Symphony from Synapses: Neocortex as a Universal Dynamical Systems Modeller using Hierarchical Temporal Memory

Fergal Byrne
HTM Theory Group, Dublin, Ireland

fergal@brenter.ie http://inbits.com

Reverse engineering the brain is proving difficult, perhaps impossible. While many believe that this is just a matter of time and effort, a different approach might help. Here, we describe a very simple idea which explains the power of the brain as well as its structure, exploiting complex dynamics rather than abstracting it away.

Just as a Turing Machine is a Universal Digital Computer operating in a world of symbols, we propose that the brain is a Universal Dynamical Systems Modeller, evolved bottom-up (itself using nested networks of interconnected, self-organised dynamical systems) to prosper in a world of dynamical systems.

Recent progress in Applied Mathematics has produced startling evidence of what happens when abstract Dynamical Systems interact. Key latent information describing system A can be extracted by system B from very simple signals, and signals can be used by one system to control and manipulate others. Using these facts, we show how a region of the neocortex uses its dynamics to intrinsically "compute" about the external and internal world.

Building on an existing "static" model of cortical computation (Hawkins' Hierarchical Temporal Memory), we describe how a region of neocortex can be viewed as a network of components which together form a Dynamical Systems modelling module, connected via sensory and motor pathways to the external world, and forming part of a larger dynamical network in the brain.

Introduction

The mammalian brain is a complex network of interconnected dynamical systems, a property repeated at every scale, down to the individual synapse and beyond. While this is also true of digital computers in the physics of components and circuits, we use engineering to design this away, building deterministic, precisely known mechanisms in successive "logical" layers, in hardware and software. Given the readily accessible concepts and tools provided by Computer Science, Applied Mathematics, Symbolic Logic, Information Theory and Statistics, it is tempting to treat the brain as an information-processing system of obscure design, whose details might be revealed by "reverse-engineering" (inferring a mechanism from knowledge of its functional behaviour). This is indeed the approach taken by many in Computational Neuroscience, Machine Learning and Artificial Intelligence.

We argue instead that Nature has taken precisely the opposite approach, reverse-engineering in another sense. From the molecular machinery inside every cell to the ecosystem of the Earth, evolution exploits the bottom-up, self-organising emergent properties of interacting adaptive dynamical systems to create and preserve structure, synergy, order and information. These systems form a kind of fractal network structure, exchanging energy, material and information.

This is the domain of Complex Systems, a branch of science which has grown in recent years from the Applied Mathematics of Dynamical Systems. In this field, each component cannot be modelled using a logical or mathematical clockwork, but instead must be treated as an entity whose future behaviour cannot be predicted, neither exactly, nor even statistically, based on observations. Surprisingly, mathematical theorems from Dynamical Systems prove that signalling between such systems can allow an agent to both model and control a complex system to some extent. The "computation" in such an agent requires neither logic nor calculation, as it emerges directly from the structure and dynamics of interacting components within the agent itself.

Computational Neuroscience has for decades investigated the role of dynamical systems and chaos in the brain, but most researchers have treated complex dynamics as an inconvenient obstacle to understanding, rather than a key aspect of neural circuit function. In addition, the field of Dynamical Systems and Chaos has only recently become a major subfield of Applied Mathematics (due mainly to the advent of high-performance computing platforms), so the key findings of the field - even when applied to neural circuits [Pecora and Carroll, 1990] - have not been made known to the bulk of workers in Computational Neuroscience. As a result, important functions of neocortex involving their ability to model dynamics in the world have been overlooked or neglected.

This paper proposes a new view of neocortex as a Universal Modeller of Dynamical Systems, in which each region learns to lock onto the dynamics of its inputs, characterise and represent their underlying causal parameters (along with their nonstationarity), and act on them through both feedback and behaviour. A detailed role is described for subpopulations of neurons in all cortical layers in each region, along with the semantics of both well-predicted and unpredicted activity in subpopulations. We present a computational model for this based on extensions of the Cortical Learning Algorithm proposed by Hawkins' Hierarchical Temporal Memory [Byrne, 2015a], and a mathematical vector-based description of the model.

We develop a multi-layer model of a region of cortex which extends HTM's Cortical Learning Algorithm (CLA) to all 6 layers of neocortex. Interlayer and inter-region communication by progressive encoding of representation-prediction anomalies is proposed as the key factor in generating self-stabilising dynamics of representation and behaviour.

NB: The ideas proposed in this paper are extensions and generalisations of the current theory of HTM as proposed by Hawkins and Numenta, which treats evolutions of temporally-varying systems as streams of snapshots of spatial data. The key novelty in this extension is that the temporal dynamics are at least as important as the sequential structure of successive spatial snapshots, and that this gives an agent the opportunity to exploit the crucial information-theoretic results of treating sensorimotor inputs as signals from a dynamical system (ie one whose underlying dynamics are compactly represented using mathematical update rules), rather than observations of a system where one can only learn sequences by rote (ie where sequences can only be recorded). The dynamical version of HTM described here is no different from the original where the signal represents non-dynamical or arbitrary sequence inputs.

1 Outline

1.1 Hypothesis: The Brain as a Universal Dynamical Systems Modeller

We begin with the overall hypothesis of the paper, which is that the neocortex is a hierarchy-like network of interconnected and coupled regions, each of which is operating as a dynamical system modelling a dynamical system. Using results from the Applied Mathematics field of Dynamical Systems and Chaos, we will show that this modelling is feasible, simple to implement and applies pervasively in all kinds of natural and social contexts. In addition, the coupled dynamical systems approach has been shown to be the only feasible approach to many real-world problems which cannot be tractably solved analytically, and we argue that evolution discovered this fact millions of years ago and exploited it in neocortex.

1.2 Overview of Relevant Results in Dynamical Systems

We introduce a number of concepts and results from Dynamical Systems theory, which provides the mathematical and information-theoretic foundation for the paper's hypothesis. In particular, we review a number of findings about the power of synchronised and reconstructed dynamical systems in extracting information of use to a computational agent.

1.3 HTM as a Universal Dynamical Systems Modeller (UDSM)

We show how HTM provides a solid computational foundation for the modelling of dynamical systems, and present a multi-layer extension of the Cortical Learning Algorithm which describes the role of populations of neurons in all layers of cortex.

1.4 Cognitive Applications of a HTM UDSM

A number of cognitive functions can be viewed as involving the controlled coupling of dynamical systems. We describe a number of such functions, including motor control in animals, language in humans, and social interaction. We speculate that this literal interpretation of cortex as a Universal Dynamical Systems Modeller will be useful in understanding and even treating a number of conditions, including dyslexia and schizophrenia.

