Biologically Plausible Cortical Hierarchical-Classifier Circuit Extensions in Spiking Neurons

by

Peter Suma

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Masters of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2017

Author’s Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Hierarchical categorization inter-leaved with sequence recognition of incoming stimuli in the mammalian brain is theorized to be performed by circuits composed of the thalamus and the six-layer cortex. Using these circuits, the cortex is thought to learn a 'brain-grammar' composed of recursive sequences of categories. A thalamo-cortical, hierarchical classification and sequence learning "Core" circuit implemented as a linear matrix simulation and was published by Rodriguez, Whitson & Granger in 2004.

In the brain, these functions are implemented by cortical and thalamic circuits composed of recurrently-connected, spiking neurons. The Neural Engineering Framework (NEF) (Eliasmith & Anderson, 2003) allows for the construction of large-scale biologically-plausible neural networks. Existing NEF models of the basal-ganglia and the thalamus exist but to the best of our knowledge there does not exist an integrated, spiking-neuron, cortical-thalamic-Core network model.

We construct a more biologically-plausible version of the hierarchical-classification function of the Core circuit using leaky-integrate-and-fire neurons which performs progressive visual classification of static image sequences relying on the neural activity levels to trigger the progressive classification of the stimulus.

We proceed by implementing a recurrent NEF model of the cortical-thalamic Core circuit and then test the resulting model on the hierarchical categorization of images.

Acknowledgements

My sincere thanks go to Dr. Chris Eliasmith for his constant support and encouragement but mostly for his friendship and for lending his intellect at key points in this process. He is the sharpest mind I have ever come across. He is truly an asset both to the field, the University of Waterloo and to this great country, Canada.

I only had the chance for a few brief encounters with Dr. Granger while attending a DARPA summit in Washington early in the course of my studies and then several follow-up emails and a few phone calls. I would like to thank him for the inspirational and motivational presentation he gave of his work at DARPA in which his passion for exploring the workings of the brain was very infectious and truly motivating. I chose to attempt to build a spiking model of his original ground-breaking thalamo-cortical model based on experiencing that presentation as it seemed to me that the combination of his work and Dr. Eliasmith's methods could very well make a significant contribution to the field one day. I hope Dr. Granger and Dr. Eliasmith have a chance to collaborate formally in the future as I accomplished only the smallest hint of what I think is possible with a large-scale integration of Dr. Granger's thalamo-cortical models built on Dr. Eliasmith's Neural Engineering Framework.

To the team at the Centre for Theoretical Neuroscience, my deepest thanks for your patience and encouragement, most especially to my good friends; Dr. Travis DeWolf and Dr. Xuan Choo; your constant cheerful direct explanations and undying patience with me as I took many side paths in an overly exuberant (and ultimately naive) attempt to integrate all of cortical theory, only to come to a very small model of one tiny piece of it, was as beneficial to me as it must have been exhausting to you both! Thank you for not giving up on me!

To Dr. Terry Stewart, of all the people I have met on this journey, Terry has the most remarkable integrated intellect and depth of knowledge for building models using Dr. Eliasmith's methods and has the ability to explain them from so many ways at once, allowing students to understand the concepts regardless of their starting perspectives. To Dr. Trevor Bekolay, if there is anyone who knows the intimate details of Nengo better than you, I have not met them.

I would also like to thank Dr. Rod Rinkus for several informal discussions about the cortex and particularly his work on sparse-distributed-representations. Rod was always kind with his time and so welcoming of my naive questions about the functioning of the cortex. Lastly, Dr. Kwabena Bohen was also so welcoming and friendly when I had the chance to visit his lab a couple of times when Chris allowed me to tag along on a ONR research trips to Stanford. Kwabena's good nature and tolerance of my beginner questions on silicon implementations of spiking circuits was very generous. In addition to my

interactions with my colleagues at the Centre for Theoretical Neuroscience, these brief interactions with Dr. ’s Granger, Bohen and Rinkus, all world-class researchers, made my introduction to the field exciting and motivating, if intimidating. I thank all of them.

Finally, a thank you to the Department of Systems Engineering at the University of Waterloo for allowing me to continue my work on my thesis part time and after several extensions as I restarted my thesis twice!

Dedication

I dedicate this thesis to my family for their support and encouragement to complete a bucket-list item of mine, studying the brain and particularly the cortex, at the expense of many hours that should have been spent with them. Thank you.

Table of Contents

List of Tables

List of Figures

Organization of Chapters

The chapters are organized as follows. For those most interested in purely the computational model, please skip directly to Chapter 6.

Chapter 1, titled “Introduction”, introduces the history of cortical exploration and discovery and the current state of the search for the algorithms and dynamics that govern cortical processing. This situates the work of Rodriguez, Whitson and Granger and their Core and Matrix Model of thalamo-cortical processing in that context, and then introduces this thesis’ work in building a spiking version of the Core circuit model.

Chapter 2, titled “Cortex and Cortical Models”, gives the anatomical and theoretical functional background needed to understand the Rodriguez, Whitson and Granger model (the Granger model) and our NEF model of it. We proceed in this chapter by discussing what we know about the cortex and then review the most prominent relevant models of the cortex.

Chapter 3, titled “Categorization”, discusses computational algorithms of categorization in general, both mathematical and neural algorithms.

Chapter 4, titled “The Granger model”, describes both the Core and Matrix Granger models and the matrix implementation by the authors.

Chapter 5, titled “The Neural Engineering Framework (NEF)”, gives an overview of the NEF and Nengo, the neural modelling software that implements the NEF. This is given in preparation for discussing our implementation of the Granger Core model in the chapter 6.

Chapter 6, titled “The NEF Granger Core circuit”, discusses the implementation of the Granger Core circuit using the NEF that is the subject of this thesis.

Chapter 7, titled “Simulations and Results”, details the model runs performed and the results of those runs.

Chapter 8, titled “Discussion and Future Work”, reviews this work in the context of cortical theory and describes some future work ideas that arise from this effort.

Devotion to the cerebral hemispheres, enigma of enigmas, was old in me... the supreme cunning of the structure of the gray matter is so intricate that it defies and will continue to defy for many centuries the obstinate curiosity of investigators. That apparent disorder of the cerebral jungle, so different from the regularity and symmetry of the spinal cord and of the cerebellum, conceals a profound organization of the utmost subtlety which is at present inaccessible.

Santiago Ramón y Cajal, 1937

All of this does not alter my belief that a new essentially logical theory is called for in order to understand high-complication automata and, in particular, the central nervous system. It may be, however, that in this process logic will have to undergo a pseudomorphosis to neurology to a much greater extent that the reverse.

John Von Neumann, 1958

1. Introduction

The human neocortex is among the least understood areas of the brain. One of the brain's most celebrated explorers, Ramon y Cajal called it the "impenetrable jungle" (Cajal, 1937). The computational processes occurring in the neocortex are an area of open research. We lack good, integrated theories of exactly what computations are occurring in the neocortex. We also do not yet understand how the cortex computationally integrates to the sub-cortical structures such as the thalamus, the striatum and the basal ganglia to produce intelligence. Nor do we have good theories that explain how goal directed behaviour emerges from these structures.

The cortex is thought to be a "vast memory of unknown capacity that contains literal and abstract representations of our life experience, as well as prescriptions of what to do about them" (Ballard, 2015, p. 63). All complex computational systems invariably employ some form of hierarchical computational abstraction (Ballard, 2015, p. 38) even if for no other reason than to perform computation on large input and output spaces using limited computational resources. Some form of encoding must be used, and it invariably requires the ability to learn and compute with these encoded abstractions organized into layers. Any theory of how cortex works must explain how the cortex learns and processes these hierarchical abstractions (Valiant, 2014).

Some of the first evidence of how hierarchical abstractions are implemented in the brain came from Hubel and Wiesel (1962) who showed that primary visual cortex neurons respond to edge-like stimuli instead of the circular receptive fields of the lateral geniculate nucleus neurons that feed them. This led them to propose that the primary visual cortex neurons were "feature combination neurons" and that the brain uses a hierarchical computational architecture at very least in the visual system, if not overall.

The cortex is composed of three classes of cells; inhibitory neurons, and excitatory neurons, and glia. Excitatory cells compose an estimated 80% of the cortical neuron cell population while inhibitory cells compose the remaining 20%. Korbinian Brodmann stained cortex and observed that cortical cell bodies are arranged in six bands or layers, numbered from I to VI starting from the surface of the cortex and moving inward (Brodmann, 1909). Vernon Mountcastle's landmark review (Mounstcastle, 1997) summarized most of the work on vertical structures in rat cortex since the discovery of vertical cylinders in the cortical layers by Lorente de No (1938). Mountcastle also was one of the first to note that the functional properties of cortex change laterally but not vertically (Mountcastle, 1957). Areas of cortex that have the same function were named columns (also known as hyper-columns in visual cortex, and segregates in somatosensory cortex) and were found to be composed of between 60 and 80 "mini-columns" (Favorov & Kelly, 1994). A mini-column is a vertical cylinder stack of the six

cortical layers containing approximately 80 to 120 neurons in total where the neurons in the mini-column all have similar or the same receptive fields. Axons and dendrites of neurons intermingle across mini-columns as well as connect within columns. The nature connectivity is still a subject of research with physical proximity being solely insufficient to explain connectivity (Kasthuri, et al., 2015) while there is recent evidence of highly non-random connectivity such as clustered connections and log-normally distributed synaptic cortical connection strengths (Song, Sjostrom, Reigl, Nelson, & Chklovskii, 2005). While widely cited, it also should be noted that it is still a subject of debate if columns, macro or mini, even exist in cortex generally.

There are many levels of physical and logical hierarchy in the cortex. The cortical sheet has a laminar structure. Inter-connected cortical columns are connected starting with incoming primary sensory columns that project to columns that process higher-level abstractions. Computation is thought to progress also in a series of hierarchically organized phases. Specifically, the circuits connecting the thalamus and the six-layer cortex are theorized to perform iterative, hierarchical-categorization inter-leaved with sequence recognition (Rodriguez, Whitson, & Granger, 2004).

One model of this hierarchical temporal processing in cortex has been advanced by Rodriguez, Whitson and Granger in a 2004 paper entitled “Derivation and analysis of basic computational operations of thalamo-cortical circuits” (Rodriguez, Whitson, & Granger, 2004). Granger was the supervisor on the paper and the common author of the papers that follow in the series (Ambros-Ingerson, Granger, & Lynch, 1990; Anton, Lynch, & Granger, 1991; Coultrip & Granger, 1992; Granger R. , 2006; Lynch & Granger, 2008; Granger R. , 2011; Chandrashekar & Granger, 2012; Rodriguez & Granger, 2016).

Granger is also the lead proponent of the theory of brain grammar that we shall be investigating. We shall refer to the model herein as “the Granger model” for brevity. The origin of the model traces back to 1990 and 1991 where the process by which the olfactory cortex would respond to stimuli and then subtract from the input the portion of the stimuli that the superficial layer of cortex responded to. The circuit would then reprocess the result. This was described in two papers (Ambros-Ingerson, Granger, & Lynch, 1990; Anton, Lynch, & Granger, 1991). The model was then generalized by Granger and Lynch (Granger & Lynch, 1991).

Granger proposes that the brain computes with a recursive grammar of sequences of categories implemented in two recurrent circuit motifs; the core and the Matrix circuits. The neural correlates of these circuits are found in the many thalamo-cortico-cortical loops in the brain. Using these circuits, the cortex is thought to learn a ‘brain-grammar’ composed of iterative, recursive sequences resulting from such hierarchical categorizations. These functions are thought to be implemented by the recurrently connected, spiking neurons of cortical and thalamic circuits. In the paper, the authors propose “a simplified anatomically based model of topographically and non-

topographically-projecting (respectively the core and Matrix circuits) thalamic nuclei and their differential connections with superficial, middle, and deep neocortical laminae" (Rodriguez, Whitson, & Granger, 2004). The thalamo-cortical, hierarchical-classification and sequence-learning circuit is referred to as the 'Core' circuit and was implemented as a linear matrix simulation in the Rodriguez et al. paper.

The model focuses on the thalamo-cortical circuits in the posterior neocortex that are driven by and reciprocally connect to the nuclei of the dorsal thalamus. Input arising from areas of cortex lower in the processing hierarchy, innervates topographically organized thalamus which in turn drives topographically organized cortical circuits.

This Core circuit is responsible for hierarchical classification of its inputs, be they sensory or higher order abstractions. The Matrix circuit is another cortico-thalamic loop in which the outputs of the Core circuit drive layer V cortical cells that in turn drive non-topographically organized thalamic circuits that learn sequences of the Core circuit's outputs. These two circuits work together to output temporal-spatial, hierarchical and sequential representations in the cortico-thalamic loops that form the brain grammar in time that Granger proposes.

The computational model of the Core and Matrix circuits built by the authors was developed in MATLAB and is a vector-based, non-spiking model. This thesis seeks to reimplement that model in spiking neurons using the Neural Engineering Framework ("NEF") (Eliasmith & Anderson, 2003). The NEF is a mathematical method characterizing how biological systems encode, into representations, quantities from the environment, and then transform those representations leading to behaviours. The NEF allows for the construction of large-scale biologically plausible neural networks.

The central question for this thesis is: can we construct a more biologically plausible version of the hierarchical-classification function of the Core circuit this using leaky integrate-and-fire neurons that performs progressive visual classification of static images? The use of spiking neurons and a categorization process driven by the neural activities themselves, rather than relying on procedural algebra in code, increases the biological plausibility of the resulting model.

We first present the NEF, specifically the parts of it relevant for our model. We then discuss implementing a recurrent NEF model of the Core circuit using the existing Nengo basal ganglia and thalamus networks, and then discuss the results of running the model to hierarchically classify single, and then multiple, MNIST digits.

The contribution of this thesis is that while existing NEF models of the many parts of the brain exist, such as the basal ganglia and the thalamus, there does not exist an integrated, spiking-neuron, cortico-thalamic-cortical circuit model implementing the Rodriguez et al. Core circuit model.

The circuit, as implemented, performs a small version of competitive feature detection, but is less complex than many of the other implementations that use deep learning for feature detection. The point of interest here lies in the other aspects of the circuit. The temporal dynamics that produce the flow of categories and sequences into the thalamo-cortical circuits. Specifically, we are interested in how these multiple representations are, at the same time, both categorized and temporally organized. In the operation of the circuit the hierarchical category representations are serialized and then the transition dynamics from both the changing categorizations and changing visual images that make up the visual scene are learned by the neurons in the circuits of layer V of cortex and the non-topographic thalamus. This process may partly explain how we represent simultaneously concepts at multiple levels (both the generic category and the specific) at the same time and are able to attend to both conceptual levels at once. In this thesis the focus is on a model that captures the recognition and categorization of hierarchies, represented as sequences of categories and sub-categories, not just static pattern classification. Rodriguez et al.'s work is among the best circuit theories for explaining these phenomena.

To situate the work of building the hierarchical classification Core circuit (Rodriguez, Whitson, & Granger, 2004) in spiking neurons in this thesis we will attempt to survey the history and current state of the theories of cortical processing, narrowed to those that propose dynamical, stateful processing architectures. The model of the Core circuit will then be discussed within the context of this larger discussion. Finally, extensions to the model will be suggested in the Future Work section that present the opportunity to test these larger more integrated theories.

2. Cortex, Thalamus and Thalamo-Cortical Models

Which way should we begin to describe a system founded on feedback loops whose ultimate functioning is theoretical? All options will lead to things being referenced before they are explained.

In trying to lay it out as simply as possible, we start with a classic overview of the cortex's layered structure and then proceed to discuss thalamic anatomy and function in detail, then describe major brain circuits, and finally discuss computational models of the cortex. From there we are in a position to review algorithms of categorization, and then the specifics of the Rodriguez, Whitson and Granger model and finally our implementation of it.

Readers who wish to skip the literature review are advised to go directly to the presentation of the Granger model Chapter 4 on page 70.

2.1. Cortex

"The cortex is the crowning achievement of brain evolution." (Cajal, 1899)

The brain is thought to have emerged from a clustering of neurons in the front end of the body of the flatworm Planaria about 560 million years ago while the cortex is thought to have emerged as a fold with six layers about 200 million years ago (Plebe & De La Cruz, 2016, p. 24). When considering the history of scientific discovery of the cortex, as with many writings on the brain, it is hard not to begin by mentioning the amazing scientific work done by Ramon y Cajal (Cajal, 1909). Cajal is famous for his astounding drawings of neurons and brain tissue but also the creation of the neural doctrine. The neuron doctrine is the concept that the neuron is the central computational unit of cognition. This doctrine has guided much of neuroscience, computational neuroscience and artificial intelligence research ever since. The first major detailed mapping of cortex occurred in the middle of the 19th century when Meynert labelled the layers he observed under his microscope starting from the outside of cortex inward based on similar morphology as layers I through V. After Meynert a series of authors repeated this kind of work, each coming to slightly different names.

2.1.1. Cortical Layers and Regions

Brodmann (1902) (*25)	Meynert (1868)	B. Lewis (1878) (*27)	Betz (1881)	Hammarberg (1895)	R. y Cajal (1902)	Campbell (1905) (*28)	Mott (1907)
I. Lamina zonalis	1. Molecular layer	1. Cell-poor layer	1. Neuroglial layer	1. Plexiform layer	1. Plexiform layer	1. Plexiform layer	1. Zonal layer
II. L. granularis externa	2. Outer granular layer (*26)	2. Small pyramids	2. Small pyramids	2 & 3. Pyramidal layer	2. Small pyramids 3. Medium pyramids	2. Small pyramids 3. Medium pyramids	2. Small, medium and large pyramids
III. L. pyramidalis	3. Pyramidal layer	3. Large pyramids	3. Medium and large pyramids		4. Large pyramids	4. Large pyramids	
IV. L. granularis interna	4. Inner granular layer	4. Inner small pyramids	4. Granular layer (*26)	4. Small irregular cells	5. Granular layer	5. Stellate cells	3. Granular layer
V. L. ganglionaris	5. Spindle cell layer	5. Ganglion layer	Spindle layer	5. Ganglion layer	6. Deep, medium pyramids	6. Inner, large pyramids	4 & 5. Inner line of Baillarger and polymorph layer
VI. L. multiformis		6. Spindle layer		6. Spindle cell layer	7. Spindle cells	7. Spindle cells	

Table 1: General basic laminar patterns according to various authors (Brodmann, 1909, p. 15)

At the beginning of the 20th century Korbinian Brodmann stained the whole of cortex and divided it into many contiguous regions of homogeneous layer arrangements based on whether the cellular layers had the same structure and appearance under staining. This produced a patchy map of cortex that today is one of the primary ways of looking at cortical areas and attempting to assign functionalities to them. Broadman, like Meynert, observed that cortical cell bodies are arranged in up to six bands or layers numbered from I to VI starting from the surface of the cortex and moving inward (Brodmann, 1909). The modern view of the layers and their proper names are shown below.

Number	Name	Contains
I	Molecular supragranular layer	Dendrites of cells from deeper layers and axons connecting with them.
II	External supragranular cell layer	Granule cells which are small and spherical.
III	External supragranular, pyramidal Layer	Many different cells types including pyramidal shaped cells.
IV	Internal Granule Cell Layer	Mostly granule cells but also pyramidals. Primary connection area of thalamic axons.

V	Infragranular Pyramidal Cell Layer	Mostly large pyramidal cells. Sends many projection s to non-specific thalamic nuclei and may be a major source of cortico-thalamic-cortico and cortico-cortico connections, feeding information hierarchically from one cortical area to another both directly and via the thalamus.
VI	Infragranular, polymorphic or Multiform Layer	Heterogeneous neural cell types, filed with axons connecting the sub-cortical white matter to cortex.

Table 2: Cortical layers names and cellular contents (Strack, 2013).

Recently a new map based on many functional MRI studies has been published, aggregating areas of cortex that are activated by similar tasks and/or physiological behaviour (Glasser, et al., 2016). It is shown below, next the Brodmann map from 1909.

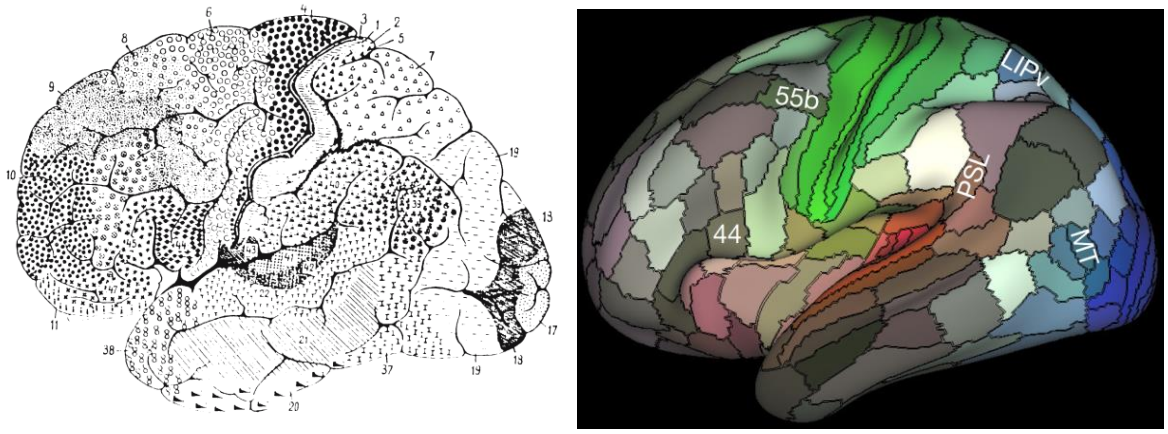

Table 3: 3D Color-coded Brodmann map of cortex on left *(Brodmann, 1909)*, new functional map on right *(Glasser, et al., 2016)* [Reprinted by permission from Macmillan Publishers Ltd: Nature copyright 2016].

It appears that Brodmann was correct (Glasser, et al., 2016). Cortex is in fact a sheet of differing functional areas correlated somewhat with macro-anatomical differences. This movement from trying to explore the cortex anatomically to functionally is today powered by new tools for scanning cortex, but also by many new theoretical models of how cortex functions.

Our goal here is to study one of the core functions that cortex performs in making sense of the world, that of grouping incoming stimulus to identify categories of things, and to classify them in related ways, to generate a hierarchical classification map of the environment and the associated mental concepts.

To begin to explore how these connected layers might perform hierarchical classification we next turn to reviewing our understanding of the vertical functional structure of cortex

in the form of the theory of cortical columns, but first we briefly review the cell types found in cortex.

2.1.2. Cortical Cell Types

The cortex is composed of three classes of cells, inhibitory neurons, excitatory neurons and astrocytes. Excitatory cells compose an estimated 80% of the cortical cell population while inhibitory cells compose the remaining 20%. There are many sub-types for each of the excitatory and inhibitory cell types. A one-millimetre square surface area of cortex (excluding primary visual cortex which has 2.5 times as many) on average in mammals contains between 100,000 and 150,000 cells composed of neurons and glia (Carlo & Stevens, 2012). In the rat, mouse, cat and monkey, across four different cortical areas, approximately 94,000 ± 10,700 were found to be neurons (Carlo & Stevens, 2012).

Excitatory neurons receive relatively sparse afferents from other excitatory cells with the prevalence of inter-connection dropping off quickly with the distance between any two such neurons. There is a ten percent (10%) chance of synapsing if they are 0.2 mm to 0.3 mm apart but only a less than 1% chance if they are less than 1 mm apart (Rodriguez, Whitson, & Granger, 2004). Excitatory neurons release glutamate while the inhibitory interneurons release GABA. Inhibitory interneurons project their axons within the local circuit whereas the excitatory interneurons can project much further.

Inhibitory interneurons typically terminate on the soma or proximal dendrite of excitatory pyramidal cells giving them powerful influence on the circuit's behaviour. Inhibitory neurons' effects also typically last longer, approximately 100-150 milliseconds whereas excitatory neurons effects last typically 15 to 25 milliseconds. Excitation is generally briefer and sparser topographically while inhibition is generally stronger and longer lasting but more topographically concentrated. The current state of research into excitatory and inhibitory neural function is though more complex than we have described. There are several subclasses of GABA receptors, including, $GABA_a$, $GABA_b$. and $GABA_c$ Some have slower, and some have faster, time constants than excitatory neurons. For the Granger model and our implementation of it, only simple inhibitory and excitatory neurons are used so we will not delve further into the subtypes here.

For our purposes, our model will include analogues of several of these cell types including the most common neuron in cortex, the excitatory pyramidal layer IV neuron, which is predominately stimulated by the spiny stellate cells of layer IV which are in turn driven by thalamic axonal projections from topographic, specific thalamus. The Granger model contains these projections. We will also model the action of the layer II and III pyramidal cells as well as the layer V pyramidals, including their inhibition inducing back projections to the thalamus.

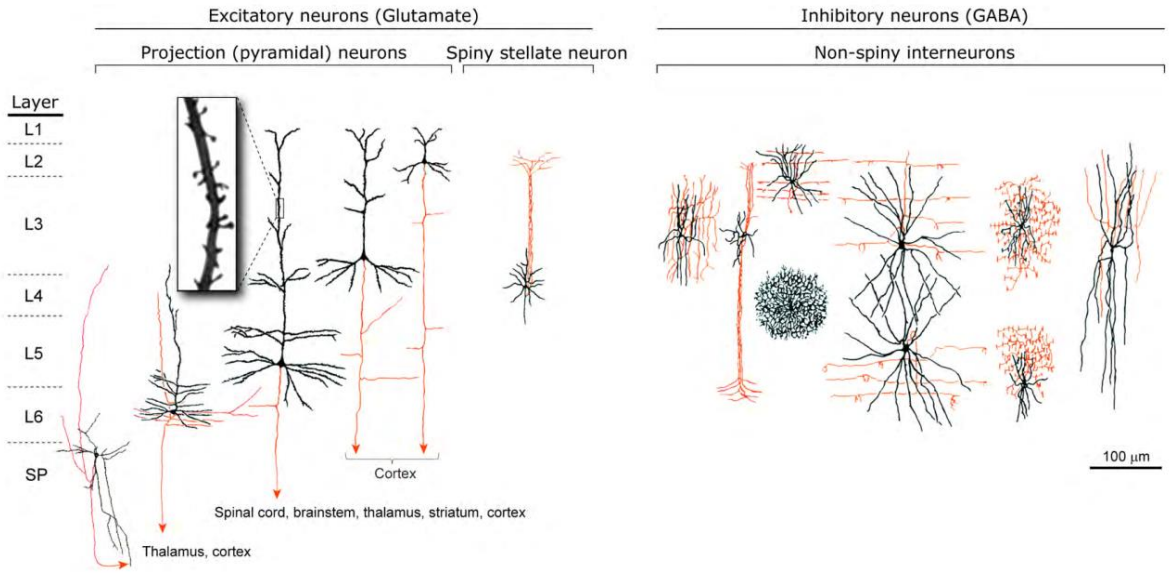

Figure 1: Major excitatory and inhibitory mammalian cortical cell types. Axons are in black and dendrites are red. Republished with permission of [The Biologists Company Inc.], from (Kwan, Šestan, & Anton, 2012), permission conveyed through Copyright Clearance Center, Inc.].

2.1.3.Cortical Macro & Mini Columns or Swirling Slabs

Vernon Mountcastle's landmark review (Mounstcastle, 1997) summarized most of the work on vertical structures in the cortex since the discovery of vertical cylinders in mouse cortical layers by Lorente de No (1938). Mountcastle's review included his physiological and anatomical work and of major contributors such as David Hubel & Torsten Wiesel (Hubel & Wiesel, 1962) and János Szentágothai, Ted Jones, and Pasko Rakic (1988). It is worth noting that topographic columns as an organizing motif for receptive fields has been found not only in mammals but also in birds (Wang, Brzozowska-Prechtl, & Karten, 2010).

Mountcastle also was one of the first to note that the receptive fields of sensory cortex change laterally but not vertically (Mounstcastle, 1997). Areas of cortex that have the same receptive fields were named "columns" (also known as hyper-columns in visual cortex, and segregates in somatosensory cortex) and were found to be composed of between 60 and 80 "mini-columns" (Favorov & Kelly, 1994). A mini-column is a vertical cylinder stack of the six cortical layers containing approximately 80 to 200 pyramidal cell bodies plus other neurons and glia.

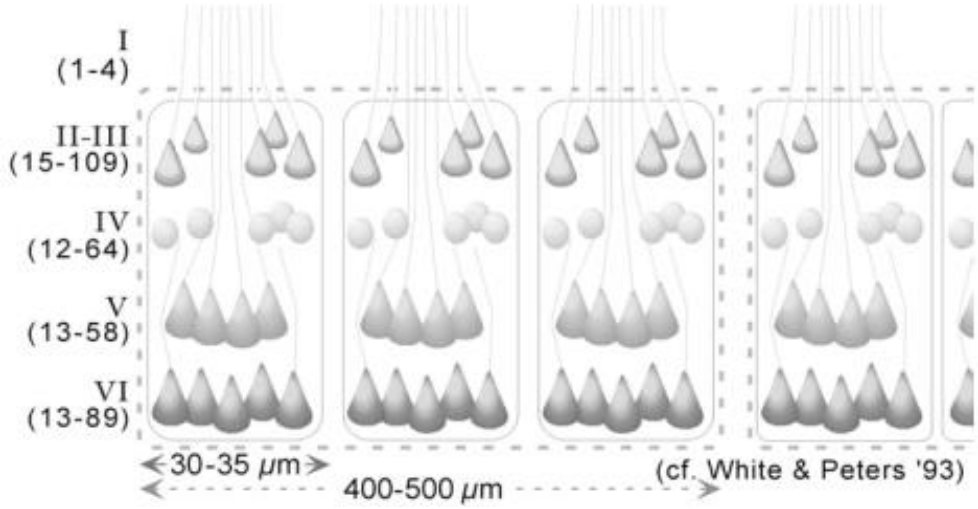

Figure 2 : Schematic of cortical mini-columns (solid lines) grouped into cortical modules (separated by dashed lines) from the Granger model showing the pyramidal neurons cell bodies; from (Rodriguez, Whitson, & Granger, 2004), reprinted with permission from MIT Press open access for academic use license. Note: cortical lateral connections not shown.

Functional mini-columns are physiologically defined by their receptive field properties being similar whereas architectonic mini-columns are defined by the fact that they have a topographical correspondence with the layout of their connected thalamic neurons (Rodriguez, Whitson, & Granger, 2004). A mini-column is approximately 50 micrometres across (L.Feldman & Peters, 1974). The cortex has about 100 million mini-columns (Krueger, et al., 2008). The axons and dendrites of neurons intermingle across mini-columns, although neurons in the same mini-column are more likely to be connected than those from different mini-columns.

It is worthwhile to note that the cortical column is a major misnomer. “Swirling parallel slabs” would be a much more accurate description; taking Hubel & Weisel’s (1972) own description of the actual slab shape they described in their landmark “ice-cube” model of visual cortex. This is supported by more modern views of the swirling, non-straight-edged receptive field delineation patterns seen in cortex by Costa and Martin (2010). The column name however, once coined, has stuck powerfully in the vernacular of students and scientists alike until the present day. To summarize this discussion, Roy (2017) quotes Horton and Adams (2005) “The columnar organization hypothesis is currently the most widely adopted to explain the cortical processing of information”. Herein we also refer to these groupings of related receptive field properties as columns as well.

