Cycle-Free Dynamics of a Cluster-Competitive Net

Arun Jagota† Anand Rangarajan§
Xin Wang‡

† Dept of Computer Science, University of California Santa Cruz, Santa Cruz, CA, jagota@cse.ucsc.edu

‡ Dept. of Diagnostic Radiology, Yale University, New Haven, CT, anand@noodle.med.yale.edu

‡ Computer Science Dept, UCLA, Los Angeles, CA, xwang@cs.ucla.edu

Abstract

In 1993, Waugh and Westervelt invented a cluster-competitive neural network that generalizes the continuous-state Hopfield model, admitting analog winner-take-all and Potts networks as special cases, and analyzed the global stability of discrete-time and continuous-time parallel dynamics for it. The discrete-time dynamics was shown to converge to fixed points or limit two-cycles. This paper presents a new discrete-time dynamics for the Waugh-Westervelt model, which generalizes both their discrete-time and continuous-time dynamics. In sequential mode the new dynamics is shown to be cycle-free. In parallel mode, the new dynamics is shown to be cycle-free under a positive-definiteness condition on a matrix derived from the weight matrix W, the neuron gains μ_k, and the dynamics step sizes α_k. Although Waugh and Westervelt also obtained a condition for cycle-free operation of their discrete-time version, our condition, which holds for our discrete-time version, is less restrictive. We also describe a preliminary form of our result, which extends the W-W result of global convergence to fixed points or two-cycles to parallel operation of our dynamics. Finally, we specialize our main result to the Potts glass, explicitly working out a milder sufficient condition for cycle-free operation in parallel mode for this important special case.

1 Introduction

For an n-neuron recurrent network, a much-studied and widely-used continuous-state symmetric neural network model is the leaky integrator model [4]. The dynamics of this model has been well studied, both in its continuous-time and in its discrete-time versions [4, 5, 8]. Recently, Waugh and Westervelt introduced a generalized version of this model in which neurons are arranged in clusters, and the overall activity in a cluster is maintained a constant [9]. This generalization is quite powerful, admitting, in addition to the original model [4], the Q-state Potts model and analog winner-take-all networks as special cases. The Potts model is a Q-state extension of the two-state spin model and is suited to Q-state versions of the associative memory and optimization applications of the spin model [9, 6]. To summarize, the W-W cluster-competitive model is a powerful mathematical generalization of the original model with potentially useful applications.

Waugh and Westervelt conducted a global stability analysis of a continuous-time and a certain discrete-time parallel dynamics for this model, showing that the continuous-time version is globally convergent to fixed points, and the discrete-time version to fixed points or to period-two limit cycles [9]. Because convergence to two-cycles is troublesome in applications of this network, Waugh and Westervelt gave a condition for cycle-free operation of their discrete-time dynamics, in terms of quantities called *cluster gains*.

This paper introduces a new discrete-time dynamics for this model, which generalizes both the continuous-time and the discrete-time dynamics of Waugh and Westervelt. Global analysis of cycle-free operation of this new dynamics is conducted. In sequential operation, this dynamics is shown to be cycle-free. In parallel operation, the dynamics is shown to be cycle-free when a certain matrix formed from W, neuron gains μ_j, and Euler-step sizes α_j of the discrete-time dynamics is positive definite.

Our cycle-free result under parallel operation is analogous to the cycle-free result of Waugh and Westervelt for their discrete-time version [9]. The condition in Waugh and Westervelt's result involves quantities called *cluster gains*, derived from neuronal gains; our condition involves Euler-step sizes. Our result therefore holds for arbitrary neuronal gains (by choosing appropriate Euler-step sizes) whereas theirs does not.

This paper also describes a preliminary form of our result which extends the W-W result of global convergence to fixed points or two-cycles to parallel operation of our dynamics.

Finally, we specialize our main result to the Potts glass, obtaining an improved (i.e., milder) sufficient condition for cycle-free operation in synchronous mode for this important special case. Our conditions are much less restrictive than those of the original W-W model.

Our results expand the applications of the W-W (in particular, Potts) models by eliminating one of its main restrictions.

2 The Waugh-Westervelt Cluster-Competitive Network

In this section, we review the W-W model in our terminology, and the relevant results from [9]. In the W-W model, the set V of all neurons is partitioned into m sets $V_1, \ldots, V_m$ called *clusters*, containing $n_1, \ldots, n_m$ elements respectively. In the continuous-time version, all neurons are updated simultaneously according to

$$\frac{dx_j(t)}{dt} = -x_j(t) + F_j(h_j(t) + B_i(t)) \tag{1}$$

where $h_j(t) = \sum_{k \in V} w_{jk} x_k(t) + I_j$ is the input coming into neuron j. Neuron j is assumed to belong to cluster V_i. Note that neuron j can get input from neurons outside its cluster V_i. The quantity I_j is a constant bias to neuron j. The quantity $F_j(.)$ is the activation function of neuron j. Through out this paper the weight matrix $W = (w_{kl})$ is assumed to be symmetric.

