4$ neurons and $p = 1$ memory from the set in Eq. (21).

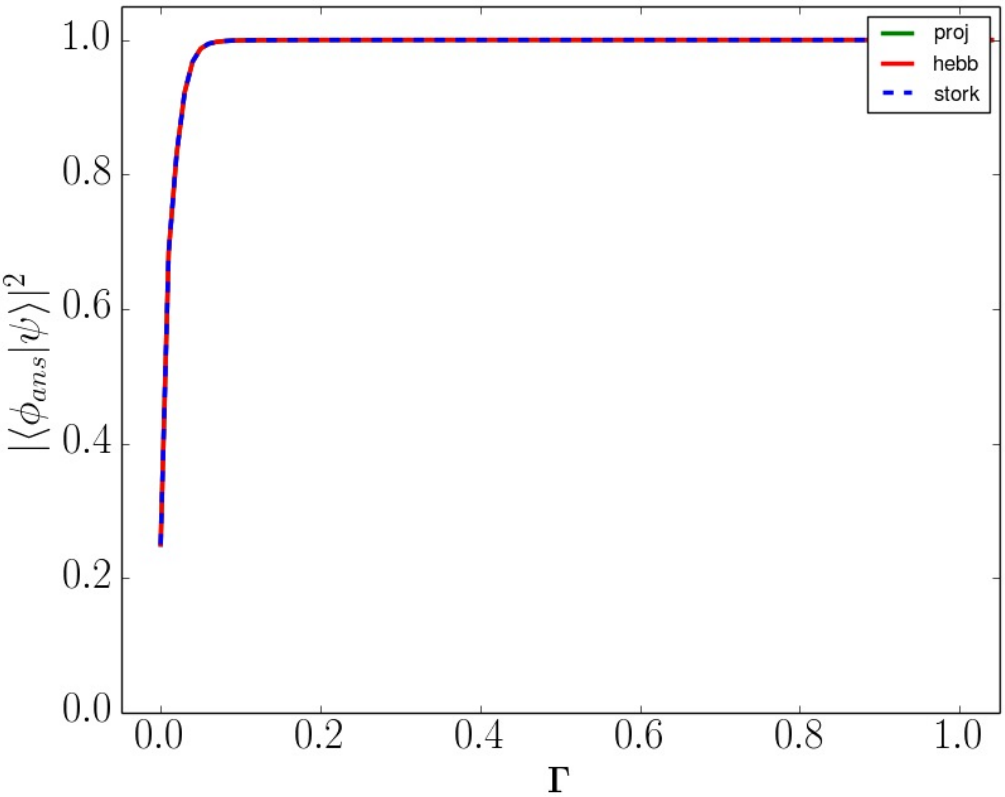

Figure 4. Recall probability with respect to applied bias Γ for $n = 4$ neurons and $p = 2$ memories from the set in Eq. (21).