A related set of hypotheses about how cortex develops a layered, columnar architecture is the ProtoMap (PM) and Radial Unit Hypothesis (RUH) which were both advanced by Pasko Rakic (1988) from Yale. The PM postulates that cells in an area of the developing cortex, referred to as the ventricular zone, contain a genetically specified map of molecular markers that determine the regional functional structure of cortical patterning as observed in the mature cortex by Brodmann (Brodmann, 1909) and the columnar architecture observed by Lorente de Nó (1934). The RUH states that each of the cells in

the protomap then migrate along routes laid out by radial glial cells and influenced by epigenetics to form the columnar map, after which the cells divide and populate the layers of the cortex in an inside out manner starting with the deep layers. Projections from the thalamus then innervate their specific targets, guided by highly selective topographic molecular maps and cortico-cortical connections that form laterally, in a process increasingly influenced by activity during development. This process culminates in the thalamo-cortical loops that the Granger model is concerned with explaining the functioning of.

Columns are very hard to distinguish anatomically. From receptive field studies of neural response to stimuli the architecture of the columns has been elucidated to some extent although it varies from area to area of cortex. Rakic's RUH model is shown next to one of Brodmann's original stains of cortex and one of Buzaki's drawings of the cellular connectome of a cortical column to put the problem of deducing function from form into context.

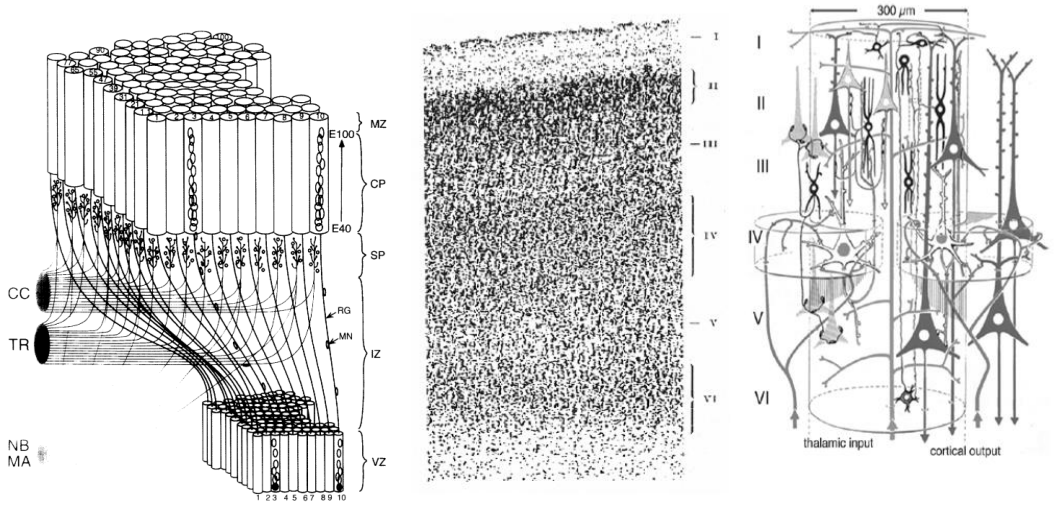

Figure 3 : Left: Schematic of Radial Unit Hypothesis of neural migration into columns, reprinted from (Rakic, 1995) with permission from Elsevier; Middle: Cortex from human foetus showing basic cortical layers after RUH development (Brodmann, 1909, p. 20); Right: Canonical cortical column layers, cell types and connections (Buzsaki & Draguhn, 2004), reprinted with permission from AAAS.

The large triangular-shaped cells in the Buzsaki diagram are the pyramidal cells, the round cells are the stellate cells. In general, the pyramidals are excitatory cells and the stellate cells are inhibitory. The connections and the dynamics through the network and how the cortical cells are connected to the thalamus are the subject of the Granger model.

'Mini-columns' and 'macro' or 'cortical' columns are organized by various large-scale features at the macro column level, along with other factors which vary vertically in the mini-column. As an example, in MT, Diogo et al. report that "cells with a similar direction of motion preference are also organized in vertical columns and cells with opposite direction preferences are located in adjacent columns within a single axis of motion

column” (Diogo, Soares, Albright, & and Gattass, 2002). These cells have been categorized horizontally by the direction of motion and then vertically by other factors.

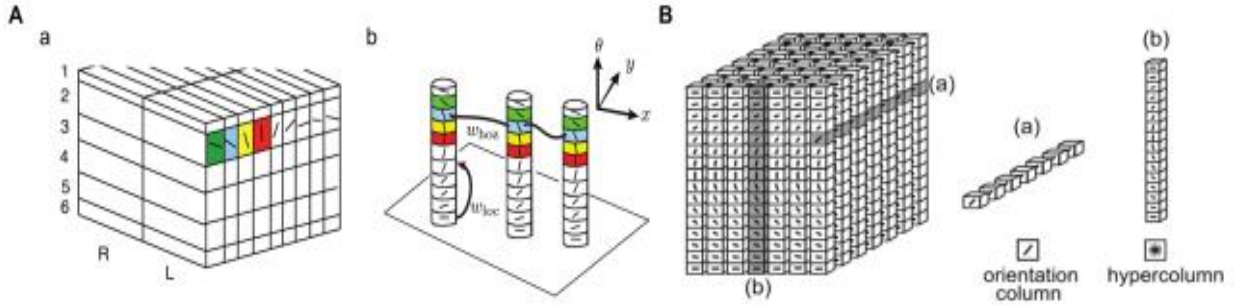

Figure 4: Orientation of motion columns, hyper-columns and layers in striate cortex of Cebus apella monkey (Roy, 2017), reprinted with permission through Frontiers in Psychology Open Access License.

Beginning with Mountcastle (1957) cortex was found to have spatially organized receptive fields; in his seminal work cat and monkey somatosensory cortex was found to have radially organized receptive fields. The Granger model theorizes that categorization is spatially encoded as well across columns.

Physical or Functional Columns?

Are these columns then a physical and corresponding functional organization? Despite the long tradition and evidence from receptive fields, the actual physical existence of columns as distinct physical gaps in the connections between cortical neurons is disputed (Horton & Adams, 2005). As Costa and Martin put it:

> *These tissue compartments appear to divide cells with different receptive field properties into distinct processing streams. However, it is unclear what advantage, if any, is conveyed by this form of columnar segregation. Although the column is an attractive concept, it has failed as a unifying principle for understanding cortical function. Unravelling the organization of the cerebral cortex will require a painstaking description of the circuits, projections and response properties peculiar to cells in each of its various areas. (Costa & Martin, 2010)*

Costa goes on to say though that whether there is a physical delineation called a column, he does not dispute that there is a functional grouping called a canonical circuit and that that is the fundamental building block of cortical computation.

For our work in this thesis this is the important point, whether the cortical column exists at all anatomically, or only in sensory cortex (where the breaks in lateral connectivity are most easily observable) or if it in fact exists everywhere, is to some extent not the point for us. The important point is that cortex has a repeating functional circuit that implements receptive selectivity in a categorized, local way with some topographical and vertical ordering in perceptual areas. So long as that is the case then the idea of theorizing and

modelling repeated topographic and vertical circuits that might implement hierarchical categorization is a valid pursuit.

Alternatively, it is possible that cortical columns are dynamic units and not necessarily fixed anatomically. The cortex is quite uniform in structure and if columns are more a functional construct they could appear anywhere in the cortical realm. As an example, the network of pyramidal axons and dendrites in cortex has been shown to have random connection points to other neuron types, pyramidal or otherwise (Kalisman, Silberberg, & Markram, 2005). Although there is random points of intersection, the functional connectivity has been shown to exhibit lognormal distributions suggesting that synaptic strengths are concentrated among much fewer synapses than a purely random connectivity would suggest (Song, Sjostrom, Reigl, Nelson, & Chklovskii, 2005).

If functional columns do exist, then perhaps they emerge on top of this random connectivity by plasticity forming the non-random synaptic strength distributions seen. We know, as discussed above, tuning curves do vary in a way that suggests functional cortical columnar organizing principles. Taken together this has been said to support the idea that the cortical column is a functional concept that emerges and is maintained dynamically over the matrix of the cortical layer architectonics (Markram, 2008). This is thought to happen, for example, by virtue of the pyramidals connecting and turning “on” the synapses that serve the computational needs of the circuit that is forming and turning ‘off’ others (Markram, 2008).

This entire functional circuit is what Rodriguez, Whitson and Granger (2004) suggest implements hierarchical categorization and whose activity in the form sequences of categorical representations forms the Brain Grammar (Rodriguez, Whitson, & Granger, 2004). This theorized circuit is the subject of this thesis. The first half of this circuit, from Thalamus to layer IV, layer II & III, then layer V and then layer VI and then back to the thalamic reticular layer is referred to by Rodriguez, Whitson and Granger as the Core circuit and the remaining circuit to non-specific thalamus is the Matrix circuit. The Core circuit is thought to perform hierarchical categorization while the Matrix circuit is thought to compute sequences over those categories from the Core circuit.

A point then which is relevant for our model, arises in Markram’s paper in relation to the Granger model; the punctuation of input from the thalamic primary sensory inputs to cortex could explain why we see evidence of what appear to be physically separate cortical columns but also much evidence for these columns being dynamically formed. If input to cortex is sent in punctuated bursts which are also topographically patterned then the plasticity mechanisms would produce a functional column the size of the “overlapping basal axonal and dendritic arbours of the neurons at the centre of the stimulation” (Markram, 2008). This is relevant to our model in that punctuated inputs, topographically and temporally separated is what you would expect from serially cascading, hierarchically ordered categorization per the Granger model. While not definitive, compared to the

original view of the cortex as more monolithic and more feed-forward; here we see evidence of temporal and spatial bursting of inputs to cortex, which as we will see below, there are plausible mechanisms in the thalamic circuits to account for such behaviour.

We discuss much more on the temporal flow of processing in thalamo-cortical circuits later after we review the relevant structure of the thalamus and the circuits that the thalamus participates in with the cortex. For the purposes of understanding the model, the general relationship between the thalamus and the cortical layers in terms of general connection flow is the most important to discuss which we do in the next section.

2.2. The Thalamus

To understand categorization in the cortex we need to first understand the functioning of the thalamus, the driver of many inputs to the cortical circuits in question. First, we examine the anatomy of the thalamus.

> *"Faced with an anatomical fact proven beyond doubt, any physiological result that stands in contradiction to it, loses all its meaning...So, first anatomy and then physiology; but if first physiology, then not without anatomy."* (Brodmann, 1909)

The diencephalon sits between the brain stem and the cerebrum, and processes a range of sensory and motor signals. The diencephalon is composed of the hypothalamus, subthalamus, thalamus and epithalamus. The thalamus (aka the dorsal thalamus) is composed of a series of nuclei together sitting atop the spinothalamic tract and reciprocally connected to many structures, including, the cortex, the basal ganglia, the cerebellum, the reticular formation, and the hippocampus. It has been implicated in a wide range of behaviours including, sensory relay of all sensory pathways except olfaction; pain sensation; the maintenance of consciousness and alertness; control of muscular movement; management of the sleep/wake state; personality and social behaviour modification; language and speech and sexual behaviours (Das, 2017). The dorsal thalamus influences what information reaches the cortex and it is this function that we are especially concerned with in the Granger model.

2.2.4. Dorsal Thalamic Anatomy

Anatomically, the dorsal thalamus is composed of two lobes, one in each hemisphere, often connected by an inter-thalamic adhesion, which is not present in all species or individuals. The lobes are composed of a series of nuclei divided internally by vertically oriented, y-shaped sheaths of white-matter axonal tracts, the internal medullary laminae. The lobes are surrounded by the external medullary laminae which are composed of

thalamocortical and corticothalamic fibres which ascend from the thalamus as parts of the internal capsule to the cortex. Three regions in each lobe of the thalamus are formed by the areas, the anterior, lateral and medial regions, all segmented by the y-shaped internal medullary laminae's.

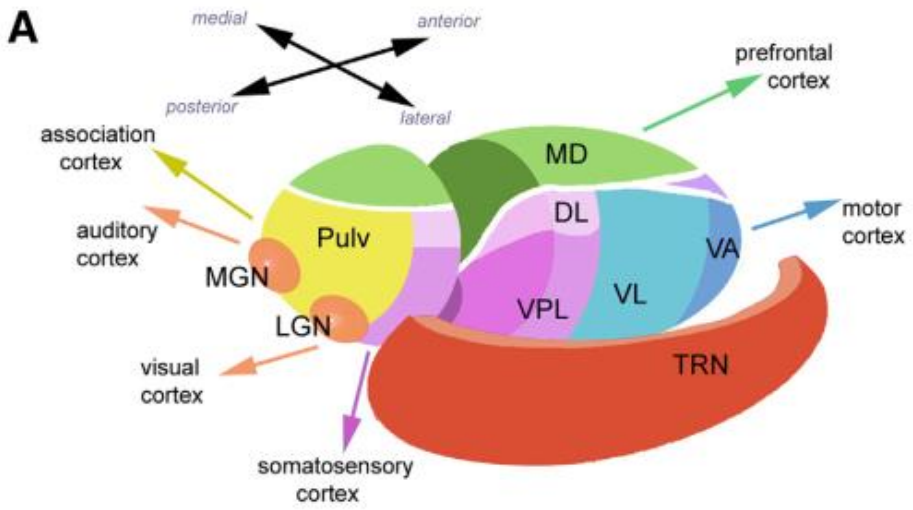

Figure 5: Main anatomical connections of the Thalamus (Prattl, et al., 2016) reprinted with permission from Elsevier.

There are approximately 50 nuclei (Jones E. G., The Thalamus, 1985) in the thalamus, the major ones are shown in Figure 5. For a long time, it was believed that no connections existed between the various nuclei of the dorsal thalamus but that has since been shown to be incorrect (Jones E. G., The Thalamus, 1985). Within each nucleus there are many inhibitory interneurons most of which are connected to the various cell types of the nuclei with dendro-dendritic synapses.

The Granger model covers the dorsal thalamic nuclei and predominately the posterior cortex's pulvinar. As shown in Figure 6, these nuclei project to the posterior cortex which houses the "association areas" found in the posterior reaches of the temporal, occipital and parietal lobes.

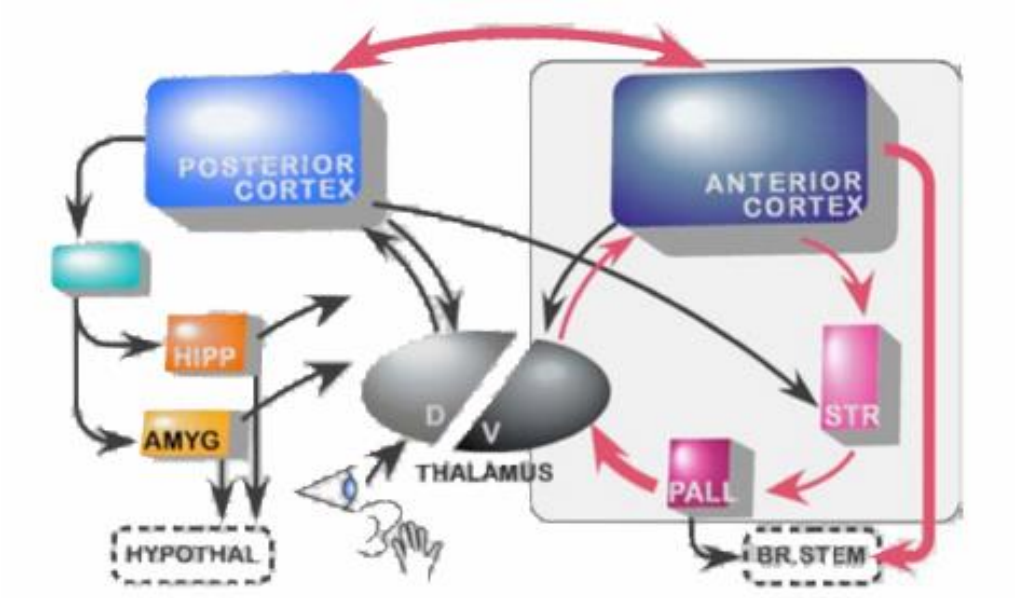

Figure 6 : Dorsal thalamus and posterior cortex connectivity, the focus of the Rodriguez, Whitson, Granger model (Granger R. , Engines of the brain: The computational instruction set of human cognition, 2006) reprinted with permission from MIT Press.

Most of the neurons in these nuclei are excitatory and utilize the neurotransmitter glutamate as do most of the neurons in the thalamus with the exception of the thalamic reticular nucleus cells (TRN) which uses GABA. The various nuclei of the thalamus are connected to all parts of the ipsilateral cortex and many parts of the sub-cortical structures as shown above. In general, all thalamic neurons receive feedback projections from the same area of cortex that they project to (Swenson, 2006).

2.2.5. Dorsal Thalamic Function

The thalamus is central to the functioning of the brain. The rhythmic EEG pattern read from the scalp is actually generated by the synchronized waves of excitatory spikes from the thalamus to the cortex (Kandel, Schwartz, & Jessell, 2000). Further supporting the idea that the thalamus is a central hub of the entire cortical brain, recent graph theoretic analysis of thalamic circuit activations suggest that the thalamus is globally connected to all the cortical regions and has the network properties needed to integrate multi-model information from these dynamically connected functional networks (Hwang, 2016).

The thalamus' prominent position between most of the incoming sensory signals from the body, the cortex, the hippocampus and the basal ganglia, originally led to it be hypothesized as being responsible for purely routing information; subsequently it has been shown that the thalamus processes and relays many kinds of information as well as helps to regulate alertness and the sleep-wake cycle (P.Tinuper, et al., 1989). The thalamus has been theorized to be involved in the maintenance of arousal, selective attention, the production of sleep spindles, and seizure production in disease states (Willis, Slater, Gribkova, & Llano, 2015).

The four basic kinds of information are thought to be processed in the thalamus are; sensory, motor, emotional memory, and alertness. All incoming sensory information (except olfaction) is relayed to the cortex via the thalamus. A major pathway of the motor system passes from the cerebellum and the basal ganglia to the primary motor cortex via the thalamus. The thalamus is also part of the Papez circuit and helps gate emotional memory information passing to the limbic cortex, and gates signals controlling the sleep/wake cycle. The signals controlling the level of alertness and attention are also thought to pass from the reticular formation through the thalamus on the way to cortex.

There have even been theories advancing the thalamus as the centre of consciousness (Jones E. G., 2002) given its central position in the brain between the reticular-activating system, which controls the state of the brain overall from alertness to sleep, and the rhythmic activity of the thalamo-basal-cortical loops which is theorized to be a major correlate of thinking in the brain.

2.2.6. Relay, Association & Non-Specific Thalamic Nuclei

The dorsal thalamic nuclei are functionally divided into three groups; the relay nuclei, the association nuclei and the non-specific nuclei.

Relay cells / Specific Nuclei. These ventral and lateral nuclei receive signals from well-defined sensory and motor pathways such as the inferior colliculus (auditory), retina (visual), medial lemniscus, as well as the spinothalamic and trigemini-thalamic tracts (somatosensory). They also project topographically to defined regions of the cortex such as the temporal, visual, and primary somatosensory cortices.

The specific nuclei include those involved in sending such primary sensory information to cortex, such as the ventral posterolateral (VPL), the ventral posteromedial (VPM), the medial geniculate (MGN), and the lateral geniculate nuclei (LGN). Other specific nuclei include those feeding back cerebellar signals, such as the ventral lateral (VL) nuclei and nuclei sending the output of the basal ganglia to cortex found in parts of the VL and the ventral anterior (VA) nucleus (Swenson, 2006).

Relay cells in specific nuclei are very important and are theorized to control the structure of cortical receptive fields (Kandel, Schwartz, & Jessell, 2000) by determining which cortical neurons are activated for incoming sensory stimuli, and then, those cortical neurons learning via plasticity to alter their receptive fields. Cortico-thalamic afferents also play a role in receptive field development, in this case for the thalamic neurons. Feedback brought by these afferents is not very strong but has been shown to influence the receptive fields of thalamic neurons (Briggs & Usrey, 2008).

Association Nuclei: These nuclei, such as the pulvinar, the dorsomedial nucleus (DM), and the anterior nucleus (Swenson, 2006) connect to the prefrontal, parietal, occipital and temporal association cortices, each projecting to many different areas of cortex. These nuclei also receive the largest inputs directly from cortex. The pulvinar projects to secondary visual areas, relaying information from the superior colliculus and the

association cortex, processing and then projecting to cortical association areas in the parieto-temporal region. The pulvinar is thought to contribute to visual perception, eye movement and attention.

Non-Specific Nuclei: The non-specific nuclei include the intralaminar nuclei and the reticular nucleus. The intralaminar nuclei are so-named as they are in the thalamic lamina. The intralaminar nuclei send diffuse projections to the cortex. Both the intralaminar and the TRN make multiple and strong inhibitory connections to all the specific thalamic nuclei. Either by direct projections or by indirect projections to specific thalamic nuclei which then project to cortex, the non-specific thalamic nuclei have the ability to affect the activity in large parts of the cortex. For this reason, their inhibitory nature and their participation in thalamo-cortico-cortical loops, the non-specific nuclei are theorized to be involved in large-scale functions like; arousal, sleep-wake cycling, and consciousness.

Classically, unlike the specific nuclei, the intralaminar nuclei were viewed as receiving their largest enervations from the cortex but receiving little direct ascending sensory afferent inputs. Some theorists including the thalamic neuroscientist, Edward G. Jones, have concluded that many thalamic nuclei connect to the cortex directly through two major types of neurons; matrix neurons (calbindin immunoreactive cells) and core neurons (parvalbumin immunoreactive cells) (Jones E. G., 1998). We discuss the core and matrix cells in more detail below, as they are a major component of our model.

2.2.7. The Reticular Nucleus (TRN)

For our model, one particular non-specific nucleus, the reticular nucleus of the dorsal thalamus (TRN), is of particular interest. The TRN is a sheet of neurons partially covering the thalamus as shown in Figure 7. The TRN regulates, filters, monitors and integrates the activity of the various thalamic nuclei exerting its influence on them through inhibitory connections.

Francis Crick theorized this in 1984 (Crick F. , 1984) when he proposed the “Attentional Spotlight Hypothesis” proposing that the TRN in particular controlled the “*attentional spotlight that simultaneously highlights all the neural circuits called on by the object of attention*”.

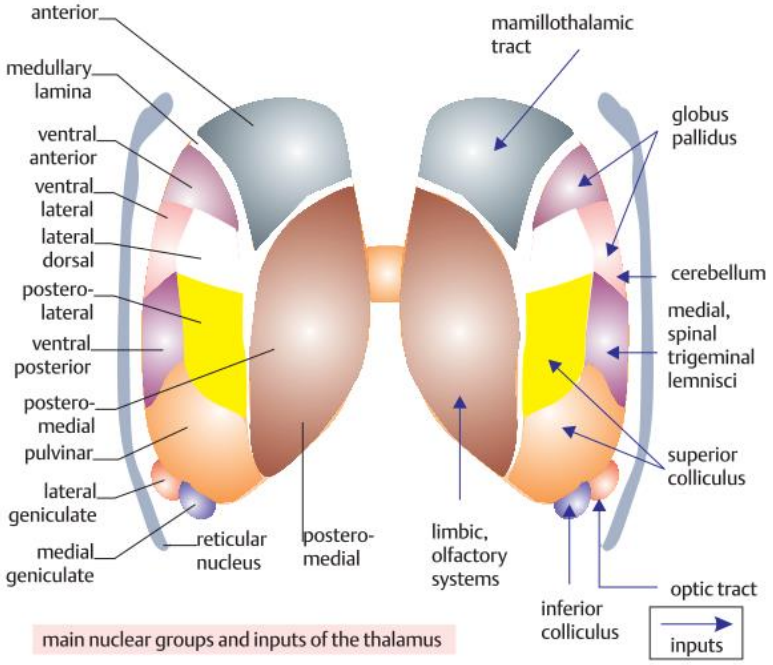

Figure 7: The thalamus, its nuclear groups with their its inputs and outputs showing the reticular nucleus (TRN) (Greenstein & Greenstein, 2000) reprinted with permission from Thieme Verlag.

Pinault explains; “*In other words, he [Crick] submitted that during either perception, or the preparation and execution of any cognitive and/or motor task, the TRN sets all the corresponding thalamocortical (TC) circuits in motion*” (Pinault, 2004). Pinault’s 2004 review states that the TRN is concerned with all or almost all functional modalities of the brain with the largest outstanding question being whether olfactory processing is regulated by the TRN. While unclear in many species, it has been shown in the rat that even olfactory inputs arrive at the TRN. The TRN has been shown to have a body centric topography in some sections and in others it is laid out according to various senses while in still others it is very heterogeneous. TRN neurons have receptive fields that are larger and in many cases more multi-sensory than the cortical neurons that their thalamo-cortical and cortico-thalamic afferents synapse to (Pinault, 2004). Taken together with the fact that the TRN connects to all of the cortex, this would appear to support the idea of the TRN being a control point for all cortical activity.

Each area of the dorsal thalamus is connected to the cortex, and most cortical areas are connected to two or more nuclei plus projecting back to the TRN. Thalamic nuclei mediated by the inhibitory GABAergic neurons of the TRN have been shown to regulate activity in specific areas of cortex, much like a modulator, turning off and on areas of ipsilateral cortex selectively both raising activity and inducing rhythmic activity found in sleep (Lewis, et al., 2015). More recently it has been shown by computational modelling

that the TRN and the dorsal thalamic nuclei together form not just a purely inhibitory filter over the output of thalamic nuclei, but rather a tuneable filter between the thalamus and the cortex. The circuit is tuneable in response to the frequency of activity and T-type calcium channel activity [(Willis, Slater, Gribkova, & Llano, 2015), (Sherman & Guillery, 2006)]. This would seem to further support the postulated TRN's and dorsal thalamic nuclei's role in the mechanisms of attention, just as Crick theorized.

As shown in Figure 9, the inhibitory GABAergic TRN neurons receive inputs from both the excitatory, glutamaturgic thalamocortical and corticothalamic axons and then project back to the thalamic principal neurons. A fluorescently labelled image of these pathways is shown in Figure 8. In this connection profile the TRN neurons have the ability to moderate the principal neurons activity based on thalamic output and cortical feedback which is a major prerequisite of the Granger model.

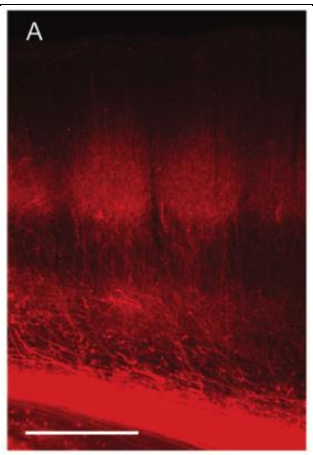

Figure 8: Fluorescently Labelled Thalamic Projections Make 2 Stops in Cortex (Wimmer, Bruno, Kock, Kuner, & Sakmann, 2010) reprinted under open access license from Cerebral Cortex.

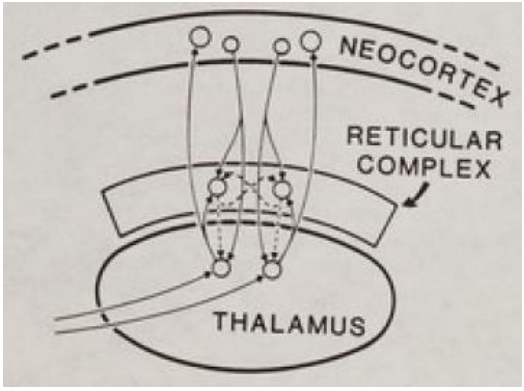

Figure 9; Main connections of the reticular complex (Crick F. , 1984) reprinted under permission from The Royal Society.

Historically it was also considered that the TRN received inputs from all of cortex but that it did not connect back to the cortex but this has since been shown not to be true although the thalamocortical projections are typically much larger than the reciprocal corticothalamic projections (Deschenes, Timofeeva, Lavallee, & Dufresne, 2005). The newer theories suggest that TRN neurons make loops with cortical neurons as well as the thalamic neurons that the TRN neurons are inhibiting, possibly further supporting the idea of the TRN as central to overall cortical activation. Indeed, it is known that TRN nuclei integrate information from diverse sensory and motor thalamic circuits thereby controlling the selection, via inhibition, of the cortico-thalamic circuits that are activated. This too is a feature of the Granger model.

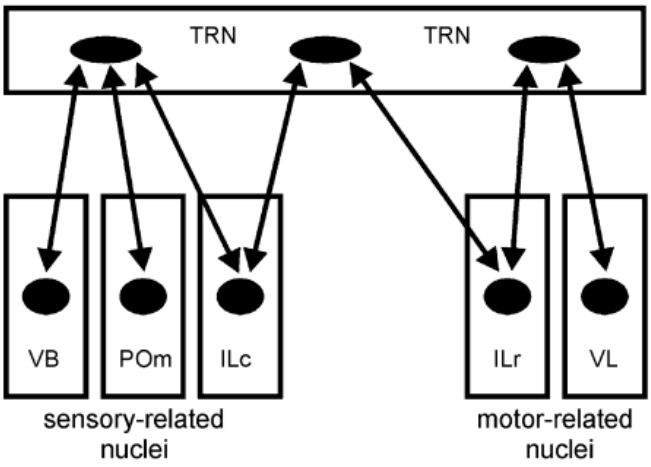

Figure 10 : TRN nuclei use integrated sensory and motor information to inhibit or activate relevant cortico-thalamic circuits (Pinault, 2004) reprinted with permission from Elsevier.

2.2.8. Thalamic Core & Matrix Cells

Two broad classes of cells found in all thalamic nuclei are the core and matrix cells. The two classes of cells have very different immunoreactive properties and very different theorized functions.

Thalamic core cells are Immunoreactive to the Ca++ binding proteins calbindin and parvalbumin. They are topographically organized both in relation to the recipients of their axons in cortex, such as layer 4 and lower layer 3, but as well with respect to their incoming sensory inputs. Whereas, matrix cells project broadly and diffusely to multiple neighboring cortical regions. They are found in both specific and non-specific nuclei (Rodriguez, Whitson, & Granger, 2004).

Core neurons target only one area of cortex each, ending in layer IV mostly and are thought to support perception by processing sensory inputs and passing them on either amplified or curtailed to cortex. This too is consistent with the Granger model. Core cells also modulate the degree of activation of their corresponding cortical area, after being inhibited or not by the TRN and then amplified or not by attention (Piantoni, Halgren, & Cash, 2016). Figure 12 shows the relationship to the core and matrix cells in a prototypical thalamic nucleus and how they connect with the TRN. Note the difference in the cortical layers to which they project.

Thalamic matrix cells are immunoreactive to parvalbumin and occur only in certain thalamic nuclei that give rise to restricted topographically or non-topographically organized projections to separate cortical regions. Matrix cells are found in all of the dorsal thalamic nuclei and project from there widely and diffusely to layer I mostly, but also to layers II and III of superficial cortex.

