# Attractor network

Chris Eliasmith (2007), Scholarpedia, 2(10):1380. doi:10.4249/scholarpedia.1380 revision #91016 [link to/cite this article]

- **Dr. Chris Eliasmith**, Center for Theoretical Neuroscience, University of Waterloo, Waterloo, Ontario

In general, an attractor network is a network of nodes (i.e., neurons in a biological network), often recurrently connected, whose time dynamics settle to a stable pattern. That pattern may be stationary, time-varying (e.g. cyclic), or even stochastic-looking (e.g., chaotic). The particular pattern a network settles to is called its 'attractor'. In theoretical neuroscience, different kinds of attractor neural networks have been associated with different functions, such as memory, motor behavior, and classification. Describing networks as attractor networks allows researchers to employ methods of dynamical systems theory to quantitatively analyze their characteristics (e.g. stability, robustness, etc.).

More precisely, an attractor network is a set of $N$ network nodes connected in such a way that their global dynamics becomes stable in a $D$ dimensional space, where usually $N>>D$. This assumes that there is no additional external input, and that *stability* indicates that the network state resides, over time, on some $D$-manifold (e.g., line, circle, plane, toroid, etc.).

Contents

# Kinds of attractor network

It is important to keep two spaces distinct when discussing attractor networks. These are:

- the network state space (which is determined by the set of all possible node states); and
- the attractor space (which is a subspace of the network state space that only includes points on an attractor).

These are not always distinguished. The subsequent discussion attempts to be clear about which is being described.

## Point attractors

The simplest attractor network is one which tends to a single stable point (fixed point) given any starting activity, and is called a 'point attractor' (see Figure 1).

A simple example of a one-dimensional point attractor network, whose fixed point is 0 is given by any network described by an equation of the form

$$\dot{x}(t) = kx(t)$$

for $\infty < k < 0$

Evidently, any *D*-dimensional damped linear system will have a single fixed point at zero. This is a somewhat trivial case, and of limited interest (unless the system is forced).

More interesting are networks with multiple fixed points (see Figure 2). If an attractor network has multiple point attractors, the set of points that results in movement to a given fixed point is called that fixed point's *basin of attraction*.

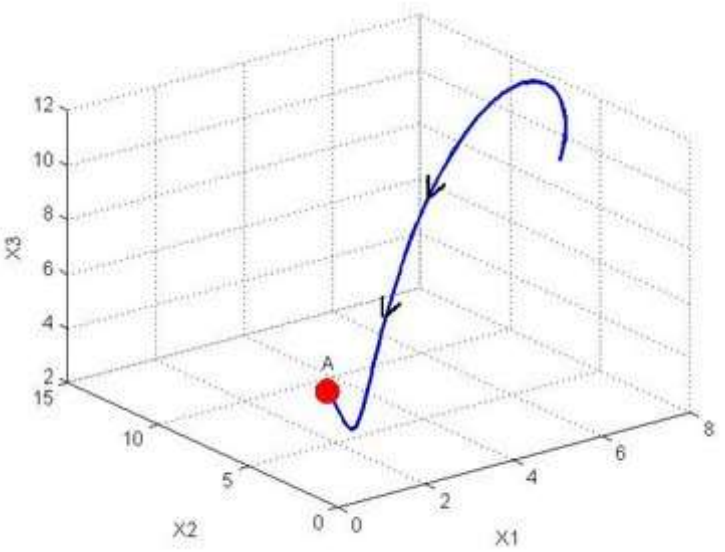


Figure 1: An example of a point attractor embedded in a 3-dimensional network. The fixed point in the network is labelled 'A'. The movement of the network in its state space over time is shown by the black arrows.

## Line, ring, and plane attractors

Line attractors are a natural extension to point attractors. Rather than the attractive states being a finite set of fixed points, it is an infinite set of points, all of which lie on a line in the state space (see Figure 3). The particular point towards which the network moves depends on the initial conditions (starting point) of the network.

As a simple example, any network with neurons $n_i$ with $i$ = 1,..., $N$ which has an attractor manifold $A$ defined by

$$A(x) = \mathbf{b} + x\mathbf{c} \text{ for } x \in \Re \text{ and } \mathbf{b}, \mathbf{c} \in \Re^N$$

is a line attractor network. Note that the attractor space itself is only one-dimensional, regardless of *N*, the dimension of the network state space.

