Self-optimization in a Hopfield neural network based on the *C. elegans* connectome

Alejandro Morales[1*] and Tom Froese[1,2]

[1]Institute for Applied Mathematics and Systems Research, National Autonomous University of Mexico, Mexico
[2]Center for the Sciences of Complexity, National Autonomous University of Mexico, Mexico
[*]Corresponding author: alejandroe@ciencias.unam.mx

Abstract

It has recently been demonstrated that a Hopfield neural network that learns its own attractor configurations, for instance by repeatedly resetting the network to an arbitrary state and applying Hebbian learning after convergence, is able to form an associative memory of its attractors and thereby facilitate future convergences on better attractors. This process of structural self-optimization has so far only been demonstrated on relatively small artificial neural networks with random or highly regular and constrained topologies, and it remains an open question to what extent it can be generalized to more biologically realistic topologies. In this work, we therefore test this process by running it on the connectome of the widely studied nematode worm, *C. elegans*, the only living being whose neural system has been mapped in its entirety. Our results demonstrate, for the first time, that the self-optimization process can be generalized to bigger and biologically plausible networks. We conclude by speculating that the reset-convergence mechanism could find a biological equivalent in the sleep-wake cycle in *C. elegans*.

Introduction

Caenorhabditis elegans is a one-millimeter-long soil worm of the nematode family. It consists of only 959 cells, of which 302 belong to the nervous system. *C. elegans* is a case study model for biology (Girard et al., 2006), thanks to its small cell count, short lifespan and rapid reproduction. The full understanding of cellular development and the function of neural system mechanisms in human biology remains an open question, which makes *C. elegans* a good alternative. *C. elegans* is the first multicellular organism whose genome has been sequenced in its entirety (Waterston and Sulston, 1998), as well as the first animal from which the complete mapping of synaptic connections, called the connectome, has been completed (White et al., 1986). Knowledge of the connectivity of a complete neural system can help to better understand genetic and molecular mechanisms in neuroscience.

Network science provides further information concerning the structure of the *C. elegans* connectome, such as the statistical and topological properties of the network (Watts and Strogatz, 1998; Varshney et al., 2011). In terms of worm's

network dynamics, simulation models have begun to play a bigger role in the experiment-theory cycle in an attempt to better understand the neural underpinnings of worm behavior (Izquierdo, 2018). *C. elegans* has been studied also in the artificial life research community (Kitano et al., 1998; Winkler et al., 2009; Hattori et al., 2012; Izquierdo and Beer, 2015; Beer and Izquierdo, 2016; Aguilera et al., 2017). However, so far these simulation models have only worked with smaller circuits, and not with the whole connectome, and so the properties of the connectome's state space remain underexplored.

In this paper we therefore go the other way: we abstract away from the worm's situated behavior, and focus on the dynamics of its whole connectome. In particular, we are interested in determining if the connectome is able to self-optimize its connectivity, namely by forming an associative memory of its attractors such that the convergence of neural states is biased toward better attractors.

The self-optimization process

The self-optimization process used in this paper is based on the work of Watson et al. (2011), using the original network model proposed by Hopfield (1982). We have two discrete values for the states in the network $s_i = \pm 1$. Such values represent the activity of a neuron. Model dynamics include asynchronous neuron state updates, calculating with the following equation:

$$s_i(t+1) = \theta \left[\sum_j^N w_{ij} s_j(t) \right] \quad (1)$$

where w_{ij} is the connection weight between neuron i and j, and θ is the Heaviside threshold function (taking values -1 and $+1$ as negative and positive arguments, respectively).