The CT model (1) differs from the original CT model [4] only in the quantity $B_i(t)$, which is a time-dependent bias—the same for every neuron in V_i—whose role is to enforce the competitive constraint that $\sum_{j \in V_i} x_j(t) = R_i$ for a given constant R_i for cluster V_i.

In the discrete-time version in [9], all neurons are updated simultaneously according to

$$x_j(t+1) = F_j(h_j(t) + B_i(t)) \tag{2}$$

where $B_i(t)$ is now chosen to enforce the constraint that $\sum_{j \in V_i} x_j(t+1) = R_i$.

Sufficient conditions on activation functions F_j to ensure that $B_i(t)$ has a unique solution are that F_j must be (i) continuous, (ii) monotone increasing (or decreasing), and (iii) for all possible values of h_j, R_i must lie within the range of the function $S_i(z) = \sum_{j \in V_i} F_j(h_j + z)$. In addition, to ensure boundedness of solutions to (1) and (2) respectively, F_j must asymptotically increase less than linearly for either very negative values of their arguments, or for very positive values of their arguments, or for both.

2.1 A Useful Special Case

The following example, taken from [9], presents a useful special case of the W-W model. This example uses the DT version (2). Set $R_i := 1$ for all i. Use $F_j(z) = e^{\gamma z}$. Then one can find $B_i(t)$ so that

$$x_j(t+1) := \frac{e^{\gamma h_j(t)}}{\sum_{k \in V_i} e^{\gamma h_k(t)}} \tag{3}$$

For $\gamma \to \infty$, this implements a $|V_i|$-state Potts model. (A Q-state Potts neuron is a generalized version of the 2-state spin neuron which takes Q real-valued inputs and produces a Q-valued output [9].) For finite γ this implements an analog winner-take-all, also called softmax, on cluster V_i. At finite γ, the r.h.s. of (3) describes the thermally averaged state of a Potts neuron at temperature $T = 1/\gamma$, much like the usual sigmoidal function describes the thermally averaged state of a spin neuron.

The choice of $B_i(t)$ that works, obviously known to the authors but not shown explicitly in [9], is easily verified to be the following. Choosing $B_i(t) = -\frac{1}{\gamma}\log\sum_{j\in V_i} e^{\gamma h_j(t)}$ gives

$$x_j(t+1) = F(h_j(t) + B_i(t)) = e^{\gamma[h_j(t)+(-1/\gamma)\log\sum_{k\in V_i} e^{\gamma h_k(t)}]} = \frac{e^{\gamma h_j(t)}}{\sum_{k\in V_i} e^{\gamma h_k(t)}} \tag{4}$$

2.2 Brief Review of Stability Analysis

It is useful to briefly review the global stability analysis of the CT version (1) and the DT version (2) from [9], as the results we obtain are analogous to these ones. Indeed our proofs closely follow the structure of the proofs in [9], and are best understood by first examining the original proofs.

Proposition 1 *The CT version (1) is globally convergent to fixed points only.*

Proof. The proof is based on the following energy function, earlier used in [4, 5]

$$L(t) = -\frac{1}{2}\sum_{k,l\in V} w_{kl}x_k(t)x_l(t) - \sum_{k\in V} I_k x_k(t) + \sum_{k\in V} G_k(x_k(t)) \tag{5}$$

where, recall that V is the entire set of neurons, and $G_k(x) = \int_{x_0}^{x} F_k^{-1}(z)dz$ with x_0 an arbitrary constant.

□

Let β_i denote the gain of cluster V_i, as defined in [9]. The actual expression for β_i is complicated; however an upper bound is the maximum gain of a neuron in cluster V_i.