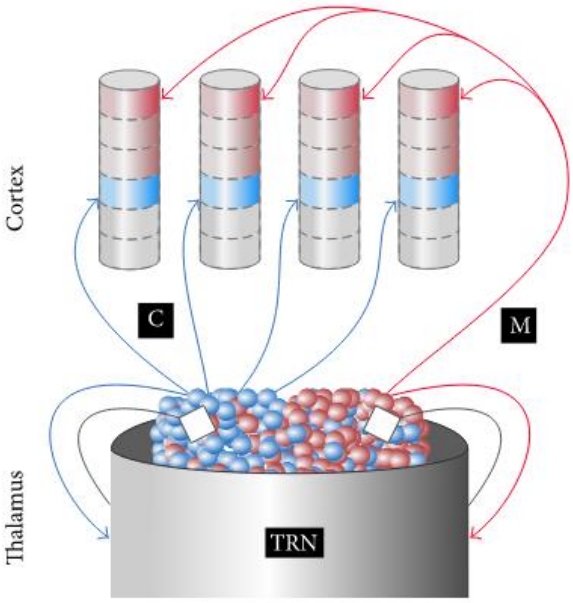

Figure 11: Thalamic Core (blue) and Matrix (red) cells (Piantoni, Halgren, & Cash, 2016) reprinted with permission from Hindwai Inc. under open access license.

Matrix cells are thought to have several functions.

(i) Matrix cells process signals from multiple cortical and thalamic areas and are theorized to bind these signals together into a single conscious percept (Jones E. G., 1998).

(ii) Matrix cells are also thought to be responsible for maintaining the state of the cortex, enabling areas that are needed to be active, and turning off areas that are not needed, doing so over the whole of the cortex especially while in sleep (Piantoni, Halgren, & Cash, 2016).

(iii) Matrix cells are thought to sparsify and orthogonalize their non-topographic inputs, learning a sparse code to represent the combinations they see in their inputs (Rodriguez, Whitson, & Granger, 2004).

(iv) They are also thought to be more responsive to specific input sequences and patterns, and not generally reactive to inputs similar to their preferred direction as the core cells are thought to be (Rodriguez, Whitson, & Granger, 2004). The matrix cells are a significant component of the Matrix circuit for sequence processing in the Granger model.

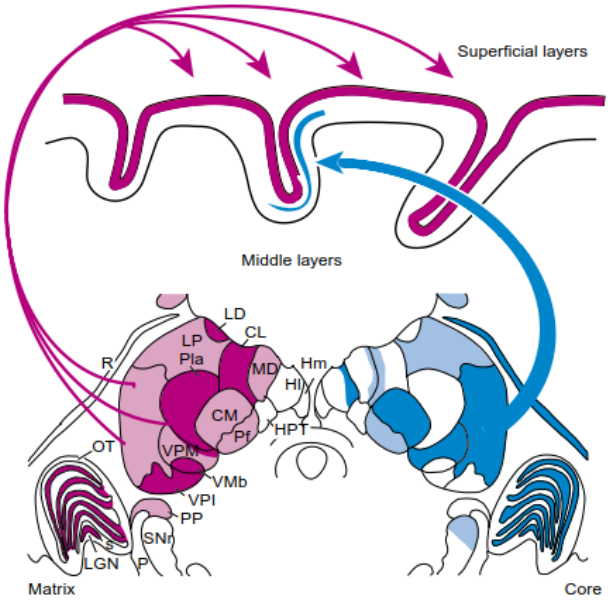

Figure 12: Relative distributions and connections of the core (blue) and matrix (red) nuclei (Jones E. G., The thalamic matrix and thalamocortical synchrony, 2001) reprinted with permission from Cell Press.

The distribution of core and matrix neurons in the dorsal thalamus is mixed throughout all nuclei, but there are more core cells in sensory relay areas of the thalamus as shown above. This would be reasonably expected if the core cells are primarily involved in gating, filtering and passing on primary sensory information, while matrix cells are involved in higher order functions of learning and regulating across the various information flows and sequences of them in the thalamic nuclei.

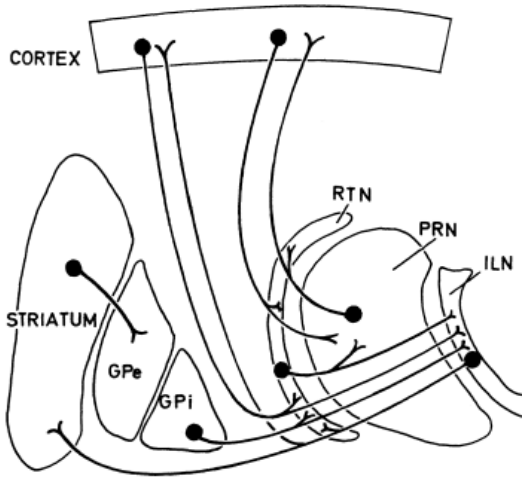

Figure 13: Basal ganglia connections to the TRN (Jones E. G., The Thalamus, 1985) reprinted with permission from Springer Verlag.

Lastly with regard to the TRN we should briefly mention that although beyond the scope of this paper and the underlying Rodriguez et al. work (Rodriguez, Whitson, & Granger, 2004), the TRN receives afferents from the basal ganglia as well. While specifically

excluded from our model, this connection in theory could supply information, perhaps including reward information, to tune the hierarchical categorization mechanism that is the subject of our model herein.

2.3. Thalamo-Cortical Circuits

Thalamo-cortical circuits are not a series of single pass systems. The system is a multi-staged set of progressive feedback loops. Sensory input is fed-forward from periphery to cortex then from cortex to the motor system. All areas of cortex have the same, approximately six-layer, structure, and all are reciprocally connected to thalamic nuclei (Jones E. G., The Thalamus, 1985).

Specifically, one of the pathways from cortex back to the thalamus is via layer 6 which is suspected to carry information, such as motor information, back to the thalamus. These layer 6 axons further branch to innervate motor areas which then feedback to higher-order thalamus cells and then back to higher-order cortex in a series of repeating cascades.

Thalamic Area	**Connected to:**
Nonspecific Thalamus	Cortex Layer I
Association Regions	Cortex Layer II & III
Higher Association Areas & Specific Thalamus	Cortex Layer V
Multiple Brain Regions & Specific Thalamus Nuclei	Cortex Layer VI

Table 4: Major thalamus to cortex projections (Maass & Markram, 2004)

One may speculate that these thalamo-cortico-cortical circuit cascades may be carrying hierarchically nested information in support of our progressive hierarchical categorization model, but there is no direct evidence for it as yet. Were this to be true, it might help to explain the general to specific, or progressive, refinement nature of much perceptive, and motor behaviour. Possibly in support of this, it has been found that the thalamus is dominated by higher-order circuits as opposed to primary input circuits. Most of the thalamus is devoted to cortico-thalamic-cortico circuit modulation according to Sherman and Guillery (2001). The main inputs and outputs of cortex and thalamus are shown below. The Granger model accounts for all of these major projections.

Cortical Layer	**Projects To / Receives from:**
I	Projects to and receives inputs from all of cortex. Layer composed of mostly axons and apical dendrites, few cell bodies (Maass & Markram, 2004).

II	Projects to and receives inputs from other cortical areas, predominately farther regions than layer III (Lamb & Contributors, 2017).
IIIa	Projects to and receives inputs from other cortical areas, closer range than layer II (Lamb & Contributors, 2017).
IV / IIIb	Projects to and receives inputs from layers above and below in the local area. Receives inputs mostly from the thalamus.
V	Major output layer for cortical columns (Guillery. & Sherman, 2011) (Douglas & Martin, 2004). Projects to subcortical regions and contralateral hemispheres. Receives inputs only from adjacent areas (Lamb & Contributors, 2017).
VI	Projects to and receives inputs from layer IV, thalamus and other sub-cortical regions specializing in different modalities such as motor areas (Thomson, 2010).

Table 5: Major cortical outputs by layer.

In summary, the thalamus was historically viewed as a feed-forward relay station which selectively gated signals reaching cortex. One theoretical synthesis across the sources above could be that the thalamus is part of a dynamic routing system where specific thalamic nuclei gate incoming signals from the periphery that feed forward to a hierarchical set of cortical areas connected by loops back through higher-order thalamic areas controlled by the non-specific thalamic nuclei. The non-specific thalamic nuclei also receive driving signals from motor and pre-motor areas as part of their pathways back to cortex where the signals both modulate cortical flow from one cortical area to the next higher cortical area, as well as mix that activity with driving input from the pre-motor and motor areas of cortex.

As to how these circuits function overall, there are several broad classes of theories, as shown in *Table 6* below, of how the cortex might be organized and function to function with the inputs from the thalamus. For a review see Van Essen (2006).

Theory	**Description**
Multiplicity of Areas Theory	Cortex is composed of many dozens of areas that differ based on architecture, connectivity, topography and function.
Distributed Hierarchical Organization	Anatomically defined hierarchy that includes 10 levels of visual processing built from feed-forward, feedback and lateral information flows. Suggests also that it may be better to consider cortex as quasi-hierarchical.
Multiple Processing Streams	A hierarchical cortex model, but with multiple processing streams at each level. Anatomically distinct yet intertwined

	compartments at early levels like V1 and V2; and physically separate dorsal and ventral streams at higher levels. More recent studies have highlighted that cross-talk between streams is extensive at multiple levels.
Dynamic routing of information	Control of information flow is highly dynamic, not fixed by anatomic connections, but routed across the various connections guided by processing dynamics such as visual attention.

Table 6: Categories of Cortical Organization and Function Theories (Van Essen, 2006)

All of these theories in one way or another posit that sensory information flows from the periphery in a feed-forward manner but then is transformed and moves, either by connectivity or by activity dependent routing, through various areas performing different functions on the information but with many areas of feedback. While there are places where the theories contradict each other, there is an intersecting set of features that can be drawn from these classes. This as well describes the flow in the Granger model.

There is certainly a multiplicity of anatomically defined areas in cortex and thalamus as has been known since Brodmann (Brodmann, 1909). The anatomical areas are quasi-hierarhical as streams of processing do exist, and the same input information is often processed in various streams, as for example, in the dorsal and ventral paths for vision. The processing of information shows marked routing in many places including two of specifc relevance to the Granger model: in the thalamus by virtue of the topogrpahic inhibition of the TRC, as discussed above; and in the lateral inhibition found within cortex.

2.3.1. Thalamo-Cortical Circuit Control Flow: Driven and/or Modulated?

Before discussing thalamic driving and modulating signals we need to introduce the general processing steps from primary sensory input in the thalamus to the cortex and back again keeping separate the specific and non-specific thalamic connections. While the full process is not yet understood, we can speculate from what is known about the overall timing of events after a sensory signal is received, the theoretical flow of which is pictured in Figure 15. The sensory signal flows from the peripheral input to the topographically organized core cells of the relevant specific thalamic nuclei to layer IV cortex that then spreads out across layers II and III and then to layers V and VI where the signal is fed back to thalamus but this time to the relevant non-specific nuclei's matrix cells, which are not topographically organized. We will discuss this much more fully later.

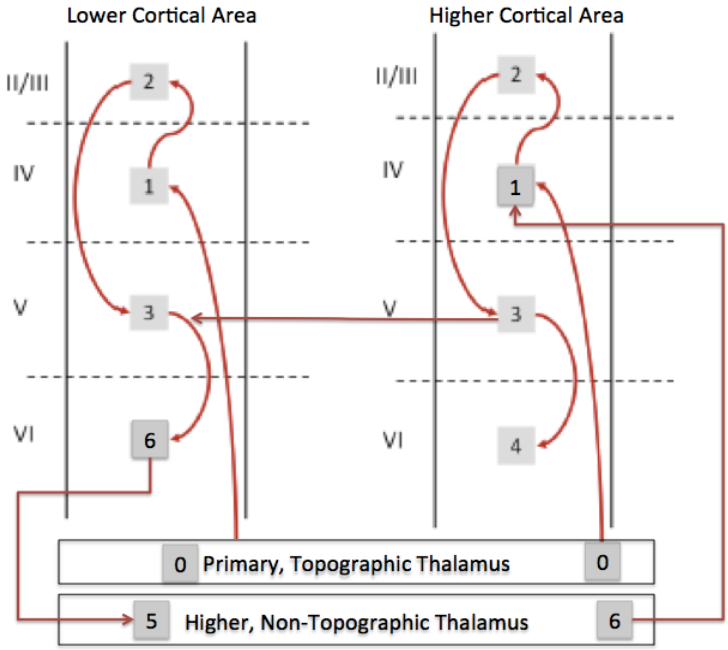

Figure 14: Cortical layer flow from Topographic and non-Topographic Thalamus, created by author from (Strack, 2013).

The stages of activity have been shown to involve “driving” sensory inputs from the specific thalamic core cells which are “modulated” by feedback from the thalamic matrix cells from higher order thalamic nuclei. This happens not just through spreading activation of higher order or associated areas of cortex, but also through modulation and shaping of the initiated cortical activity (Sherman & Guillery, 2006). For our discussion of the Granger model this general flow is enough, but the reader should know that there are many more connections known than the ones discussed herein, as one example, there are more ascending connections to cortex than just through thalamus (Felleman & Essen, 1991).

The flow of information is also shaped once it has started, driven from primary sensory flows and then cascading through the thalamic-cortico-cortical connections. The connections to and from thalamic relay cells and cortex also become differentiated by whether the connection is driving a particular information flow (passing the information from the sensory periphery in a feedforward way), or is modulating a flow once started.

The questions of whether or not first-order sensory, topographic thalamic nuclei are drivers of the circuits only, and whether higher-order thalamic nuclei, previously seen as having a modulating role (Sherman & Guillery, 1998),are also driving the thalamo-cortical circuits, are, like so much of cortical theory, both open questions. Some authors, Sherman included (Sherman S. M., 2005), see modulatory information flow in the cortical signals being passed back to thalamus from layer IV as shown in Figure 15. It is interesting to note that the modulation mechanism found is the low-threshold Ca2+ conductance which also causes the cells to shift from tonic to burst mode and back. In the Granger model,

these cells both drive and modulate the flow of sensory information. More recently it has been theorized that both the sensory and the feedback circuits act as both drivers and modulators, possibly driven to achieve a dynamic balance. How the balance of driving and modulating is controlled is not known (D'Souza & Burkhalter, 2017).

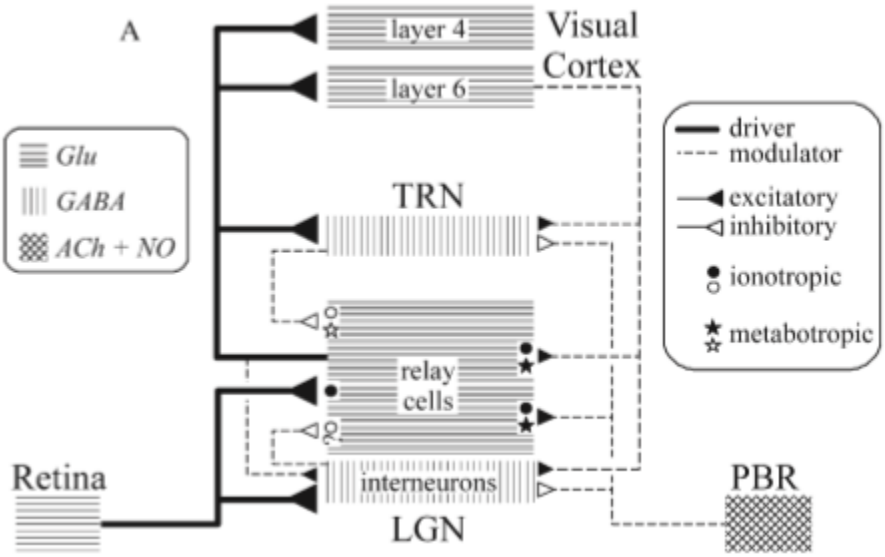

Figure 15: Driver and modulatory connections of the thalamic relay cells (Sherman S. M., Thalamic Relays and Cortical Functioning, 2005) reprinted with permission from Elsevier.

2.3.2. Thalamo-Cortical Circuit Control Flow: Lower to Higher Cortex and Vice-Versa

Guillery sees evidence that higher-order thalamic circuits are in fact driving the thalamo-cortical circuits by adding ascending motor inputs to the signals sent back to cortex from higher-order thalamus (Guillery, 2006). The information content of these motor signals is not known. It is theorized by Rolls that these circuits are implementing attractor networks with both bottom-up and top-down inputs (2016). For our investigation, the Granger model also suggests both bottom-up and top-down flows are both are involved.

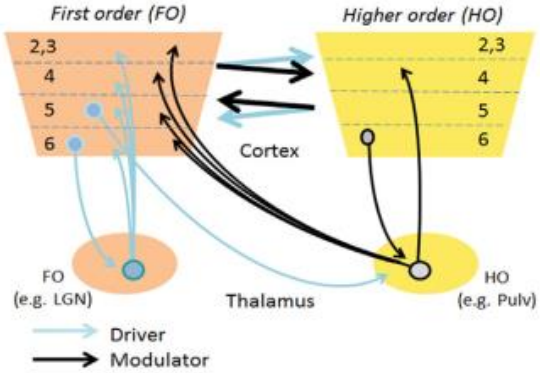

Figure 16: Driver and modulator cortico-thalamic circuits (Prattl, et al., 2016) reprinted with permission from Elsevier.

The regional cascade from lower to higher areas of cortex and thalamus can clearly be seen in the findings from Guillery's study (2006) and as illustrated by Sherman in Figure 17 (2005). In general, Sherman found that the higher order thalamic nuclei received their modulating inputs from layer VI of cortex and lower and higher areas of cortex communicated with each other both through the thalamus and also directly using cortico-cortico connections. The Granger model and our model both contemplate both direct cortical and indirect cortico-thalamic-cortical, connections in accordance with this.

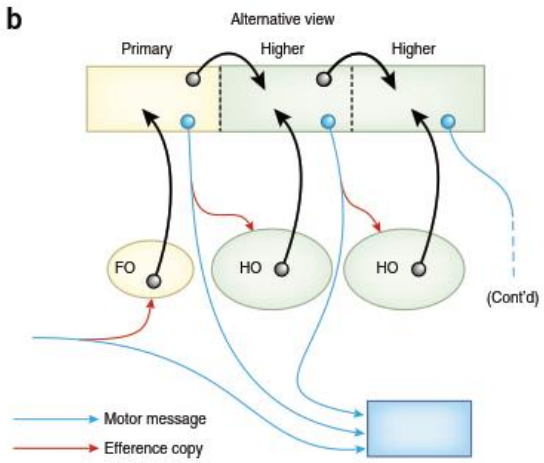

Figure 17 : Thalamo-Cortical lower to higher area connections with motor efference copies (Sherman S. M., 2016) reprinted with permission from Macmillan Publishers Ltd.

Sherman added to his model in a subsequent paper (2016) writing that the cortical areas uniformly convey an efference copy of the layer V motor output signals to sub-cortical areas whose branching axons innervate the thalamic higher-order nuclei as they pass through the thalamus. This step adds support for theories that see cognition as an outgrowth of the processing of movement such the Confabulation theory of cognition. Both the Granger model and Confabulation theory proposes a grammar for cognition. In Confabulation theory, the "grammar" is composed of a series of orchestrated movements of cognitive neural modules. We will discuss Confabulation theory in more detail later on.

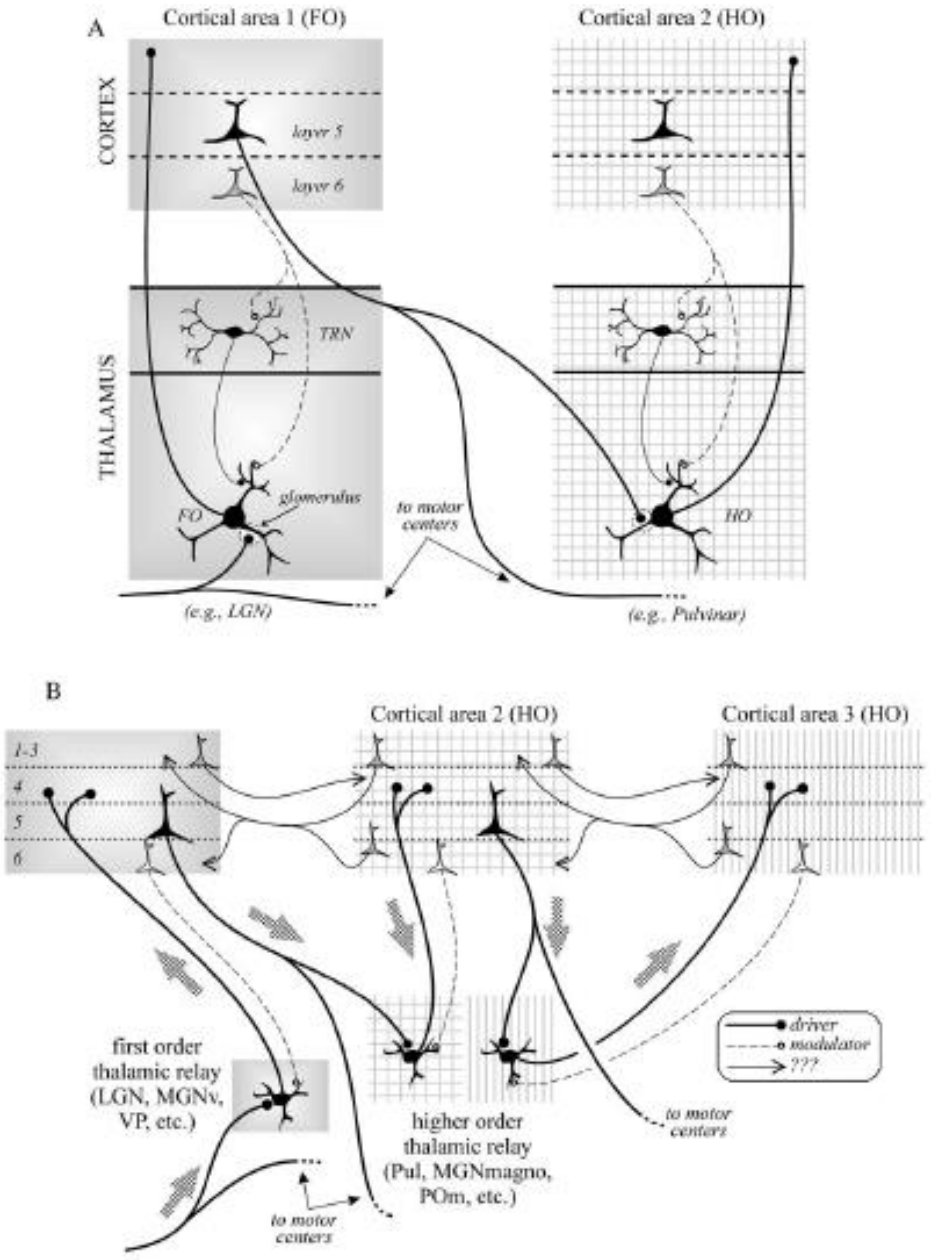

Figure 18 : Cascade of multiple areas of cortex to primary and higher order thalamic nuclei. (Sherman S. M., 2005) reprinted with permission from Elsevier.

Further support for the idea of dynamically formed networks forming dependant on the processing of the network's inputs and internal states (via TRN inhibition and cortical feedback) was published recently in 2017. It has been confirmed that at least in the case of the dorso-medial nuclei (DMN) and the prefrontal cortex (PFC), the DMN and the TRN together mediate the interconnection of functional cortical connectivity (Schmitt, et al., 2017) forming functional networks dynamically as needed.

We see this as a significant finding that supports the general move toward dynamic functional circuit models of cognition, centred on dynamically formed cortico-thalamic-cortical networks. The challenge remains to understand what is flowing across those

networks. We explore that, once we complete our survey of the relevant anatomy, in the section on cortical models which will include Granger's model of the "brain's grammar" (Granger R. , 2006) implemented in the Granger model.

2.3.3. Progressive Categorization via SVM's in the TRN

Another controversial model which points to thalamo-cortical loops as the source of categorization and specifically lends support to the Granger model's focus on the TRN's core role in categorization is described in a 2009 IEEE Conference paper (Jandel, 2009).

The model suggests that thalamic nuclei, the TRN, and cortex participate in an oscillatory pattern classifier whereby incoming stimuli arrive at the relevant thalamic nuclei and are classified by the biological equivalent of a support-vector-machine (SVM) like kernel function computed in the dendritic trees of the TRN nuclei, using all inputs from the incoming stimuli. The cortex is an unstable associative memory for support vectors that are delivered to the TRN and thalamic nuclei by the relevant cortico-thalamic projections (Jandel, 2009). The burst mode of the thalamic nuclei is theorized to align the inputs to the cortex and the TRN so that the pattern detection can occur on all inputs presented simultaneously. This also is so that the combined pattern separation and matching abilities of the TRN and cortical networks can cycle many times over the sustained-input, slower thalamic nuclei cycle, all the while minimizing interference.

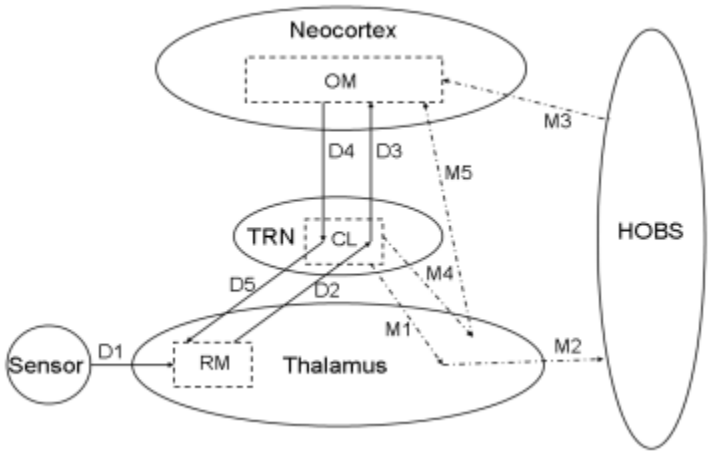

Figure 19: Jandel's Support Vector Machines theory of thalamo-cortico-cortical pattern separation and matching. Legend: Higher order brain regions (HOBS), Cortical support vector memory (OM), TRN kernel function calculation circuits (CL), Thalamic input memory (RM) (Jandel, 2009) reprinted with permission from IEEE.

Jandel's theory lends support for the progressive categorization function theorized in the Granger model implemented in the model's thalamo-cortical-thalamic circuits. In the Granger model, as we will see below, progressive hierarchical categorization of inputs is done in an iterative fashion which would require a similar buffering of the inputs while the thalamo-cortical-thalamic networks compute the classification and then the thalamic matrix cells communicate the results to higher order areas of cortex, while the concept

that the support vectors are maintained in the thalamic reticular nucleus seems compatible with the Granger model.

2.3.4. The TRN, Thalamic Relay Bursts: An Attention & Categorization Mechanism?

The Granger model asserts that the thalamo-cortical loops are gated by the TRN. The TRN bursting mechanism is already thought to be part of attention and sleep/wake cycle. Burst firing of neurons has been suggested to wake-up circuits while tonic firing has been offered as the ongoing information processing mode once a set of neurons are awake, not blocked by the thalamus. It is possible that bursting is also a mechanism by which currently inactive thalamo-cortical categorization loops are activated from tonic lower frequency firing to gamma-band firing, thus engaging the associated cortical circuits into a now activated thalamo-cortical loop in response to the specific stimuli as the Granger model postulates.

The reticular activating system (RAS) projects inputs to the thalamus from the midbrain reticular formation (MRF). The RAS is thought to mediate the states of sleep, being awake and being highly attentive and alert by controlling the state of thalamic and cortical excitation (Moruzzi & Magoun, 1949). When the brain is in REM sleep or awake the EEG signal is formed from low-voltage, fast-bursting, desynchronized oscillating potentials.

When the thalamic relay neurons are in tonic mode, the EEG is not synchronized but when they are in burst mode the EEG is synchronized. When the RAS is stimulated, the slower cortical rhythms are suppressed, including slow cortical waves (0.3–1 Hz), delta waves (1–4 Hz), and spindle wave oscillations (11–14 Hz) (Hum, Steriade, & Deschenes, 1989), the higher frequency waves, such as gamma (20 – 40 Hz) (Steriade, Dossi, Pare, & Oakson, 1996), are seen to be more active but desynchronized.

These faster oscillations are controlled by bursting of thalamic relay neurons caused when the TRN cells are hyperpolarized (with membrane potential in the -70 mV range) by strong sustained inhibition from their connected TRN cells, and then stimulated by incoming spikes. If the TRN cells are not hyperpolarized and their membranes are at their resting potential of -60 mV at the time they receive incoming spikes, their response is very different. They then respond with tonic firing rates of 25 Hz to 100 Hz (Crick F. , 1984). The burst firing sends a synchronized wave of excitatory post-synaptic potentials to all cortical neurons connected to the relay neurons, releasing their bursts at the same time (Kandel, Schwartz, & Jessell, 2000).

This RAS mechanism is thought to be part of the mechanism of attention, when highly activated the higher frequency waves of the gamma band dominate and are thought to be correlated with concentrated awake thought, among other states (Burlet, Tyler, & Leonard, 2002). Further evidence that the TRN is directly involved in attention comes from studies of the axonal connectivity and synaptic architecture of the connections from

the pre-frontal cortex and the TRN (Zikopoulos & Barbas, 2006). Axons from the pre-frontal cortex ostensibly associated with higher order thoughts than the primary sensory pathways, make connections that are more diverse and involve several kinds of synapses, compared to the localized, ordered and mono-synaptic connections made by sensory corticothalamic afferents (Zikopoulos & Barbas, 2006).

2.3.5. Toward Categorization

The discussion above serves to bring together a high-level view of the overall biological circuit architecture that the Granger model is based on. We will return to a more detailed version of the dynamics in the Granger cortico-thalamic circuit model in our discussion of categorization models of the Rodriguez et al. (Rodriguez, Whitson, & Granger, 2004) paper. Before we do so, we give a high-level account of the circuits of the cortex as a whole, and then review the most prominent models of inter-cortical circuitry. After this we will have reviewed the thalamo-cortical anatomy, the major circuits the cortex participates in and the circuits of the cortex. We will then specifically discuss categorization and the Granger model's circuits and dynamics.

2.4. Major Networks of the Brain

The exact functioning of the circuits of the brain overall are not known. However, there are many studies of various circuits with varying evidence for each hypothesized individual circuit. One large work attempting to bring together a series of over 200 neuroanatomical studies with the goal of determining the anatomical basis of cognition was done by Solari et al. (2009). The study attempts to create a canonical set of circuits linking the cortex and the major sub-cortical structures; including the hippocampal formations, the thalamus, the basal ganglia, the basal forebrain and the metencephalon, as well as the spinal cord and the cerebellum. Solari's work was done under the supervision of Dr. Robert Hecht-Nielsen, the creator of Confabulation theory.