In order to be able to test the connectome's convergent states with respect to its original connectivity, we differentiate two parts of w_{ij}, w_{ij}^O represents the original configuration of the edges at the beginning of the process, and w_{ij}^L store the Hebbian learning changes. Both make up the cur-

rent weights of the network:

$$w_{ij} = w_{ij}^O + w_{ij}^L \qquad (2)$$

In the Hopfield network model, a node i satisfy a constraint posed by its interaction with node j if $s_i s_j w_{ij} > 0$. While actual interaction with neighbors determines the state update s_i, system energy represents the degree to which internal constraints with the original weight configuration, w^O, remain unsatisfied. It is calculated as follows:

$$E = -\sum_{ij}^N w_{ij}^O(t) s_i(t) s_j(t) \qquad (3)$$

Hebbian learning should be applied to all system connections (i.e. the change in weight, $\Delta w_{ij} = \delta s_i s_j, \delta > 0$) to increase the attractor variation and reinforce learning.

$$w_{ij}^L(t+1) = w_{ij}^L(t) + \delta s_i(t) s_j(t) \qquad (4)$$

The algorithm repeatedly goes through the following sequence of steps: (a) arbitrary assignment of states for the neurons (reset), (b) convergence of the network for a certain time period, most frequently resulting in an attractor, and (c) application of Hebbian learning.

Methods

The connectome. We ran the self-optimization algorithm on the most recent version of the *C. elegans* connectome produced by Jarrell et al. (2012). The database contains hermaphrodite neural system information (males arise infrequently, at 0.1%), such as synaptic direction, type of synaptic connection (synapse or gap junction), and the number of synapses between A and B neurons. We translated the connectome into a directed multigraph, with neurons as nodes and synapses as edges. This representation allows multiple synapses and gap junctions between the same two neurons, which naturally occur in the worm neural system.

Only 282 neurons belonging to a large somatic nervous system are taken into account. We did not consider pharyngeal neurons because they belong to another independent neural system (Albertson and Thomson, 1976; White et al., 1986). Neurons were arbitrarily assigned binary activation states $(-1, 1)$. In the neural network under study, chemical synapses are modeled as single-directed links between neurons (for example, $A \rightarrow B$ indicates that neuron A is presynaptic to neuron B, and B is postsynaptic to A). Gap junctions are represented in the connectome as double-linked neurons (if two neurons, C and D, have a gap junction between them, there are two links: $C \rightarrow D$ and $D \rightarrow C$). There are 5,611 connections in total. 62.5% of the total connections are chemical synapses, while 37.5% are gap junctions.

We assigned the number of synapses between neurons as the weight of each edge normalized in the interval $(0, 1)$.
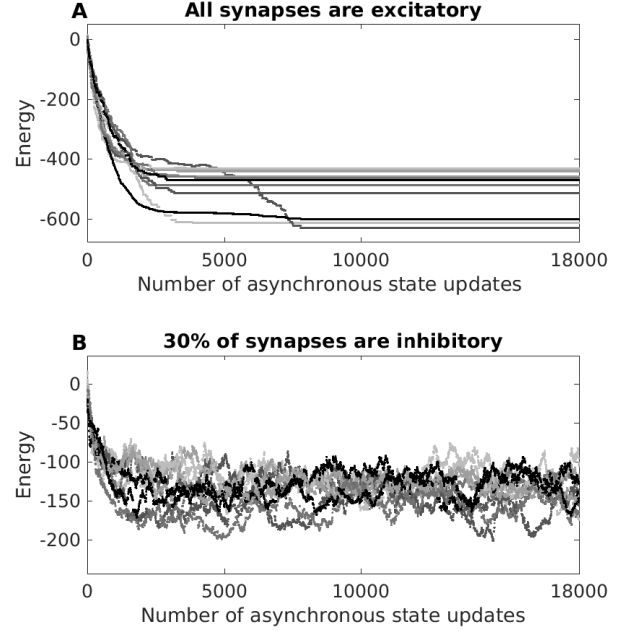

Figure 1: Illustrative examples of network state convergence without self-optimization. We show network energy after successive state updates for 10 independent convergences. **A** Scenario without inhibitory connections (negative weights). **B** Scenario with 30% inhibitory connections (negative weights). The addition of inhibitory connections increases the difficulty of coordinating neural activity across the connectome. Note the different y-scales when comparing panels A and B.