Given such a manifold definition, we can write the dynamics of the state space variables (e.g., in 3 dimensions)

$$\dot{x_1} = 0$$

$$\dot{x_2} = -(x_2 - (mx_1 + b))$$

$$\dot{x_3} = -x_3$$

The first dimension allows the system to be stable on the line, the second defines the shape of the line, and the third collapses higher dimensions exponentially quickly to the subspace of interest. To get behavior similar to that shown in Figure 3, $m = 3$ and $b = -18$ .

The line may also not be straight as defined above. If the ends of the line meet, the attractor is often called a *ring attractor*.

If the attractor is allowed to be *D*-dimensional, where 1<*D*<*N*, the network is said to implement a plane attractor instead of a line attractor.

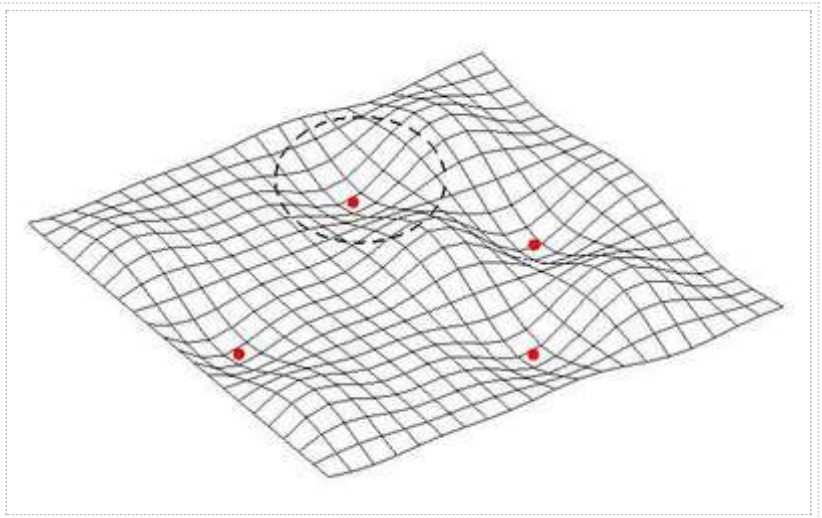

Figure 2: The energy landscape of a network with multiple point attractors (e.g. a Hopfield network). Fixed points are shown as red dots. A sample basin of attraction is shown as a dotted circle.

Notably, when implemented in a neural network, each of these kinds of attractors is usually *approximated*. That is, most neural network implementations consist of a number of point attractors organized to approximate a line, ring, or plane. As a result, an important part of understanding attractor networks is to characterize the goodness of the approximation and/or the number and organization of distinct points in the attractor space.

## Cyclic attractors

Point, line, and plane attractors all result in a network settling to a given point in state space (which may depend on initial conditions), which it does not move from without external input. However, it is possible to have a set of states that a network continuously and repeatedly traverses, which is called a *limit cycle*. Networks that have these kinds of attractors are called *cyclic attractors* (see Figure 4).

As a simple example, any network that has an attractor manifold defined by a simple harmonic oscillator will be a cyclic attractor

$$\ddot{x}(t) = -Ax(t)$$

Again *x* is defined over the attractor space and is usually of much lower dimension than the number of nodes in the network (here it is one-dimensional).

## Chaotic attractors

Like cyclic attractors, chaotic attractors (aka *strange attractors*) have stable manifolds that are continuously traversed. However, the manifolds are generally of fractional dimension, and can thus be non-repeating, though bounded. A common example of a chaotic attractor is the Lorenz attractor defined by

$$\dot{x}_1 = A(x_2 - x_1)$$
$$\dot{x}_2 = x_1(B - x_3) - x_2$$
$$\dot{x}_3 = x_1 x_2 - C x_3.$$

When the constants are set to $A$=10, $B$ = 32, $C$ = 8/3 the attractor is chaotic (see Figure 5). Any network that has attractor space dynamics described by these equations (within a certain parameter range) implements a chaotic attractor in its (much higher dimensional) state space.

A simple way to construct arbitrary chaotic attractors directly in the network state space is to generate randomly connected networks of nodes (Sompolinsky et al., 1988).