2 Hypothesis: The Neocortex is a Universal Modeller of Dynamical Systems

The field of Applied Mathematics was transformed in the 1960s and '70s by dramatic discoveries in Dynamical Systems and Chaos. Long avoided by mathematicians over a century after Poincaré, major progress became possible using newly available computer hardware, software and methods. In the late '70s and early '80s, important mathematical results by Takens [1981] and others showed that the dynamics of many real-world dynamical systems could be directly and accurately modelled based only on simply observing measurements of the real-world systems over time. This mathematical property allows an agent to characterise an observed system and perform short-range forecasting of its future behaviour, without having access to the underlying causal mechanics or latent parameters of the system.

Our hypothesis is that the structure of a region of neocortex has evolved to use a time series of sensory (and sensorimotor) inputs to generate a dynamical analogue of the system generating the sensory inputs, to forecast its short-term future, to identify stable and slowly-changing characteristics of the system which indicate its hidden controlling parameters or state, and to model the nonstationarity of the system state.

We begin with a very brief review of results from Dynamical Systems Theory. This is followed by a brief description of the HTM model of neocortex (a full description of HTM is found in Byrne [2015b]). We then combine the two concepts and describe the key structures in laminar cortex which build modellers of natural, cognitive, linguistic and social dynamical systems.

3 Dynamical Systems and Chaos

A Dynamical System is a mathematical model whose dynamics are characterised by express update rules, typically differential equations in continuous systems, or difference equations in discrete time (a comprehensive survey of Dynamical Systems is Strogatz [2014]). The study of dynamical systems began with the advent of calculus, and many simple systems have been studied ever since. Simple Harmonic Motion, which is an idealised approximation of small oscillations of a simple pendulum or spring, is the archetypical dynamical system introduced in high school physics.

Simple Harmonic Motion (SHM) is a type of periodic motion where the restoring force is directly proportional to the displacement and acts in the direction opposite to that of displacement [Wikipedia, 2015]. The restoring force F_{net} is given by:

$$F_{net} = m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -kx$$

where m is the mass, x the displacement and k is a constant (called the *spring constant* in the case of a spring). This system has a closed form solution (a formula for x as a function of t) given by:

$$x(t) = c_1 \cos(\omega t) + c_2 \sin(\omega t) = A \cos(\omega t - \varphi)$$

Such systems can be solved exactly using calculus, and most systems studied by engineers and scientists are similarly straightforward to understand and reason about. However, outside the boundaries of approximation used for SHM, even simple springs are no longer subject to perfect analysis.

For these reasons, Engineering and Science have historically avoided the problem of analytically insoluble dynamical systems, usually by approximating the real system using something akin to SHM, a process called linear approximation. With the advent of computers in the mid-20th century, however, researchers have been able to study the dynamics of nonlinear and complex systems, and since the 1960s this has become a primary focus of Applied Mathematics.

In the seminal study which triggered the revolution in Dynamical Systems, Lorenz [1963] described a simple model of atmospheric convection, using three coupled differential equations:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x), \tag{1}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y, \tag{2}$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z. \tag{3}$$

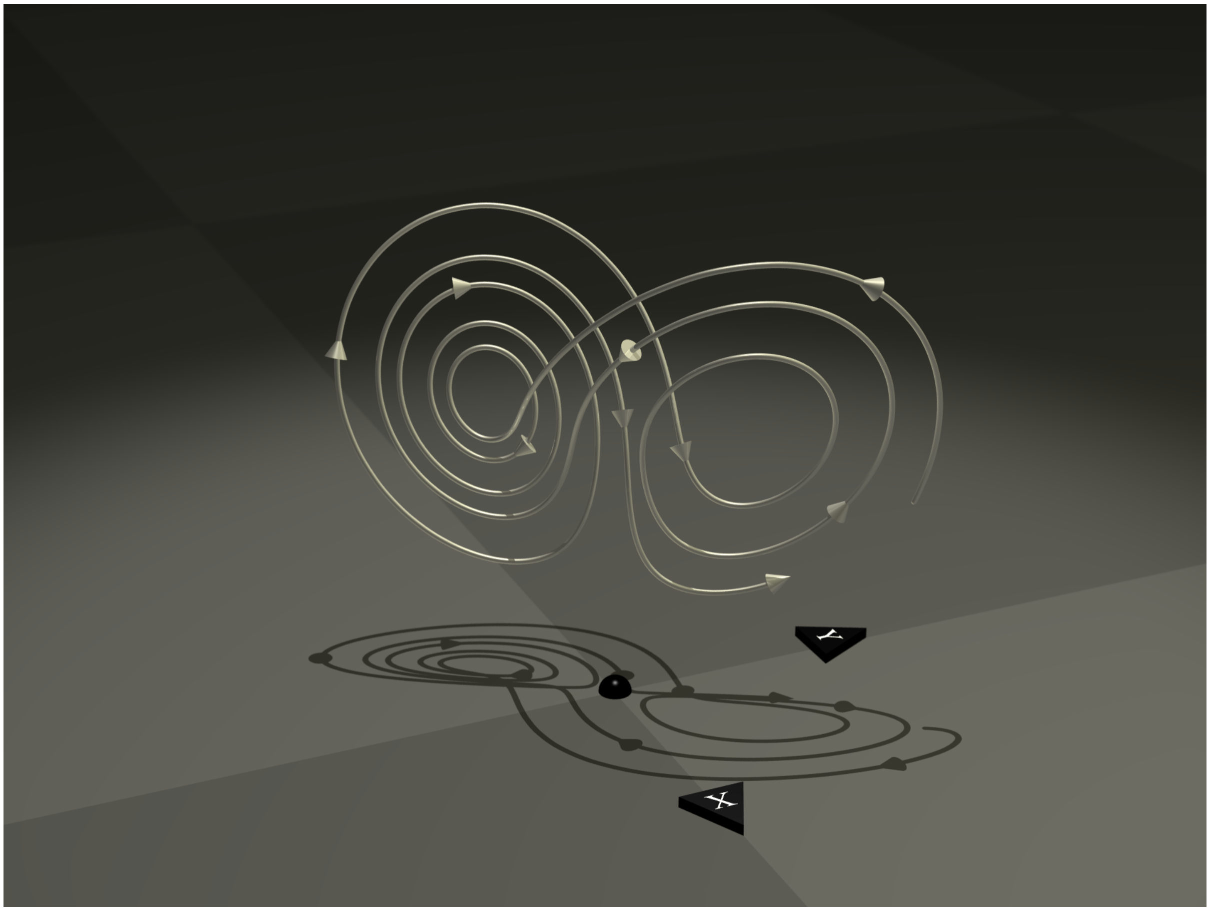

Figure 1: A visualisation of Finite segment of a trajectory of Lorenz's equations, computed by numerical integration and rendered as a metal wire. Parameter values are classical. Conical beads indicate the direction of travel. The black hemisphere marks the origin, and directions of the x and y axes are as indicated. The base of each axis triangle is 20 units from the origin. The z axis points upward; the highest point on the wire is approximately 45 units above the x-y plane. The shadow is cast by a spotlight high on the z axis. [Wikimedia, 2006]

.