Jarrell et al. (2012) estimate the functional strength of synaptic interactions with the resulting number. Both links in gap junctions are assigned the same weight, and values vary between 1 and 81 (5.07 is the average weight per link). We clip to 1 the 15 high-weight values, which we determined with an arbitrary cut-off of weights greater than 44, before normalization. Edges with higher weight values are scarce (only 34 connections has a higher value than 44). Reduction of this outliers broadens the space-state exploration during the self-optimization; in other words, the range value of local attractors is wider.

Since the connectome does not contain information concerning inhibitory or excitatory connections, we tested two separate scenarios. First, a scenario in which all synapses are excitatory, and second, a scenario in which a percentage of synapses is inhibitory. For the second scenario we arbitrarily selected 30% of the edges and assigned negative weights to them, based on the fact that 30% is a percentage of inhibitory synapses that has been proven optimal in the
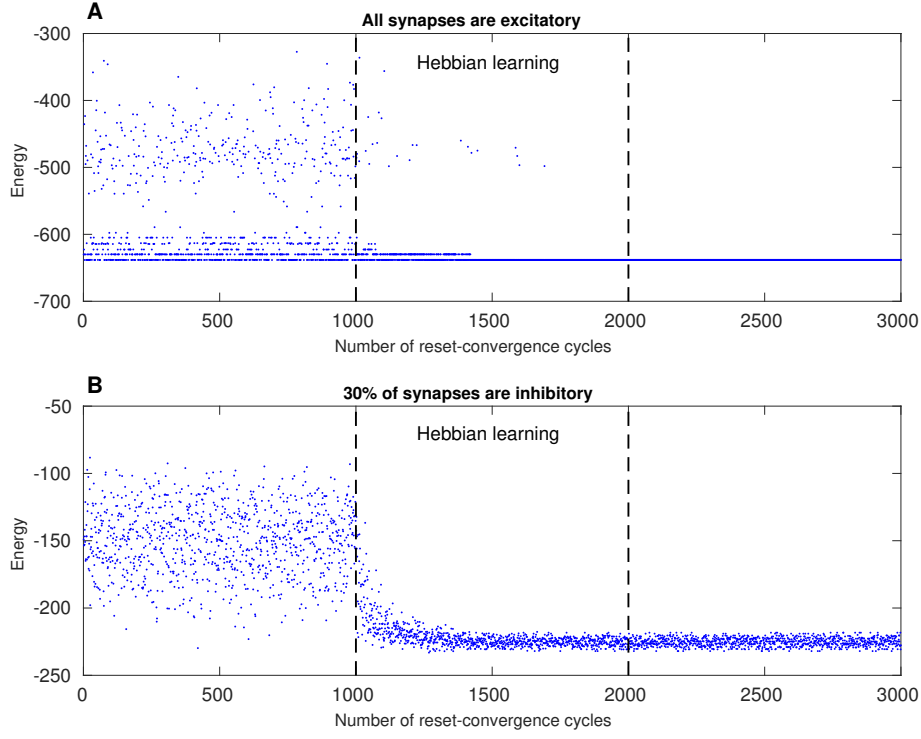
Figure 2: Illustrative examples of self-optimization in two different scenarios. **A** without inhibitory connections (negative weights) at the beginning of the process, and **B** with 30% inhibitory neurons (negative weights). Each panel shows the energy of the neuron states after 1,000 reset-convergence cycles during three distinct phases: before learning (1-1,000), during the self-optimization process (1,001-2,000), and after learning (2,001-3,000). Examples of state convergences of the phase before learning can be seen in Figure 1. Self-optimization can be observed in both scenarios. However, the processes can be distinguished in that better attractors tend to be found without inhibitory synapses, but inhibitory synapses preserve a greater diversity of attractors.

simulation of multi-task learning processes, at least for mammals (Capano et al., 2015).