Proposition 2 *The DT version (2) is cycle-free for sufficiently small cluster gains β_i.*

Proof. The proof is again based on the energy function L, given by (5). □

Proposition 3 *The DT version (2) is globally convergent to fixed points or to period-two limit cycles.*

Proof. The proof uses the following energy function, used in [5], and a similar one used in [1]

$$E(t) = -\sum_{k,l\in V} w_{kl}x_k(t)x_l(t-1) - \sum_{k\in V} I_k[x_k(t) + x_k(t-1)] + \sum_{k\in V}[G_k(x_k(t)) + G_k(x_k(t-1))] \tag{6}$$

□

All these proofs use the following key idea [5, 9]: for any k and any numbers x and y

$$G_k(x) - G_k(y) \le G_k'(x)(x-y) \tag{7}$$

with equality holding only when $x = y$. This inequality holds because the functions G_j are concave up, since F_j are monotonically increasing. It may be verified by a simple geometric argument. The proof of Proposition 2 uses a stronger version of inequality (7), which holds only when the neuron gains are bounded. Let μ_j denote the gain of neuron j, i.e. the maximum value of F_j'. Then, inequality (7) may be generalized to [5]

$$G_k(x) - G_k(y) \le G_k'(x)(x-y) - \frac{1}{2\mu_k}[x-y]^2 \tag{8}$$

which may be verified using Taylor's approximation theorem and noting that $G_k''(z) \ge 1/\mu_k$ for any z.

3 A New Discrete-Time Version

This paper introduces a new discrete-time dynamics for the W-W model. Neurons are updated according to

$$x_j(t+1) = (1-\alpha_j)x_j(t) + \alpha_j F_j(h_j(t) + B_i(t)) \tag{9}$$

where $B_i(t)$ is chosen, as in (2), to enforce the constraint that $\sum_{j \in V_i} x_j(t+1) = R_i$. The quantities α_j are newly introduced parameters taking values in $(0, 1]$. If all neurons are updated simultaneously according to (9), the dynamics is said to be *parallel.* If only one neuron is updated, according to (9), at any one time, the dynamics is said to be *sequential.* In sequential dynamics, the neuron chosen for updating is assumed to be selected *fairly*: in any infinite sequence of chosen neurons, each neuron appears infinitely-often.

The new DT model (9) is based on an Euler discretization of the CT model (1), admitting the latter model as a special case when $\alpha_j \to 0$ for all $j \in V$, and the DT model (2) as a special case when $\alpha_j = 1$ for all $j \in V$. Neuron states are kept bounded because $\alpha_j \in (0, 1]$.

This new model is inspired by an analogous model studied in [8], for the CT and DT versions of the original non-competitive model [4], and shares similar motivations. These are:

The W-W cluster-competitive model inherits the associative memory and optimization applications of the base model (see [3, 4, 2]), in extended form. For example, the Q-state Potts model, a special case of the W-W model, has applications to Q-state associative memory problems and to Q-state optimization problems [6].

In computer simulations, in computational applications, or in specialized implementations of the CT model (1), it is the DT version (9)—and not (2)—that serves as the most natural discrete-time approximation to the CT one, because by controlling α_j one may approximate the CT model arbitrarily well.

As reviewed in the previous section, the DT model (2) admits 2-cycles during its operation. Such 2-cycles are trouble-some from the point of view of the associative memory and optimization applications of this model (see the analogous argument in [8]). In both these application areas, one wishes to retrieve fixed points only [2]. Though 2-cycles may be eliminated from (2) by appropriate choice of cluster gains [9], it is more natural to eliminate 2-cycles by choosing the step sizes α_j sufficiently small in the DT model (9).

First we describe a preliminary version of an extension of Proposition 3, which holds for the DT version (9). The statement includes a forward reference to a condition in the proof.

Proposition 4 *With choices of the neuron gains μ_k and step sizes α_k that satisfy (11), the DT version (9) is globally asymptotically convergent to fixed points or to period-two limit cycles.*

Proof. The proof is based on an appropriate generalization of the energy function E, given by (6), used in the analogous proof of Proposition 3. The generalized energy function, with a somewhat similar one described in [8] in an analogous situation, is

$$E2(t) = -\sum_{k,l \in V} w_{kl} x_k(t) x_l(t-1) - \sum_{k \in V} I_k[x_k(t) + x_k(t-1)] + \sum_{k \in V} \alpha_k [G_k(y_k(t)) - G_k(y_k(t-1))] \tag{10}$$

where $y_k(t) = x_k(t-1) + \frac{\Delta x_k(t-1)}{\alpha_k}$ and $\Delta x_k(t-1) \equiv x_k(t) - x_k(t-1)$. E2 reduces to E when $\alpha_k = 1$ for all k.