The quantities x, y and z are abstractions of the state of the convecting fluid (here x is proprtional to the intensity of the convective motion, y to the temperature difference of the ascending and descending currents, and z to the departure of the vertical temperature profile from linearity). While abstract, these quantities exactly describe the state of motion of the entire system as a function of time. We can visualise the system by drawing plots of the trajectories of points $(x(t), y(t), z(t))$ in the *phase space* of the system (see Figure 1).

The deterministic evolution of the Lorenz system depends on the choices of σ, ρ and β, which are together known as the system's *control parameters*. Lorenz studied the system with $\sigma = 10$, $\rho = 8/3$ and $\beta = 28$, which places the system in a *chaotic* regime, in which only short-term forecasts of the state are possible, and the exact trajectory of the system exhibits *sensitivity to initial conditions*.

Floris Takens [1981], basing his study on results going back to Whitney [1936], proved that a system such as Lorenz' could be reconstructed in all important details using only a time-series of a single measurement from the system. A system with manifold dimension m can be reconstructed simply by plotting, for each time t, the vector of time-delayed mesaurements $(x(t), x(t-\tau), x(t-2\tau), ...x(t-k\tau))$ of $k > 2m+1$ observed values x. Sauer [2006] showed that Takens' Theorem holds in less restricted systems, for either k separate measurements at time t or for a set of k time-delayed measurements (see 2).

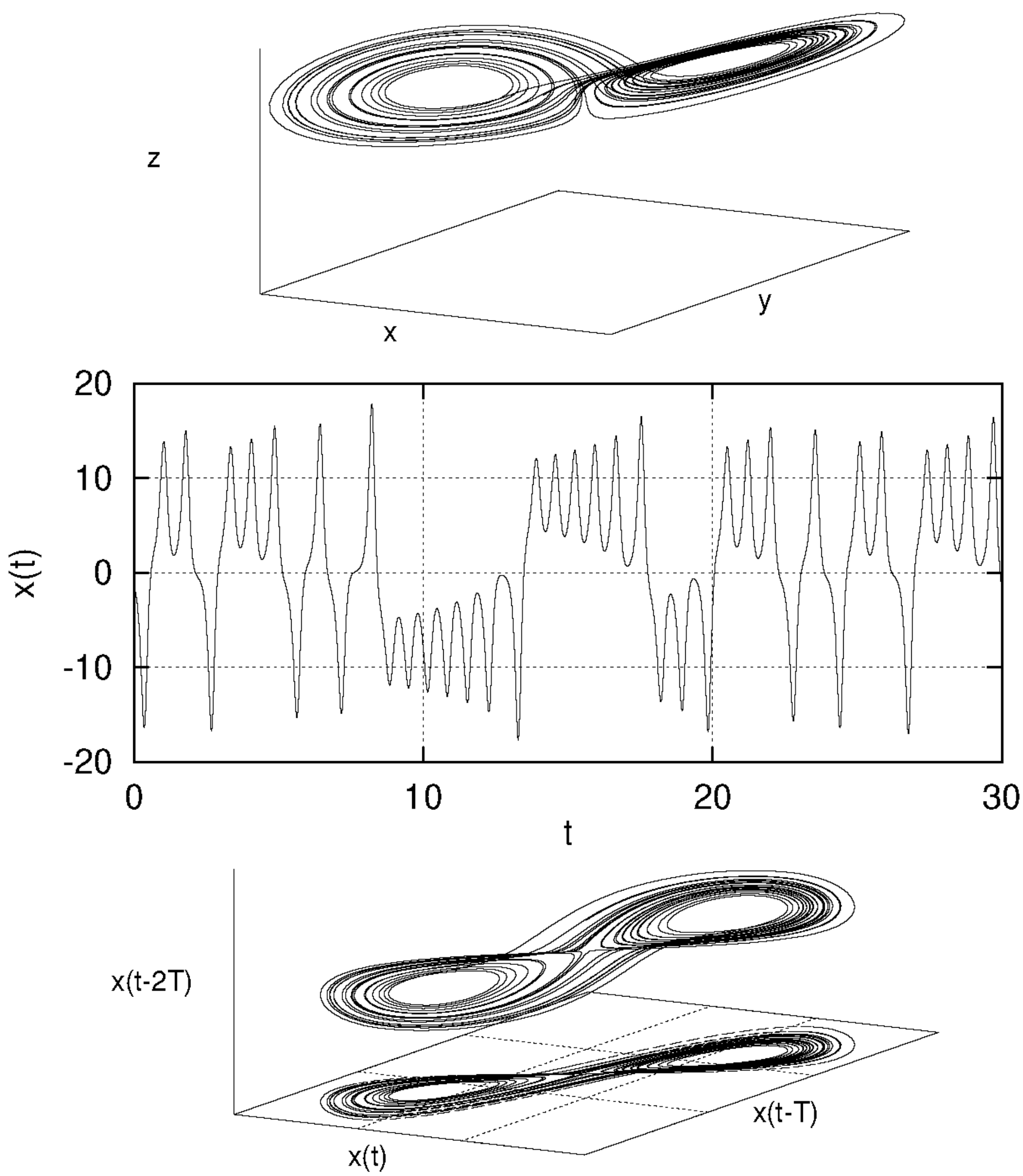

Figure 2: Attractor Reconstruction of the Lorenz Attractor: (top) the Lorenz Attractor in 3D; (centre) a plot of the observed time series $x(t)$; (bottom) reconstruction of the attractor by plotting $(x(t), x(t-\tau), x(t-2\tau))$. From Sauer [2006]

As can be seen in the figure, the reconstruction is not visually identical to the original system. It is, however, topologically identical in an important sense called *diffeomorphism*. This means that key analytic properties of the observed system, in particular its divergence characteristics, are replicated in the reconstruction. Predictions made using the replica have the same properties as the near-term future of the observed system.

This fact about well-behaved systems such as Lorenz' has been empirically observed to hold in all kinds of natural and real-world phenomena, even in systems where Takens' Theorem cannot be shown to stricly hold. We claim that Nature has taken advantage of this fact about reconstruction of complex latent dynamics in neocortex, in order to perform computations about the real world without access to the underlying mathematics or machinery for solving the differential equations.

An agent exploiting this aspect of Dynamical Systems Theory can perform the following functions:

1. using very simple computations on the signal, the agent can generate a simulacrum of the observed system which preserves the essential properties of that system;

2. by modelling or learning the transitions between points in the reconstructed space (alternatively the vector field in the tangent space of a point), the agent can run a simulation forward in time to perform forecasting;

3. by making multiple predictions of the transition vectors, and comparing the resulting predictions with actual trajectories, the agent can remain locked on to a nonstationary system;

4. by monitoring and tracking the correct predictions of transitions, the agent can characterise the controlling parameters (such as σ, ρ and β in Lorenz' system), and track those parameters over time if the system is nonstationary.