The neural network. Once the connectome was mapped to a neural network, extra zero-weighted edges were added to make a complete directed graph so that Hebbian learning could add extra connections by changing their weights from zero to non-zero values. In other words, we added previously non-existing connections between neurons in two directions, and those extra edges represent potential connections, which may be needed for the removal of constraints regarding interactions during the process (constraints posed in terms of the satisfaction equation, $s_i s_j w_{ij} > 0$) . There is an initial total of 5,611 connections in the network (from the connectome), which after adding extra edges rises to 80,213. These extra edges do not have anatomical equivalents in the connectome, but they could be conceived of in terms of functional rather than structural connectivity. We measured and reported the energy in terms of the original weight configu-

ration (5611 connections from the connectome). When the learning phase ends, the constraint satisfaction percentage is also based in this 5611 connections.

In a multigraph schema we must consider an identifier (k) for each multiple edge that shares the same direction. In light of this, the weight sum of all edges with the same direction was added to the state update equation (Eq. 1), rewritten as follows:

$$s_i(t+1) = \theta \left[\sum_j^N \left(\sum_k w_{ijk} \right) s_j(t) \right] \qquad (5)$$

A multigraph schema is also reflected in the way we calculate system energy and Hebbian learning (Eq. 3 and 4, respectively), because we iterate above all edges including the multiple ones that share the same direction:

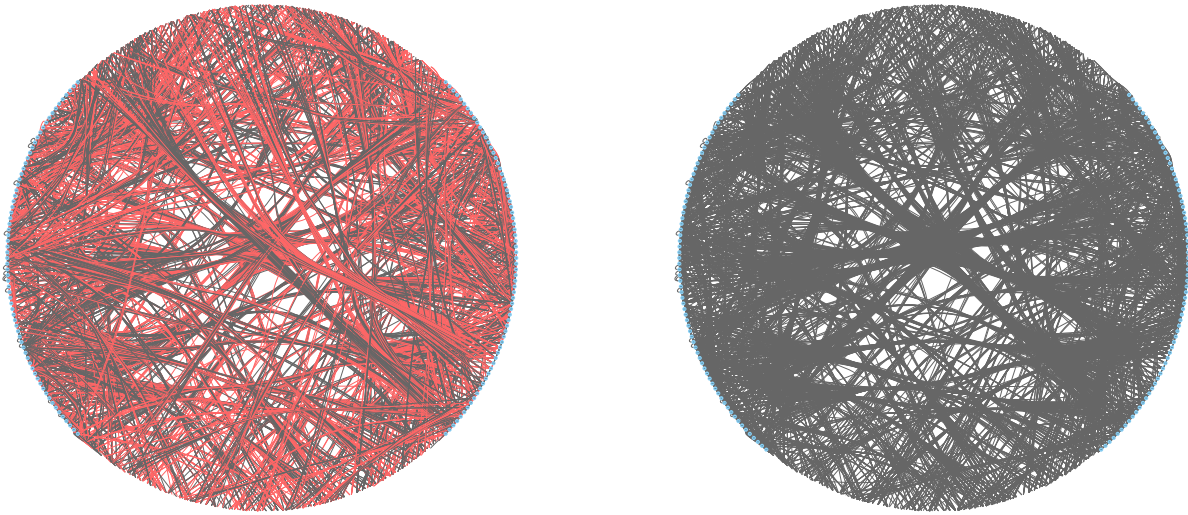$$E = -\sum_{ijk}^N w_{ijk}^O(t) s_i(t) s_j(t) \qquad (6)$$

Figure 3: Visualization of the *C. elegans* connectome, consisting in 282 neurons with 5,611 excitatory connections, in a circle layout with a topological bundling procedure. Black connections represent edges that satisfy the constraints that were specified by the networks original weight configuration. Red ones represent edges that do not satisfy these constraints. We present two illustrative examples. The figure on the left shows the connectome with arbitrary states at the beginning of the self-optimization procedure, and the figure on the right shows the connectome with converged states after learning. Complete constraint satisfaction can be observed on the right side.

$$w_{ijk}^L(t+1) = w_{ijk}^L(t) + \delta s_i(t)s_j(t) \qquad (7)$$

In summary, the network resulting from the connectome data differs from a traditional Hopfield neural network in the following ways:

- Asymmetrical connections are permitted.