We reproduce the central diagram in Figure 20 as it serves as a good orientation to the brain's major networks. An interactive version of the diagram can be found at http://www.frontiersin.org/files/cognitiveconsilience. The results correlate well with the canonical model of cortex and the anatomy that the Granger model is based on. As well, it serves as a good orientation before we discuss our model and other models of the flow of processing in the neocortex in more detail. Unlike with pure structural anatomy, once we attempt to understand theories of the function of the brain and cortex, we venture into the dynamics of thought, an area with only a few seemingly cogent theories of how it all works and even less consensus in the research community.

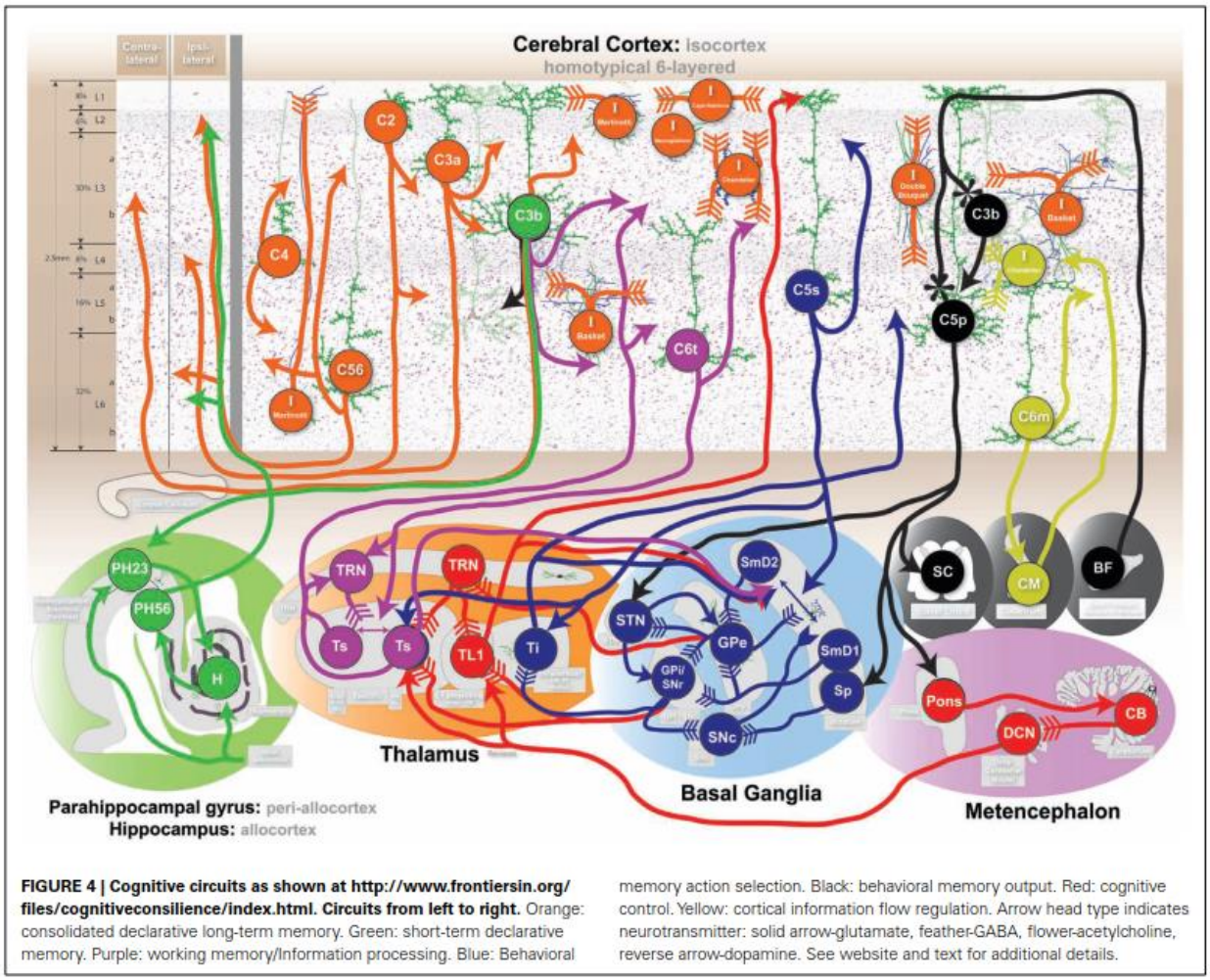

FIGURE 4 | Cognitive circuits as shown at http://www.frontiersin.org/files/cognitiveconsilience/index.html. Circuits from left to right. Orange: consolidated declarative long-term memory. Green: short-term declarative memory. Purple: working memory/Information processing. Blue: Behavioral memory action selection. Black: behavioral memory output. Red: cognitive control. Yellow: cortical information flow regulation. Arrow head type indicates neurotransmitter: solid arrow-glutamate, feather-GABA, flower-acetylcholine, reverse arrow-dopamine. See website and text for additional details.

Figure 20 : Major circuits of the human cortex and sub-cortical structures from (Solari S. , 2009), (Solari & Stoner, 2011), reproduced with permission under Open Access license from Frontiers in Neuroanatomy.

Solari performed his review and summary of all the neuroanatomical papers to provide an intellectual substrate on which to rest his supervisor's Confabulation theory. An overarching theme of this theory is that cognitive behaviours arose from evolution's repurposing of motor circuits for higher order thought. Confabulation suggests that all thoughts are composed of modules which are implemented as nuclei of neurons and their circuits in the brain. According to the theory, each module specializes in learning and processing one attribute of one perceived object or cognitive concept. Modules are composed of many groups of neurons that come to represent symbols relating to the attribute they are focused on. A symbol is like a basis function or a primitive of knowledge whether that be motor or cognitive knowledge. The neurons that represent symbols within the modules connect to symbols in other modules by Hebbian-learned connections. In Confabulation theory, to think is to 'contract' the network of cognitive "symbol-muscles" so-formed. A 'contraction' causes the modules with the highest support for them, the largest incoming excited links to it, to be activated while all others are suppressed. The result is winning module is said to represent the stimulus.

This has strong parallels to our subject model herein. This is similar to the Winner-Take-All (WTA) algorithm, where the neuron with the most excitation 'wins' by crossing a summed activity level and then immediately causing inhibition of the competing neurons. WTA is the basis upon which the Granger model is built and is used therein in the same way, to select the most likely explanation from a set of prior-learned competing representations held within other thalamo-cortical circuits. Note that WTA, as with Confabulation, can be used across individual neurons, nuclei or whole circuits. For our purposes, Solari's anatomical and functional mapping is the best high level early map to orient ourselves as we work toward modelling our iterative categorization circuit in an anatomically grounded way.

Solari proceeds to develop the functional assignments of each circuit in his model based on literature reviews of recordings and functional-deficit-through-lesion studies. The authors segregated the findings from the works they reviewed into seven major functional networks. An easier to follow, less-anatomic, information-flow version of Figure 20's anatomical circuit diagram is shown in Figure 21.

We summarize, at a high level, the circuits in order to situate our circuit of interest, the Cortico-Thalamo-Cortical Working Memory Circuit, which is coloured Purple. This circuit corresponds to the Core circuit in the Granger model. What follows is a summary of Solari's model (2009) to give an overview in which to situate the Granger model within the brain's major networks.

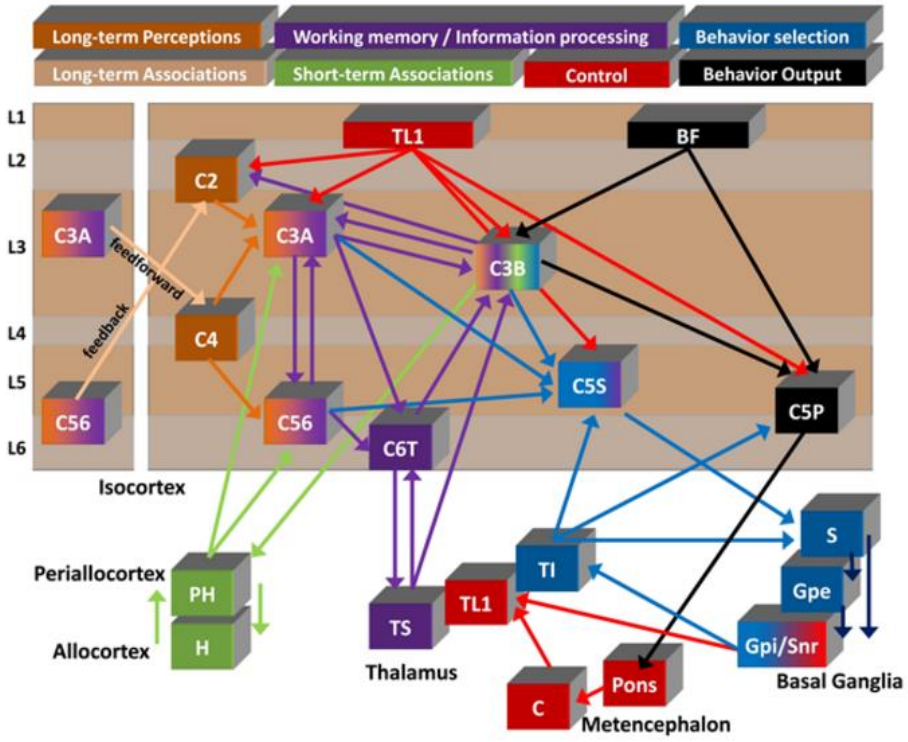

Figure 21: Information flow among theorized seven major networks fo the brain (Solari S. , 2009), reproduced with permission under Open Access license from Frontiers in Neuroanatomy.

2.4.1. Corticocortical Circuit : The consolidated long-term declarative memory circuit (Orange)

This circuit is thought to implement long-term memory storage in isocortex between both hemispheres of cortex, connected by the corpus callosum. The major connected nuclei are listed underneath in the order of traversal:

a. layer 4 cortical pyramidal neurons, which form a feedforward input system to cortex;
b. layer 2 pyramidals which function as a cortico-cortical feedback input system;
c. layer 3a pyramidals function as a corticocortical feedforward output system;
d. layer 3b functions as stable invariant perceptual representations in the cerebral cortex that are then used by many of the other circuits such as the short-term memory circuit which associates these long-term memories with short-term ones;
e. cortical layer 5/6 pyramidal neurons are viewed as implementing the linking of memories via cortico-cortical connections predominately of movement and behaviour related information; and,
f. interneurons; such as the Basket, Chandelier, Double Bouquet and Marinotti cells; 'synchronize information processing and facilitate excitatory competition through localized vertical and horizontal inhibitory projections'.

2.4.2. Cortico-hippocampal-cortical Circuit: The short-term declarative memory circuit (Green)

The declarative memory that utilizes the parahippocampal gyrus and the hippocampus forms the short-term memory system. The circuit connects representations in the hippocampus to representations in cortex, forming an indexed association of memories along spatial and temporal dimensions. The parahippocampal gyrus' layers 2/3 and 5/6 interestingly form associations between the short-term contents of this system and the long-term representations in layer 3b's association networks.

2.4.3. Cortico-basal ganglia-thalamocortical circuit: the action and behaviour learning circuit (Blue)

Actions and behaviours are learned via this network which connects the basal ganglia, the thalamus and cortex. The primary cortical inflows and outflows of to this circuit arrive and leave via cortical layer 5. Specific thalamus 'drives the convergent and re-entrant selection of the cortical layer 3b and cortical layer 6t perceptual representations in cortico-thalamic oscillations. The interlaminar thalamus excites and inhibits the output of the lower layers of cortex to select the right thalamo-cortico-cortical circuits for behavioural output. The circuit then flows to the nuclei comprising the basal ganglia include:

i. the sub-thalamic nucleus stops actions already commenced in the intralaminar thalamus;
ii. the globus pallidus internal segment and the substantia nigra pars reticulata perform precise action triggering in the intralaminar thalamus and ventral thalamus via disinhibition;
iii. the substantia nigra pars compacta, which computes a differential dopamine signal (a signal of the unexpected part of any reward sensed) directing the striatum to start and stop learning action sequences;
iv. the globus pallidus external segment which seems to direct the negative learning and stopping of action sequences under the direction of cortex layer 5 afferents;
v. the Striatum, which is thought to compute dopamine-based learning signals. Further support for the idea that these striatal-cortical loops learn task to reward associations also comes from Granger and his co-author (Chandrashekar & Granger, 2012) who show evidence for the existence of support vector machine like computations occurring in cortico-striatal loops. For more information see the Cortico-Pontine Circuit description below.

2.4.4. Cortico-pontine circuit: Cortically learned behaviour output and/or skill learning circuit (Black)

This circuit involves the flow of behavioural cognitive process (in non-motor areas' layer 5 cortex) and movement process (in motor areas' layer 5 cortex) information.

a. This circuit begins in the basal forebrain which sends its cholinergic projections to layer 3b of cortex which then sends afferents to layer 5p of cortex.
b. Layer 5p then sends afferents to:
    i. the subtahalmic nucleus, thought to be involved in; preparing desired output actions by its interactions with the intralaminar thalamus, and/or stopping actions already begun (via raising inhibition in the intralaminar thalamus);
    ii. the striatum patch, which receives a dopamine-based learning signal via layer 5p to reinforce the start and stop action mappings that the striatum's matrix D1 and D2 receptor cells learn; and lastly;
    iii. the pons which sends the cortical layer 5p output to the cerebellum which refines and stabilizes it before sending to the motor areas or (according to the author's theory that cognitive processes arose from the same machinery as muscular movements) back to cortex and/or sub-cortical areas as refined cognitive behaviours.

2.4.5. Thalamo-cortical projections circuit: The cognitive control circuit (Red)

This circuit controls the flow of information to and from the cortex and the basal ganglia, possibly to mediate goal-directed cognition. It incorporates;

- Thalamic reticular nucleus, which gates the flow of information through the thalamo-cortico-cortical-striatal networks, selecting the right flow of information for the current cognitive state.
- As before, the specific thalamus, which drives the convergent and re-entrant selection of the cortical layer 3b and cortical layer 6t perceptual representations in cortico-thalamic oscillations.
- Thalamic projections to layer 1 of cortex appear to select the areas of cortex (columns or micro-columns perhaps) that are needed to process the current cognitive state.
- As above, the globus pallidus internal segment and the substantia nigra pars reticulata perform precise action triggering in the intralaminar thalamus and ventral thalamus via disinhibition.
- As above, the globus pallidus external segment seems to direct the negative learning and stopping of action sequences under the direction of cortex layer 5 afferents.
- The pons integrates cortical layer 5p output to the cerebellum.
- The cerebellum appears to stabilize sequences for both movements and thoughts
- The deep cerebellar nuclei then transmit the cerebellar output to the ventral thalamus where specific thalamic circuits and layer 1 projecting circuits transmit the combined signals to the cortex.

2.4.6. Cortico-claustral-cortico circuit: The cortical information flow circuit (Yellow)

Crick & Koch, as cited by Solari, credit the claustrum with the integration of conscious perceptions (2005) as evidenced by the wide connections that the claustrum makes to all of the cortex, and predominately to the inhibitory chandelier interneurons. The chandelier cells therefore appear to 'synchronize information processing and facilitate excitatory competition through localized vertical and horizontal inhibitory projections enabling cortical information processing' (Solari & Stoner, 2011, p. 156). The layer 6m neurons coordinate with layer 5s cortical neurons to integrate actions suggestions and transmit some form of processed information about them to the claustrum. The claustrum's output appears to be the inhibition layer 4 cortical inhibitory chandelier interneurons, possibly filtering down the choice of available circuits that are relevant to the planned actions.

2.4.7. Cortico-Thalamo-Cortical: The working memory circuit (Purple)

Finally, we come to our circuit of interest. The "Cortico-Thalamo-Cortical, Working Memory Circuit", which implements working memory in the thalamo-cortico-cortical loops mediated by TRN activity. This circuit is consistent with the Granger model, which is a dynamical model of the functioning of this circuit. As above, the thalamic reticular nucleus

gates the flow of information through the thalamo-cortico-cortical-striatal networks, selecting the right flow of information for the current cognitive state. As before, specific thalamus 'drives the convergent and re-entrant selection of the cortical layer 3b and cortical layer 6t perceptual representations in cortico-thalamic oscillations' (Solari & Stoner, 2011). Layer 3b functions as stable invariant perceptual representations in the cerebral cortex that are then used by many of the other circuits such as the short-term memory circuit which associates these long-term memories with short-term ones.

One observation not present in the Granger model is Solari's view that 'cortical layer 6t appears to function in conjunction with cortical layer 3b and specific thalamus to facilitate cortico-thalamocortical oscillations'. The Granger model does not make predictions for the role of oscillations.

2.5. Cortical Models

We now review several of the prominent models of cortical circuitry.

2.5.1. Common Traits Inspired Generalized Models

Studies of many areas of cortex and thalamus have found significantly similar architectonics among thalamo-cortical areas concerned as summarized below (Rodriguez, Whitson, & Granger, 2004). Based on these similarities across most of the thalamic-cortical range, many theorists have proposed various general cortical-thalamic architectonic and computational models.

Cortical anatomy
- Number of neurons per layer remains relatively constant across different cortical regions
- Excitatory (pyramidal) and inhibitory cells occur at roughly a 5:1 ratio.
- Pyramidal cells emit long axons with distant targets, as well as local collaterals.
- Inhibitory cells have a radius of axonal arborization restricted to only local targets.
- Inhibitory cell axons target cell bodies, proximal dendrites, and initial axon segments of local pyramidal cells, thus powerfully influencing the responsivity of excitatory cells.
- Neurons are arranged in layers of superficial (II-III), middle (IV) and deep (V-VI) cells.
- Neurons are vertically organized into pyramidal cell modules containing ~200 neurons.
- Columns of ~200 pyramidal cell modules are spatially localized within topographically defined thalamo-cortico-thalamic projections.

Thalamocortical anatomy
- Thalamic matrix cells project broadly and diffusely to L.I, contacting apical dendrites of neurons in II, III and V.
- Core thalamic cells project with preserved topographic organization to layers III and IV.
- Middle layer cells (layer IV) project apically to suprajacent layers II-III.
- Superficial layer cells (II-III) project basally to subjacent deep layers (V-VI).
- Layer V cells project to motor targets and, via collaterals, to thalamic matrix.
- Layer VI projects topographically to core, to overlying nucleus reticularis, and to layer IV.

Physiology & plasticity
- Synapses from thalamic core to L.IV are plastic during development, and exhibit little or no plasticity in adults.
- Synapses of superficial (II-III) and deep layer (V-VI) pyramidal cells potentiate in adults.
- Reciprocal excitatory-inhibitory connections within layers yield lateral inhibition.
- Excitatory post-synaptic potentials (PSPs) are brief (~15-20 msec).
- Inhibitory PSPs are roughly an order of magnitude longer (~100-150 msec).

Figure 22: Architectonic and physiological similarities across many regions of thalamo-cortical areas (Rodriguez, Whitson, & Granger, 2004), reproduced with permission from MIT Press.

The commonality of these traits has inspired the search for a canonical cortical model.

2.5.2. The Canonical Cortical Model: The Starting Point for the Granger model

One of the more recent, yet classically inspired models of cortical functional architecture comes from Costa and Martin (Costa & Martin, 2010) who, although disputing the existence of the physical cortical column, still point to the classic Douglas and Martin canonical circuit (1991) model as the best candidate for defining the canonical cortical functional circuit. Costa and Martin then modify it to generalize it even more as shown below in Figure 23 (Costa & Martin, 2010). They argue that their "canonical microcircuit respects the known connectivity of the neocortex, and it is flexible enough to change transiently the architecture of its network to perform the required computations".

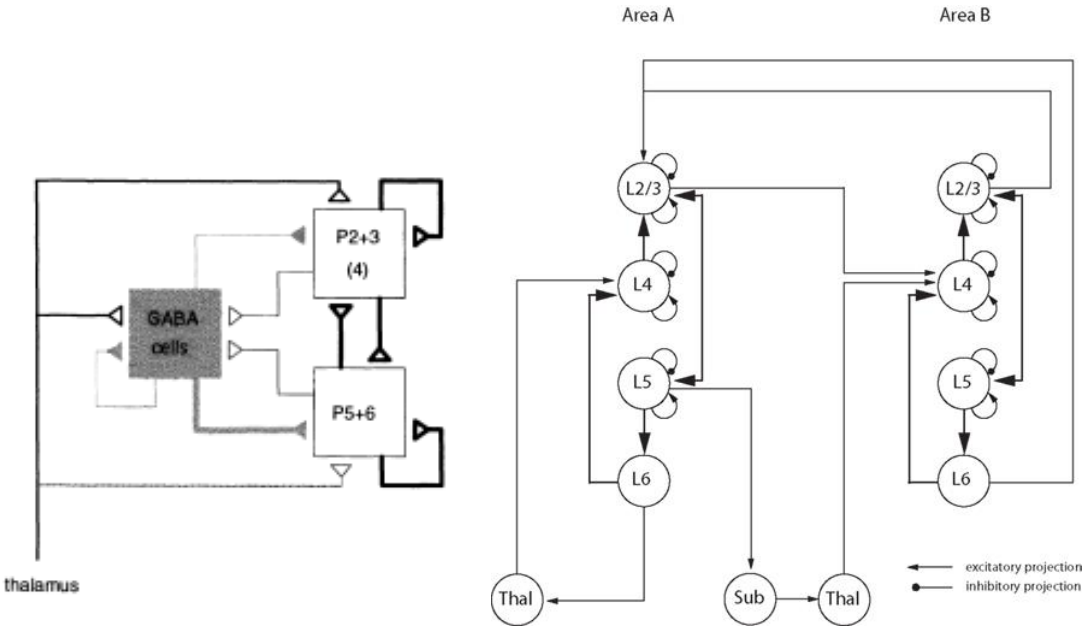

Figure 23: Costa & Martin's (right (2010) reproduced with permission from Frontiers in Neuroanatomy under Open Access license) generalization of the classic Douglas and Martin (left) canonical cortical microcircuit (Douglas & Martin, 1991) from (Ulinski, Jones, & Peters, 1999) reproduced with permission from Springer.

In this circuit, we see the classic circuit composed of these stages;

(1) thalamus feeding topographic sensory information to layer 3b and 4, commonly thought to only be layer 4 but mostly layer 3b in fact (Jones & Burton, 1976); and recently shown to also be fed into to layer 5;
(2) and then the connected pyramidals in layers II and III are activated, which in turn,
(3) activates layer V pyramidals;
(4) which then drive sub-cortical structures and non-specific thalamic cells which then feedback to the layer II & III pyramidals in the originating column;
(5) and as well project to layer IV of higher cortical areas and then,

(6) back-projections from those higher cortical areas in layers II & III feedback to the originating column's layer II and III pyramidals; which then,
(7) activate connected layer VI cells that project back to thalamus.

This model is the starting point for the Granger model. The Granger model adds a few important connections to the above. Rodriguez, Whitson and Granger (2004) cite the physiological evidence for these connections from the following studies (Conley & Diamond, 1994; Bourassa & Deschenes, 1995; Deschenes, Veinante, & Zhang, 1998; Rouiller & Welker, 1991).

(1) the projections from layer V back to thalamus go to the matrix cells; and secondly,
(2) this projection to the core cells passes through the TRN whereas the projection to the matrix cells does not, and finally,
(3) the projection from the layer VI cells go to the thalamic matrix cells.

These last few additions are key components upon which the Granger model 's iterative hierarchical categorization mechanism is reliant.

2.6. Dynamic Cortical Computation with Time Models

We have seen various authors' theories of the columnar and/or functional canonical structure of the cortex which attempt to account for the horizontal and, in many areas of sensory and motor cortex, topographical features of the receptive fields found in cortex. But what of the layers and how does the computation evolve over time? What theories are there about how the layers of cortex function and how do those relate to our destination herein to uncover the hierarchical categorization mechanisms of the thalamo-cortico-cortical loops found in cortex?

In general, it is fair but unfortunate to say, that really no one yet knows why the cortex appears to be organized into columns (or interleaved slabs for that matter) and layers. The functional significance is theoretical still (Hawkins, Ahmad, & Cui, 2017). The Granger model is largely a model about how sequential hierarchical categorization and object identification occurs through the interaction of the cortical layers and the thalamus. We will review some of the more prominent and relevant models in this section and comment on areas that significantly support the Granger model.

Three are many theories of how cortex functions. All to date are based on theoretical grounds and not on any form of direct observation other than multi-cell recordings. We have not yet been able to record from enough neurons in all the related areas of the thalamus, the cortex, the hippocampus and the basal-ganglia simultaneously in order to develop a complete data-driven picture of the changes of state in all areas that the cortex. Therefore, no fully evidence-grounded theory of cortical operation yet exists. While this can be disheartening, the theories advanced so far are instructive, and we will review a few of the more well-known ones in trying to situate our core hierarchical categorization circuit model. The following summary of the available models was not able to be located

in the literature. We believe that this summary is of value in painting the landscape of where the literature has been and where our work is within the cortical computational theory that the Granger model is part of.

2.6.1. Hebbian Spike Timing Dependant Plasticity (STDP) Generated Cell Assemblies: How the Network is Formed

One of the original theories of neural organization is Hebb's famous plasticity model, later referred to as his 'fire-together, wire-together' (Löwel & Singer, 1992) theory.

In Hebb's words;

> *'When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased' (Hebb, 1949).*

Hebb's theory has become the neural wiring foundation over which much of modern neuroscience's theories of cortical function and cortico-thalamic function rest. Many believe it gives the basic mechanism which networks come to represent the statistics of the world. The complex distributed functional representations of the world in cortex have not so far been able to be explained relying purely on the local Hebbian learning rule however. There is much more to be discovered as we attempt to connect the dots from neurons with synapses forged from temporal coincidence to functioning models of how thalmo-cortical loops learn complex hierarchically-organized, categorical representations the world.

How, on top of this network of neurons wired by simulataneous firing, does cognition operate? This requires theories that explain both the function of the signals, the role of the layers, and the dynamics of the whole, as well as characterizations of how the sum total generates behaviours. We list the models in a rough order below based on how much of the needed theory they address, the simpler ones to the more all encompassing.

2.6.2. Calvin's Inbox/Outbox Office Model: High Level Layer Functionality

Calvin (1995) provides one of the simplest ways of looking at the functions of the cortical layers inspired by Peters and Yilmaz's layered model (1993). This model captures a surprising amount of the consensus of the canonical models from Douglas and Martin's (1991) to subsequent models like that proposed by Rolls (2016). The basic idea is that layer IV is the "inbox", layers II & III are the "inter-office communications" and layers V & IV are the "outbox" sending cortical outputs to the sub-cortical structures and the body. While this is nowhere detailed enough, it is a useful general orientation when starring into the complexity ahead.

2.6.3. Abele's Synfire Chains: Putting Limits on Dynamics in Random Networks

Abele proposed that neurons are organized in chains of pools of neurons, such that the neurons in one chain send feed forward excitations to the next pool of neurons and so on (Abeles, 1982).

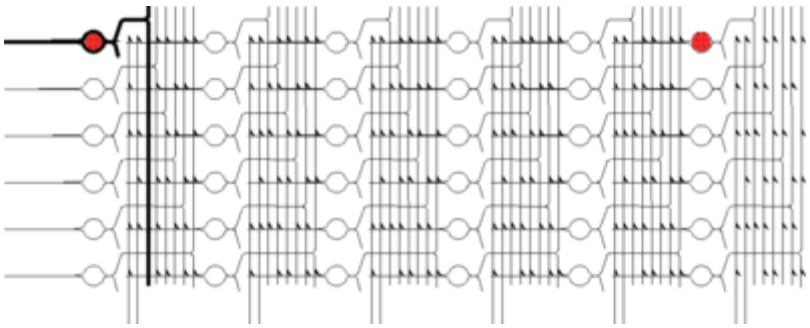

Figure 24: Synfire chains. Note neurons propagating axons in a feed-forward direction. (Scholarpedia, 2017) reproduced with permission from Scholarpedia under the Creative Commons license.

Abeles argued that networks of neurons that converge and diverge in randomly connected networks using integrate-and-fire neurons cannot transmit information reliably and repeatably using an asynchronous spiking arrangement without large local inhibition and even then, the activity will spread beyond the local inhibitory network. He reasoned that this activation, if for no other reason than the timing of inhibition and excitation across a diverse network, cannot be controlled exactly or contained, and hence the network will saturate or be driven to full inhibition (Abeles, 1991). He also showed that asynchronous excitation through a randomly connected network results in spiking that is larger than the total sum of its inputs at each converging neuron over time, guaranteeing that the network will always go to full saturation if there is no inhibitory mechanism. He further commented that since the inhibition cannot contain the saturating excitatory forces, unless it moves to full inhibition, there must be something else going on to maintain the network's activity between saturation and full inhibition. He postulated that these dynamics were basic properties of all complex neural networks. The solution he came to for the dynamic control was that the neurons could fire synchronously or asynchronously, but that only the synchronous mode was stable in the face of these limitations. He therefore defined synfire chains as the case where one population fires synchronously and then one time-step later the next group of neurons fires.

He proposed additional constraints to ensure stability. First, the synchronous firing needs to be truly synchronous or at least squelched during the non-integrate-and-fire windows. So tight windows of where a cell will take input are required and then the cell should not take input until the next window. Second, the cells connections need to be strong enough

to ensure synchronous firing for those cells that should fire. Lastly, the network on each synchronous volley must be in the right starting condition (Abeles, 1991).

It is interesting at this point to speculate, especially in light of the later models of cortical function that have come along, on the consequences of this view.

Temporally Limited Functional Circuits Manage Saturation: Perhaps Abeles is both right and wrong regarding the need for synchronous firing across the whole of any randomly connected network like the cortex. Perhaps you can have asynchronous firing in functionally delimited networks such as those formed by basal-ganglia-thalamo-cortico-cortical networks (BGTCC Networks) where the thalamic specific cells are controlled by the reticular nuclei's inhibitory cells. The claustrum's wide integration across cortex may provide more functional control, as may the strong chandelier and other inhibitory interneuron connectivity in the cortex, all of which may be combined with strong oscillatory windows where spiking can occur. This may functionally activate the circuits with exactly the conditions Abeles claims are fundamental to mitigate saturation and full inhibition, allowing the circuits to operate reliably between these extremes. In this view the brain creates temporally contained circuits wherein asynchronous (or close to synchronous) firing conditions can occur in those selected circuits, the results of which are captured in the recurrent resonance of the BGTCC Networks, which is then provided at the next time-step to the next incarnation of individually active BGTCC Networks.

Burst Mode Helps Synchronize Inputs: It also interesting to recall the burst firing mode found in thalamic nuclei and speculate that synchronized burst mode could be used at the initiation of dynamic circuit formation to both synchronize, reset and perhaps align the firing of the then dynamically selected networks. This is similar to the discussion above that the burst mode "awakens" the right circuits.