Following [9] in a similar situation, the one-step change $\Delta E2(t) = E2(t+1) - E2(t)$ is

$$\Delta E2(t) \quad = \quad -\sum_{k \in V} [\sum_{l \in V} w_{kl} x_l(t) + I_k] \Delta_2 x_k(t) + \sum_{k \in V} \alpha_k [G_k(y_k(t+1)) - G_k(y_k(t-1))]$$

Once again, following [9], we subtract and add

$$\sum_{i=1}^{m} B_i(t) \sum_{j \in V_i} \Delta_2 x_j(t) = \sum_{i=1}^{m} B_i(t)[\sum_{j \in V_i} x_j(t+1) - \sum_{j \in V_i} x_j(t-1)] = \sum_{i=1}^{m} B_i(t)[R_i - R_i] = 0$$

giving

$$\Delta E2(t) \quad = \quad -\sum_{k \in V}[\sum_{l \in V} w_{kl} x_l(t) + I_k + B_i(t)]\Delta_2 x_k(t) + \sum_{k \in V} \alpha_k [G_k(y_k(t+1)) - G_k(y_k(t-1))]$$

Now from (9), $\sum_{l \in V} w_{kl} x_l(t) + I_k + B_i(t) = G'_k(x_k(t) + \frac{\Delta x_k(t)}{\alpha_k}) = G'_k(y_k(t+1))$. Using this, and inequality (8) gives

$$\begin{aligned}
\Delta E2(t) \quad &\leq \quad -\sum_{k \in V} G'_k(y_k(t+1))\Delta_2 x_k(t) + \sum_{k \in V} \alpha_k \{G'_k(y_k(t+1))[\frac{\Delta_2 x_k(t)}{\alpha_k} - (1-\alpha_k)\frac{\Delta_2 x_k(t-1)}{\alpha_k}] - \\
&\qquad \frac{1}{2\mu_k}[y_k(t+1) - y_k(t-1)]^2\} \\
&= \quad -\sum_{k \in V}[(1-\alpha_k)G'_k(y_k(t+1))\Delta_2 x_k(t-1) + \frac{1}{2\mu_k \alpha_k}\{\Delta_2 x_k(t) - (1-\alpha_k)\Delta_2 x_k(t-1)\}^2]
\end{aligned}$$

Therefore if

$$-\sum_{k \in V}[(1-\alpha_k)G'_k(y_k(t+1))\Delta_2 x_k(t-1) + \frac{1}{2\mu_k \alpha_k}\{\Delta_2 x_k(t) - (1-\alpha_k)\Delta_2 x_k(t-1)\}^2] \leq 0 \qquad (11)$$

then $\Delta E2(t) \leq 0$. Assuming (11) holds, when is $\Delta E2(t) = 0$? By inspecting where the inequality (8) is used in the proof, we see that this happens only when $\Delta_2 y_k(t) = 0$, which in turn implies that $\Delta_2 x_k(t) = (1-\alpha_k)\Delta_2 x_k(t-1)$. For $\alpha_k = 1$ this implies that $\Delta_2 x_k(t) = 0$ where as for $\alpha_k < 1$ this implies that $\lim_{t \to \infty} \Delta_2 x_k(t) = 0$. In either case, asymptotic convergence to fixed points or two-cycles is implied. □

When is the condition (11) of Proposition 4 satisfiable? First note that the second term inside the summation is always nonnegative. Hence, when $\alpha_k = 1$ for all k, (11) holds, showing that the proof gives the result of [9] in this case. When $\alpha_k < 1$, the first term might possibly have a negative sign. However, for any fixed $\alpha_k \geq 0$ the first term has a finite lower bound, since the variable x_k is bounded (implying that $G'_k(y_k(t+1))$ is bounded). Therefore, for fixed α_k's, if all the μ_k's are sufficiently small, (11) holds again. Similarly, since α_k appears in the denominator of the second term in the summation, for fixed μ_k's if all the α_k's are sufficiently small, (11) holds once again.

We now show that two-cycles are absent if the DT version (9) is operated sequentially.

Proposition 5 *If $w_{jj} > -(2-\alpha_j)/(\alpha_j \mu_j)$ for all j, the DT version (9) is globally convergent to fixed points under sequential operation.*

Proof. Use the energy function (5). For the DT version (9), the one-step change $\Delta L(t) = L(t+1) - L(t)$ is

$$\Delta L(t) = -\frac{1}{2}\sum_{k,l \in V} w_{kl}\Delta x_k(t)\Delta x_l(t) - \sum_{k \in V}(\sum_{l \in V}(w_{kl} x_l(t)) + I_k)\Delta x_k(t) + \sum_{k \in V}(G_k(x_k(t+1)) - G_k(x_k(t)))$$