- Self-connections are permitted.

- Since there are synapses and gap junctions between the same neurons within the connectome, multiple directed connections are permitted.

The first two connectivity restrictions were already lifted by previous models of self-optimization (Zarco and Froese, 2018a; Froese and Manzanilla, 2018), but this is the first time that the process is tested with a multigraph.

The self-optimization algorithm repeatedly goes through the following sequence of steps: (a) arbitrary assignment of states for the neurons (reset), (b) convergence of the network for a certain time period, most frequently resulting in an attractor, and (c) application of Hebbian learning.

The number of steps for (b) has been adjusted to $18,000$. We observe this to be an adequate quantity to ensure either stability or convergence in each cycle. We fixed the learning rate δ ($\delta = 0.00001$) for all experiments.

Results

Overall, the results demonstrate the feasibility of self-optimization on the connectome. We report the results of both the scenario that starts with all positive weights, and the scenario with 30% negative weights.

We first explored the attractor dynamics of the network without self-optimization. Figure 1 A,B shows the network energy after successive neuron state updates. The network always reaches an attractor before the next state reset in the scenario with all excitatory synapses (Panel A). Unsatisfied network constraints also tend to decrease in the scenario with inhibitory synapses, but the network does not always reach an attractor (Panel B). This shows that the addition of inhibitory synapses increases the difficulty of neural coordination.

We then explored the network's self-optimization capacity. The experiment shown in Figure 2 A,B consisted in three stages. First, we set an initial weight configuration taken from the connectome and normalized (only positive values in A and 30% negative values in B) and performed 1000 reset convergence cycles without Hebbian learning. Then, we applied the self-optimization process using 1000 reset-convergence cycles. Finally, we apply the network 1000 cycles without Hebbian learning using the configuration obtained by the self-optimization process. This graph illustrates a tendency to decrease global energy in both A and B, although a global attractor is not reached in B.
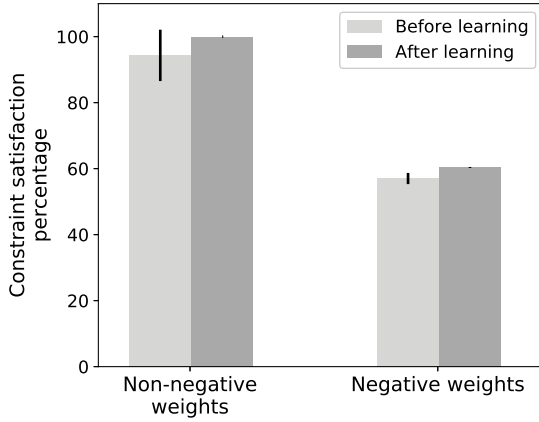
Figure 4: Percentage of constraint satisfaction before and after self-optimization (Hebbian learning), measured at the end of a reset-convergence cycle. For each case we consider the average of 1,000 cycles. A t-test was used for each data set (without negative weights and 30% negative weights) to see if the difference between before and after learning is significant. We obtain a p-value $<<< 0.05$ both for non-negative weights and 30% negative weights, indicating a highly significant improvement.

To illustrate the size and complexity of the connectome, we present in Figure 3 a diagram which shows the connectome structure with a circular neuron layout where the edges of different colors represent the satisfied and unsatisfied connections.

Figure 4 shows the percentage of constraint satisfaction before and after the self-optimization process. We applied a t-test to each data set to compare before and after learning.

Discussion

In the present work, we have tested the dynamics of the connectome of *C. elegans* with a Hopfield network and applied the self-optimization procedure in a multigraph structure. The results demonstrate that self-optimization can be generalized to this bigger and biologically more plausible topology. However, it is an open question to what extent this success has implications for understanding the operation of the nervous system of the worm. We respond to this open question by considering three issues.