5. by combining prediction with measurement, the system can distinguish between the chaotic signal and non-chaotic inputs, including noise and any signal using the chaos as a carrier signal.

6. one agent can communicate arbitrary information to another by using such coupled dynamical systems; see Marcus and Williams [2008] for an overview of Symbolic Dynamics.

7. an agent can control a dynamical system such as an arm, leg, an entire body, or a body-tool assemblage by sending controlling signals to the driving inputs of that system; see Ott [2006] for a guide to this.

We will describe a subset of these functions.

3.1 Reconstruction of an Observed System

Real-world systems found in nature, in society, and in economics are difficult or impossible to study using the standard tools used for centuries since Newton and Leibniz. In most cases, we have access neither to the internal dynamics (the differential or difference equations governing the update rules), the control parameters which vary the global properties of the system, nor the exact values of the observables of the system. Engineers typically address these problems by modelling the system approximately using simpler, linear dynamics and tools for estimating their governing parameters and measurement error.

Several decades of research into the information theory and mathematics of such systems has shown that there is a simpler and in many ways more powerful method: *reconstruction* of the dynamics using signals from the observed system [Takens, 1981, Sauer, 2006]. The essence of the strategy is that sufficient information about the observed system is encoded in its signal to allow its reconstruction in an observing system using very simple computations.

A classical Dynamical System such as Lorenz' can be seen as a *vector field* or *flow* in its phase space, where the value of the vector field at each point is effectively a velocity vector indicating how a "particle" at the point will move, tracing out a trajectory over time (see 1). When the system is chaotic, no long-term forecast of its position can be made, because nearby trajectories diverge exponentially quickly as a function of time. However, for sufficiently short timescales, nearby trajectories can be shown to remain within a required volume. This means that, despite uncertainty due to measurement noise (or noise inherent in the system), the near-future evolution of the system can always be predicted within finite bounds.

The key result of Takens' Theorem is that this property is preserved in the reconstructed system (in fact this is the main practical use of the theorem). Formally, the two systems are *diffeomorphic*, which simply means that they have the same topological properties.

Informally, the key insight is that the reconstructed system is to all intents and purposes *the same thing* as the observed one, but represented in a form fully accessible to the observer, without needing a mathematical model of the original dynamics of the real system.

3.2 Learning to Predict using Reconstruction

An agent can record a long time series of observed signals from a Dynamical System using this method, recording for each point in the reconstructed space the transition vector observed from that point to its successor. Over time, this allows the agent to estimate the flow field in the neighbourhoods of more and more observed points in the signal. If the vector field is sufficiently smooth, this allows the agent to estimate the successor of a new data point by combining (in some simple way) the transition vectors of its near-neighbours in the recording.

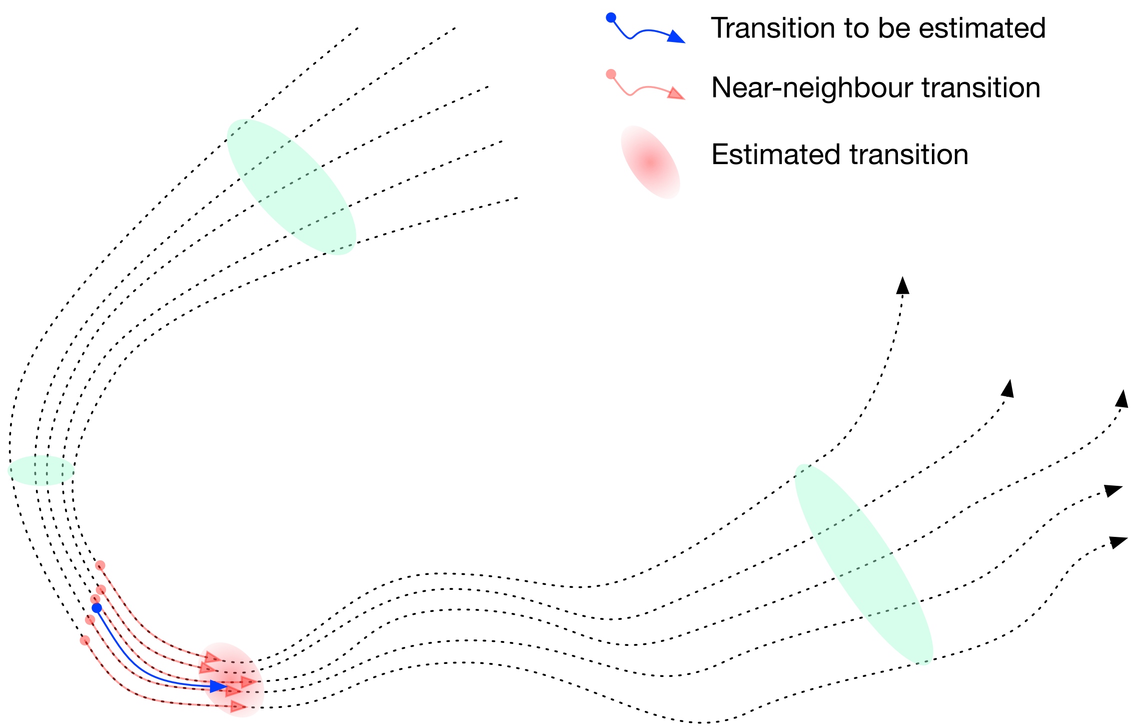

Figure 3: Prediction of future position using near-neighbour transitions. Five trajectories in recontructed phase space are shown, which get close together at bottom left. A new data point (blue dot) is presented, and we choose the closest points on the nearby trajectories (red dots) to form an estimate (red cloud) of the future position of the new input.

.

We have described a somewhat naïve method for prediction of a transition, which involves recording every transition at every point in the past, searching over them for every new point, choosing some number of near neighbours, and calculating the estimate from the chosen transition vectors. This procedure can be dramatically optimised using appropriate machine learning methods, as we shall demonstrate using HTM.

3.3 Multiple Predictions and Non-stationarity

In real-world systems, underlying control parameters may change over time, a condition known as *non-stationarity*. For example, as we grow our bones change length, our muscles strengthen, and our weight changes. This changes the dynamics of the resulting system from one *regime* to another. An agent may discover and characterise this non-stationarity by making multiple predictions, and observing which of them is fulfilled.