Like before, subtracting and adding $\sum_{i=1}^{m} B_i(t) \sum_{j\in V_i} \Delta x_j(t) = 0$ gives

$$\begin{aligned}\Delta L(t) &= -\frac{1}{2}\sum_{k,l\in V} w_{kl}\Delta x_k(t)\Delta x_l(t) - \sum_{k\in V}(\sum_{l\in V}(w_{kl}x_l(t)) + I_k + B_i(t))\Delta x_k(t) + \sum_{k\in V}(G_k(x_k(t+1)) - G_k(x_k(t)) \\ &= -\frac{1}{2}\sum_{k,l\in V} w_{kl}\Delta x_k(t)\Delta x_l(t) - \sum_{k\in V}(G'_k(x_k(t) + \frac{\Delta x_k(t)}{\alpha_k}))\Delta x_k(t) + \sum_{k\in V}(G_k(x_k(t+1)) - G_k(x_k(t))\end{aligned}$$

Using (8) and rearranging gives

$$\Delta L(t) \le -\frac{1}{2}\sum_{k,l\in V} w_{kl}\Delta x_k(t)\Delta x_l(t) - \sum_{k\in V}[G'_k(x_k(t) + \frac{\Delta x_k(t)}{\alpha_k}) - G'_k(x_k(t+1))]\Delta x_k(t) - \sum_{k\in V}\frac{1}{2\mu_k}[\Delta x_k(t)]^2$$

Now using the fact that $G'_k(y_0 + y_1) - G'_k(y_0) \ge y_1/\mu_k$ for any y_0, y_1 we get

$$[G'_k(x_k(t) + \frac{\Delta x_k(t)}{\alpha_k}) - G'_k(x_k(t+1))] \ge \frac{(1-\alpha_k)\Delta x_k(t)}{\alpha_k\mu_k}$$

for all k. Therefore

$$\Delta L(t) \le -\frac{1}{2}\sum_{k,l\in V} w_{kl}\Delta x_k(t)\Delta x_l(t) - \sum_{k\in V}\frac{2-\alpha_k}{2\alpha_k\mu_k}[\Delta x_k(t)]^2$$

If at time t only neuron k is chosen to update its state (and all the others remain unchanged), then

$$\Delta L(t) \le -\frac{1}{2}w_{kk}[\Delta x_k(t)]^2 - \frac{2-\alpha_k}{2\alpha_k\mu_k}[\Delta x_k(t)]^2 \le 0, \tag{12}$$

with equality holding only when $\Delta x_k(t) = 0$. Hence, any sequential trajectory tends to some fixed point. □

Proposition 5 holds even if the neuron gains μ_k are unbounded, provided that $w_{kk} \ge 0 \; \forall k$.

The next proposition provides a condition that eliminates 2-cycles from the DT version (9), under parallel operation.

Proposition 6 *In parallel operation, the DT version (9) is globally convergent to fixed points if the matrix*

$$W + (2I - A)A^{-1}M^{-1} \quad \text{is positive definite,} \tag{13}$$

where $A = diag(\alpha_k)$ and $M = diag(\mu_k)$ are diagonal matrices formed by α_k and μ_k.

Proof. Use the energy function $L(t)$ used in the proof of Proposition 5. ¿From (12), the one-step difference $\Delta L(t)$ of $L(t)$ along any trajectory $x(t)$ satisfies

$$\Delta L(t) \le -\frac{1}{2}\Delta x(t)^\top(W + (2I - A)A^{-1}M^{-1})\Delta x(t). \tag{14}$$

where $\top$ denotes the transpose operation. Condition (13) gives $\Delta L(t) \le 0$, with equality holding only when $\Delta x(t) = 0$. Hence the parallel dynamics is globally convergent to fixed points. □

Proposition 6 holds even if the neuron gains μ_k are unbounded, with the condition (13) replaced by the more restrictive condition "W is positive definite". It is instructive to compare Proposition 6 with Proposition 2, the analogous cycle-free result of Waugh and Westervelt. Their proof is a special case of ours. In particular their version of (14) is a special case of ours, with $\alpha_k = 1$. In this case, choosing the neuron gains μ_k to be sufficiently small allows their version of (14) to hold. The benefits of our result is that we do not need to restrict μ_k—rather we can trade-off the neuron gains μ_k with the step sizes α_k.

Next we prove an stronger version of Proposition 6 that holds for an important special case: the Potts model of Section 2.1. Recall that this model uses the exponential neuron transfer function with the unit simplex constraint (4). This Potts model is frequently used in solving optimization problems such as TSP [6] graph partitioning [6], quadratic assigment [7] and pairwise clustering [9]. This model may also be used to solve optimization problems on probability vectors (e.g. maximum likelihood computations).