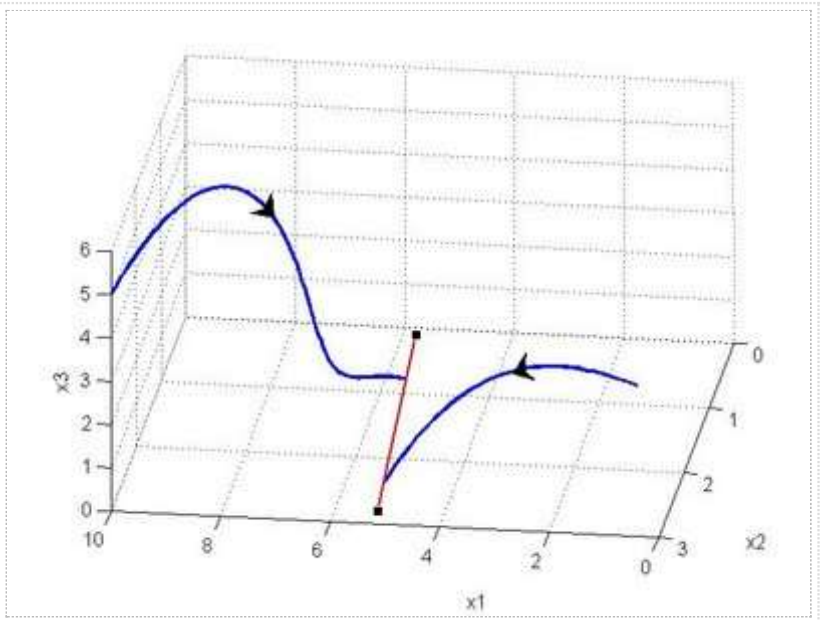

Figure 3: An example of a line attractor embedded in a 3-dimensional network. Two possible trajectories from different starting points are shown in blue. The attractor manifold is shown in red.

## Biological interpretations of attractor networks

Because neural networks can implement any nonlinear dynamical system, they can implement any attractor network. Of greatest interest to computational neuroscientists is determining which attractors are relevant for understanding information processing in biological systems.

Stable, persistent activity has been thought to be important for neural computation at least since Hebb (1949), who suggested that it may underlie short-term memory. Amit (1989), following work on attractors in artificial neural networks, suggested that persistent neural activity in biological networks is a result of dynamical attractors in the state space of recurrent biological networks. This seminal work resulted in attractor networks becoming a mainstay of theoretical neuroscience. Often, these biologically inspired models have adopted non-biological nodes (e.g., sigmoid response functions, or *rate neurons*). However, more recent work has relied largely on spiking models, though of varying degrees of biological plausibility. In addition, there is increasing evidence that many brain areas act as attractor networks (e.g., Wills et al. 2005).

### Point attractors

Following Hopfield's (1982) work (see Hopfield network), many biologically inspired models have taken the fixed points of a network to represent *memories* encoded in the system. Such memories have been mapped on to biological function in a number of ways, including:

- associative memory
- content addressable memory (pattern completion)
- categories
- noise reduction

Point attractor networks are usually considered in two regimes. In the first, input to the network is used to change (i.e., learn) the connection weights such that the network is in a determinate, stable state after the input is removed. In the second, learned weights remain fixed, and the network is probed with both familiar and unfamiliar input. Familiar inputs result in an expected output (i.e., a trained fixed point), and unfamiliar input results in the output whose basin of attraction the unfamiliar input is in.

If the state during and after a given input are different, then the network is said to act as an associative memory (associating the input with the subsequent fixed point after input is removed). If the state during and after input is the same, the network will usually act as a *content addressable memory*, and can perform *pattern completion*. This occurs because the basins of attraction for distinct fixed points will tend to vary smoothly between points. As a result, similar patterns will tend to similar fixed points. In this way, such networks are also often said to *categorize* their inputs, with one category for each possible fixed point. Similarly, this kind of behavior has been characterized as *noise reduction*, since, if the input is a noisy version of a familiar input, it will often result in the fixed point associated with the original, familiar input.

A significant amount of effort has been spent on quantitatively characterizing the storage capacity of these networks, and developing learning rules for them.