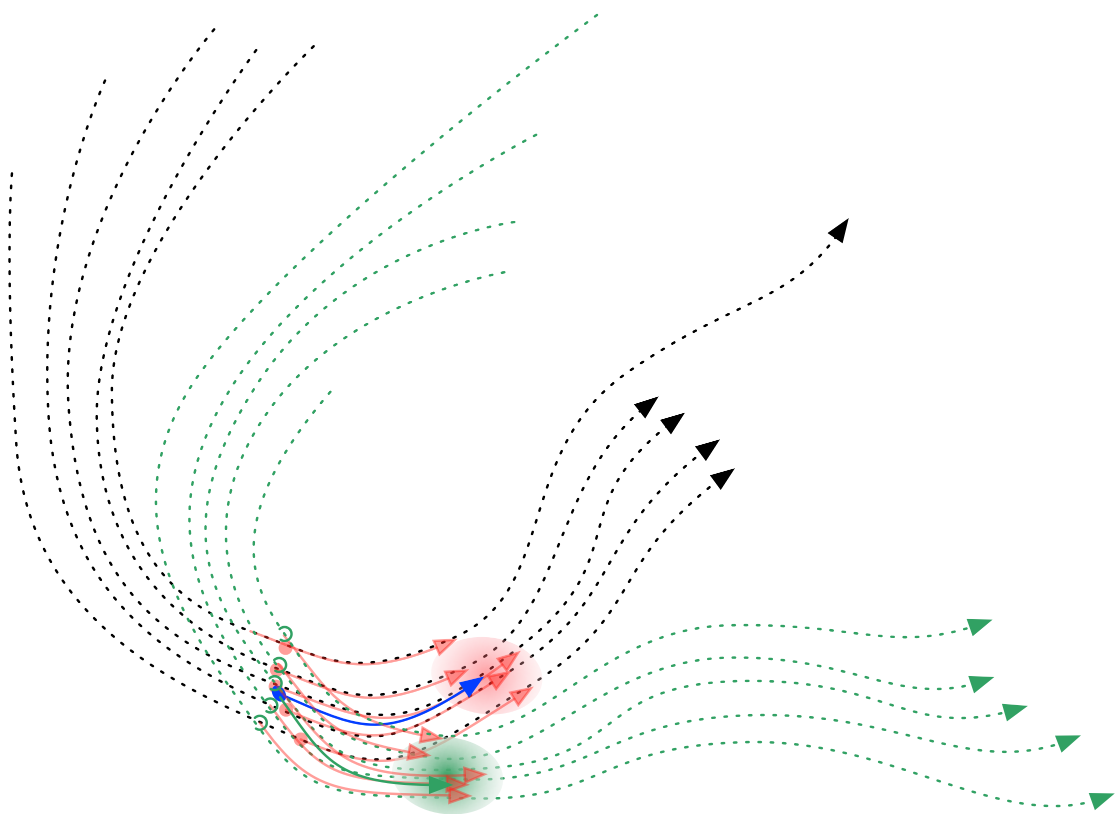

Figure 4: Multiple Predictions. Trajectories for two slightly different dynamical systems (black and green dashed lines). Given a new data point (blue dot), the agent can make two distinct predictions (red and green clouds), by using near-neighbours from both classes (red and green dots). The system may be non-stationary, ie the control parameters may have shifted from the black regime to the green, in which case the next input will distinguish between regimes.

.

3.4 Prediction Error and Correction Vectors

Looking again at the multiple prediction visualisation (4), consider a situation where the agent, unsure of which regime is in play (which neighbours to trust), chooses to predict *both* red and green clouds, and the next input is in fact the green one. In this case, we could consider the green arrows as *correctly predicting* and the red as incorrect. In this case, we can view the vector from blue to green estimates as a *prediction error vector* or alternatively as a *correction vector*. This information can be used by a learning system to make better predictions in future (if the green regime remains in place), to assist in classifying the regime, to adjust the influence of the neighbours, and so on. We shall see that this aspect of the dynamical modelling is important in the HTM-based system.

3.5 Driven Dynamical Systems

Reconstruction of dynamical systems is a very interesting and useful concept for working with complex dynamical systems.

A related but crucial property of Dynamical Systems is where a signal from one is used as an input to the controlling equations of the other, a process known as *driving*. The dynamics of the driven system are altered in a way which merges information about the driving system with its own states and dynamics.

This process is the key to interaction between agents and their environments. Sensory information is a signal which drives the dynamics of primary sensory cortex, interacting in turn in complex ways with other parts of cortex, and also with the dynamical systems of the body via motor signals.

3.6 Multiple Systems, Synchronization and Causality

These ideas can be extended to model multiple systems, allowing the agent to determine and detect the nature of the relationships between different objects, systems of objects, and sensorimotor modalities of single systems. The study of the relationships between different systems involves the concept of *synchronization*, defined in Brown and Kocarev [2000] as follows:

“Two dynamical systems $\mathbf{x}$ and $\mathbf{y}$ are *synchronized* with respect to the properties g_x and g_y if there is a time independent mapping $\mathbf{h} : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ such that $\|\mathbf{h}(\mathbf{g}(\mathbf{x}), \mathbf{g}(\mathbf{y}))\| = 0$ holds on all trajectories.”

An alternative definition from Pikovsky et al. [2003] is “the appearance if a statistical relationship between the observables of two systems due to mutual interaction”.

Note that in the case of an observed system and its reconstruction, there is a trivial synchronization whereby the observed signal and the reconstructed vector of lagged signal values are connected directly in a *master-slave* synchronization.

The most interesting kind of synchronisation is *Generalised Synchronization*, defined originally in Rulkov et al. [1995], which involves a functional relationship between the variables of $\mathbf{x}$ and $\mathbf{y}$, ie. $\mathbf{x}(t) = \phi(\mathbf{y}(t))$.

Attempts to use reconstruction to characterise the relationships between putatively synchronized dynamical systems have had limited success [Kato et al., 2013], but recent progress has been made using a number of relatively simple innovations. In particular, Chicharro and Andrzejak [2009] and Sugihara et al. [2012] have developed methods which use ranking of near-neighbours in phase space to overcome difficulties in handling noise, tranposing between systems of different dimension, and normalising the volumes occupied by neighbourhoods in each system. Chicharro and Andrzejak define their L-index by:

$$L(X|Y) = \sum_{i=1}^{N} \frac{G_i(X) - G_i^k(X|Y)}{G_i(X) - G_i^k(X)}$$

where $G_i(X) = \frac{N}{2}$ and $G_i^k(X) = \frac{k+1}{2}$ are the mean and minimal mean rank, respectively, for k nearest neighbours and N data points. $G_i^k(X|Y)$ is the mean rank of k neighbours of X_i using the k neighbours of Y_i to choose indices. The index is thus a dimensionless measure of the volume increase of a neighbourhood in X when using points from Y to define neighbours. The authors demonstrate the robustness of this method for both artificial (Lorenz) and natural systems.

Interestingly, this use of ranking is also a property of the activation-ordered formation of Sparse Distributed Representations in HTM (see Byrne [2015b] and Section 4).