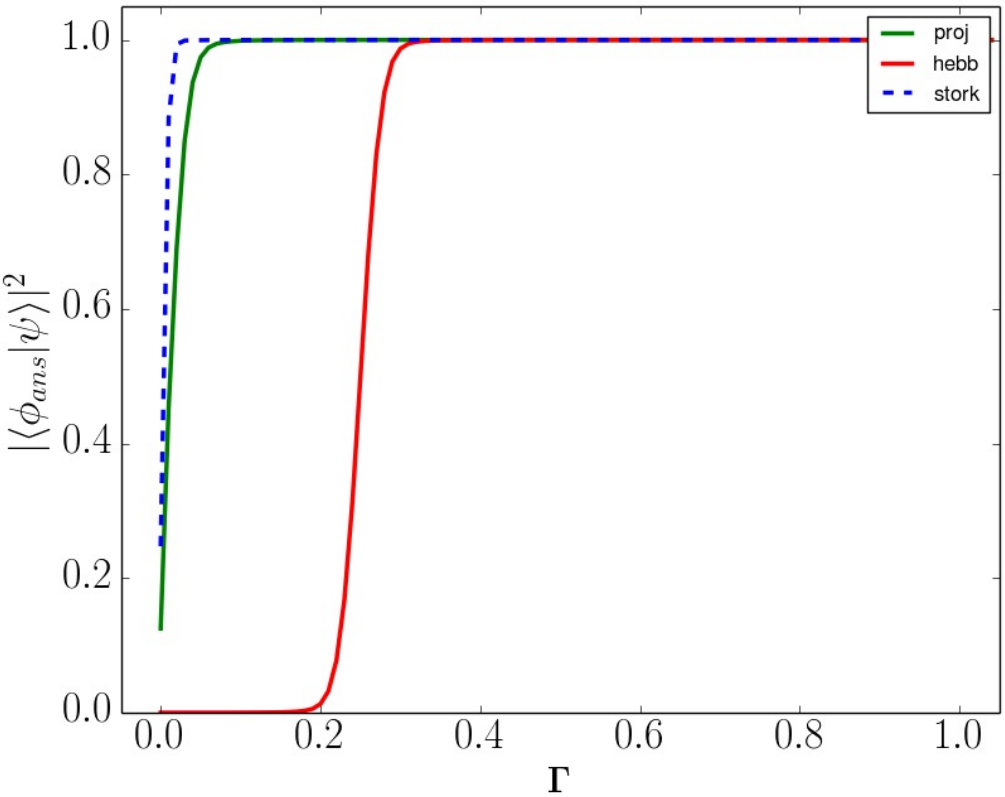

Figure 5. Recall probability with respect to applied bias Γ for $n = 4$ neurons and $p = 3$ memories from the set in Eq. (21).

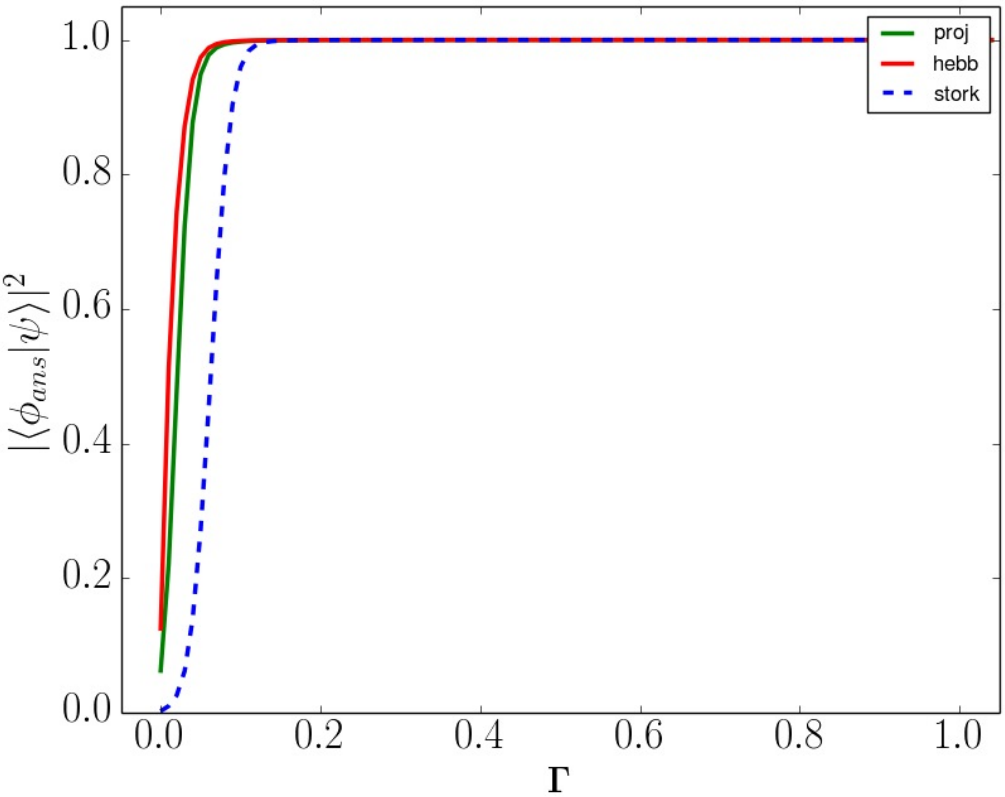

Figure 6. Recall probability with respect to applied bias Γ for $n = 4$ neurons and $p = 4$ memories from the set in Eq. (21).