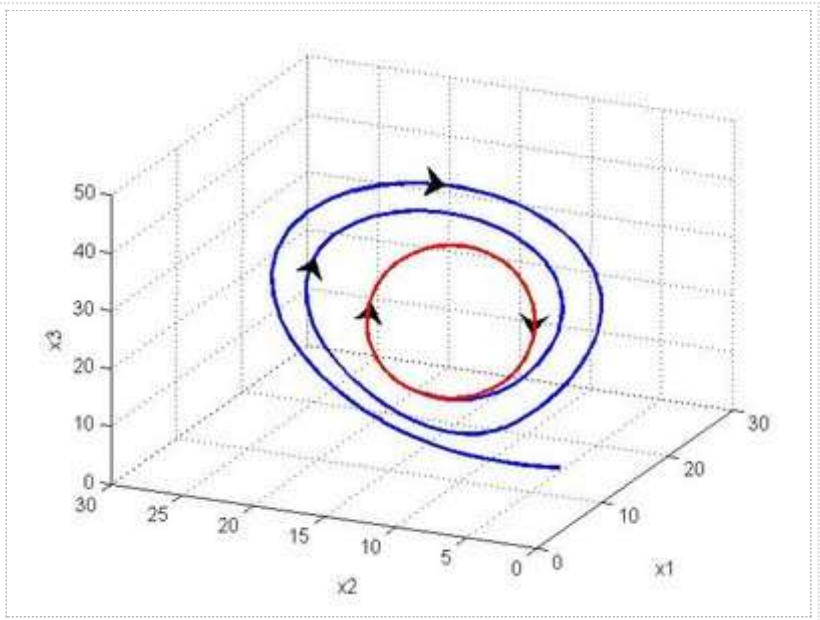

Figure 4: An example of a 2-dimensional cyclic attractor embedded in an $N$-dimensional network. An example trajectory is shown in blue. The attractor manifold is shown in red. This manifold is repeatedly traversed in the stable state.

## Line attractors

A special class of line attractors has been extensively explored in the context of oculomotor control. Of particular interest has been the activity of the nucleus prepositus hypoglossi in the brain stem, which is involved in the control of horizontal eye position across a wide variety of species, including fish and humans. These specific line attractors are called *neural integrators*. The terms 'line attractor' and 'neural integrator', while often used interchangeably, describe the difference between the network state space and the attractor space. In addition, the class of line attractors is much broader than those which act as integrators (line attractors only act as integrators if the driving term is along the attractor manifold, otherwise only the projection of the driving term onto the manifold is integrated).

These networks are called 'integrators' because the low-dimensional variable (e.g., horizontal eye position) $x(t)$ describing the network's output reflects the integration of the input signal (e.g., eye movement velocity) $v(t)$ to the system. That is,

$x(t) = \int v(t)$ or $\dot{x}(t) = v(t)$

Notice that in the low dimensional space, with no input, the output will remain constant. This means that a network which can be described by this equation will display persistent activity with no input, thus acting as an attractor (assuming it is stable to small perturbations away from this subspace).

The line attractor has more recently been used to describe experimental results on decision making tasks.

It should also be noted that such biological networks seldom act exactly as characterized by these idealizations. For instance, eye position integrators in most individuals have a centripetal drift (which varies with age) with no input, which suggests there is a slight slope to the attractor, and that the attractor is technically a point attractor. Other individuals have integrators with other kinds of drift, which can be indicative of various diseases or damage.

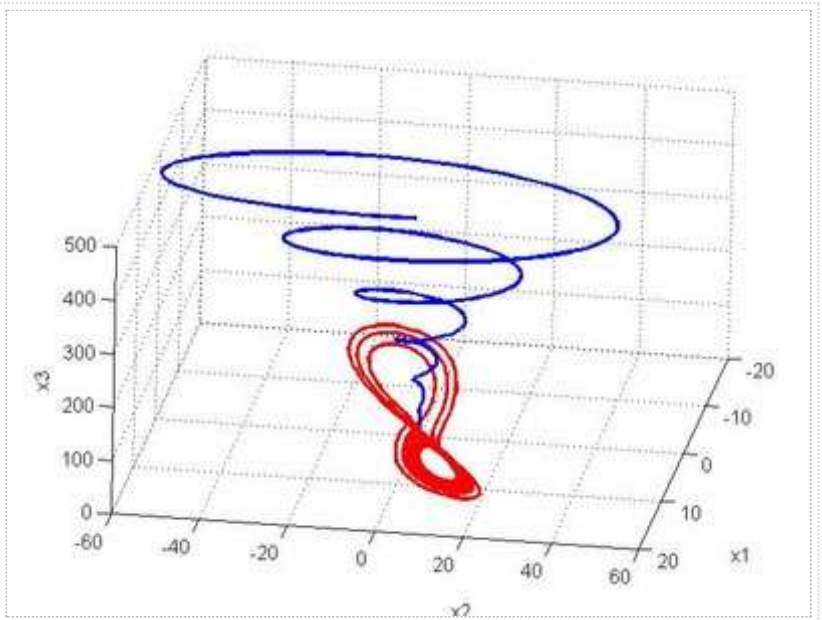

Figure 5: An example of a 3-dimensional chaotic attractor (Lorenz attractor) embedded in an *N*-dimensional network. An example trajectory is shown in blue. The attractor manifold is shown in red. This manifold is bounded, though may not be repeatedly traversed in the stable state.