First, is the connectivity required for the self-optimization process biologically plausible? We showed that under the right conditions the connectome network presents a tendency to optimize its own connectivity. It does so through Hebbian learning-based removal of constraints during repeated exploration of its state space. In the original Hopfield neural network model, this was achieved by restricting connections in such a way that the network state would nec-

essarily always converge on point attractors, namely by ensuring that connections are symmetric and that there are no self-recurrent connections. The restriction to point attractors and symmetric, non-recurrent connectivity is biologically implausible. An important step is therefore to show that self-optimization can still be achieved on an asymmetric, multiedged, recurrent network, which can give rise to a richer set of dynamics. We achieved this step with the present work.

Second, is the reset-convergence cycle required for the self-optimization process biologically plausible? In particular, what could be the mechanism that periodically relaxes the constraints of normal functioning, which permits the network to explore its state space? Following speculations by Woodward et al. (2015), in our model the neuron state resets could be interpreted biologically as a sleep-like state of *C. elegans*, especially when the worm is quiescent. *C. elegans* presents a quiescent behavior during lethargus, a sleep-like stage that occurs during larval development (Nelson and Raizen, 2013). Moreover, this sleep-like state also occurs during satiety and after exhaustion. Nevertheless, the way to define a wake-sleep cycle in the worm remains controversial. Our model suggests that it would be worthwhile to look for it, and to test its relationship to network self-optimization, perhaps akin to sleep-dependent learning.

Third, is the percentage and distribution of inhibitory connections biologically plausible? Again, we are limited by the lack of more detailed biological information. We introduced negative weights in the second scenario of the self-optimization process in an exploratory manner. The results suggest that their presence restricts the coordination between neurons. Such a restriction of neural coordination across the whole connectome could be biologically desirable, for instance by helping to avoid problems arising from excessive neural synchrony, such as in disorders like epilepsy. Another advantage is that there is an increased diversity of better attractors found at the end of self-optimization with inhibitory connections. In this way we overcome worries raised by Zarco and Froese (2018b) that convergence on single attractors could be limiting for applying the self-optimization process in cognitive robotics, which typically require the possibility to switch between multiple attractor configurations.

Future work

Further improvements to our current model can be made, especially because the real neurons of the worm tend to activate in a continuous rather than binary manner. In future work, we will therefore test the self-optimization connectome model with the dynamics of a continuous time recurrent neural network (CRTNN), which should in principle be possible (Zarco and Froese, 2018b). It would also be interesting to see what happens if we do not provide extra virtual or functional connections to the worm's connectome, and restrict Hebbian changes to the original anatomical connec-

tome only. Another possibility is to analyze neural coordination in terms of local clusters of neurons that are anatomically related (Nonet, 1999), rather than measuring success in terms of neural coordination across the entire connectome. Finally, further mathematical analysis of our model can be made, for instance by taking inspiration from the techniques employed by other Hopfield neural network models of sleep (Fachechi et al., 2019).

References

Aguilera, M., Alquézar, C., and Izquierdo, E. J. (2017). Signatures of criticality in a maximum entropy model of the *C. elegans* brain during free behaviour. In Knibbe, C., Beslon, G., Parsons, D., Misevic, D., and Rouzaud-Cornabas, J., editors, *ECAL 2017: The Fourteenth European Conference on Artificial Life*, pages 29–35, Lyon, France. MIT Press.

Albertson, D. G. and Thomson, J. N. (1976). The pharynx of *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 275(938):299–325.

Beer, R. D. and Izquierdo, E. J. (2016). Propagation of rhythmic dorsoventral wave in a neuromechanical model of locomotion in *Caernohabditis elegans*. In Gershenson, C., Froese, T., Siqueiros, J. M., Aguilar, W., Izquierdo, E., and Sayama, H., editors, *ALIFE 2016: The Fifteenth International Conference on the Synthesis and Simulation of Living Systems*, pages 544–545, Cancún, Mexico. MIT Press.