Nested Rhythms Manage Learning: Lastly, brain rhythms could also play a large part also in temporally aligning inputs. Here we speculate that another benefit of oscillatory rhythms at the higher levels of cortex is to create a temporal context to control learning. The alpha rhythm, found in upper cortex, has been shown in some structures to interact with faster rhythms, entraining and amplitude-limiting them. This may provide a context within which some cortico-thalamic circuits can learn while others are inhibited from learning. This may help solve the learning-stability problem at the heart of the yet to be uncovered mechanisms in cortex (see section below on Stephen Grossberg's ART model for a fuller account). The formation of learned hierarchical categorizations in cortical thalamic networks as in the Granger model would be aided by such a mechanism as we shall discuss later.

2.6.4. Friston's Free Energy Model:

Friston's Free Energy Model (Fiston, 2010) states that any self-organizing system attempts to minimize its free energy in order to resist the push to disorder. This is how self-organizing systems resist the push to increased entropy. For predictive systems entropy is equivalent to surprise, in the sense that given the distribution of expected states of a self-organizing system, a state which is unexpected is an increase in the system's entropy. Formally, surprise is equal to the negative log-probability of the states, so states that are improbable have large surprise.

Friston developed his Free Energy Model starting with the Principle of Efficient Coding (Barlow, 1961) which states that the brain optimizes the Shannon mutual information (Shannon, 1948) between an incoming sensory signal and the brain's representation of the signal. This was then generalized and formalized under the Infomax theory (Linsker, 1990) which develops this into a general learning algorithm that optimizes a function transforming the sensory input signals into the internal representation held in the corresponding neural network, such that the transformation function maximizes the mutual information while minimizing redundancy and complexity in the encoding. Friston then extends the InfoMax principle adding to it the minimization of surprise, computed via gradient descent, making it a more comprehensive continuous learning model for use in neural networks. Friston's extension of the Infomax model is formulated as a gradient descent optimization of the parameters for the generative predictive model over the paired sensory input vector and cortical state representation vectors.

Applying this to a cortical model, Friston's model attempts to minimize the difference between the predictions made by cortex and the actual sensory data of the world through positing a specific set of neurons in the brain that track the prediction error and that the cortex self-organizes to minimize this error using his free energy formula (see (Fiston, 2010)). Feedback in his model is purely inhibitory in that the top-down prediction is subtracted from the bottom-up input to generate the error signal. While this last point has been used as a fault in the model given the long-range feedback projections from cortex are largely excitatory, the excitatory cortico-thalamic projections are in fact used to excite inhibitory TRN neurons, making the effect inhibitory at an information processing level.

This prediction error then arises in the model as the difference between the top down and bottom up activities of the cortical columns. This is plausible given our previous discussion of layer I's spreading of activation from the bottom-up cortical columns initially activated by specific thalamic afferents to their layer IV pyramidal neurons, after their neurons are activated by direct sensory inputs to columns, which have been linked to the sensory ones through the nonspecific thalamic networks. This could be one form of anatomical correlate of Friston's proposed top-down signals.

Friston's model therefore is in agreement with the Granger model in three major ways:

1. The coding of the differences between each successive category are via the interaction of the cortical pyramidals and the thalamic nonspecific cells.
2. Friston's inhibitory subtraction of the inputs from the predictions is the same as the Granger model which subtracts a first round of general representations from the incoming signal by activating the TRN inhibitory suppression over parts of the specific thalamic incoming signal and then re-processes the residual specific thalamus signals through the cortex again to generate the next level in the hierarchical matching process.
3. An extension of Friston's Infomax theory, Coherent Infomax, is even closer to the Granger model in that Coherent Infomax maximizes the mutual information between the sensory inputs. The contextual inputs are represented in the path history and the current state of the neural network up to the current time. The implementation of the resulting context, path and efficiency sensitive optimization function in a learning rule for artificial neural models has been correlated to neurophysiological readings on real synapses lending support for the idea that real neurons do optimize their input response and information encoding in context sensitive ways that include the path history (Phillips, 2012). The Granger model also uses an encoding maximization, and context and path sensitive implementation of the transformation function from inputs to hierarchical representations.

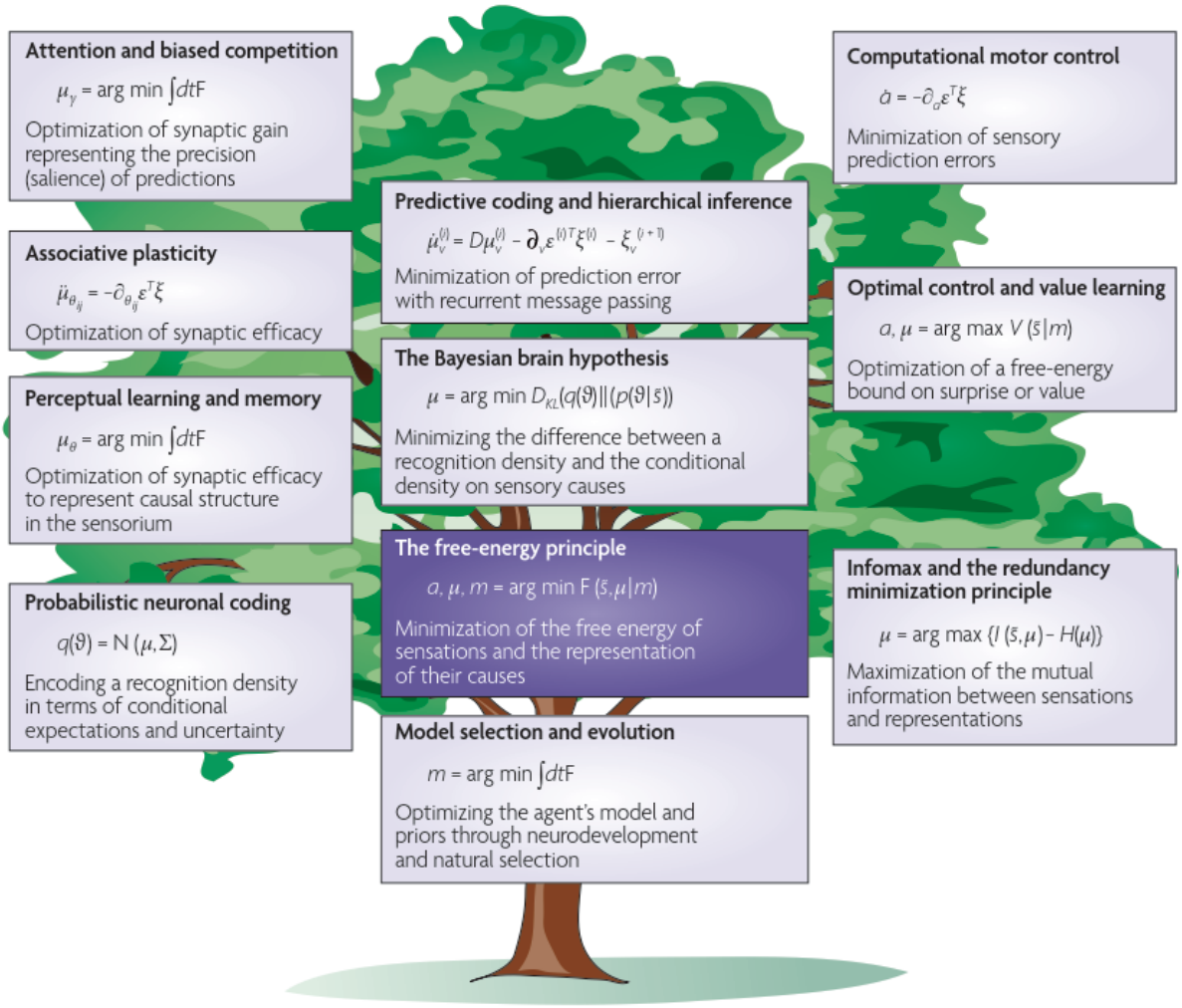

Figure 25 : Friston's free-energy model situated among the other predictive / optimization algorithms of sensory encoding (Fiston, 2010) reproduced with permission from Springer Nature.

There are many such predictive/optimization models for encoding and predicting sensory inputs to neural representations. Friston's Figure 25 shows many of them. We will not discuss the others suffice to say that the details are in Friston's papers and that the captions under each give the basic idea of what is being optimized and/or predicted for each. They all try to generate efficient encodings that maximize the learning of the probability distribution (whether conditional or joint) from sensory inputs to neural representations and predictions thereof.

2.6.5. Grossberg's Adaptive Resonance Models

Steven Grossberg's wide ranging Adaptive Resonance Theory (ART) is at its heart a set of neural models that attempt to explain how humans can learn stably over lifetimes and fuse their learning at multiple levels into a sense of the world, self, and consciousness. The model has been developing for many years. We only scratch the surface and call out the areas where it is in agreement or not with the Granger model.

The initial ART model itself is a cognitive and neural model of how the brain 'attends, categorizes, recognizes, and predicts objects and events in a changing world' (Grossberg, Adaptive Resonance Theory, 2014, p. 1). The model attempts to explain the fundamental stability/plasticity dilemma, as to how the brain can remain plastic and continue to learn while maintaining stability. The original model was then extended to the LAMINART model which added support for the multiple cortical layers and then again extended to the SMART model which added the support for spiking and showing how multiple cortical areas interact with specific and non-specific thalamic nuclei to implement a solution to the stability / plasticity dilemma.

The core proposal is that top-down predictions in the thalamo-cortical system focus attention on salient cues, filtering extraneous ones out, while attention creates self-normalizing biased-competition amongst the activated neural circuits. When a good match between inputs and predictions occurs, a synchronous resonant state is created that enables adaptive learning of the bottom-up categorized inputs and the associated top-down predictions. This improves the context sensitive match between the inputs and the predictions for the next time they are called upon. The focusing of learning in context-sensitive ways avoids the contamination of the inputs across to other circuits to which the inputs are not relevant or are out of the appropriate context. When combined with forgetting within the right contexts, the model claims to be the most widely applicable model to a range of cognitive behavioural phenomena.

The SMART model is shown in Figure 26. The model is somewhat different than the Granger model. The SMART model's flow is as follows from (Grossberg, 2008);

1. Bottom-up input arrives at first-order thalamic matrix cells.
2. First order thalamic matrix cells provide nonspecific excitatory priming to layer 1 (blue dashed line) which turn on the corresponding layer 5 cells that then respond to layer 2 & 3 inputs.
3. Layer 5 then activates the pulvinar in second order thalamus.
4. Layer 4 of first order cortex now receives inputs from both (i) a direct LGN input and (ii) a layer 6 modulatory on-centre/off-surround input that contrast-normalizes the layer 4 activity pattern via the recurrent [layer 4, to 2/3, to layer 5, to layer 6, to layer 4] feedback loop. The contrast normalization circuit is a key proposition that enables all learning to be done over time on a normalized basis allowing for better pattern recognition and for incremental learning on the normalized representation resulting in less noise.
5. The cortico-cortico pathways from V1 layer 2/3 to V2 layer 6 and layer 4 supports the divisive contrast normalization process.
6. The cortico-thalamic pathway from V2 layer 5 to the pulvinar then projects to V2 layers 6 and 4. The layer 6 to 4 pathway here in V2 also provides divisive normalization.

7. Cortico-cortical feedback from the now-activated V2 layer 6 projects to V1 layer 1 where it activates the layer 5 apical dendrites. The layer 5 cells activate the layer 6 to 4 pathway. This delivers a top-down expectation to the LGN.
8. The thalamic TRN cells are linked by gap junctions which synchronize activations across the two thalamocortical circuits that are activated while the bottom-up stimuli are being processed.
9. The nonspecific thalamic nuclei are receiving convergent bottom-up excitatory input from specific thalamic nuclei and TRN inhibition. The nonspecific thalamic nuclei project their resulting activity to layer 1 where they regulate mismatch-activated reset and hypothesis testing.
10. The specific thalamus in V1 matches feedback from the higher V2 area by comparing the (i) cortico-cortical feedback from layer 6 of the higher cortical area V2 which terminates on layer 1 of the lower cortical area v1 and (ii) cortico-thalamic feedback from layer 6 which projects to specific thalamus and the TRN.

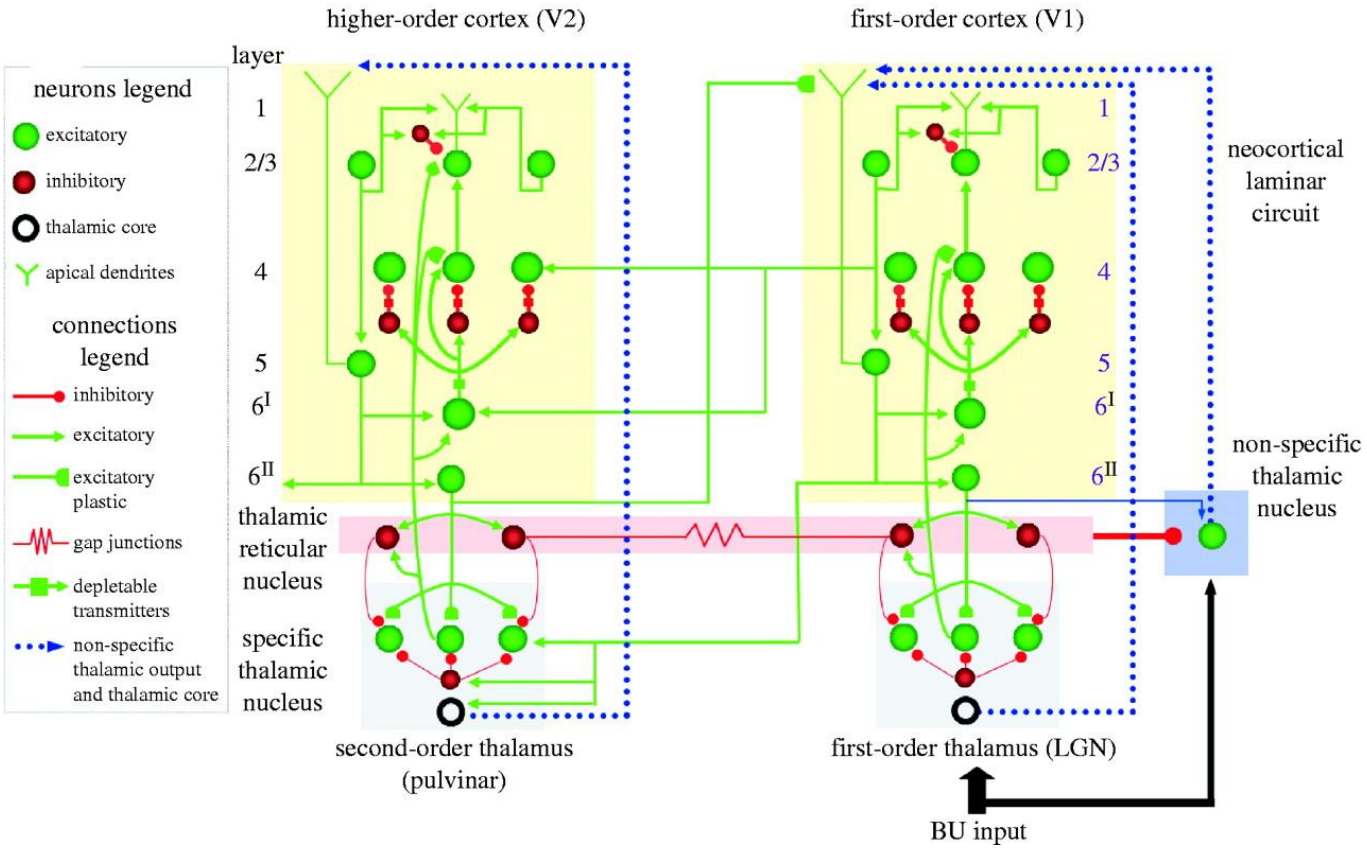

Figure 26: Cortical and subcortical predictive dynamics and learning during perception, cognition, emotion and action. (Grossberg, 2008) reproduced with permission from Elsevier.

The SMART model is a well-developed, large and complex model, which has been matched to many anatomical and functional studies. The SMART model and the Granger model have several commonalities:

i. they both utilize a cortico-thalamic-cortical loop structure with multiple feedback paths; and
ii. they share specific thalamic input which is progressively matched and netted out to focus processing on the residual input as the iterations of the feedback loop progress; and
iii. they agree that spreading activation of related cortical areas via layer 1; and
iv. they propose a series of higher order nonspecific thalamic loops that relate the inputs to higher order processing.

The general flow appears to concur with the Granger model and SMART offers many attractive extensions beyond Granger but it less explicit in the direct mechanics of how hierarchical temporal categorization results from the circuit as the Granger model is.

2.6.6. Jeff Hawkins' Cortical Layer Model

Jeff Hawkins advanced his Hierarchical Temporal Memory (HTM) theory in his book "On Intelligence" (Hawkins & Blakeslee, 2004). The goal of HTM was to produce an object recognition system which enables classification of visual objects over time. The more recent commercial goals are to predict temporal signals. The core neuroscientific advance that Hawkins' models are working on making is to uncover how the cortex implements predictive learning of temporally varying signals. The neocortex's anatomy and dynamics have been compared favourably to the functions implemented in the model, if not as much as the anatomical reality (George & Hawkins, 2009).

Hawkins' overall premise is that at its root the neocortex uses a hierarchical sequence memory for inferring and then storing causes of events in the world as received via the sensory inputs to the brain. Algorithms that attempt to emulate the cortex must create a hierarchical, spatio-temporal model of the sequences in the world and their relationships. Time is crucial to this process. This focus on time is a shift from many of the efforts at modelling cortical algorithms where most focus more (as of the date of the HTM's creation) on the feedforward structure of the networks, with neurons largely being integrate-and-fire and not containing much in the way of sequence memories, thus resembling more batch-mode processing.

One of the key elements of HTM, especially the newer versions (Hawkins, Ahmad, & Cui, 2017), is that the model proposes that the pyramidal neurons of the cortex use their proximal dendrites to predict when they should become activated (depolarized). Once activated, they can fire and learn based on whether they recognize any one of many sequences that signal the presence of their learned preferred direction vector in their inputs. The Granger model utilizes purely thalamic activation of cortical pyramidals to implement this context selection mechanism, but the functional goal is similar.

HTM builds this model of the world by pooling or categorizing both spatial and temporal aspects of the input it receives. The initial incarnations of HTM were based on Bayesian belief propagation principles whereas the more recent versions of the model use

interacting vectors representing cortical columns and layers, with vectors for each of the various components of cortical neurons (proximal dendrites, distal dendrites, soma integration models and axons). All of these vectors interact according to a set of heuristics modelling the activation of neurons to constrain the learning and prediction contexts (solving the learning-stability trade-off, similar in concept to how Grossberg approaches the problem, and in theory how the cortex solves it).

For our purposes, we are not going to reproduce the math of Bayesian belief networks or the implementation in George and Hawkins original paper instead, we are going to focus on what the HTM model (both the original base model and the more current model applied to 3D memory) say about the functions computed in the cortical layers and then relate that to the Granger model. In the original model (George & Hawkins, 2009) the following functions are theorized for the various cortical layers:

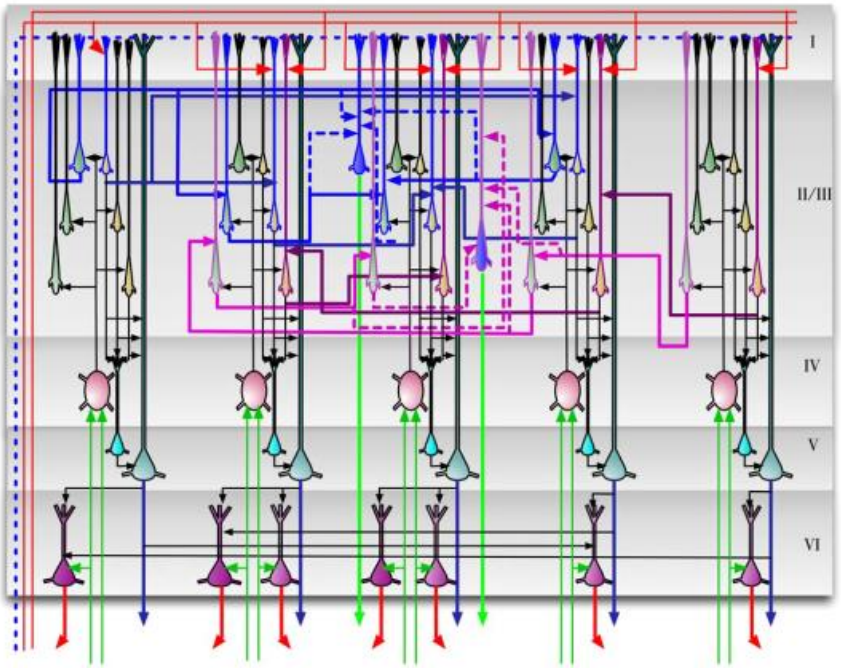

Figure 27 : HTM original cortical Bayesian Belief model. See text for cortical layer descriptions. (George & Hawkins, 2009) reproduced with permission from PLOS Computational Biology under their Open Access license.

1. Layer 1 a broadcast layer for feedback and timing information.

2. Layer 2/3 pyramidal cells store the sequence memories, pooled over the spatio-temporal structure of the sequences of inputs and incorporating the feedback received into the memories. The layer has pyramidal cells that;
    a. calculate the feed-forward Markov chain sequence of states;
    b. project the Markov chain information to higher level cortical areas; and
    c. compute sequences based on feedback where the top-down feedback constrains the assignment of events into the sequences already in cortex.
3. Layer 4 is:

d. the primary input layer (recall what we said though above in the anatomy section, that it is technically layer 3b that is the input layer); and also,
e. the place where the computation of the feed-forward probability of each sequence over the set of all coincident sequences occurs.

4. Layer 5 computes the belief over all coincident patterns, so given a set of coincident patterns, layer 5 determines which of the encoded beliefs is the one the system takes as the explanatory one to activate. In order to do this layer 5 implements the calculation of duration which is a major part of the model. The whole of HTM's rest upon the assumption that the cortex is storing sequences, and that means that the elements of a sequence are not the only things being stored, but that the timing between the elements is somehow stored. Hawkins and George propose that this happens all across cortex using different methods, but that a governing point of storage for the durations for sequence recognition, storage and recall is in layer 5.

   Hawkins and George believe that the neurons themselves cannot account for arbitrary delays on the scale needed to encode all sensory events. As such they find evidence for the neural circuit implementation of delays in the interaction of the non-specific thalamo-cortical projections interacting with their assumption that layer 1 implements a background timing signal to coordinate all cortical calculations, rather like a general system clock and an instance specific offset off of the system clock to time each element of a sequences commencement. The second source of delay they point to is due to the recurrent connections from layer-5 through the matrix thalamus to the apical dendrites of layer 2/3 and layer 5.

   One interesting point to note about layer 5 is that it is possible that layer 5 is the source of timing for sequence memory in the cortex. The argument is only theoretical, but because this will be relevant to our discussion of the Granger model we will mention it here. There are two kinds of pyramidal cells in layer 5; intrinsic bursting (IB) and regular spiking (RS). IB neurons get dendritic feedback from layer 1 predominately while RS neurons get feedback from layer 4. RS cells send their outputs to IB cells largely. IB cells send their outputs to subcortical areas including to non-specific thalamus and to motor areas. The authors postulate that the RS cells learn sequence associations and that the IB cells learn the timing between the sequence associations where the IB cells then send the resulting bursts to non-specific thalamus which then fires a cascade of bursts to cortex which activate the associated circuits in the sequence in the learned order. This is unproven but correlates well with the Granger model's role for nonspecific thalamus implementing the sequence linking each stage of the hierarchical categorization process.

5. Layer 6 receives inputs from layer 5, including the belief predictions about the next symbols from the layer RS cells, and computes the feedback messages from areas higher in the logical hierarchy to the lower areas. Layer 6 is considered here, as with all the models, as the primary source of cortico-cortical feedback. Connections from a lower (child) logical column to a given column's layer 6 neurons means that the child column's Markov chain is an element in the higher-level node's coincident pattern. The authors also point out that some layer 6 neurons are implicated in attention and that this would be direct result of their model as well. Selective activation of the layer 6 cells in the Granger model could also create a focusing effect supporting attention, although that is not claimed in the model.

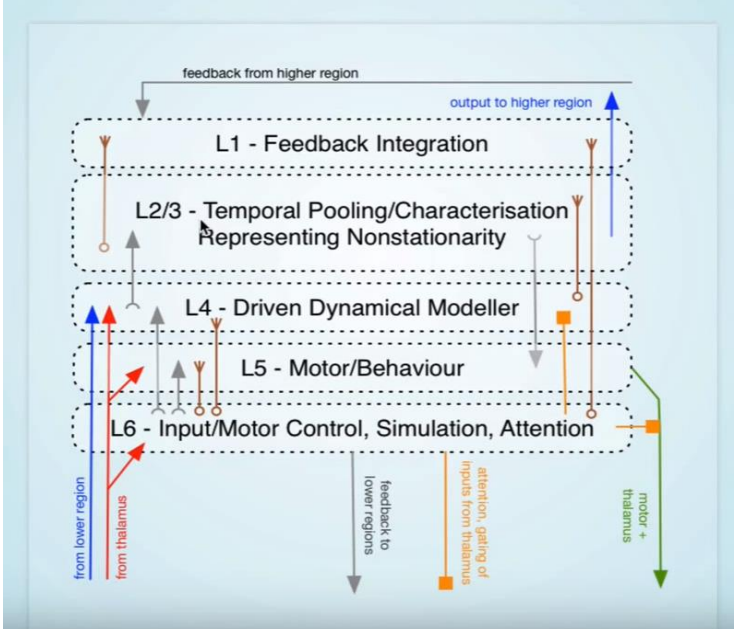

Figure 28 : HTM Updated Multilayer Cortical Layer Model (Byrne, 2015) reproduced with permission from arxiv.org open access.

Just recently, Hawkins et al. have updated the HTM Model and added several new features, the most striking of which is that they have proposed that cortical layer 6 encodes not just the belief of what is next in the RS cells and the timing of that belief in the IB cells but that it also encodes the location in space where the sequence of sensory data is coming from in an allocentric way. There is no such proposal in the Granger model.

2.6.7. Randall C. O'Reilly's LeabraTi Continuous Time Neural Model for Sequential Action Learning

Randall O'Reilly's (2014) revision to his 1996 PhD Thesis' Leabra model (O'Reilly R. C., 1996), Leabra TI, continuous time neural model for sequential action, adds temporal integration (TI) to Leabra and postulates that the core function of cortex is learning by

computing the difference between expectations and actual outcomes. Expectations are processed (minus phase) first and then outcomes (plus phase) are processed and the differences are learned across all of cortex. This form of learning they name ‘predictive learning’ and describe it more fully as predictive, auto-encoder, error-driven, temporal learning. It is based on temporal difference learning between from the original Boltzman machine formulation (Ackley, Hinton, & Sejnowski, 1985) and learns off the differences from the minus to the plus phase. The Leabra TI model then attempts to explain how the thalamo-cortical systems perform the representation of this temporal context and how that supports the predictive learning processes.

This is similar to recurrent networks with a driving rhythm to coordinate time with a learning step between oscillations. It is different than the models above where the learning is done between the top-down and bottom-up flow of information. Importantly the Granger model has both forms of learning, temporal and top-down/bottom-up spatial, which we take to be circumstantial support for its greater biological plausibility.

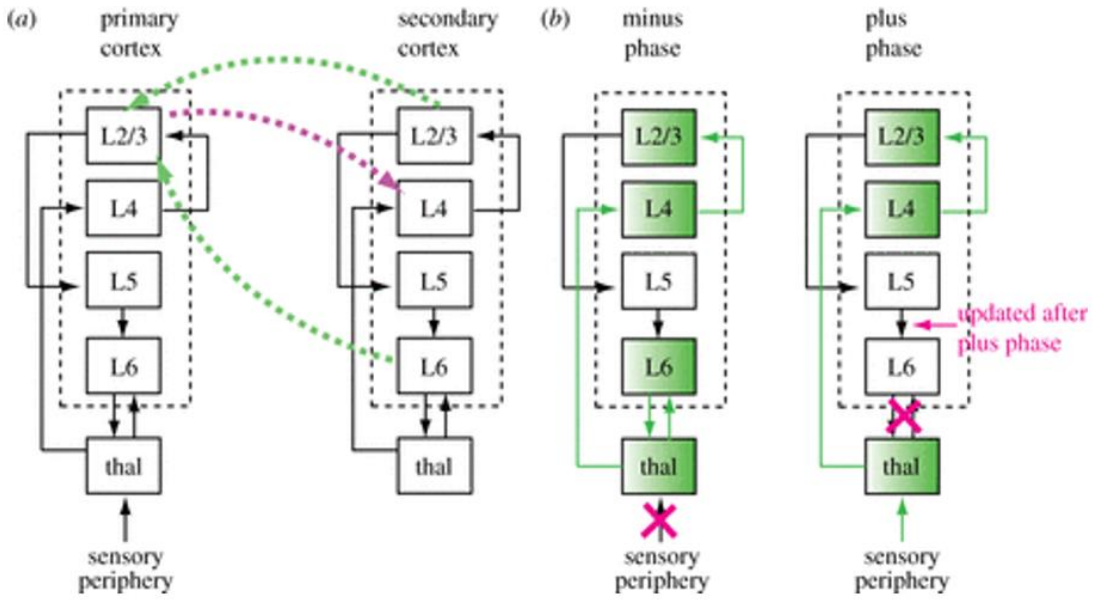

Figure 29 : Randall O'Reilly's Continuous time neural model for sequential action (Kachergis, Wyatte, O'Reilly, Kleijn, & Hommel, 2014) reproduced with permission from the Royal Society.

The basic idea is that the 10 Hz alpha rhythm, prominent in posterior cortex, creates a series of windows within which the previous oscillation’s context (including a prediction of the next oscillation’s context) are compared to the current oscillation’s inputs and processing state. Error-driven learning is then applied to the difference and the cortex’s predictive circuits are updated so that they make better predictions going forward.

The recurrent loop of context and prediction representation is our now familiar (as in the Granger model) thalamo-cortical loop with layer 5b being the generator of the intrinsic alpha frequency which then projects its outputs to layer 6 regular spiking neurons which then send their outputs to the thalamus, which finally sends its outputs back to the input

layers and 6. This recurrent loop sustains the context throughout the processing in cortex. In more detail, the model postulates that the cortex operates in the following cycle of perception and learning from O'Reilly et al.'s 2014 paper (O'Reilly, Wyatte, & Rohrlich, 2014):

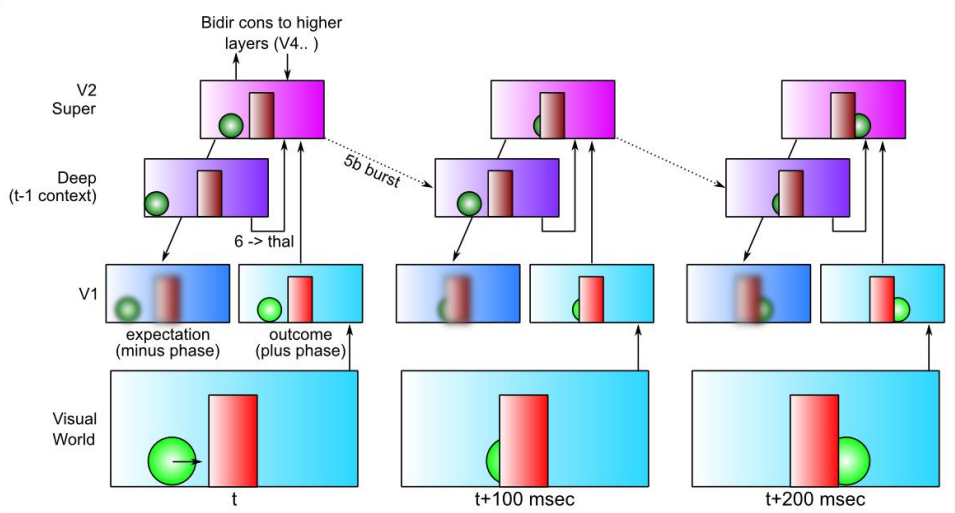

Figure 30 : An abstraction of what the Leabra TI model learns through time. Note the updated expectations in V1 over time. (O'Reilly R. , 2016) reproduced from arXiv.org open access.