## Ring attractors

Since the mid 1990s, ring attractors have been proposed as a model of the rodent head direction system (Zhang 1996). This network, which includes several regions of the rodent limbic system, indicates the current head direction of an animal, and receives velocity information as input. As a result, its function seems to be to integrate the velocity command to determine (head) position, just like the neural integrator in the oculomotor system.

There are two main differences between the head direction and oculomotor system. The first is that head direction is a cyclic variable. As a result, the attractor is a ring, rather than a line in the network state space.

The second difference is that the neural representation is different in the two cases. In the oculomotor integrator, neurons in the population monotonically (either positive or negative) change their firing rate with the represented variable (eye position). The represented value in the population is thus taken to be the weighted mean of each neuron's response. In the head direction integrator, neurons have a 'preferred' head direction at which they reach their maximum firing (and firing decreases symmetrically on either side of the maximum). The represented value of the head direction is taken to be determined by the mean of the function determined by population firing rates. In essence, the head direction system contains a 'bump' of activity, the center of which indicates the current best estimate of head direction.

## Plane attractors

While the oculomotor neural integrator as usually studied is sensitive to only one direction of eye movement, eyes in most animals have more degrees of freedom. As a result, an extension of the neural integrator to two dimensions results in a plane attractor embedded in the neural state space.

A more general interpretation of plane attractors is as function attractors: i.e., attractors for which stable points are functions of some underlying variable. This interpretation is important because networks that have sustained Gaussian-like bumps of activity have been found in various neural systems, including the head direction system, frontal working memory areas, visual feature selection areas, arm control systems, and path integration systems. In each case, the stable state is a 'hill', 'bump', or 'packet' of activity which is most naturally interpreted as a function of some underlying variable(s) (see Figure 6).

Function attractors and plane attractors are mathematically equivalent for smooth functions with finite degrees of freedom because any such function can be written as a point in a vector space (whose axes are an orthonormal basis of the function space). If we construct an appropriate plane attractor in that vector space, it will be indistinguishable from a function attractor in the original function space (e.g. Figure 6 is a 25 dimension plane attractor, there plotted as a two-dimensional function attractor).

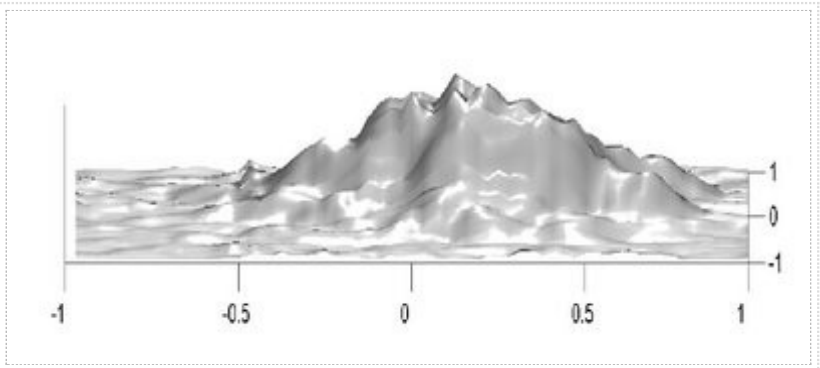


Figure 6: An example of a 2-dimensional function/plane attractor. The bump is smoothed spiking activity of neurons whose receptive fields are ordered by their spatial sensitivity. As a result, the bump is a function of 2-dimensional spatial co-ordinates. However, it can also be interpreted as a *point* on a plane attractor in a 25-dimensional vector space. Adapted from Conklin & Eliasmith (2005)

## Cyclic attractors

Cyclic attractors are natural descriptions of repetitive biological behaviors such as walking, swimming, flying, or chewing. In short, because cyclic attractors can describe oscillators, and many neural systems exhibit oscillatory behavior, it is natural to use cyclic attractors to describe oscillatory behavior in neural systems. The interpretation of repetitive biological behavior in terms of oscillators is at the heart of most work on central pattern generators (CPGs). However, the characterization of these oscillators at the network level as attractors is less common, although it as been done both for swimming and walking.