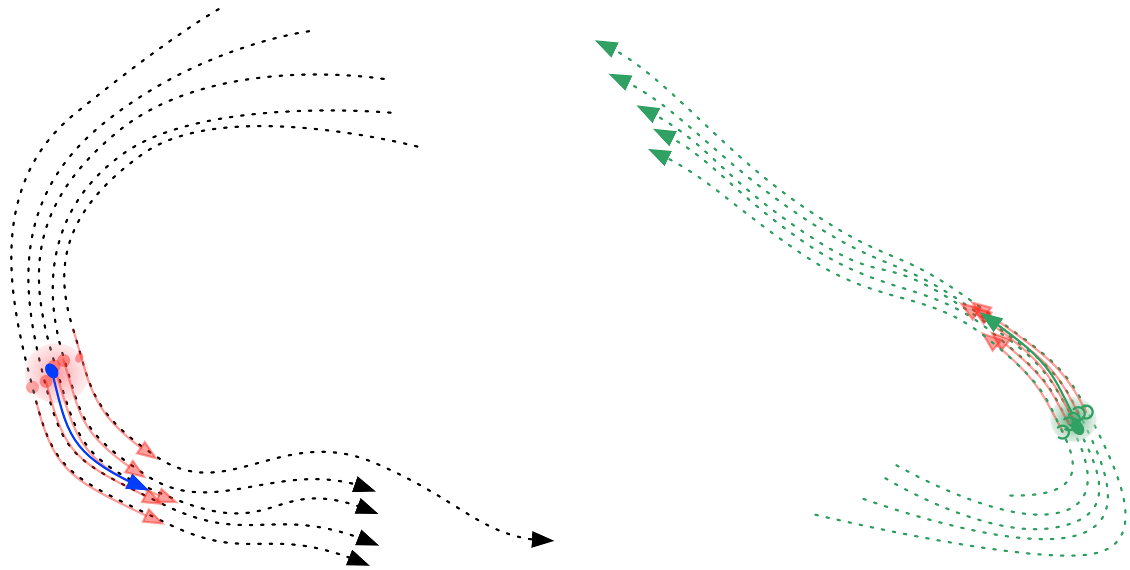

Figure 5: Convergent Cross Mapping [Sugihara et al., 2012] and L-index [Chicharro and Andrzejak, 2009]. If the system X at left is driven by Y at right, then a neighbourhood of a point Y_i can be used to identify candidate points in X which will occupy a volume around X_i.

.

4 HTM as a Universal Dynamical Systems Modeller (UDSM)

Dynamical Sysems Modelling using classical point-neuron networks (usually known as Artificial Neural Networks or ANNs) has been studied in the past (see Masulli et al. [1999] for example), but only in a feedforward context where the ANN is used only to read out the state, forecast evolution, and anomaly profile of the observed system. In these models, the ANN is not itself a dynamical system in a complex network.

Here, we develop a multi-layer model for HTM which describes how a region of neocortex learns to model and interact with dynamical systems, whether external or internal to the agent.

4.1 Prediction-Assisted Cortical Learning Algorithm - Review

Hierarchical Temporal Memory and the Prediction-Assisted Cortical Learning Algorithm (paCLA) are described in considerable detail in Byrne [2015b]. In essence, a series of sensorimotor inputs, each represented as a sparse distributed representation vector, is projected into a layer's output space of cell activations, forming an SDR representing each input. Pattern Memory is the learning (via synaptic adaptation in proximal dendrites) of this spatial mapping. In addition, the layer uses recurrent distal connections to learn and predict transitions between output SDRs, a process called Transition Memory. The output of the layer is a combination of orthogonal vectors, one formed from highly-predictive cells, the other signifying the difference between prediction and reality.

paCLA models L4 of cortex as just described. A second layer (L2/3) can use this L4 output to learn to form a more stable representation of a sequence or cycle of lower-layer SDRs, a process called Temporal Pooling. L2/3 thus outputs a series or sequence of SDRs which characterises the succession of higher-level meta-states in the input stream. This output is passed up the hierarchy for further processing.

We now extend this simple, feedforward sensory processing model to an entire region of 6 layers. We begin by connecting the form of signals in HTM, the Sparse Distrbuted Representation, with the concepts from the section on Dynamical Systems.

4.2 Sparse Distributed Representation (SDR) as High-Dimensional Embedding

We've already seen from the mathematics of Takens [1981], Sauer [2006] and Whitney [1936] that a reconstruction or synchronization of dynamic systems can be achieved as long as sufficient spatial and/or temporal dimensionality is represented by an agent using the observed signal. This is certainly true in HTM when considering the representational capacity of SDRs of typical sizes (see Ahmad and Hawkins [2015]).

As also mentioned in Byrne [2015b], k-Winner-Take-All (kWTA) representations are highly efficient universal function approximators (via Maass [2000]). In addition, Chicharro and Andrzejak [2009] identified the kWTA near-neighbour representation as optimal for automatic analysis of the causality relationships of putatively coupled dynamical systems.

SDRs derived from sensorimotor input clearly unite both of these factors of dimensionality and efficient kWTA functional representation.