Proposition 7 *For the exponential neuron $F_j(z) = e^{\gamma z}$, and the unit simplex constraint $R_i = 1$ for all clusters i, in parallel operation, the DT version (9) is globally convergent to fixed points if the matrix*

$$W + \frac{2A^{-1}}{\gamma} \text{ is positive definite,} \tag{15}$$

where $A = diag(\alpha_k)$ is a diagonal matrix formed by α_k.

Proof. Use the energy function $L(t)$ given by (5). ¿From the discrete-time update equation (9), the one-step difference $\Delta L(t)$ of $L(t)$ along any trajectory $x(t)$ satisfies

$$\begin{aligned}\Delta L(t) &= -\frac{1}{2}\sum_{k,l\in V} w_{kl}\Delta x_k(t)\Delta x_l(t) - \sum_{k\in V}[\sum_{l\in V}(w_{kl}x_l(t)) + I_k + B_i(t)]\Delta x_k(t) + \sum_{k\in V}[G_k(x_k(t+1)) - G_k(x_k(t))] \\ &= -\frac{1}{2}\sum_{k,l\in V} w_{kl}\Delta x_k(t)\Delta x_l(t) - \sum_{k\in V}[G'_k(x_k(t) + \frac{\Delta x_k(t)}{\alpha_k})]\Delta x_k(t) + \sum_{k\in V}[G_k(x_k(t+1)) - G_k(x_k(t))]\end{aligned}$$

We now use an improved version of inequality (8), which is also due to Waugh and Westervelt [9]:

$$\sum_{k=1}^{V_i}[G_k(x_k(t+1)) - G_k(x_k(t))] \le \sum_{k=1}^{V_i} G'_k(x_k(t+1))\Delta x_k(t) - \frac{1}{2\beta_i}\sum_{k=1}^{V_i}(\Delta x_k(t))^2$$

where the cluster gain β_i is obtained via the inequality

$$\frac{1}{\beta_i}\sum_{k=1}^{V_i}(\Delta x_k(t))^2 \le \sum_{k=1}^{V_i} G''_i(x_k(t))(\Delta x_k(t))^2$$

Given the definition of $F_k(z)$ in (4), we find that $G_k(x) = \frac{1}{\gamma}(x\log x - x)$ (up to an arbitrary additive constant). From this, we get $G''_k(x_k(t)) = \frac{1}{x_k(t)\gamma}$. β_i is obtained from

$$\frac{1}{\beta_i} = \min_{x_k(t),\Delta x_k(t)} \frac{\sum_{k=1}^{V_i}\frac{(\Delta x_k(t))^2}{x_k(t)\gamma}}{\sum_{k=1}^{V_i}(\Delta x_k(t))^2}$$

subject to the constraints $\sum_{k=1}^{V_i} x_k(t) = 1$, $\sum_{k=1}^{V_i}\Delta x_k(t) = 0$. As shown in Appendix A, this gives us a value of $\beta = \frac{\gamma}{2}$ (independent of the cluster index i). Using this value of β, and the above-mentioned W-W inequality, the change in energy is

$$\Delta L(t) \le -\frac{1}{2}\sum_{k,l\in V} w_{kl}\Delta x_k(t)\Delta x_l(t) - \sum_{k\in V}[G'_k(x_k(t) + \frac{\Delta x_k(t)}{\alpha_k}) - G'_k(x_k(t+1))]\Delta x_k(t) - \sum_{k\in V}\frac{1}{\gamma}[\Delta x_k(t)]^2$$

Now

$$G'_k(y_0 + y_1) - G'_k(y_0) \geq (\min_y G''_k(y))y_1 \geq y_1/\gamma$$

where the first inequality holds for any y_0, y_1 and the second follows from the fact that the domain of G'_k is restricted to the unit simplex $[0, 1]$. This yields

$$[G'_k(x_k(t) + \frac{\Delta x_k(t)}{\alpha_k}) - G'_k(x_k(t+1))] \geq \frac{(1-\alpha_k)\Delta x_k(t)}{\alpha_k \gamma}$$

for all k. Therefore

$$\Delta L(t) \quad \leq \quad -\frac{1}{2} \sum_{k,l \in V} w_{kl} \Delta x_k(t) \Delta x_l(t) - \sum_{k \in V} \frac{2}{2\alpha_k \gamma} [\Delta x_k(t)]^2$$