1. Every 100 ms cortex integrates its prior activities that were captured to the current point across the deep layers of all of cortex generating a set of predictions that are integrated across all of cortex.
2. An alpha burst wave then propagates from the periphery inward across cortex driving the plus-phase of error-driven learning in comparison to the prior prediction which serves as the minus-phase. The alpha waves propagation is controlled by the layer 6b bursting dynamics and as it passes over the deep layers of cortex it updates the deep layers predictions for use in the next predictive phase. This enables the following predictions to integrate all prior context.
3. This integration of prior context to the current state of upper cortical layers and to the predictions made in the lower cortical layers creates a spatio-temporal representation of the learned predictions within the context of all cortical columns giving a systematic representational and predictive structure of the events in the world.

O'Reilly's model is shown in Figure 30. Note the expectation in V1 as it updated at each time step to a better prediction of the motion of the bar relative to the cylinder. This prediction is hypothesized to be learned in the layer 5b bursting neurons which is different to, but not contravening of, the Granger model, where the primary source of sequence learning, storage and hence prediction, is within the thalamic matrix cells.

2.6.8. Edmund Roll's Multiple Layered Attractors

We see that many models support the idea of both a flow of hierarchical categorical representations and their learned predictions thereof in thalamo-cortical loops. Such dynamics in large connected spiking neural networks. shaped by the various forms of inhibition discussed, and requires attractor states in order to maintain stable representations even for the short tens of millisecond scale processing times seen in cortex. One large body of work that gives a strong account of how these attractor states may be created by the anatomy and functioning of the cortex is given in (Rolls, 2016).

Rolls covers virtually all that is known about the cortex biologically. It is an amazing work. There is far too much knowledge for us to delve into here, but we wish to point out his overarching theory about the function of the cortical layers. He theorizes that the cortex is composed of three major computing motifs as follows:

1. Layer IV is the inflow of sensory and cortico-cortical input to each area.
2. Layer I is the back-projection layer from higher cortical areas to lower areas as well as the area of spreading activation from the amygdala and the hippocampal system.
3. Layers II & III are association cortex that implements a discrete, slow-long, auto-associative, hierarchical-categorization attractor network map across all the columns.
4. Layer V stores the stored dynamics of the concepts in the layers II & III attractor map.
5. Layers V and VI implement a continuous attractor map together.
6. Layer VI bridges those maps to its motor input and output memories.
7. The thalamus participates in this system by gating sensory input in the topographic nuclei and implementing intelligent routing and selection of sequences of action in the non-specific nuclei and their activity circuits formed dynamically with layer V.

This is both consistent with several of the models discussed above, such as O'Reilly's (if you posit that Rolls' layer 5 attractor maps are O'Reilly's layer 5b expectation learning networks). The Rolls model is also consistent with the Granger model exactly in terms of the above general functions. Rolls does not give his thoughts on the exact contents of the layer 2/2 or layer 5/6 attractor maps though. For an engaging read, and the most direct route to the core of his functional model, see Chapter 18 in (Rolls, 2016).

2.6.9. Deep Learning for Categorization

It is now important to mention deep learning here both to position the Granger model properly as well as to ensure that there is no confusion when we state that we are in search of a more biologically plausible categorization algorithm, that we are not saying that we are in search of simply categorization in of itself.

Incremental error-driven learning has proven very useful at learning in many algorithms designed to learn the structure of the world (O'Reilly, Wyatte, & Rohrlich, 2014). Starting from the original backpropagation algorithm (Rummelhart, Hinton, & Williams, 1986), to

the autoencoder model, to the later support vector machines (Cortes & Vapnik, 1995) and more recently deep learning (Krizhevsky, Sutskever, & Hinton, 2012). Currently convolutional, feedforward, sparse-representation, backpropagation networks with batched learning are the best performing networks for categorization. These networks are not, however, recurrent nor are they able to learn temporally continuous input like biologic networks are (though they do use discrete time slicing as an approximation in the case of LSTMs).

Currently, deep learning more than suffices for many categorization tasks, but the Granger model serves a different purpose: it is a proposal of how the earliest stages of cortical dynamics over time implement categorization as a foundational dynamic and learning process to support full cognition and consciousness. It attempts to explain temporal sequence learning in a way that integrates with other functions of brains such as dynamic spatio-temporal prediction. The Granger model is a set of postulates that have strong biological support and are part of a system that suggests mechanisms for performing of these advanced functions.

2.6.10. Thalamo-Cortical Resonant Functional Circuits Communicating Via Nested Oscillations?

> *"The mechanism for the emergence of correlation, synchronization, or even nearly zero-lag synchronization among two or more cortical areas which do not share the same input is one of the main enigmas in neuroscience. It has been argued that nonlocal synchronization is a marker of binding activities in different cortical areas into one perceptual entity"*
>
> (Olshausen, 2013)

We know that biological categorization involves multiple sensory perceptions. We also know that the brain has many nuclei of neurons and those many nuclei are further divided into layers and columns. Yet we experience one percept of things in the world. We know that, at some level, the activities of these various neurons and areas must be bound somehow together. We now review several mechanisms of binding neural activities at various levels to elucidate possible ways in which this coherence of percept might correlate with coherence of neural activity.

As we have discussed above, in the section on synfire chains, if we accept the premise the networks of asynchronously firing integrate-and-fire neurons would move toward global saturation or a paucity of activity (or at the least they cannot reliably reproduce activity patterns due to propagating inconsistent activity offsets), then networks of neurons that fire in an aligned manner may require some form of inhibitory isolation from the rest of the connected network for each of the various dynamically formed interacting circuits. [We say they "may" require this because from Spaun (Eliasmith, et al., 2013) we know that it is in fact possible to do binding without such hypothesized inhibitory isolation methods. Which is biologically correct? We do not know yet.] They may also require some form of temporal control to implement the alignment between these dynamically formed

circuits to reliably propagate and process information. Under this view then, in order to coordinate the outputs of these integrated functional subnetworks, the brain would need mechanisms for temporal alignment and spatial segregation to control which circuits are active at any time.

Within thalamo-cortical circuits there are several well-known frequency bands of activity: delta range at 1 to 4 Hz; the theta range at 4 to 12 Hz; and the gamma range at 30 to 40 Hz or higher. Resonance among the thalamo-cortical column circuits is often pointed to

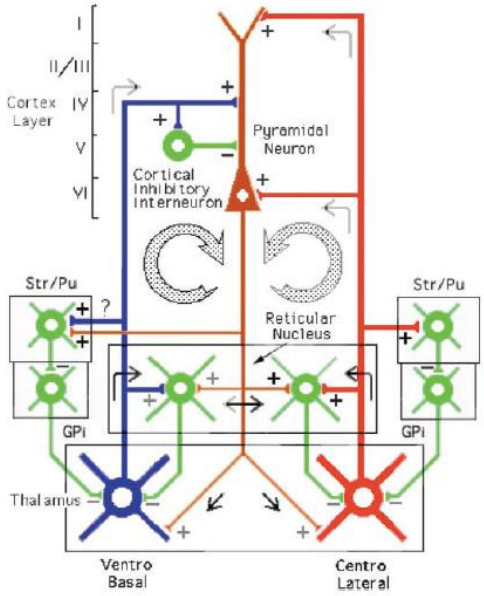

Table 7: Temporal binding via cortical coincidence detection of specific and nonspecific thalamocortical inputs: A voltage-dependent dye-imaging study in mouse brain slices (Llinas, Leznik, & Urbano, 2002) reproduced with permission of PNAS under open access policy for academic use.

as being the basis for distributed neural populations to synchronize into a functional circuit across both topological and rhythmic ranges. Llinas et al. model the mouse thalamo-cortical circuits arising from the specific thalamic nuclei, which provide the sensory content that is bound to representations of internal context and relations from the non-specific circuits' outputs.

He posits that recurrent, resonant gamma-range activity in specific and non-specific thalamo-cortico-basal-gangliar-cortical circuits connected to the ventral and central thalamus is the mechanism behind cognitive binding (Llinas, Leznik, & Urbano, 2002).

In theory, this binding combined with integrations (such as the summation of activities within higher-order, winner-take-all circuits in higher-order cortical areas) could create a recurring temporal window for the faster circuits to be summed over to implement a two-level processing. The Granger model has exactly such a structure where the incoming activity from the specific thalamus enters cortex to layer 4 and propagates feedforward upward to layers 2 and 3, while the layer 4 signal also travels downward where interactions with layer 5/6 and non-specific thalamus result in inhibitory feedback from the thalamus and therefore feedback subtraction of the recognized parts of the inputs. While

there is no claim within the Granger paper that the feedforward path operates at the alpha frequency nor that the feedback path operates at the gamma frequency, it is worth further study to assess the possibilities as a means of understanding the model and the biology better. The Llinas' model, mentioned above, has the same lower feedback pathways in visual cortex oscillating at the gamma range as we reported above.

Llinas states that thalamo-cortical resonance is the 'basic coinage of cognition' (Llinas, Leznik, & Urbano, 2002) supporting global cognitive binding. This is broadly in agreement with the Granger model in that the overall layer models are the same and the incoming sensory information arises from the specific circuits, while the nonspecific circuits learn the flow between the cortical circuits representing the next categories in the learned hierarchy.

As to mechanisms for resonating thalamo-cortical circuits to coordinate information exchange with each other beyond the ones directly connected, there is a growing body of evidence that the coupling of activities may be coordinated on several levels by the interaction of the multiple neural oscillations at varying frequencies. The mechanisms of the actual oscillations are still largely unknown, but the prevailing view is that the neural circuits in-vivo by virtue of their activity within their excitatory / inhibitory connectivity (Bosman, Lansink, & Pennartz, 2014) generate rhythmic neural activities as they do in artificial circuits. The brain rhythms most commonly thought of in this kind of description are the field fluctuations measured at the scalp, which arise from the combination of neural spiking, ionic flows as well as below-threshold fluctuations in the intra- and inter-cellular currents, which then sum. Here though we are referring more specifically to the actual neural spiking of connected networks of neurons in various parts of the brain.

Activity oscillating at different frequencies in different circuits has been shown to interact in various ways. One type of interaction is when slower oscillating activity constrains the amplitude of faster oscillating circuits, in what is called phase amplitude coupling (PAC), which is a form of the general class of theorized cross frequency coupling (CFC) mechanisms. As an example of studies that narrowed the coordination via oscillation down to a specific set of responsible cells, Dann et al. found in primates that fronto-parietal functional networks were coordinated by a small group of frontal neurons oscillating at in the beta range (Dann, Michaels, & Schaffelhofer, 2016).

Higher frequency activity, like the gamma band at 40 Hz approximately, has been shown to be a more local activity frequency in the lower cortical layers 4, 5 and 6 of cortex. It has also been shown that lower frequency activity, such as the alpha frequency at approximately 10 Hz, entrains larger areas of cortex and is found predominately in the upper cortical layers 1, 2 & 3. (Canolty & Knight, 2010). Gamma has been shown in visual cortex to arise in layer 4 and propagate to the lower cortical layers, and is, associated with feedback processing. In the same study, alpha has been shown to arise in layer 4 and propagate upward to the higher layers, while being associated with feedforward processing (Kerkoerlea, Selfa, Dagninoa, Gariel-Mathisa, & Poorta, 2014).

There is also evidence for oscillatory information communication across layers. Laminar studies have shown that oscillation frequencies are compartmentalized within specific

layers (Bosman & Aboitiz, 2015), and this is theorized to be evidence that rhythmic synchronization of spiking activity enables efficient inter-areal dynamic information transfer among areas of the brain (Fries, 2005). Compartmentalized oscillations within layers may also contribute to segregation of feedforward and feedback processing (Bosman, Lansink, & Pennartz, 2014).

PAC has been theorized to enable the coordination of information in time as the alternating phase of the lower frequency oscillation provides a window to organize the dynamics of the faster oscillating circuit, giving a spatio-temporal organization to the integration of the information from both circuits (Canolty & Knight, 2010). Studies on simulated networks of attractors with nested gamma oscillations within theta (2 to 5 Hz) have even shown that the nesting of the oscillations can drive the attractor networks to encode and retrieve precisely timed spatio-temporal firing patterns, implementing spatio-temporal memory (Hermann, Lundqvist, & Lansner, 2013).

While the Granger model does not make frequency predictions, it is interesting to offer the possible integrating theory that as the categorization activity flows upward from specific thalamus to layer 4 as in the Granger model, it may do so by eliciting alpha frequency activity in layers I, II and III, which then propagates downward the activities within the layers 5 and 6 attractor networks. The interaction of these frequencies may also be a form of matching of categorization to the dynamic predictions stored in layer 5 (as discussed above in Hawkins latest theories). Possibly, various combinations of predictions are thereby filtered through and the one matching best to the bottom-up sensory inputs and the top-down alpha frequency is selected, resulting in a set of resonating and therefore bound circuits and representations. The Granger model, while not explicitly supporting such a theory, is not in contravention of it.

2.6.11. Possible Neurobiological Coordination and Communication in the Granger model

In theory, taken together this synchrony to form dynamic cortico-thalamic circuits and PAC to communicate between the higher order representations in layers 2/3 and the supposed gamma-mediated representations in layers 5/6, would seem to be a plausible set of mechanisms to participate in the neurobiological implementation of the Granger Core and Matrix circuits. This theoretical biological implementation of the model could form the target hierarchical categorization and sequence learning model. When also combined with the theorized spreading activation of cortical areas through layer I, it would seem to lend support to Granger's view that the brain computes sequences of categories forming a brain grammar composed of the activity over time in multiple thalamo-cortico-cortical functional circuits. The layer II and III attractors would implement the higher order concepts-memory through learning, pattern completion and prediction of categories. The hierarchical connections from cortical area to area would implement the hierarchical relationships amongst the learned categories, and the integration of multi-sensory information into the definitions of those categories.

A recently published modelling study of simulated populations of gamma oscillating excitatory and inhibitory interconnected neurons (cortical column analogue) driven by two alpha-oscillating simulated neurons (thalamic pulvinar analogue) show that the phase difference between the two alpha driver populations modulates the synchrony of the cortical column analogue populations (Quax, Jensen, & Tiesinga, 2017). This study is a model, and not evidence in-vivo, but evidence of similar effects in humans exists and lends support to the concept of phase controlled activity synchrony and information exchange. A 2014 study of human cortico-thalamic coherence showed direct evidence of thalamic theta controlling activity in cortex and controlling thalmo-cortical theta-beta phase-amplitude coupling (Malekmohammadi, Elias, & Pouratian, 2015).

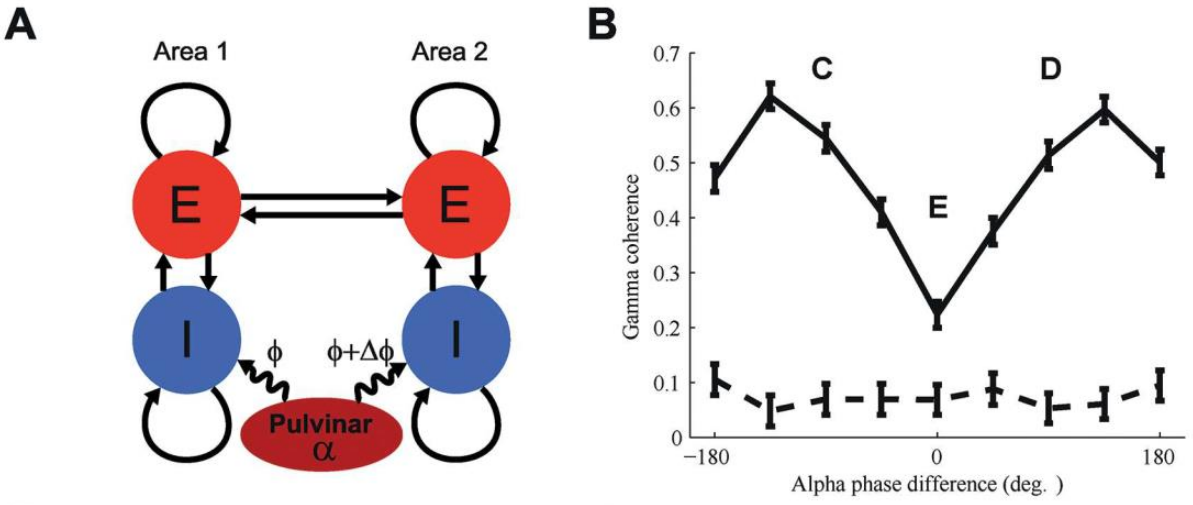

Figure 31: Alpha-phase modulated gamma synchrony model. (A) network model where E= excitatory population, I = inhibitory population; B - graph showing gamma synchrony modulated buy the phase difference between the two alpha populations rhythms (Quax, Jensen, & Tiesinga, 2017), reproduced with permission from PLOS Computational Biology under open access license.

These models are at least suggestive of thalamo-cortical binding within a sequential process, which evolves over several cycles driven by the circuit design, and possibly mediated by alpha/gamma nesting, to bind an incoming percept between the cortical column representations and thalamic core cell representations. While not confirmatory, these are the same mechanisms that the Granger model relies on: an incoming sensory flow; a multi-stage cortical processing step; feedback to the thalamic non-specific nuclei; and spreading activation of related cortical areas over a timescale involving several sequential loops of processing.

The Granger model adds something significant to these models: the suggestion that this process is happening at ever increasingly specific frames of reference, progressively, hierarchically categorizing its inputs against already learned patterns. We have discussed many theories and their commonalities as to how hierarchical, categorical representations could be implemented in the brain. We now need to have a theory as to how seriality could emerge in neural networks at various time scales.

2.6.12. Dynamic Functional Cortico-Thalamic-Cortical Networks and Consciousness

Lastly, we mention that some authors go further and lay the whole of consciousness, not simply cognition, at the feet of an integrated model of thalamo-cortico-cortical functional circuits. Edward G. Jones (2002) attributes conscious events to the functioning of these cortico-thalamic-cortical networks.

> *"Interactions of focused corticothalamic axons arising from layer VI cortical cells and diffuse corticothalamic axons arising from layer V cortical cells, with the specifically projecting core relay cells and diffusely projecting matrix cells of the dorsal thalamus, form a substrate for synchronization of widespread populations of cortical and thalamic cells during high–frequency oscillations that underlie discrete conscious events"* (Jones E. G., Thalamic circuitry and thalamocortical synchrony, 2002).

Many authors have proposed that synchronous activity among cortical areas is the source of cognitive binding, while others conclude thalamo-cortical resonance is the minimum required level of coherence. For a summary of the various authors views see (Llinas, Leznik, & Urbano, 2002). A 2006 paper that has received significant citations (over 500) and caused significant debate is Lamme's "Towards a true neural stance on consciousness." (Lamme, 2006). Lamme makes a sweeping argument that the study of consciousness needs to focus on the neuroscience components far more than the personal experience of consciousness that has been the subject of much of the research on consciousness for far too long. He reviews many studies, such as lesion deficit studies, and concludes that taken together they strongly suggest that consciousness can only occur, and is likely due to, the recurrent processing along the ventral visual stream to frontal cortex and back again. For our purposes here, a proper discussion is far beyond our scope, but it is relevant to point out that all of the areas pointed to in Lamme's theory pass through the thalamo-cortical loops that are the subject of the Granger model. The thalamo-cortical loops resonance that has been discussed above, combined with the longer inter-area and inter-layer binding by oscillations theories offers a possible architecture for further study, to host Lamme's recurrent processing along the visual stream in concert with the cortical columnar computations within the areas that he points to. Interconnecting Lamme's overall argument for consciousness with Granger's machinery for computing the grammar of thought is tempting. As with all reaching statements without proof, there is value in offering a theory for study, but nothing of certainty can be said about it until further evidence emerges. We find great value, if for nothing other than our own proto-theorizations, in the process of trying to see how top-down theories might connect with Granger's more bottom-up circuit mechanics.

3. Categorization

As mentioned in the introduction, Granger proposes that the brain computes with a recursive grammar composed of sequences of categories that are themselves generated from two recurrent circuit motifs, the core and the Matrix circuits. The neural correlates of these are found in the many thalamo-cortico-cortical loops in the brain. Having reviewed the anatomy of the cortex, the thalamus and the thalamo-cortical loops, and also reviewed some of the more prominent theories on how the cortex process information (possibly support the Granger model's temporal grammar formed from the dynamics of thalmo-cortical processing) we now turn our attention to the core function of such a grammar as proposed by Granger: hierarchical categorization. We begin by reviewing the algorithms used for categorization, then discussing the Granger model in detail, and finish describing our spiking neural implementation of the Granger model's Core circuit. Categorization is:

> *the mental operation by which the brain classifies objects and events. This operation is the basis for the construction of our knowledge of the world. It is the most basic phenomenon of cognition, and consequently the most fundamental problem of cognitive science.* (Cohen & Lefebvre, 2005)

Clustering or unsupervised nonparametric categorization is the process of grouping unlabelled data into collections based on similarity such that the instances in each collection are more like each other than they are to the members of any other collection. Biological categorization systems naturally evolve under two optimization criteria imposed by the costs and structure of the biological world; (i) categorization mechanisms evolve to provide as much information as possible for the least biological effort, and (ii) the maximum information is achieved if the structure of the categories created matches the structure of the world and the sensory data obtained from it (Rosch & Lloyd, 1978).

3.1. An Historical Overview of Categorization Methods

Categorization dates back to the Greek classical period where rigid categories were manually assigned. Conceptual clustering came along later where clusters are assigned based on rules that can assign different levels of application to different categories.

Modern classification either relies on fuzzy set theory, or prototype theory. In fuzzy set theory objects are assigned based on mathematical membership functions allowing for complex partial belonging, in prototype theory, a given exemplar is supplied, which potential members of the given class are compared to, and then assigned to the category with the highest congruence between the instance and the exemplar.

The Granger model is similar to a k-means clustering that builds up a set of category exemplars over time. The category exemplars are created by a closest match to the strongest stimulating features that the neural circuits representing them have been

assigned. This assignment happens hierarchically through neural plasticity brought about by a competitive winner-take-all activation process. This allows the model to create categories through best-matching and competition, without prior exemplars, even starting from random states for all categories. The Granger model works this way.

It must be noted that the idea that exemplars are learned at all is a subject of debate, a recent article suggests that human categorization stems from innate exemplars (Branan, 2010).

3.2. The Implementation of Categorization in the Brain

There are many variations and combinations of these and other methods, at many levels of performance and biological plausibility discussed in the literature. The brain's ability to make categories of objects and concepts has always been a core area of study in cognitive science and computational neuroscience. The brain gives animals a distinct advantage using categorization because it creates compactness of representations, allows translation of learning from one instance to the next, and, generality of learning. Organisms only see a fraction of the possible instances of stimuli, but must identity, predict, and learn to both react to, and influence, the full spectrum of instances of objects, environmental features, and other creatures. Categorizing stimuli and concepts facilitates these functions (Plebe, 2011).

The physical implementation of categorization in cortex is a subject of much research. The environment, all things in it, and their likely next actions must be identified from dispersed and disjoint sensors. Modelling the continuum of reality is in some ways an act of continuous categorization. There are several areas of the brain that have been implicated in specific forms of categorization for various purposes. Some examples are:

i. a well-studied area of cortex, the ventral temporal cortex (VTC), implements part of visual categorization;
ii. the striatum has been implicated in the categorization of the components of objects, and theorized to work in concert with the prefrontal cortex in relation to task execution (Antzoulatos & Miller, 2011).

Because much of neuroscience is consumed with the brain's processes for turning sensory input into an internal semantic representation, much of neuroscience is about categorization. Our focus here is on how the incoming stimuli are categorized by neural dynamics and how those dynamics are implemented in the cortical columns and thalamic nuclei. While the Granger model does not speak to the specific layout of the neurons across cortex that implement categorization, it does theorize that categories are distributed spatially across cortex, for sensory representations, mirroring the spatial arrangement of the core cells of the thalamus. The model proposed by Huth et al. agrees at this most general level.

3.3. Spatio-Temporal Categorization in the Granger model

Does such a topographic, spatial organization extend to the representations of complex, higher order categories such that the Granger model 's topographic and layer-based category encoding could find potential support there as well?

Spatial organization of categorical representations in cortex has been found in several areas. As an example, a review from 2017 gathers evidence that the categorical receptive fields of the VTC are spatially arranged. The pattern is organized by convergent clustering and divergent topological ordering as well as super-position into a nested spatial hierarchy across the cortical expanse. In other parts of cortex, such as V1 and the entorhinal cortex, columns stack different receptive fields, where in VTC they differ across the cortical sheet.

The idea that the brain stores a set of significant exemplars is in line with the design concepts from the Granger model. It still interesting to attempt to grasp the full extent of the problem in mapping the massive dimensions of the real world into the storage system of the brain. It becomes ever clearer how much the brain must rely on categorization as a means of compression for storage, processing, comprehension, prediction, interaction and measurement of the world and its dynamics.

The receptive fields, down to the resolution possible with a voxel, were estimated to be storing between 35 and 50 dimensions of the stimulus space (Grill-Spector & Weiner, 2014). This likely means that few of the actual instances of the possible real-world exemplars can be stored simply because although there are many neurons in the brain, the permutations of perspectives, lighting, deformations, transformations and types of real world stimuli is larger. The full dimensions of this stimulus space are unknown. Likely this space would be composed of all manner of sensed properties, plus derived ones, varying across multiple time scales. It is reasonable to conclude that the cortex stores not the full breadth of its experience but rather a selected, reduced set of meaningful exemplars.

fMRI experiments of human subjects categorizing images have been used to train ResNet networks to predict cortical activations, which were then validated by fMRI scans. From such trained ResNets, the structure of the cortical category storage was inferred. The results suggest that the cortical category maps are distributed, over-complete, property-based representations filed away into nested hierarchical categorization maps (Wen, Shi, Chen, & Liu, 2017). In so far as the Granger model is concerned, the distribution of receptive fields is modelled across the artificial cortical sheet in using a similar topographic layout which seems very biologically plausible given the anatomical and simulation evidence.

Four functions are possibly sufficient to create categories in a self-organizing system: (i) coincidence detection of stimuli; (ii) coding that coincidence for future retrieval with just a partial set of the stimuli; (iii) the ability to both learn the associations and refine them over time as well as to increase the aversion of mutually exclusive associations; and finally (iv) a mechanism for scoring how well a given stimulus matches the already learned

categories by predicting which categories the stimulus will elicit and then refining the predictions over time (Plebe, 2011). This combination of self-reinforcement combined with constrained competitive compensation scored by predictive accuracy may be enough to give rise to the cortical maps seen above. The Granger model is consistent with this mechanism.

3.4. Cortical Categorization Model Algorithms

Categorization models began in the psychological and cognitive sciences with conceptual category models divorced from cortical and neural mechanisms that implement them (Plebe, 2011). Plebe (2011) gives a excellent review of the evolution of cortical categorization algorithms. The rest of this section on cortical categorization algorithms follows his summary closely. For our purposes, this will allow us to review the history of categorization algorithms quickly but with sufficient depth so that we may then move on to discuss cortical dynamics and the temporal processing of categorization. We then situate the Granger model's hierachical, temporal, progressive categorization model within the context of the above discussion on thalamo-cortico-cortical circuits and categorization algorithms.

3.4.1. Von der Malsberg's Visual Circuits Model

Von der Malsberg's visual circuits model (Plebe, 2011) is defined as

$$\frac{\partial y_i}{\partial t} = -\alpha y_i + \overrightarrow{k_\iota} \cdot f(\overrightarrow{y_\iota}) + \overrightarrow{w_\iota} \cdot f(\overrightarrow{z_\iota}) + x_i \qquad (1)$$

$$\frac{\partial w_{i,j}}{\partial t} = \eta f(y_i)\left(f(z_j) - w_{i,j} \sum_{l \epsilon z_i} f(z_l)\right) \qquad (2)$$

where y_i = the neuron activity in layer IV, Z_i = the neuron activity in non-layer IV connected to neuron yi, $w_{i,j}$ = the connection strength between z_j and x_i, $\overrightarrow{k_\iota}$ = the gaussian kernel function modulates activities near i, $\overrightarrow{w_\iota}$ = the vector of all connections of neurons connecting to i, $\overrightarrow{y_\iota}$, $\overrightarrow{z_\iota}$ = the activations of neurons that connect to i, f = the sigmoid non-linearity, and η = the learning rate.

Von der Malsberg introduced the first model of how visual cortex neural circuits might self-organize. von der Malsberg's system of differential equations implements Hebb's learning principle between neurons as the mechanism of self-organization.

The basic idea is that layer IV pyramidal neurons' activities are sigmoidal, like neural spiking dynamics and saturation and the neurons in a cortical layer are locally excitatory and then laterally inhibitory beyond the immediate vicinity, falling off in a 2-dimensional Gaussian shape. The weights are then adjusted based on these activity dynamics using a Hebbian mechanism in the lower equation to learn these dynamics over time.

The relevance for our discussion is that this was among the first models to try to capture self-organizing category behaviour (in visual cortex V1). Also, it recognized that a

combination of local excitation and distal competition was required to do so and that these forces interacted with each other producing topographically organized receptive fields.

3.4.2. Kohonen Topographic Self-Organized Maps (SOMs)

Kohonen (1982) popularized the concept of topographical self-organization, making it simpler and more explicit in his famous Self-Organizing Maps (SOM) model. The model starts with M nodes that are analogues for neurons in a cortical sheet. The neurons have preferred direction vectors and connection weights to all the other neurons.

Kohonen's self-organizing map equations model (Kohonen, 1982) is defined as

$$w = arg\ min_{i\varepsilon\{1,\ldots,M\}}\{\|\vec{v} - \overrightarrow{x_i}\|\} \quad (3)$$

$$\Delta\ \overrightarrow{x_i} = \ \eta e^{-\frac{\|\vec{r}w - \overrightarrow{r_i}\|^2}{2\sigma^2}}(\vec{v} - \overrightarrow{x_i}) \quad (4)$$

Where $\vec{v}$ = the input vector, M = the number of neurons in the SOM, w = the winning neuron at each training iteration, η = the learning rate, $\overrightarrow{r_i}$ = a two-dimensional coordinate location in the SOM for neuron i using 0:1 as the range for both dimensions, $\overrightarrow{x_i}$ = the preferred direction vector of neuron i, and σ = the amplitude of the neighborhood used when updating.