Capano, V., Herrmann, H. J., and de Arcangelis, L. (2015). Optimal percentage of inhibitory synapses in multi-task learning. *Scientific Reports*, 5:9895. doi:10.1038/srep09895.

Fachechi, A., Agliari, E., and Barra, A. (2019). Dreaming neural networks: forgetting spurious memories and reinforcing pure ones. *Neural Networks*, 112:24–40.

Froese, T. and Manzanilla, L. R. (2018). Modeling collective rule at ancient teotihuacan as a complex adaptive system: Communal ritual makes social hierarchy more effective. *Cognitive Systems Research*, 52:862–874.

Girard, L. R., Fiedler, T. J., Harris, T. W., Carvalho, F., Antoshechkin, I., Han, M., Sternberg, P. W., Stein, L. D., and Chalfie, M. (2006). Wormbook: the online review of *Caenorhabditis elegans* biology. *Nucleic Acids Research*, 35:D472–D475. doi:10.1093/nar/gkl894.

Hattori, Y., Suzuki, M., Soh, Z., Kobayashi, Y., and Tsuji, T. (2012). Modeling of the pharyngeal muscle in *Caenorhabditis elegans* based on FitzHugh-Nagumo equations. *Artificial Life and Robotics*, 17(2):173–179.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8):2554–2558.

Izquierdo, E. J. (2018). Role of simulation models in understanding the generation of behavior in *C. elegans*. *Current Opinion in Systems Biology*, 13:93–101.

Izquierdo, E. J. and Beer, R. D. (2015). An integrated neuromechanical model of steering in *C. elegans*. In Andrews, P., Kaves, L., and Doursat, R., editors, *ECAL 2015: The Thirteenth European Conference on Artificial Life*, pages 199–206, York, UK. MIT Press.

Jarrell, T. A., Wang, Y., Bloniarz, A. E., Brittin, C. A., Xu, M., Thomson, J. N., Albertson, D. G., Hall, D. H., and Emmons, S. W. (2012). The connectome of a decision-making neural network. *Science*, 337(6093):437–444.

Kitano, H., Hamahashi, S., and Luke, S. (1998). The perfect *C. elegans* project: An initial report. *Artificial Life*, 4(2):141–156.

Nelson, M. D. and Raizen, D. M. (2013). A sleep state during *C. elegans* development. *Current Opinion in Neurobiology*, 23(5):824–830.

Nonet, M. L. (1999). Visualization of synaptic specializations in live *C. elegans* with synaptic vesicle protein-GFP fusions. *Journal of Neuroscience Methods*, 89(1):33–40.

Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H., and Chklovskii, D. B. (2011). Structural properties of the *Caenorhabditis elegans* neuronal network. *PLOS Computational Biology*, 7(2):1–21.

Waterston, R. and Sulston, J. E. (1998). The Human Genome Project: Reaching the Finish Line. *Science*, 282(5386):53–54. doi:10.1126/science.282.5386.53.

Watson, R. A., Mills, R., and Buckley, C. L. (2011). Global adaptation in networks of selfish components: Emergent associative memory at the system scale. *Artificial Life*, 17(3):147–166.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440.

White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society, Biological Sciences*, 314(1165):1–340. doi:10.1098/rstb.1986.0056.

Winkler, D. A., Burden, F. R., and Halley, J. D. (2009). Predictive mesoscale network model of cell fate decisions during *C. elegans* embryogenesis. *Artificial life*, 15(4):411–421.

Woodward, A., Froese, T., and Ikegami, T. (2015). Neural coordination can be enhanced by occasional interruption of normal firing patterns: A self-optimizing spiking neural network model. *Neural Networks*, 62:39 – 46.

Zarco, M. and Froese, T. (2018a). Self-modeling in Hopfield neural networks with continuous activation function. *Procedia Computer Science*, 123:573–578.

Zarco, M. and Froese, T. (2018b). Self-optimization in continuous-time recurrent neural networks. *Frontiers in Robotics and AI*, 5:96. doi:10.3389/frobt.2018.00096.