This can be rewritten in matrix notation as

$$\Delta L(t) \leq -\frac{1}{2} \Delta x(t)^\top (W + \frac{2A^{-1}}{\gamma}) \Delta x(t). \tag{16}$$

where $\top$ denotes the transpose operation. Condition (15) gives $\Delta L(t) \leq 0$, with equality holding only when $\Delta x(t) = 0$. Hence the parallel dynamics for this specialized case is globally convergent to fixed points when the matrix $W + \frac{2A^{-1}}{\gamma}$ is positive definite. □

Let us now examine how the condition of Proposition 7 is an improvement over the previous results. For simplicity let $\alpha_j = \alpha$ for all j. For this Potts model, the sufficient condition of Waugh and Westervelt for their dynamics is that $W + 2I/\gamma$ is positive definite [9]. The sufficient condition given by Proposition 7 is that $W + 2I/(\alpha\gamma)$ is positive definite [9]. Thus we may make this matrix positive definite not only by controlling γ but also by controlling (decreasing) the step size α.

The Proposition 7 condition is also a major improvement over the Proposition 6 condition (which holds for the general case). The main improvement is that the latter condition is expressed in terms of the neuron maximum slope μ which can be much larger than γ. (Since $F_j(h_j) = e^{\gamma h_j}, F'_j(h_j) = \gamma e^{\gamma h_j}$. Furthermore, since h_j is the *field* coming into neuron j, not the external value in $[0, 1]$, the value of h_j, though bounded, can be large. Thus $F'(.)$ can be large.) A secondary improvement is due to the fact that β, the cluster gain, is not γ but $\gamma/2$.

4 Conclusion

We have derived sufficient conditions for convergence of a new discrete-time dynamical system which generalizes Waugh and Westervelt's [9] cluster-competitive network. The new discrete-time dynamics yields, at its extremes, Waugh and Westervelt's discrete- and continuous-time dynamics. The new sufficient condition for cycle-free operation is milder than the original one the Waugh and Westervelt discrete-time network.

This is especially important in optimization problems where the Waugh and Westervelt conditions are known to be often overly restrictive. For instance in attempting to solve the Maximum Clique problem via a Potts encoding (based on the so-called Motzkin-Strauss quadratic program for Maximum Clique [10]) we found that the original W-W dynamics did not converge at all. The new dynamics, with sufficiently (but not excessively) small step sizes not only always converged, but also yielded good solutions in preliminary experiments. A thorough evaluation of the new dynamics to the Maximum Clique problem via this encoding is being addressed in [10].

References

[1] E. Goles, F. Fogelman-Soulie, and D. Pellegrin. Decreasing energy functions as a tool for studying threshold networks. *Discrete Applied Mathematics*, 12:261–277, 1985.

[2] J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation.* Addison-Wesley, 1991.

[3] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA*, 79, 1982.

[4] J.J. Hopfield. Neurons with graded responses have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences, USA*, 81, 1984.

[5] C.M. Marcus and R.M. Westervelt. Dynamics of iterated-map neural networks. *Physical Review A*, 40(1):501–504, 1989.

[6] C. Peterson and B. Söderberg. A new method for mapping optimization problems onto neural networks. *International Journal of Neural Systems*, 1:3–22, 1989.

[7] P. Simic. Constrained nets for graph matching and other quadratic assignment problems. *Neural Computation*, 3:268–281, 1991.

[8] X. Wang, A. Jagota, F. Botelho, and M. Garzon. Absence of cycles in symmetric neural networks. In D.S. Touretzky, M. Mozer, and M. Hasselmo, editors, *Proceedings of Ninth Neural Information Processing Systems Conference*, Cambridge:MA, 1996. MIT Press.

[9] F.R. Waugh and R.M. Westervelt. Analog neural networks with local competition. I. dynamics and stability. *Physical Review E*, 47:4524–4536, June 1993.

[10] A. Jagota, M. Pelillo and A. Rangarajan A new deterministic annealing algorithm for maximum cliques (in preparation), 1998.

A Proof of the cluster gain inequality

We have the following optimization problem:

$$\frac{1}{\beta} = \min_{z_k,\epsilon_k} \sum_{k=1}^{V_i} \frac{\epsilon_k^2}{z_k \gamma}$$
$$\text{subject to } \textstyle\sum_{k=1}^{V_i} z_k = 1, \ \sum_{k=1}^{V_i} \epsilon_k = 0, \text{ and } \sum_{k=1}^{V_i} \epsilon_k^2 = 1. \tag{17}$$

Using the method of Lagrange parameters, this can be converted to

$$\min_{z_k,\epsilon_k} \sum_{k=1}^{V_i} \frac{\epsilon_k^2}{z_k \gamma} + \lambda \left(\sum_{k=1}^{V_i} z_k - 1 \right)$$
$$\text{subject to } \sum_{k=1}^{V_i} \epsilon_k = 0, \text{ and } \sum_{k=1}^{V_i} \epsilon_k^2 = 1. \tag{18}$$

In (18), λ is a Lagrange parameter enforcing the constraint $\sum_{k=1}^{V_i} z_k = 1$.