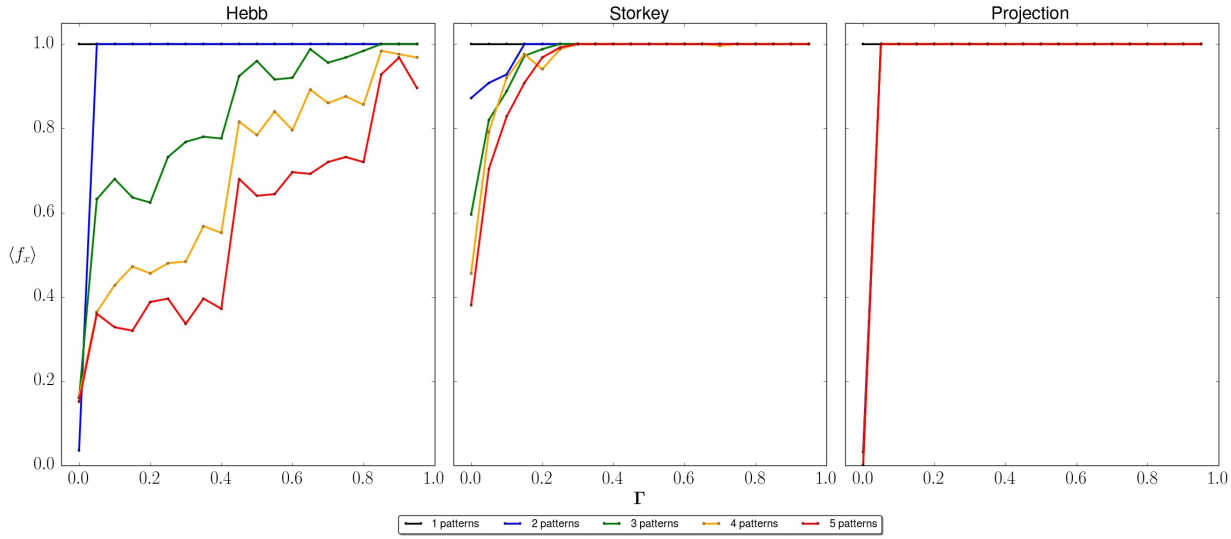

Figure 7. Recall success rate when the input is Hamming distance 0 away from a memory stored using $n = 5$ neurons.