## Chaotic attractors

It has long been suggested that chaos or chaotic attractors may be useful for describing certain neural systems. For example, Skarda and Freeman (1987) hypothesize that the olfactory bulb, before odor recognition, rests on a chaotic attractor. They suggest that the fact that the state is chaotic rather than merely noisy permits more rapid convergence to limit cycles that aid in the recognition of odors. While these kinds of information processing effects themselves are well-documented (e.g., there a number of practical control problems that can be more efficiently solved if a system can exploit chaotic attractors effectively), the existence of chaos in neural systems is the subject of much debate, and is extremely difficult to verify experimentally.

# Using attractor networks

As the complexity of computational models continues to increase, attractor networks are likely to form important sub-networks in larger models. This is because the many clear information processing abilities of attractor networks (e.g., categorization, filtering noise, integration, memorization, etc.) makes them good candidates for being some of the basic building blocks of large-scale brain models.

One significant challenge to understanding how such networks can be exploited by larger biological systems is to determine how attractors can be controlled. Control may amount to simply moving the network state to another point on an established attractor, or it may demand completely changing the kind of attractor the network is implementing on-the-fly (e.g., changing from a point to a cyclic attractor).

A general approach to the implementation and control of attractor networks in spiking neuron models can be found in Eliasmith (2005).

# References


Amit, D. J. (1989). Modeling brain function: The world of attractor neural networks. New York, NY: Cambridge University Press.

Conklin, J. and C. Eliasmith (2005). An attractor network model of path integration in the rat. Journal of Computational Neuroscience. 18: 183-203

Eliasmith, C. (2005). A unified approach to building and controlling spiking attractor networks. Neural Computation. 17(6): 1276-1314.

Hebb, D. O. (1949). The organization of behavior. New York, NY: Wiley.

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

Skarda, C. A. and W. J. Freeman (1987). How brains make chaos in order to make sense of the world. Behavioral and Brain Sciences 10, 161–195.

Sompolinsky, H., A. Crisanti and H.J. Sommers (1988). Chaos in random neural networks. Phys. Rev. Lett. 61: 259–262.

Wills, C. L., F. Cacucci, N. Burgess J. O'Keefe, (2005). Attractor dynamics in the hippocampal representation of the local environment. Science, 308 (5723): pp. 873 – 876.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. Journal of Neuroscience 16, 2112–2126.</math>


**Internal references**


- John W. Milnor (2006) Attractor. Scholarpedia, 1(11):1815.
- Edward Ott (2006) Basin of attraction. Scholarpedia, 1(8):1701.
- Valentino Braitenberg (2007) Brain. Scholarpedia, 2(11):2918.
- Olaf Sporns (2007) Complexity. Scholarpedia, 2(10):1623.
- James Meiss (2007) Dynamical systems. Scholarpedia, 2(2):1629.
- Keith Rayner and Monica Castelhano (2007) Eye movements. Scholarpedia, 2(10):3649.
- John J. Hopfield (2007) Hopfield network. Scholarpedia, 2(5):1977.
- Jeff Moehlis, Kresimir Josic, Eric T. Shea-Brown (2006) Periodic orbit. Scholarpedia, 1(7):1358.
- Philip Holmes and Eric T. Shea-Brown (2006) Stability. Scholarpedia, 1(10):1838.


# External links

- A unified approach to building and controlling spiking attractor networks (http://watarts.uwaterloo.ca/~celiasmi/Papers/eliasmith.build%20and%20control%20spiking%20attractors.nc.pdf)

- Wikipedia Attractor entry (http://en.wikipedia.org/wiki/Attractor)

Sponsored by: Eugene M. Izhikevich, Editor-in-Chief of Scholarpedia, the peer-reviewed open-access encyclopedia
Reviewed by (http://www.scholarpedia.org/w/index.php?title=Attractor_network&oldid=18622) : Anonymous
Reviewed by (http://www.scholarpedia.org/w/index.php?title=Attractor_network&oldid=23562) : Alessandro Treves, SISSA, Trieste, Italy
Accepted on: 2007-10-18 13:23:07 GMT (http://www.scholarpedia.org/w/index.php?title=Attractor_network&oldid=23645)

Categories: Recurrent Neural Networks | Computational Intelligence | Neural Networks

*"Attractor network" by Chris Eliasmith is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Permissions beyond the scope of this license are described in the Terms of Use*