Carrying out the minimization of (18) w.r.t. z alone keeping ϵ and λ fixed, we get $z_k = \frac{|\epsilon_k|}{\sqrt{\lambda}}$ and $\lambda = \left(\sum_{k=1}^{V_i} |\epsilon_k|\right)^2$. Substituting the expression for z above in (18), we find that we have to minimize the following objective function:

$$
\begin{aligned}
\frac{1}{\beta_i} &= \min_{\epsilon_k} \frac{\left(\sum_{k=1}^{V_i} |\epsilon_k|\right)^2}{\gamma \sum_{k=1}^{V_i} \epsilon_k^2} \\
\text{subject to } & \sum_{k=1}^{V_i} \epsilon_k = 0.
\end{aligned} \tag{19}
$$

Using a result from Appendix B, we get $\left(\sum_{k=1}^{V_i} |\epsilon_k|\right)^2 \geq 2 \sum_{k=1}^{V_i} \epsilon_k^2$ when $\sum_{k=1}^{V_i} \epsilon_k = 0$. From this, we obtain $\beta = \frac{\gamma}{2}$ (independent of the cluster index i).

B Proof of the inequality $\left(\sum_i |x_i|\right)^2 \geq 2 \sum_i x_i^2$ when $\sum_i x_i = 0$

We are required to show that $\left(\sum_i |x_i|\right)^2 \geq 2 \sum_i x_i^2$ where the $x_i,\ i \in \{1, \ldots, N+1\}$ are constrained such that $\sum_{i=1}^{N+1} x_i = 0$.

Since $\sum_{i=1}^{N+1} x_i = 0$, $x_{N+1} = -\sum_{i=1}^{N} x_i$. Hence

$$
\begin{aligned}
& \left(\sum_{i=1}^{N+1} |x_i|\right)^2 \geq 2 \sum_{i=1}^{N+1} x_i^2 \\
\Rightarrow\ & \left(\sum_{i=1}^{N} |x_i| + |\sum_{i=1}^{N} x_i|\right)^2 \geq 2 \left[\sum_{i=1}^{N} x_i^2 + \left(\sum_{i=1}^{N} x_i\right)^2\right] \\
\Rightarrow\ & \left(\sum_{i=1}^{N} |x_i|\right)^2 + \left(\sum_{i=1}^{N} x_i^2\right) + 2 \sum_{i=1}^{N} |x_i| |\sum_{i=1}^{N} x_i| \geq 2 \sum_{i=1}^{N} x_i^2 + 2 \left(\sum_{i=1}^{N} x_i\right)^2 \\
\Rightarrow\ & \sum_{i=1}^{N} x_i^2 + \sum_{ij,\ i \neq j} |x_i x_j| + 2 \sum_{i=1}^{N} |x_i| |\sum_{i} x_i| \geq 2 \sum_{i=1}^{N} x_i^2 + \left(\sum_{i=1}^{N} x_i\right)^2 \\
\Rightarrow\ & \sum_{ij,\ i \neq j} |x_i x_j| + 2 \sum_{i=1}^{N} |x_i| |\sum_{i} x_i| \geq \sum_{i=1}^{N} x_i^2 + \left(\sum_{i=1}^{N} x_i\right)^2.
\end{aligned} \tag{20}
$$

Since $\sum_{i=1}^{N} |x_i| \geq |\sum_{i=1}^{N} x_i|$, it is sufficient to show that

$$
\begin{aligned}
& \sum_{ij,\ i \neq j} |x_i x_j| + 2 \left(\sum_{i=1}^{N} x_i\right)^2 \geq \sum_{i=1}^{N} x_i^2 + \left(\sum_{i=1}^{N} x_i\right)^2 \\
\Rightarrow\ & \sum_{ij,\ i \neq j} \left(|x_i x_j| + x_i x_j\right) \geq 0
\end{aligned} \tag{21}
$$

which is always true. Therefore, $\left(\sum_{i=1}^{N+1} |x_i|\right)^2 \geq 2 \sum_{i=1}^{N+1} x_i^2$ when $\sum_{i=1}^{N+1} x_i = 0$.