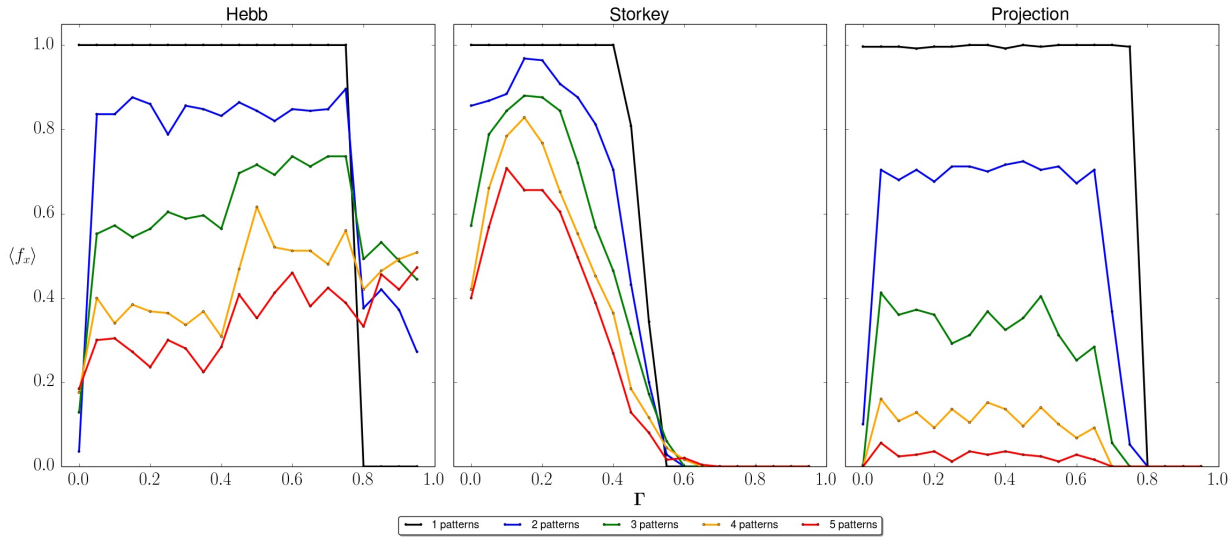

Figure 8. Recall success rate when the input is Hamming distance 1 away from a memory stored using $n = 5$ neurons.

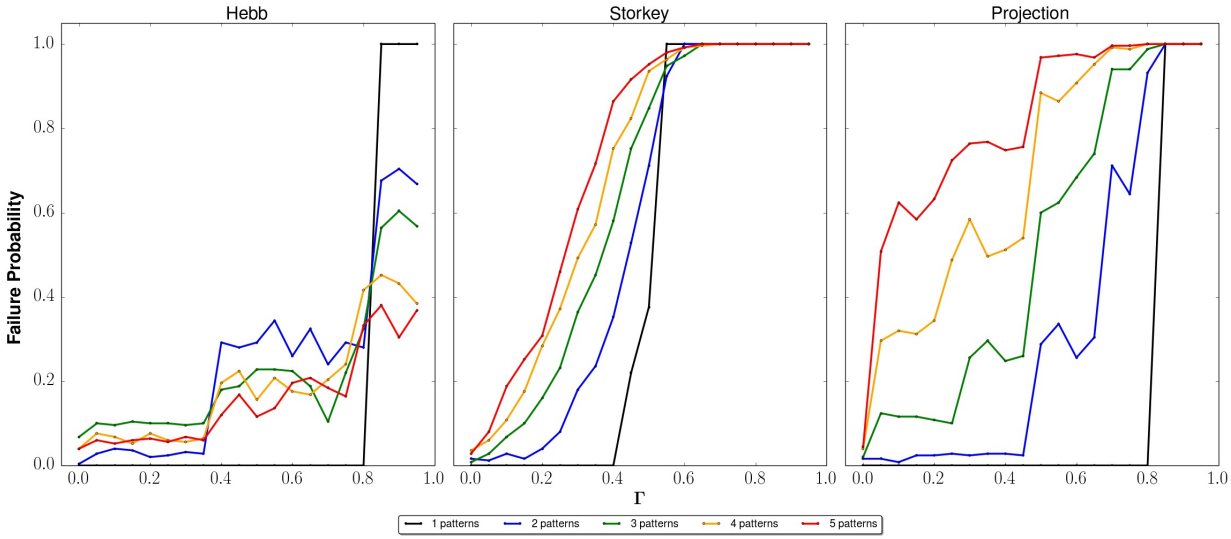

Figure 9. Recall failure rate when the input is Hamming distance 1 away from all stored memories with $n = 5$ neurons. The failure rate increases with Γ because the input state has formed a fixed point in the network.