4.3 Multilayer Flow Model of HTM

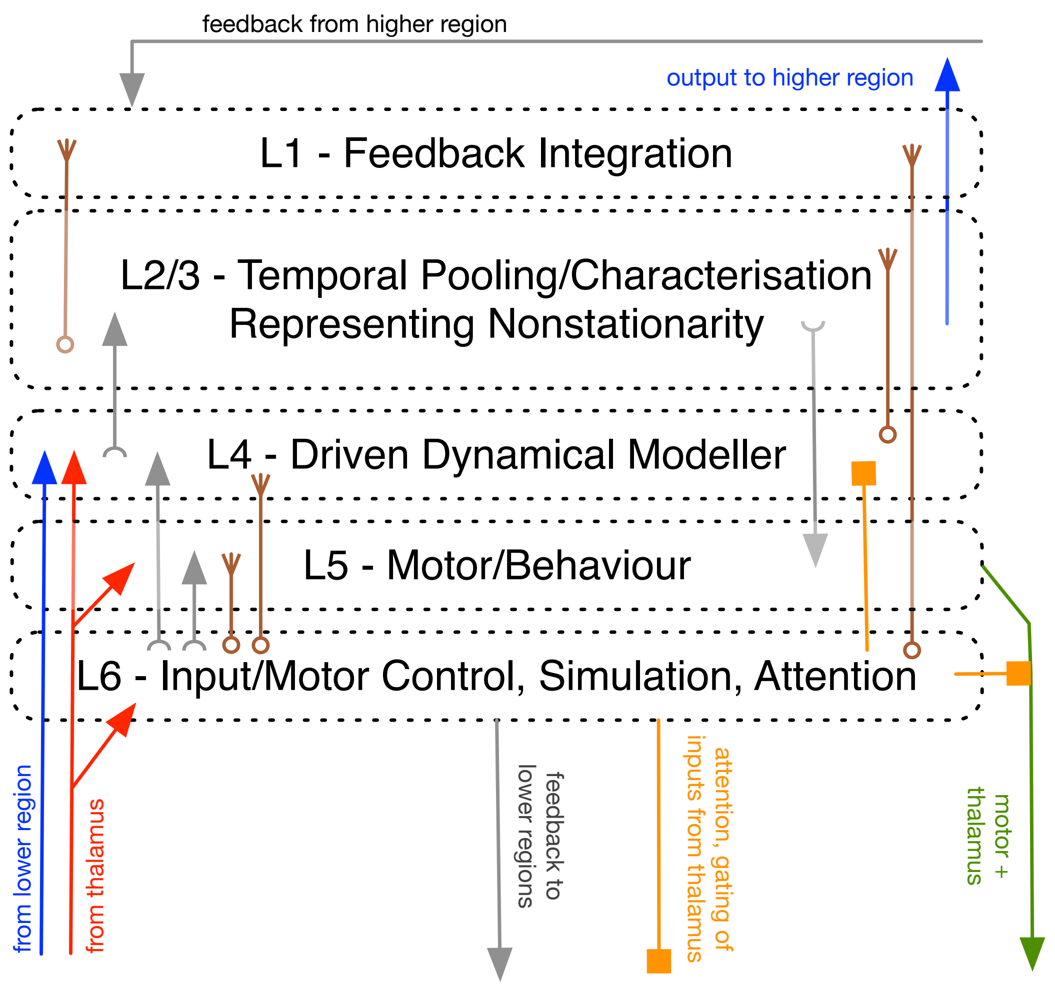

Figure 6: Flows of inputs and control signals in Multilayer HTM. Sensory (red) inputs from thalamus and (blue) from lower regions flow to L4 and L6. L4 models inputs and transitions, L2/3 temporally pools over L4, passing its outputs up. L5 integrates top-down feedback, L2/3 outputs, and L6 control to produce motor output. L6 uses gating signals (orange) to co-ordinate and control inputs and outputs, and can execute simulation to run dynamics forward.

.

A region of neocortex is itself a network of dynamical circuit systems, with flows of signals entering from sensory and motor sources, from other regions of cortex in the network, from subcortical structures, and from relay nuclei in the thalamus. Within the region, a specific network of connections directs the interactions between cell populations in the various layers. The content, sparseness and dynamics of these signals dynamically interact with the inherent learned dynamics of each circuit, and gating circuits control whether and how signals travel in the network. The outputs of the region include signals up and down the hierarchy and into the cortical network, gating signals controlling information flow in and out of the region, interaction with subcortical structures, and motor behaviour.

4.4 Layer 4 as a Transition Memory Modelling Input Dynamics

Layer 4 of neocortex is the primary recipient of afferent sensorimotor inputs to neocortex, as well as the main recipient of corticocortical "feedforward" signals in hierarchy. We claim that the purpose of Layer 4 is to represent its input in a high-dimensional form, in order to reconstruct the dynamics of its inputs. Using the Prediction-Assisted Cortical Learning Algorithm (paCLA) described in Byrne [2015b], L4 can represent each input as a high-dimensional vector, and can learn the transition vectors predicted by the history of inputs to the layer.

Importantly, a layer of cortex using paCLA represents each input as a sorted sequence of action potentials, ranked by activation arising from a combination of predictive and actual sensory inputs, in a fashion ideally modelled using the mathematics of Chicharro and Andrzejak [2009]. This allows L4 to represent its perception as a combination of predicted dynamics and a correction vector of bursting columns, as illustrated in Subsection 3.4.

Layer 4 is effectively "plotting" its inputs in a high-dimensional space of SDRs, thereby fulfilling the first key representational requirements of the mathematics of reconstruction of dynamical systems. Secondly, L4 learns to predict the future evolution of its inputs using the transition vectors represented by its distal dendrites. It thus models both the instantaneous state (position in phase space) and the vector field representing its future trajectory.

4.5 Layer 2/3 Temporal Pooling Models Control Parameters and Nonstationarity

As described in Byrne [2015b], a subpopulation of neurons in Layer 2/3 of cortex performs a function known as Temporal Pooling, representing sets of successively predicted SDRs in Layer 4 as a single, stable output. This TP SDR can be seen as a kind of dynamical symbol for the regime currently experienced by L4, which encodes the particular latent control parameters of the system L4 is modelling.

It's important to note that Temporal Pooling is itself a dynamical phenomenon, depending as it does on the extent to which the L4 neurons can predict transitions in their inputs. If the observed system is nonstationary, the transitions in L4 will shift to track new trajectories on altered attractors, represented by different neurons in L4, and so the L2/3 representation will itself transition to represent the new regime. Any prediction errors in L4 are passed to L2/3 separately, causing L2/3 to partially cease pooling over the old transitions and gradually begin pooling over new ones. This can itself be viewed as shifting in L2/3 of its SDR from one basin of attraction to another.

Layer 2/3 can learn sequences of such stable, temporally pooled representations, and incorporate contextual and top-down feedback to model the higher-level structure of the system or environment being observed.

This dynamical perspective of L4-L2/3 function operates in conjunction with the more sequence-centric and chunk-centric view originally proposed for Temporal Pooling by Hawkins [2014]. Which function predominates depends on the kind of inputs being processed by the region.

4.6 Layer 5 Dynamically Models Motor Interactions between the Region and the World

The primary function of Layer 5 is to produce motor output, via signals sent to subcortical motor centres. This happens in every region of cortex, not just in so-called motor cortex (which is primarily dedicated to large limb movements). The reason is that each region can use local information to interact directly with its source of sensory information - the system it's modelling.

Layer 5 is the main integration layer in neocortex. Its primary pyramidal neurons are the largest in the region, with an active apical trunk passing up through Layers 4, 3 and 2, apical dendrites in Layer 1, and basal dendrites in Layers 5 and 6. A L5 neuron thus has access to all sources of information on the dynamical states of the region, and its purpose is to use this information to affect the system of interest to the region by sending motor signals.

L5 neurons project their axons to synapse onto cells in subcortical motor centres. This is guided initially by genetic control (to send the axons to the correct target area), but then the precise alignment of individual axons to target motor centre cells is based on activity matching. For example, primary visual cortex (V1) neurons target the Superior Colliculus, a centre using retinal inputs to control eye movements [Triplett et al., 2009].

In addition, L5 axons split in two as they leave the region. The second branch ascends hierarchy in the cortex and informs higher regions of the motor/behavioural decisions made by a region. This signal complements that generated by L2/3, giving other areas of the brain a full sensorimotor picture of this region's dynamics.

L5 cells learn to form the longest link in the feedback signalling loop between the brain and the world. This is perhaps why it has the broadest access both to fast-changing sensory signals and every layer in its region.