An input vector is computed at each time step and the preferred direction vectors then the weights of all neurons are updated according to the equations shown. The result is a map of the input space topographically organized with the continuum of preferred direction vectors in the neurons. The top equation selects the winning neuron by finding the neuron that has the least absolute difference between its' preferred direction vector and the current input vector. The lower equation then updates the preferred direction vectors of all neurons based on how close topographically they are to the winning neuron.

SOM's implement an instance of the Winner-Take-All (WTA) mechanism that is very popular in many models trying to capture the overall effect of lateral inhibition in the cortex. The downsides of SOM, as with van de Malsburg's model, are that the WTA mechanism at this high-level abstracts away all the effects of the non-uniform nature of the lateral inhibition in cortex (Plebe, 2011). In SOM the shape of the neighbourhood inhibition function can be chosen, but it is uniform. In cortex, each and every inhibitory and excitatory connection is created or pruned and tuned across layers, not just in a two-dimensional plane.

3.4.3. Laterally Interconnected Synergetically Self-Organizing Map (LISSOM)

To add more realistic lateral connection topologies, the Laterally Interconnected Synergetically Self-Organizing Map (LISSOM) was developed (Sirosh & Mikkulainen, 1997). It added lateral inhibitory and excitatory connections to each neuron in addition to the input vector (Plebe, 2011).

$$x_i^{(k)} = f\left(\frac{\gamma_{\mathrm{A}}}{1+\gamma_{\mathrm{N}}\vec{I}\cdot\vec{v}_{r_{\mathrm{A}},i}}\vec{a}_{r_{\mathrm{A}},i}\cdot\vec{v}_{r_{\mathrm{A}},i} + \gamma_{\mathrm{E}}\vec{e}_{r_{\mathrm{E}},i}\cdot\vec{x}_{r_{\mathrm{E}},i}^{(k-1)} - \gamma_{\mathrm{H}}\vec{h}_{r_{\mathrm{H}},i}\cdot\vec{x}_{r_{\mathrm{H}},i}^{(k-1)}\right)$$

where $\overrightarrow{vi}$ = the input vector, r_E = the circular radius of excitatory neurons lateral connections, γ = connection strength scalars, r_H = the circular radius of inhibitory neurons lateral connections, $\overrightarrow{xi}$ = neuron i's preferred direction vector, A = matrix of neural activities, H = matrix of inhibitory connection weights, E = matrix of excitatory connection weights, $\vec{x}_{r_{\mathrm{E}},i}^{(k-1)}$ = the lateral excitatory connections vector, $\vec{x}_{r_{\mathrm{H}},i}^{(k-1)}$ = Lateral inhibitory connection vector, $x_i^{(k)}$ = activation of neuron x_i at timestep k.

In LISSOM the weights are updated by a series of Hebbian equations (not shown, see (Plebe, 2011) for details). The model has been used by Plebe to implement a series of systems that are able to emulate early word learning and real-time object recognition using uni- and multi-modal sensory inputs. One of LISSOM's contributions is that it explicitly models lateral inhibitory and excitatory connections contributing to its biological plausibility.

LISSOM does not resolve what the activity of a single neuron comes to represent (Plebe, 2011). Does the cortical stack represent layers of neural networks the implement hierarchical maps resulting in activities of single neurons in the top most layers representing specific things in the real world? These conceptual neurons have been called "grandmother neurons", alluding to the existence of a single neuron, high in the logical cortical stack, that responds only when one sees their grandmother.

(5)

Hinton (Hinton, McLleland, & Rumelhart, 1986) and Churchland (Churchland, A Neurocomputational perspective: The Nature of Mind and the Structure of Science, 1989) have proposed that neural activities across many cortical columns represent meaning, a concept that has become known as any of the following: sparse distributed representations (Rinkus, 1995), distributed coding, population coding, vector coding and state space coding. Many papers have shown evidence of population coding in cortex (Plebe, 2011).

The central idea in population coding is that neurons that fire at the same time collectively represent the meaning of a higher-level concept, and that each individual neuron does not in and of itself represent a real-world concept. In this regard, all forms of population coding share similarities with the NEF, which the computational model of this thesis is built with. Natively, the NEF is not topographic but our modelling work takes NEF vectors and treats them as topographic neural ensembles representing categories. While still a long way from the cortex, incorporates more of the factors we see in biology.

3.4.4. Toward Dynamic Categorization Models

None of the models account for the hierarchical, spatio-temporal categorization that the Granger model proposes. On the other end of the spectrum, biophysical models of neural networks are progressing but as yet do not encapsulate categorization algorithms.

In our model built with Nengo using the NEF methods, we will use spiking dynamics, recurrent connections, WTA mechanisms and recurrent explaning-away to propose how spiking neural networks can implement hierarchical categorization.

4. The Granger Model

Granger argues that brains compute with grammars that encode time structured relations. The structure of this grammar is an extensive hierarchical tree of sequences of categories, of sequences of categories, and so on (Granger R. , 2011). For a visual example of the categorization of the image of a flower see Figure 34.

Similar theories have been advanced that the brain has a symbolic basis to its computation. Some argue more for a series if programs built of symbols (Ballard D. H., 2015), Eliasmith argues that semantics are computed using a set of semantic pointers that are manipulated directly at the higher cognitive levels, and expanded to the fuller meaning when needed (Eliasmith, 2013). Eliasmith's model is not inconsistent with an implementation in biology via a combination of; Granger's model of the emergence of categories in cortical layers II and III (holding perhaps Eliasmith's higher order concepts), or with Rolls and Hawkins theories that the lower layers encode the dynamics and specifics of such higher-layer attractor-based concepts.

4.1. Two High-Level Abstraction Circuits: Core & Matrix in Posterior Cortex

The Granger model posits that there are two distinct interoperating circuits that implement the cognitive behaviour of categorization. These circuits are both thought to be formed from neurons in posterior cortex and their associated thalamic circuits. Granger refers to these two circuits as the Core circuit and the Matrix circuit. The Core circuit is responsible for categorizing inputs from the thalamus. The Matrix circuit learns the sequential flow of the categories. The Core and Matrix circuit models use several features of cortical functional anatomy such as: different time courses for excitatory and inhibitory neurons, and different axonal arborisations for the pyramidal neurons versus the interneurons. These core-circuit and matrix-circuit patterns hold for the primary sensory areas found in posterior cortex, such as the VPM/VPL, LGd., MGv, Pom, LP/Pul, and for the MGm projections to layer I of somatosensory, visual, and auditory cortices. They are also evident in a wide array of thalamic nuclei, intralaminar and non-intralaminar thalamic nuclei.

Neither the Core nor the Matrix circuit model covers the myriad of inhibitory neurons in the cortex beyond a few significant ones. Hawkins' latest model (Hawkins, Ahmad, & Cui, 2017), mentioned previously, goes into more detail, especially regarding layer 5 and seems compatible with the Granger model. The Granger model focuses on explaining categorization in the posterior cortex and not the anterior cortex, whose circuits connect to the ventral thalamic nuclei and involve the striatal complex. We should mention that Chandrashekar and Granger have published more recent work showing a modified version of the categorization circuit in those areas (Chandrashekar & Granger, 2012), which is beyond the scope of this thesis.

4.2. Processing Flow

The overall flow in the Granger model is that sensory information flows from the periphery to the core cells then to the middle layers of cortex, to superficial layers and then to deep layers. The topographic Core circuit organizes stored memories into similarity-based hierarchies using hierarchical clustering. The non-topographic and diffuse, Matrix circuit uses a hash-like storage system to store time-varying sequences chains. See Figure 32 below for a schematic of both circuits.

The Core circuit extends from the topographic peripheral sensory inputs of the thalamic core cells to topographically-organized layer IV and then apically to layers II & III and to layers V & VI with reciprocal topographic feedback from layer VI to the originating core cells in thalamus.

In the Matrix circuit, the TRN overlays the thalamic nuclei and their matrix cells which project non-topographically to layer I after receiving their strongest driving projections from cortical layer V, but without the TRN intervening in the circuit.

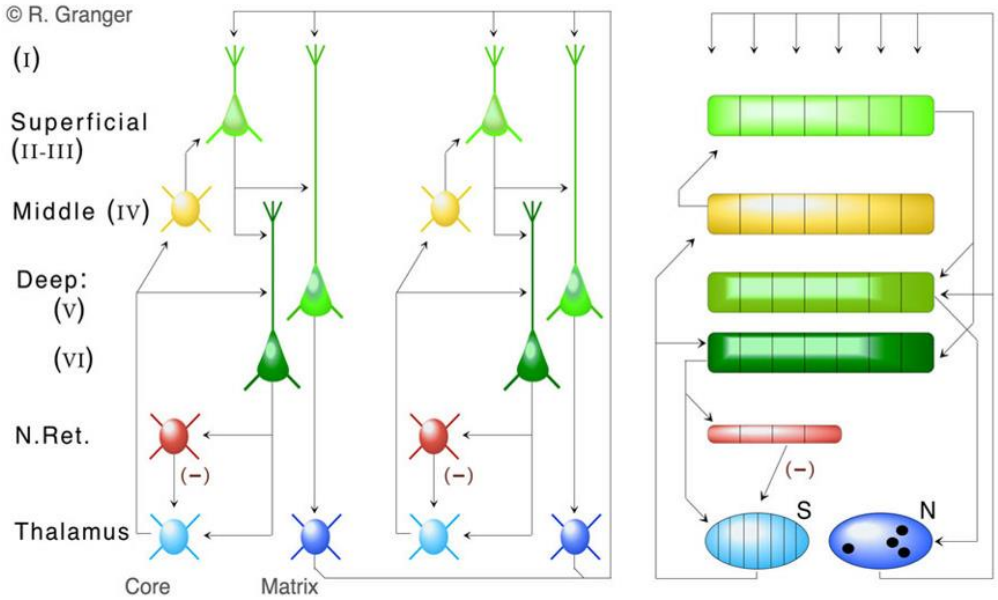

Figure 32: Granger model's Core and Matrix circuits. Neural circuit on left, matrix simulation model on right. (Rodriguez, Whitson, & Granger, 2004), reproduced with permission from MIT Press open access for academic use license.

4.3. A Sparse, Spatial, Temporal Code

The Granger model postulates that all cortex is involved in widespread synchronous activity where thalamo-cortical circuits encode information on the basis of sparse, fast-gamma neural activities that are punctuated by rhythmic excitation rising from the thalamus, targeting largely the longer-acting inhibitory cells of the cortex. Granger also theorizes that the pattern of changing excitation from the lower layers of cortex and the thalamus contains information relevant to the categories learned by the cells of the higher cortical layers.

The Rodriguez et al paper (Rodriguez, Whitson, & Granger, 2004) further offers that there may be a flow to the information sent. The first volley of activity, in a rhythmic bursting mode, contains information about the stimulus, making initial identification easier. The activity that follows, sent in tonic mode, is thought to contain further information about the stimulus. This progressive flow of information is the core of the hierarchical categorization. The major claim of the model is that the same circuity implements a process over time of information transfer, categorizing the stimuli into a series of hierarchical percepts.

4.4. Time Resolution

In the Granger model, the sparse, spatial code that the cortex uses operates at the maximum speed of the excitatory-post-synaptic-potentials (EPSP's), which is in the range of 10 to 15 milliseconds, limiting the precision of thalamo-cortical computation to this resolution. Granger offers that it is likely that the sums and differences between the various circuits spiking at gamma range in the lower cortex are integrated and compared by the circuits, and that possibly alpha activity may also contain information. For our implementation of the Granger model, the simulation runs with a granularity of no less than 10 ms, the shortest time constant we have employed.

4.5. Memory Storage

In the Rodriguez et al. paper, the authors suggest that, in the superficial (Layers II & III) and deep layers of cortex (layers 5 & 6), NMDA-based long-term potentiation (LTP) is the mechanism that stores the memory resulting from the hierarchical categorization process. They suggest that there are likely other processes at work, but the one that has the time course, the capacity, and the duration to support the functions they describe is NMDA-mediated LTP. In layer IV of cortex and in the thalamus, they point to the overwhelming volume of glutamatergic synapses and the presence of NMDA-limiting gene expressions to conclude that the glutamatergic synapses are the focus of learning that occurs in those areas. The Granger model does not address the possibility that the resonance in the thalamo-cortical columns bound to frontal cortex may be a form of short term memory as well, as is theorized in some of the models discussed previously.

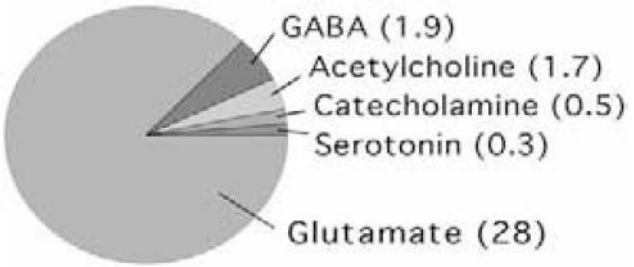

Figure 33: Relative percentage pie-chart of the neurotransmitters in layer IV cortex and thalamic nuclei. Numbers are pico-moles per microgram, (Rodriguez, Whitson, & Granger, 2004) reproduced with permission from MIT Press open access for academic use license.

4.6. Simplified Circuit Wiring

The Granger model makes a series of nested feedback loops as shown in Figure 32. To simulate the circuit, Rodriguez et al. made a series of simplifying assumptions about the wiring of these loops, while attempting to preserve the sequential, hierarchical categorization behaviour of them. The most significant set of assumptions is the complete preservation of topographic ordering through the whole of the Core circuit. This is a large simplification over the biology, but one necessary for model tractability. It is theorized that a major part of the axonal projections on these paths are in fact direct, as the model makes them to be, but not all.

4.7. Computing a Hierarchical Cluster in the Granger model

The cortical columns in layers II & III compete to learn stimulus patterns to represent. The strongest parts of the stimuli are responded to first. For any ongoing stimulus event, the strength of the incoming signal has been filtered by top-down expectations arising from previous activity. Therefore, the most salient aspects of the incoming stimuli are not only those that have the highest contrast (or some other property), they are the ones that are predicted from the past stimuli. Circuits representing stimuli that were predicted from the previous stimuli will have the higher activities and if the current stimuli match the prediction, those predictive circuits will be the most active, and most likely to be strengthened by Hebbian learning.

The largest, most useful categorical features will therefore be the ones sent through first by the thalamus to the cortex. This initial volley arrives from specific thalamus to layer IV and then layer II & III, which then activates the “winning” cortical columns in associated layer II & III pyramidal cells. These pyramidals then activate their corresponding layer VI and layer V pyramidals. The layer VI pyramidals topographically signal the specific circuits of the thalamus, for which the TRN inhibits the original incoming cells, causing a subtraction of the parts of the incoming signal that was recognized. While this is happening, the ‘winning’ layer II & III pyramidals activate their corresponding layer V pyramidals, which in turn activate their connected thalamus non-specific cells, and together they compute a predication of the sensory stimuli to next arrive. This prediction is then fed to higher order cortical columns via the nonspecific thalamus. See the left-hand side of Figure 32.

In Figure 34, the authors diagram the logical flow of categories over time as the input stimulus is classified. In their analogy they construct a hierarchical classifier for images of flowers where the number of petals, the colour of the petals (dark or light edges), the petal size and/or the petal orientation, determines the major and minor categories. The categories recognized are on the right. On the left the states of the cortical and thalamic layers are shown as the incoming stimulus is processed and changed. The stimulus is processed such that the combined activities over time represent the sequence of categories and then the sub-categories in the superficial layers of cortex. The deep layers learn the sequences of categorical representations (the activities of the superficial layer neurons as discussed above). The deep layers use these learned category sequences

(Granger's brain grammar) to predict the category sequences in future, thereby aiding future recognitions via top-down feedback.

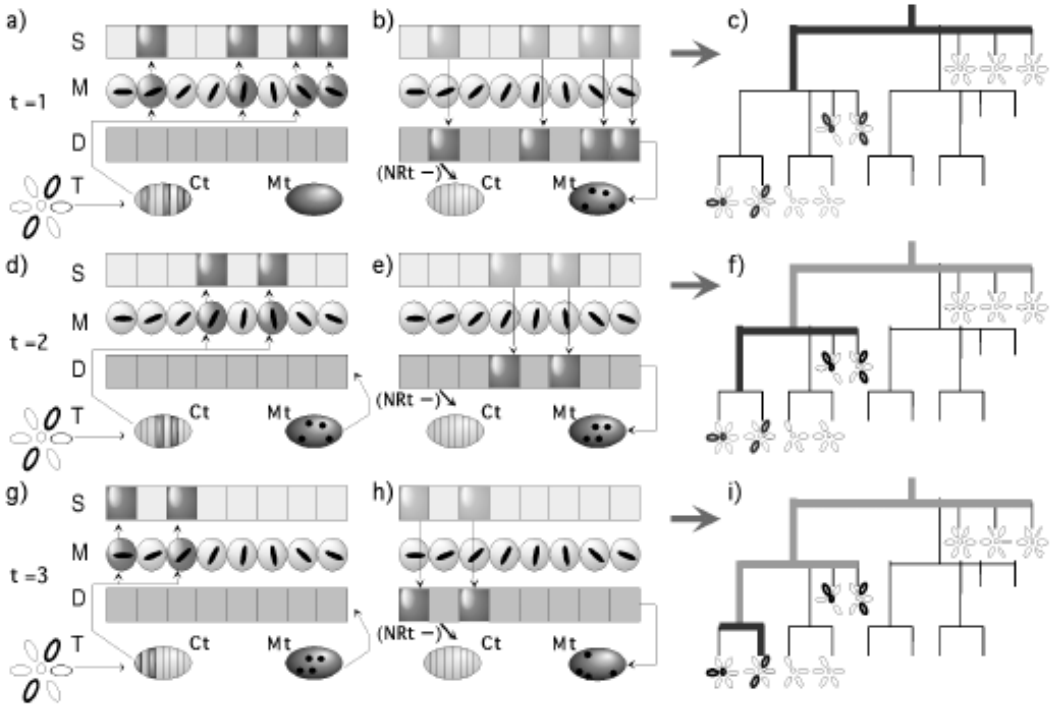

Figure 34: Diagram of the sequential functioning of the core (Ct) and matrix (Mt) thalamic circuits of the Granger model. Superficial cortex (S), middle cortex (M), deep cortex (D) and TRN (Nrt)). See model description in text for details. Time steps; t=1, t=2, t=3. (a) t=1, stimuli flows topographically from Ct to M to S; (b)S to D then topographic inhibition in Nrt to Ct, (c) response to this point is the same for stimuli in same cluster, (d) t=2 remaining core cells not inhibited win the WRT process and repeat process, (e) deep layers now learn the sequence of superficial layer responses, (f) second superficial layer response represents the sub-cluster, (g), (h) and (i) repeat the process of d,e, and f. From (Rodriguez, Whitson, & Granger, 2004), reproduced with permission from MIT Press open access for academic use license.

This topographic and temporal activity flow of columns being activated based on the most salient components of the inputs can be used to sub-categorize the sensory inflow, and are the neural correlates of the process of iterative, hierarchical categorization in the Granger model.

4.8. What does the feedback to Layers II & III of the originating column do?

One of the interesting high-level problems that remains is that while we might have a good theory for the feed-forward output of layer V and VI, we don't have a good theory of the functional value of the feedback to layers II & III from nonspecific thalamus related to the column that originated the sensory input.

One possible explanation comes from again considering the Granger, Rolls and Hawkins models. Granger theorizes layer V to contain the core storage of the progressive categorization sequences, Hawkins theorizes that layer V is the source of the predictions of the next stimuli to be received, which is similar to Rolls' theory that layer V contains a

competitive attractor map of the 'physics' of the things represented in cortex. The models are all quite similar at this level of abstraction.

The layer V projection to nonspecific thalamus, that then projects to higher-order cortex, is consistent with each of the theories with slight variations. The projection to higher-order cortex also feeds back to layers II & III and layer V of the originating columns. The feed-forward aspect is likely activating the next columns in the prediction, and deactivating those that are not consistent with the sensory inputs received. The feedback to the originating columns could be a signal reinforcing the accuracy of the prediction. Or, if the computation in layer V and non-specific thalamus were to yield only predictions that failed to be sustained (given that as Rolls theorizes these layers are maps of attractors) then we must see the activity for each of these scenarios as happening not just in the Granger model's feed-forward and feed-back flows, but also through a series of lateral winner-take-all competitions within each completive attractor map layer. This process would be happening both in the attractor map that Roll's theorizes in layers II and III and the one operating in layers V and VI.

This may be the purpose to the feedback from nonspecific thalamus to the originating column's layers II & III. The feedback would be present to continue to activate the column's inputs, which we know tapers off after the initial bursts, as discussed previously. So, if the prediction was activated and then it was not suppressed in the attractor map in layer V, the original column continues to receive activity to sustain its participation in the ongoing categorization and percept. If not, then the column's activity falls and another column, whose prediction better explains the sensory inflow, is activated and participates in the creation of the new categorization and percept.

4.9. Core Loop Algorithm Implementation

The algorithm to compute the model using matrices is shown for both circuits here. For the Core circuit computes a series of categorizations of its input and over time trains the weights in simulated layer I, II, III to build an iterative hierarchical set of feature detectors therein for the categorization of the input.

i. The maximum dot-product of the input activity vector (X) and the layer I synaptic weight matrix (W) determines the winning layer II/III cortical column set C for the current X. Multiple cortical columns may win if the they all have the same dot-product with X.
ii. The set C is the category for this iteration of the hierarchical categorization of this instance of X.
iii. The winning weight vectors all have a portion of the difference between their current state and the input vector X added to them to move them closer to X, the rate they move is called the learning rate (k).

iv. The portion of the input now 'explained away' is the new adjusted W after learning. The mean of this categorized portion of X is now subtracted from X, leaving only the 'uncategorized' portion of X.

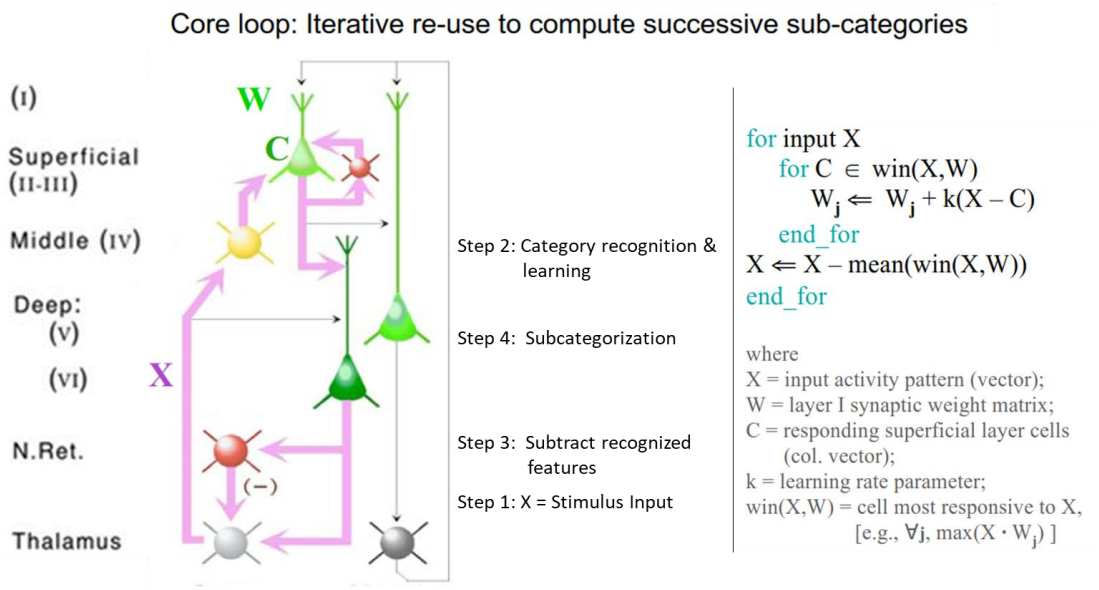

Figure 35: Granger model Core loop. Adapted from (Rodriguez, Whitson, & Granger, 2004) reproduced with permission from MIT Press open access for academic use license.

v. The now uncategorized X is fed back into the process to be 'sub-categorized'. At each iteration of the outer loop a new sub-category is generated in the changing values of C.

vi. Once X is depleted by repeated subtraction, there is no more input to process from this X and another input arrives or the circuit goes quiet.

4.10. Matrix Loop Algorithm Implementation

The matrix loop learns the sequences of categories generated by the core loop combined with other thalamic inputs and can reinstate them from partial stimuli information. It is a high capacity sequence storage and recall system (Rodriguez, Whitson, & Granger, 2004). Recall from our discussion about matrix cells in the thalamus that they are thought to sparsify and orthogonalize their inputs, so a matrix loops' responsibility in the Granger model is to learn this sparse, orthogonal code. The Granger model proposes that this is done via a hierarchically deepened hash code. This loop is different from the core loop in that it produces very specific responses for similar outputs, it is trying to store a very accurate code of sequences of categories, including learning their transition dynamics. This is a different and larger representational space than that learned by the core circuit, but one that requires more precise storage and recall.

The Granger matrix circuit performs its function in three basic steps across the connected thalamo-cortical circuits (note the circuit diagram only shows one instance of the circuit) to generate this hash code, as follows:

i. Feedforward input is sent from the layer II / III cells to layer V.
ii. A hash is performed (using layer V and the matrix nuclei of the thalamus) of the feedforward inputs combined with the result of the last time-step's hash function output.
iii. The output of the matrix thalamus cells is fed back to layers I, II, III and V via the thalamo-cortical fibers.

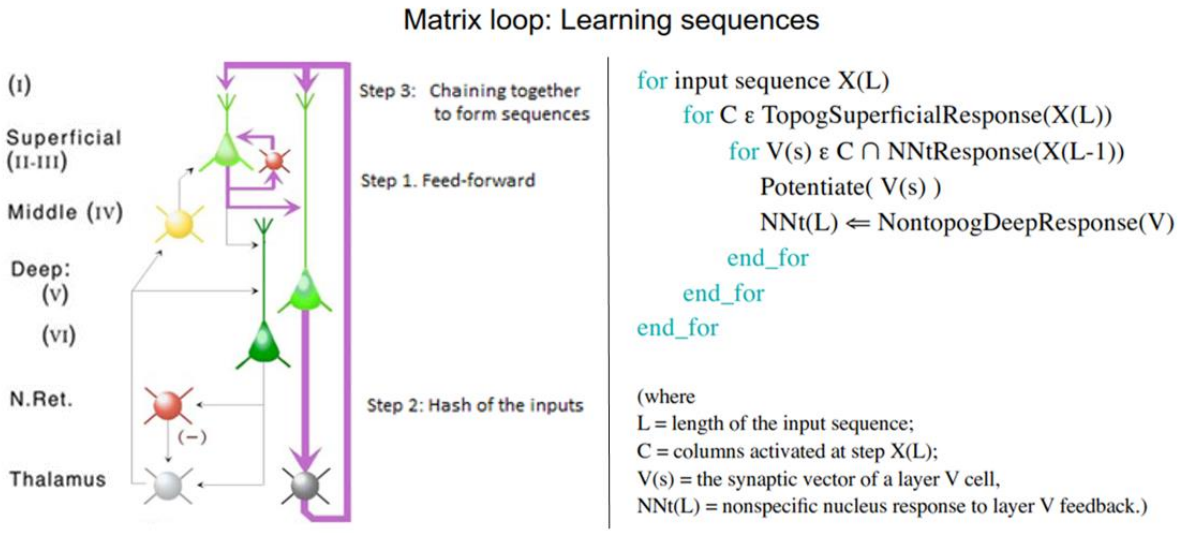

Figure 36: Granger model Matrix Loop. Adapted from (Rodriguez, Whitson, & Granger, 2004) reproduced with permission from MIT Press open access for academic use license.

To perform the actual hash, the Granger model uses a space efficient hash algorithm similar to the Bloom filter commonly used in spelling correction systems, where a series of hash function outputs map onto a fast lookup table. In our view, the Granger model's Matrix circuit implementation of this algorithm is likely not very biologically accurate given that the reality of a spatio-temporal neural implementation is much more likely to be implemented as a neural network with dynamically learned associations. However, the overall contribution of illuminating the higher level biologically plausible steps in the thalamo-cortical system for learning real-time sequences are more the point of the work, and this matrix algorithm remains an area of further study. We do not model the matrix loop, only the core loop in this work.

The inter-operation of the core and matrix loops in the Granger model produces a data-store within cortex and thalamus of sequences of stimulus categories. For more details on this circuit and the physiological basis for it, beyond the summary below see (Rodriguez, Whitson, & Granger, 2004). Our implementation focuses only on the first part of the Granger model, the Core Circuit.

5. The Neural Engineering Framework (NEF)

5.1. Representational Space versus Neural Activity Space

How do stimuli (such as light levels, colour frequencies of light, sound frequencies, pressure levels, etc …) become translated from the outside world into electrical and chemical states in brain so they can be computed with? Brains use sensors that respond to the properties of the world. Neurons that monitor these sensors (or monitor sensory-neurons) generate spikes at various rates and times to encode these values from the world within their networks.

The NEF is a set of mathematical methods for translating real-world values to spiking activities of the neurons representing them. The NEF allows us to build complex circuits that simulate these processes in the brain. The NEF was designed to enable neural models that are; biologically-plausible, large-scale, recurrent and realistic using spiking neuron models. The NEF is a synthesis of theories of neural coding, neural computation, dynamical systems and control theory (Eliasmith, 2007).

In the NEF, the space of values the neurons are representing in their spiking activities is called the representational space. The values in this space are human-readable: they are values in a stimulus space, such as the intensity of various frequencies of light for vision. In the NEF and its implementation in Nengo, representational space values are vectors of a user-specified length. In our implementation of the Granger model, the representational space values are the values of the intensity of each of the pixels for the images of the digits we are trying to categorize.

Neural activity space refers to the activities of the neurons that result once a representational space stimulus, a vector, is input to an NEF ensemble of neurons. In the NEF, these activities are represented as a vector of neural firing activities with one element per neuron in the ensemble, where each element is a 1 if the neuron has fired and a 0 otherwise. These are computed in the NEF at variable time scales, as specified by the time constants specified for the chosen neuron model.

5.2. Mapping Between NEF Spaces

The NEF maps representational space to activity space using decoders and encoders. Encoders are weight matrices that incoming representational space vectors are multiplied with to yield the corresponding neural inputs. Encoders are dimensioned by the number of neurons in the ensemble and the number of dimensions that are being represented by the neurons. In Figure 37, the basic mathematics of the NEF are shown from stimulus to neural ensemble responses. We will now walk through the framework in order of its major flow from encoding to decoding.