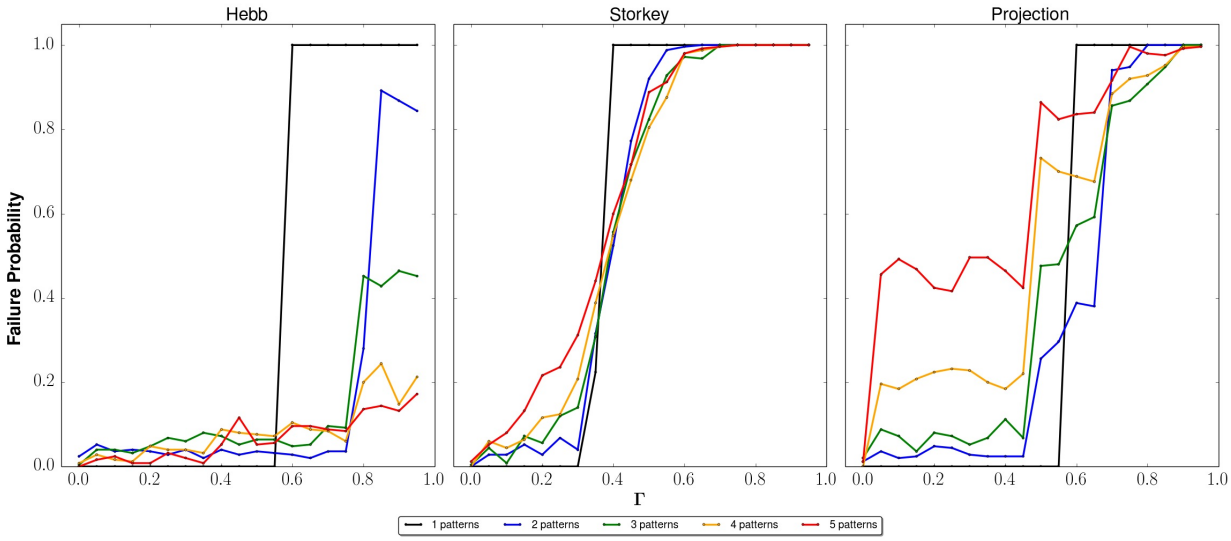

Figure 10. Recall failure rate when the input is Hamming distance 2 away from all stored memories with $n = 5$ neurons. The failure rate increases with Γ because the input state has formed a fixed point in the network.

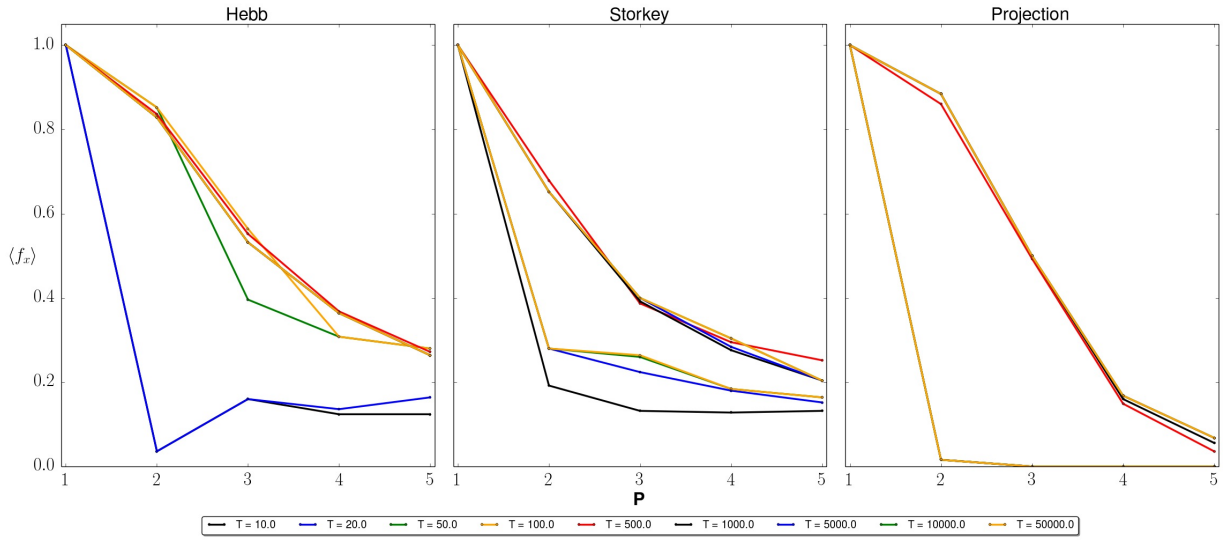

Figure 11. Recall success rate with increases in annealing time T. In all cases, small values of T suppress the success implying the adiabatic conditions has not been met. However, the recall success converges for sufficiently large values of T, an indication that the adiabatic condition is satisfied.