4.7 Layer 6 Coordinates, Controls and Simulates Signalling and Activity

Layer 6 is the most complex of all cortical layers. Its numerous interconnected circuits together perform key "operating system-like" functions which coordinate and control the activity, communications and dynamics of the region:

1. L6 processes the same inputs as L4, but uses extra information from L5 (and thus from L1-L2/3-L4) to control its dynamics.
2. L6 uses projections to thalamus to gate inputs to the region, deciding how much "reality" the region sees, and controlling attention.
3. L6 projects strongly to L4, allowing it to provide artificial inputs to the region for forward simulation.
4. L6 projects feedback signals to lower regions, providing their L2/3 with top-down context, assisting lower-level sensory interpretation.
5. Feedback to lower regions from L6 also provide top-down signals controlling finer-grained L5 motor output in lower regions (unfolding motor sequences)
6. L6 gates the activity and motor output of L5, providing control over whether actions really happen.
7. Some L6 circuits form a dynamical "control system" which interacts with and controls how each of the above systems evolve.

Combining all these circuits, it is clear that L6 is the conductor of operation of the region, deciding dynamically what region-level functions are carried out, coordinating with other regions to perform a larger-scale task.

4.8 Learning and Meaning in Dynamical HTM

While the above is a plausible model for neocortex, it somewhat begs the question of how this highly complex fractal network structure of interacting dynamical systems can be made to work. The answer is actually quite simple.

In early development, the gross circuit layout described above is established using well-understood genetic mechanisms. The formation of specific cellular connections and synapses between neurons takes place, however, through activity-based learning (as described for V1 neurons in Triplett et al. [2009]).

Motor and sensory modelling is learned through activity and behaviour, in a process described by Warren [2006] and Freeman [2002] as the Action-Perception cycle.

Freeman [2003, 2004] describes meaning in neural systems as constructed from the interaction of incoming information with the dynamics of cortex (itself the learned product of its history of interaction with the world).

References and Notes

Subutai Ahmad and Jeff Hawkins. Properties of Sparse Distributed Representations and their Application to Hierarchical Temporal Memory. arXiv:1503.07469 [q-bio.NC], Jul 2015. URL `http://arxiv.org/abs/1503.07469`.

Reggie Brown and Ljupčo Kocarev. A unifying definition of synchronization for dynamical systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 10(2):344–349, 2000.

Fergal Byrne. Hierarchical Temporal Memory including HTM Cortical Learning Algorithms. Revision of Hawkins and Ahmad, 2011, Oct 2015a. URL `http://bit.ly/htm-white-paper`.

Fergal Byrne. Encoding Reality: Prediction-Assisted Cortical Learning Algorithm in Hierarchical Temporal Memory. *arXiv preprint arXiv:1509.08255*, Sep 2015b. URL `http://arxiv.org/abs/1509.08255`.

Daniel Chicharro and Ralph G Andrzejak. Reliable detection of directional couplings using rank statistics. *Physical Review E*, 80(2):026217, 2009. URL `http://dx.doi.org/10.1103/PhysRevE.80.026217`.

Walter J Freeman. The limbic action-perception cycle controlling goal-directed animal behavior. In *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on*, volume 3, pages 2249–2254. IEEE, 2002.

Walter J Freeman. A neurobiological theory of meaning in perception part i: Information and meaning in nonconvergent and nonlocal brain dynamics. *International Journal of Bifurcation and Chaos*, 13(09):2493–2511, 2003.

Walter J Freeman. How and why brains create meaning from sensory information. *International journal of bifurcation and chaos*, 14(02):515–530, 2004.

Jeff Hawkins. New ideas about temporal pooling. Wiki Page, Jan 2014. URL `http://bit.ly/temporal-pooling`.

Hideyuki Kato, Miguel C Soriano, Ernesto Pereda, Ingo Fischer, and Claudio R Mirasso. Limits to detection of generalized synchronization in delay-coupled chaotic oscillators. *Physical Review E*, 88(6):062924, 2013.

Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 2015/09/26 1963. doi: 10.1175/1520-0469(1963)020⟨0130:DNF⟩2.0.CO;2. URL `http://dx.doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2`.

Wolfgang Maass. On the computational power of winner-take-all. *Neural Computation*, 12 (11):2519–2535, 2015/09/19 2000. doi: 10.1162/089976600300014827. URL `http://dx.doi.org/10.1162/089976600300014827`.

B. Marcus and S. Williams. Symbolic dynamics. 3(11):2923, 2008. URL `http://www.scholarpedia.org/article/Symbolic_dynamics`. revision 143845.

Francesco Masulli, Riccardo Parenti, and Léonard Studer. Neural modeling of non-linear processes: relevance of the takens–mané theorem. *International Journal of Chaos Theory and Applications*, 4(2-3):59–74, 1999.

E. Ott. Controlling chaos. 1(8):1699, 2006. URL `http://www.scholarpedia.org/article/Controlling_chaos`. revision 91167.

Louis M. Pecora and Thomas L. Carroll. Synchronization in chaotic systems. *Phys. Rev. Lett.*, 64:821–824, Feb 1990. doi: 10.1103/PhysRevLett.64.821. URL `http://link.aps.org/doi/10.1103/PhysRevLett.64.821`.

Arkady Pikovsky, Michael Rosenblum, and Jürgen Kurths. *Synchronization: a universal concept in nonlinear sciences*, volume 12. Cambridge university press, 2003.

Nikolai F Rulkov, Mikhail M Sushchik, Lev S Tsimring, and Henry DI Abarbanel. Generalized synchronization of chaos in directionally coupled chaotic systems. *Physical Review E*, 51(2): 980, 1995.

T. D. Sauer. Attractor reconstruction, 2006. URL `http://www.scholarpedia.org/article/Attractor_reconstruction`. Revision 91017.

Steven H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press, 2014.

George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *science*, 338(6106):496–500, 2012.

Floris Takens. *Detecting strange attractors in turbulence*. Springer, 1981. URL `http://bit.ly/takens-1981`.

Jason W Triplett, Melinda T Owens, Jena Yamada, Greg Lemke, Jianhua Cang, Michael P Stryker, and David A Feldheim. Retinal input instructs alignment of visual topographic maps. *Cell*, 139(1):175–185, 2009.

William H Warren. The dynamics of perception and action. *Psychological review*, 113(2):358, 2006.

Hassler Whitney. Differentiable manifolds. *Annals of Mathematics*, pages 645–680, 1936.

Mrubel Commons Wikimedia. A solution in the lorenz attractor rendered as a metal wire to show direction and 3d structure, 2006. URL `https://commons.wikimedia.org/wiki/File:Lorenzstill-rubel.png#/media/File:Lorenzstill-rubel.png`. File: `Lorenzstill-rubel.png`.

Wikipedia. Simple harmonic motion — wikipedia, the free encyclopedia, 2015. URL https://en.wikipedia.org/w/index.php?title=Simple_harmonic_motion&oldid=679436256. [Online; accessed 26-September-2015].