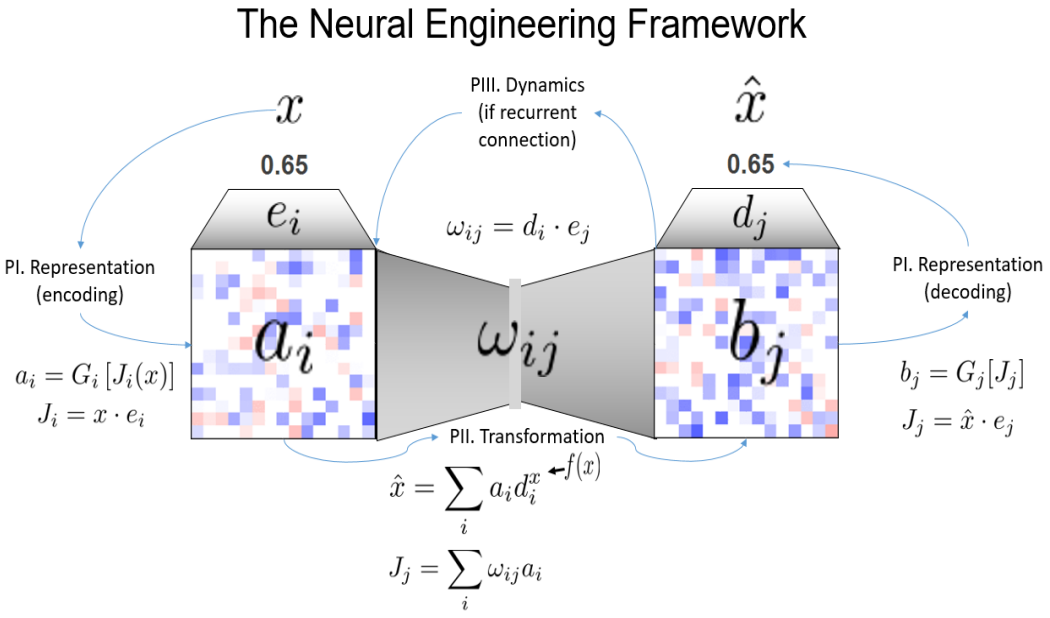

Figure 37: The mathematics of the Neural Engineering Framework (NEF). Adapted from (Eliasmith & Anderson, Neural engineering: Computation, Representation, and Dynamics in Neurobiological Systems., 2003).

The NEF proposes three basic principles of biological neural computation: representation, transformation and dynamics.

5.3. Representation

Referring to the representation step (PI) in Figure 37, representation of stimuli and internal states in the activity of neurons is accomplished in the NEF by mathematically mapping from representational or stimulus space to neural space and back again. The dot product of the stimulus vector with the given neuron's preferred direction vector (x) is fed through the neural non-linearity. The neural non-linearity is a property of the chosen neuron model. In the NEF this nonlinearity is referred to as the G(x) function. The default neural model used in the NEF is the LIF model as shown in Figure 39, although many spiking neuron models can be used.

The LIF model produces the canonical spike. If the sum of the currents arriving at the neuron's inputs (the equivalent of its dendrites) crosses the neuron's set spike threshold, a spike is generated. This causes it to respond increasingly as the input current increases. The way in which the LIF neuron responds when it is stimulated with a constant current is by constant spiking. First the LIF neuron integrates the current, then rests during its refractory period and then it spikes again. This results in spiking at a rate determined by its' response curve function's properties. The NEF represents neural spiking as a series of delta functions at the spike times generated by the chosen neuron model.

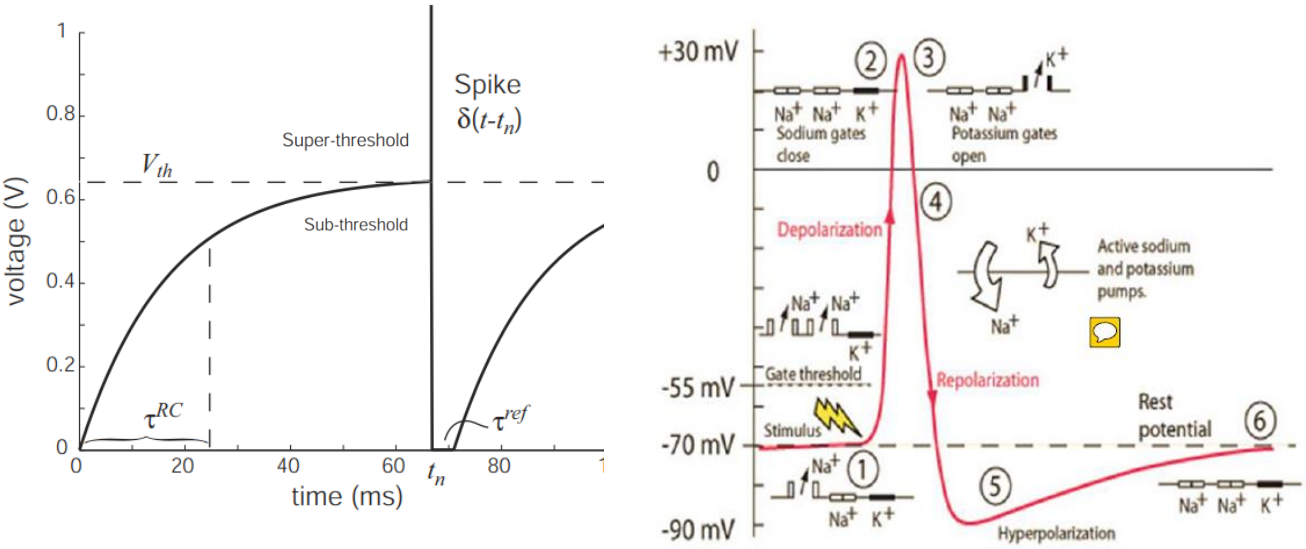

Figure 39 : (Left) Standard LIF neuron model from (Eliasmith & Anderson, 2003). The neuron model is the non-linearity that the NEF G(x) function models. When the input current exceeds V_{th} the neuron spikes. After a refractory period, it begins to integrate its inputs again. (Right) Biological spiking neuron model (Charand, 2005). The LIF model does not incorporate 2,3,4 or 5.

The preferred direction vector (PDV) or encoder of a neuron is the point in representational space that the neuron responds maximally to. A neuron's response curve is driven by the dot product between the input and its PDV, which is called its tuning curve. In the NEF when a group of neurons are created that together represent values in

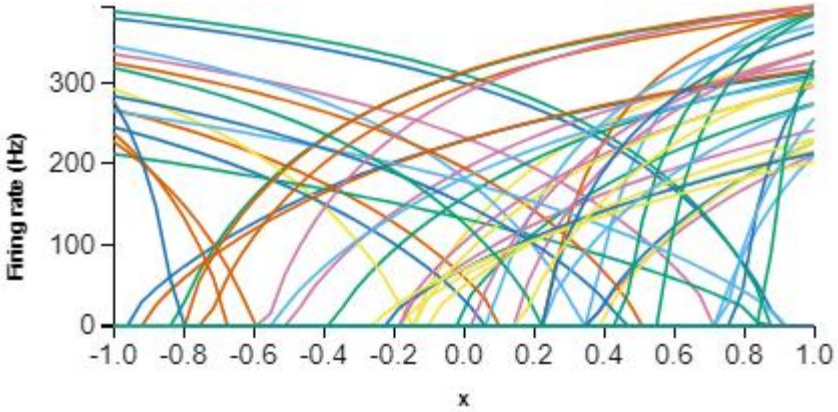

Figure 38: Randomly chosen tuning curves for an ensemble of neurons in the NEF. Generated with Nengo. x = the dot-product between the stimulus value in representational space and the neuron's PDV or encoder.

representational space they are collectively called an “ensemble”. In the NEF, the tuning curves for each of the neurons in the ensemble are chosen randomly, as shown in Figure 38.

5.3.5. NEF Encoding

At any point in time, the vector of activities of all the neurons in an ensemble in response to a stimulus represent values in representational space. The process of going from a representational space stimulus to neural activities is referred to as encoding. By then taking weighted combinations of the activities of the ensemble’s neurons to a given stimulus the representational space value encoded can be recovered from the neural activities. This is much like a set of basis functions being used to represent a signal.

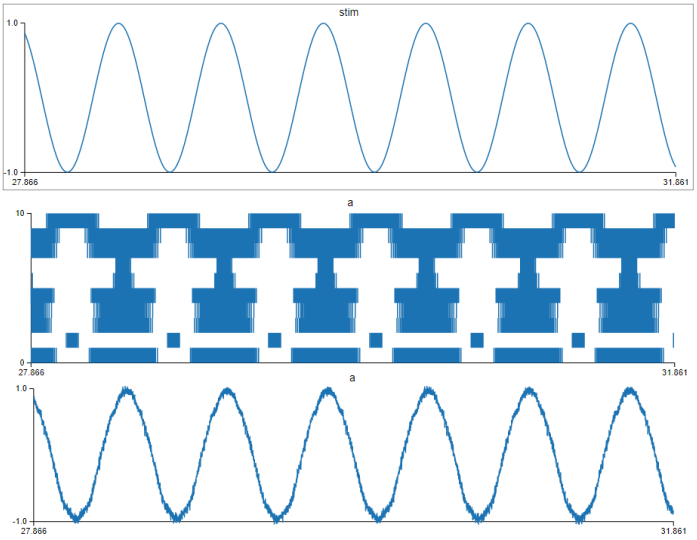

Figure 40: Stimulus sin(x) signal over time (top graph) and the firing of each of the 10 neurons (one row per neuron, middle graph) that are encoding the sin(x) signal as it changes these are the neural activities with represent sin(x) in the neural ensemble. The bottom graph shows the decoding of the neural activities to reconstruct the estimate of sin(x). Generated with Nengo.

In Figure 38 each line plot is one neuron’s tuning curve. At a specific x value (imagine a vertical line at x=0.2 intersecting each curve at some firing rate value) each neuron will fire at a given rate producing a vector of firing rates which collectively represent the value x, this is referred to as a nonlinear encoding x. Nonlinear because the tuning curves are not linear in x, and the spiking outputs are not linear in time. In the NEF, encoding of a stimulus into a neural spike train is formally defined as:

$$\sum_n a^x{}_i\left(t-t_n\right)=G_i\left[\alpha_i\left(e_i\bullet x(t_n)\right)+J^{bias}\right] \qquad (6)$$

where t_n, t = time ($t_n < t$), n = the index to the time interval (divided into n intervals), i = the index into the neuron population, $x(t_n)$ = x is the stimulus vector indexed by time t_n, e_i or $\tilde{\phi}$ = the encoders (the PDV’s) for neuron i, $(e_i \bullet x(t_n))$ = the dot-product between $\tilde{e}_i$

and $x(t')$ for neuron I, $\sum_n a^x{}_i(t-t_n)$ or δ = the spiking activity of the neuron i from time t_n to time t (divided into n time slices), in response to stimulus x, $G_i[\ ]$ = the neural nonlinearity function, α_i = a scalar constant that converts units in of the stimulus x to neural current units, J^{bias} = a constant input current that models the background firing rate found in most neurons.

5.3.6. NEF Decoding

To reverse, or decode, this representation in spikes back to the domain of the input signal, x, we need to multiply the firing rates vector (aka the activities vector) by an appropriately chosen matrix that will reverse the encoding process. In the NEF the decoders, *d*, is computed by solving for the vector that gives the best least-squares approximation of the stimulus signal x.

$$\hat{x}(t_n) = \sum_{in} a^x{}_i(t-t_n) * d_i(t) \tag{7}$$

where the symbols have the same meanings as in NEF encoding equation above, except the following: $\hat{x}(t_n)$ = the estimate of $x(t_n)$ reconstructed from the neural activities, and d or ϕ = the decoders for the ensemble of neurons (derived as explained below).

The result of the computation of the encoders and decoders is a mapping of a stimulus in representational space to neural activities in a reversible way.

5.4. Transformation

Referring now to Figure 37, the decoders are found by solving Equation 7 with the system of equations below for d using least-squares minimization over the error between f(x) and $\hat{x}$. For the ensemble to hold the representation of x, as opposed to a function of x, we set f(x) = x. If we wish to have the ensemble represent transformations of x, we substitute x with f(x), the desired function of x, and then solve for the decoders. The decoders will then implement the desired transformation when they are applied to the ensembles' activities at decoding. The equations for solving for transformational decoders in the NEF (Eliasmith & Anderson, 2003) are:

$$d^{f(x)} = \Gamma^{-1}\Upsilon \tag{8}$$

$$\Gamma_{ij} = \sum_x a^x{}_i a^x{}_j \tag{9}$$

$$\Upsilon_j = \sum_x a^x{}_j f(x) \tag{10}$$

In the NEF and Nengo, when communicating between ensembles using decoded connections (as opposed to a full connection weight matrix, which passes the currents directly), the decoder is applied to the activities of the source ensemble and then the

encoder is applied for the receiving ensemble. This has the effect of implementing the standard neural connection weight matrix across the connection. If needed, a full weight matrix that computes the same transformation between all pairs of neurons in the two ensembles can be computed from the decoders and encoders by multiplying the decoders from the source ensemble by the encoders from the receiving ensemble as follows:

$$w_{ij} = d_i e_j \tag{11}$$

For example, the identity function encoding and then decoding shown in Figure 40 is implemented in Nengo in this way. Nengo contains the NEF algorithms such that the decoders are automatically computed that represent the identity function in the ensemble of neurons. When the population is fed an input signal, sin(t), it generates in its decoded output, an estimate of sin(t).

5.5. Dynamics

Solving for a decoder that implements the desired transformation function does not allow that transformation function itself to change as a function of simulation time, and so cannot implement dynamics like those found in neural systems (Eliasmith & Anderson, 2003). In order to implement dynamics in the NEF, modern control theory is used as the foundation of Principle III. Modern control theory describes how a dynamic output is generated by a system being driven by an input that changes over time.

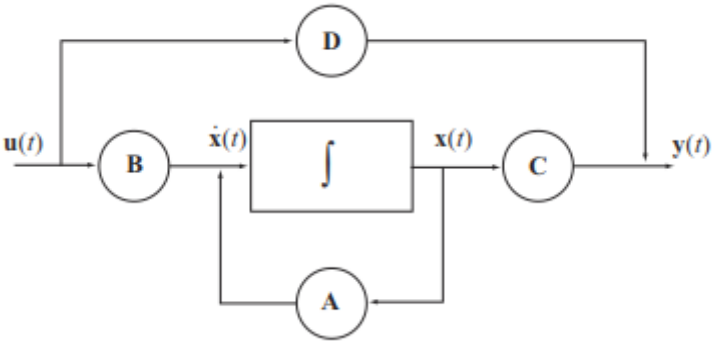

Figure 41 : Standard time-invariant, linear system block diagram (Eliasmith & Anderson, Neural engineering: Computation, Representation, and Dynamics in Neurobiological Systems., 2003), reproduced with permission from MIT Press.

The standard characterization of time-invariant, linear systems used in control theory is shown in Figure 41. The output of the system and the change in its input is described by the following system of differential equations (Eliasmith & Anderson, 2003):

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{12}$$

$$y(t) = Cx(t) + Du(t) \tag{13}$$

where u(t) = the input signal, A = the time invariant, recurrent (feedback) control matrix; B = the time invariant, input transform control matrix, C = the time invariant, output

transform control matrix, D = the time-invariant, input feed-forward transform control matrix, $\dot{x}(t)$ = the change in the input signal, and $y(t)$ = the output signal from the system after all transforms and feedback have been applied.

In neural systems, (Eliasmith & Anderson, 2003) point out that these control blocks are connected in series, modelling the layers of the neural networks, such that the system can be simplified by condensing D into the output transform, C, from the previous ensemble and by then combining the resulting C into the input transform of the following ensemble, leaving one input transform B' and one recurrent transform A' in the simplified version of the model shown in Figure 42.

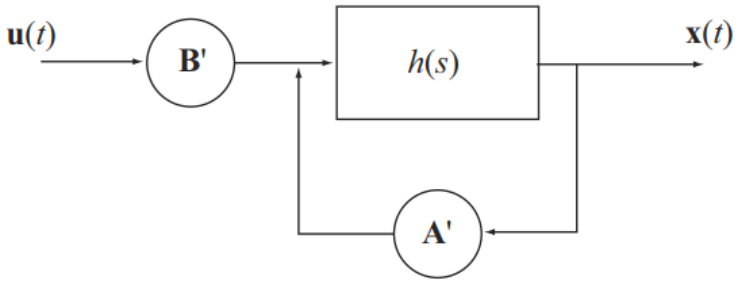

Figure 42 The NEF neural time invariant control system block diagram (Eliasmith & Anderson, 2003), reproduced with permission from MIT Press.

Below, in the simplified version of the model, A' and B' are matrices that allow the intrinsic dynamics of the neurons, captured by h(s), to be used to compute the same dynamics as a standard control system. The control equations simplify to Equation 14 shown below:

$$\dot{x}(t) = A'x(t) + B'u(t) \tag{14}$$

5.6. Integrated NEF Dynamical Neural Population Model

Taking into account dynamics, the NEF modifies the expression for computing the encoding of x into activities by incorporating the control model into the encoding. Here we substitute x(t) for the output of our simplified control model in Figure 42 resulting in the dynamics form of the NEF encoding equation (Eliasmith & Anderson, Neural engineering: Computation, Representation, and Dynamics in Neurobiological Systems., 2003):

$$\sum_n a^x{}_i\left(t - t_n\right) = G_i\left[\alpha_i\left(e_i \bullet h_i(t) * \left[A'x(t) + B'u(t)\right]\right) + J^{bias}\right] \tag{15}$$

where $h_i(t)$ = the synaptic dynamics function, $h(t) = \frac{1}{\tau}e^{-t/\tau}$ and τ is the neuron's synaptic time constant.

The decoders for this dynamic form of the NEF encoding are solved for in the usual way by minimizing the squared error between the originating signal and the estimate produced by our now revised encoding.

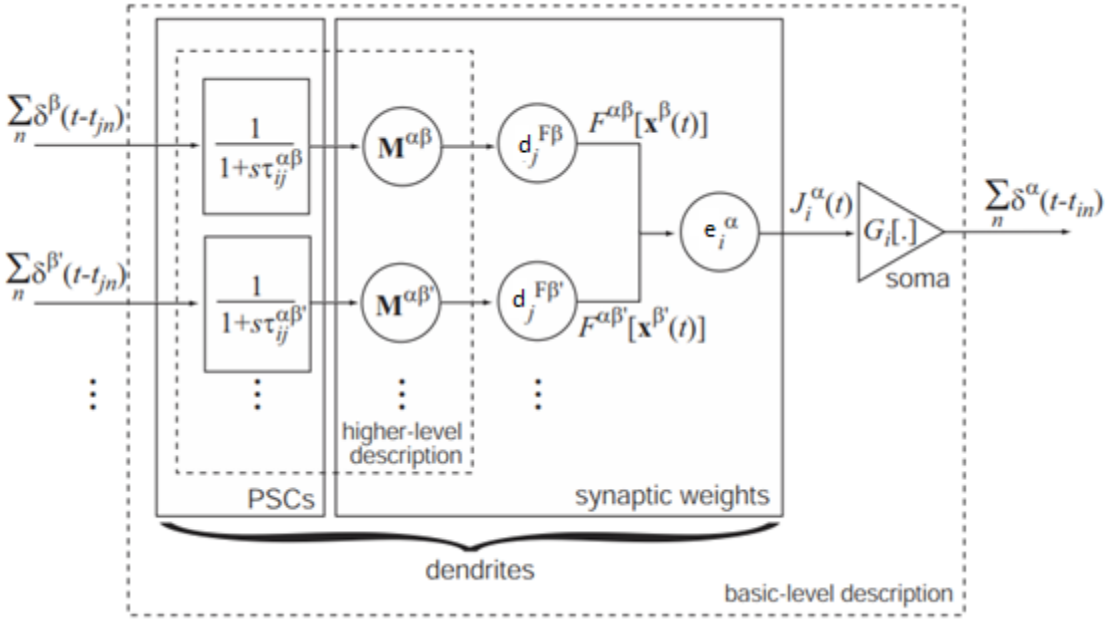

Figure 43: Neural Engineering overview (Eliasmith, 2013, p. 228). Note: symbols are different in the diagram above than in the text herein, see alternate symbols defined in tables for the NEF equations above for their meanings, reproduced with permission from MIT Press..

Figure 43 shows the generic neural populations model used in the NEF built from the equations above. Using standard control conventions per (Eliasmith & Anderson, 2003), boxes are transformation functions, circles are multiplications, and intersecting lines are summations. The input population here is β and the output population is α.

Spike trains arrive from the left from population β to a series of neurons in population α (the two rows of arrows starting from the left). The spike trains are transformed by the post-synaptic current, then the control matrices' combined effects are computed by the matrix M (both input and dynamic/feedback control, if any, are combined into M in this diagram), the resulting current is then decoded by the decoders, d, which were computed to implement the transformation function F. The signal is then passed through the encoders, e, for the neuron α. All of this happens in the zone from the axons of β and the dendrites of α. The resulting encoded signal generates the somatic current, the *J*() function, which adds the J^{bias} current to it. The somatic current then drives the neuron's nonlinear response function G[] which generates the output spike trains from α. Note that the box labelled "synaptic weights" delineates the three components that the weights are composed of; the transformation matrix M, the encoders ϕ, the transformation function F and the encoders $\tilde{\phi}$.

Having covered the basic three NEF principles; representation, transformation and dynamics, we now give a high-level overview of their implementation in software to understand how we built the NEF version of the Granger model.

5.7. Nengo NEF Entities

Nengo (Bekolay, et al., 2013) is a neural simulation package that implements the NEF. Nengo implements the NEF components as a series of classes. The reference implementation of Nengo is built with Python, where these Nengo objects are Python classes. For a more complete list and description please see the Nengo documentation at http://nengo.ca/documentation. Note that there are many other NEF objects in Nengo, for a comprehensive treatment of how to develop neural models with Nengo please see the website (www.nengo.ca) and Dr. Eliasmith's books (Eliasmith & Anderson, 2003; Eliasmith, et al., 2013). The Python class prototypes are shown below to illustrate the syntax as an introduction only. For a quick primer on the parameters in these class headers see:

https://pythonhosted.org/nengo/frontend_api.html#nengo-objects .

The Nengo code in Figure 44 shows simple example of using a Nengo node, an ensemble and a connection to illustrate NEF encoding and decoding of the identity function, with sin(t) as the input. For examples of using NEF ensemble arrays and all the other Nengo objects please see https://pythonhosted.org/nengo/examples.html .

```
2
3  import nengo
4  import math
5
6  model = nengo.Network()
7  with model:
8
9      def sx(t):
10         return math.sin(t*10)
11
12     stim = nengo.Node(output=sx, label="input")
13
14     a = nengo.Ensemble(n_neurons=50, dimensions=1, label="Decoded Representation")
15
16     nengo.Connection(stim, a, label="Input->Neurons")
17
```

Figure 44 : Nengo code for the sin(t) encode, decode example in Figure 40.

5.7.1. NEF Nodes:

Class nengo.Node (output=Default, size_in=Default, size_out=Default, label=Default, seed=Default)

Nodes are sources of inputs or destinations of output in a model. They can take a function and take its input to be then communicated across a connection from the node to an NEF

ensemble or NEF ensemble array (defined below). They can also be used as an output, capturing simulation results.

5.7.2. NEF Neurons:

Neurons are software implementations of various artificial neural models used as the basic unit of neural circuit simulation in the NEF. The default model used is an implementation of the leaky integrate-and-fire (LIF) model. Many other types can be used as detailed in the Nengo documentation. NEF LIF Neurons respond to a subset of a particular range of stimuli values (e.g.: -1 to 1, or -10 to 10, 0.4 to 1….).

5.7.3. NEF Ensembles of neurons:

Class nengo.Ensemble (n_neurons, dimensions, radius=Default,
encoders=Default, intercepts=Default,
max_rates=Default, eval_points=Default,
n_eval_points=Default, neuron_type=Default,
gain=Default, bias=Default, noise=Default,
normalize_encoders=Default, label=Default,
seed=Default)

Ensembles are collections of NEF neurons of some designated quantity (e.g.: 100, 200, 10,000…). An ensemble of neurons can represent a vector of values (in representational space – see definition herein). For a given representational error level, the number of neurons needed to represent representational spaces with large ranges of values and/or large dimensional vectors increases.

5.7.4. NEF Ensemble Arrays

Class_nengo.networks.EnsembleArray (n_neurons, n_ensembles,
ens_dimensions=1, neuron_nodes=False, label=None,
*seed=None, add_to_container=None, **ens_kwargs)*

NEF ensemble arrays are arrays of ensembles. As an example, they are used when large numbers of ensembles are needed where each ensemble is used to represent separately the elements of the array. The difference between representing a vector in an ensemble array versus using one multi-dimensional ensemble is that in while the array keeps the dimensions separate, one per ensemble (as one way to implement it), the single multi-dimensional ensemble's encoders and decoders capture the interactions among the array elements. When transformations are used from connections in the single multi-dimensional ensemble implementation, these interactions can affect how the transformations are applied versus the purely separate dimensions of the ensemble array version. Depending on the desired results, this can be everything from essential to undesirable. In the case of learning connections between ensembles, the decoders learned from multi-dimensional ensemble inputs across a learning connection (see

below) will capture the interactions between the dimensions being represented, whereas using array of ensembles, without special additional efforts, will not.

5.7.5. NEF Connections

Class.nengo.Connection (pre, post, synapse=Default, function=Default, transform=Default, solver=Default, learning_rule_type=Default, eval_points=Default, scale_eval_points=Default, label=Default, seed=Default, modulatory=Unconfigurable)

Connections are information links from a source ensemble, ensemble array or a node to a receiving ensemble, ensemble array or node. They run between these objects communicating information from one entity to another. There are two kinds, decoded connections and neuron-type connections. Decoded connections apply the decoders before passing on the information, whereas the neuron-type connections pass individual neural activity directly from the source entity to the receiving one. Connections can have transformation matrices specified when they are constructed, in which case the transformation matrix is used to solve for the decoders needed to implement them on the source ensemble.

5.7.6. NEF Learning Connections

Class nengo.connection.LearningRule (connection, learning_rule_type)

Learning can also operate across these same transformation connections to minimize some error signal. There are two types of NEF learning rules we use in our model. Voja and hPES learning. The symbols have the following meanings in the learning formulas below.

Symbol	***Meaning***
i	Input neuron.
j	Output neuron.
k	Summation counter, cycles across all input neurons to neuron i.
x	State vector encoded by the neurons.
Δw_{ij}	The change in the weights from input neuron i to output neuron j.
η	Learning rate.
a_i	Neural activity of neuron i.
e	NEF neural encoders.

n	Time period.
E	Error signal. The difference between the supervised learning example vector (the target representation) and the current vector the neural activities are encoding.
α_j	The scaling factor for neuron j. A constant tuning gain specific to each neuron.
$x \bullet y$	The dot-product between the vectors x and y.
$E[a_j / c]$	The expected value of the neural activity (a_j) of neuron j over all possible input patterns, divided by a chosen scaling constant c.

1) ***Voja Learning:*** Donald O. Hebb's theory of synaptic plasticity (Hebb, 1949) states, loosely, that neurons that fire together wire together. Oja's Rule is the single neuron case of Sanger's Rule. Voja is the vector case of Oja's Rule (Voelker, Crawford, & Eliasmith, 2014) which was implemented using the NEF in Nengo to learn encoders.

a) ***Hebb's Rule:*** Hebb's Rule sets the weight between an input neuron and an output neuron to be the product of their activities. This is the original form of unsupervised learning. A very simple formulation of the concept is shown below. The problem is that the weights can grow unbounded creating stability problems for a large network of neurons learning with such a rule, especially if the activity is recurrent. Hebb's learning rule equation is:

$$\Delta w_{ij} = \eta a_i a_j \tag{16}$$

b) ***Oja's Rule:*** Oja's Rule attempts to remedy this problem by normalizing the weights to the zero to one range. The weight update is now computed based on the product of the output from neuron i to neuron j, but neuron i's output has been reduced by the weighted output of neuron j. Oja's learning rule equation is (Oja, 1982):

$$\Delta w_{ij} = \eta a_j \left(a_i - a_j w_{ij} \right) \tag{17}$$

c) ***Voja's Rule:*** The NEF implementation of Voja's Rule uses activities to drive the change in weights. In the NEF, it is encoders we update to change the weights between neural ensembles. The encoder update rule in vector form is shown below. Here we change the encoders by the product of the activities of the input neuron i and the difference between the output neuron's vector and the existing encoders. For the full derivation of Voja's Rule and the NEF implementation see (Voelker, Crawford, & Eliasmith, 2014). The NEF Voja learning rule equation is (Applied Brain Research Inc., 2017):

$$\Delta e_i = \eta a_i \left(x - e_i \right) \tag{18}$$

In all the cases above, the main idea is that when neurons are simultaneously active by virtue of both responding strongly to the current state of the network, their connection grows stronger. Neurons learning under any of these rules move their internal preferred direction vectors closer to their inputs.

2) ***hPES Learning:*** When neural systems are to be trained starting with examples of what is to be learned to guide the learned weights, error-driven or supervised learning can be used. In Nengo, one particular form of an error-driven learning rule is the Homeostatic Prescribed Error Sensitivity (hPES) (Bekolay, Kolbeck, & Eliasmith, 2013) learning rule. The hPES rule is a novel spiking learning rule derived in part from the Bienenstock, Cooper, Munro (BCM) (Bienenstock, Cooper, & Munro, 1982) learning rule. The hPES rule is shown below. Here the change in the weights is computed as the dot-product between the neuron's PDV, scaled by the neuron's gain and learning rate, and the error between the neuron's PDV and the supervised learning vector we are trying to minimize scaled by the neuron's activities. The hPES learning rule equation is (Eliasmith, 2013):

$$\Delta w_{ij} = \alpha_j a_i [\eta_1 e_j E + \eta_2 a_j (a_j - E[a_j / c])] \qquad (19)$$

hPES is a very powerful learning rule and the authors assert that it improves on previous methods of learning such as backpropagation, self-organizing maps and deep-learning by implementing separate training and inference phases, reduces the need for many layers with supervised and unsupervised learning and offers a spiking implementation. This is offered by hPES using synaptically local information, making it a more biologically plausible learning rule. The derivation of hPES is found in the authors paper (Bekolay, Kolbeck, & Eliasmith, 2013). For a concise explanation of the hPES rule see Appendix D.4 in (Eliasmith, et al., 2013).

For more information on neural learning in Nengo see the documentation on learning connections at https://pythonhosted.org/nengo/examples/learn_communication_channel.html.

5.7.7. NEF Networks: Product, Basal Ganglia & Thalamus Networks

In Nengo, a basic organizing model is the NEF Network. NEF Networks are derived from the Nengo.Networks base class. Networks are objects that wrapper networks of NEF objects such as ensembles, nodes, and connections. They make reuse and functional separation easier, and can be used in analogous ways to biological neural nuclei or layers as a higher order organizing principle in large neural systems.

We make use of several of Nengo's pre-built networks in our implementation of the Granger model, such as the Product, Basal Ganglia, and Thalamus networks. The Product network computes the dot-product between its inputs. The Basal Ganglia network takes in a number of dimensions on its input connection (often composed of a dimension from each input NEF object it has been connected to) and its output connection has the same number of dimensions with only the one corresponding to the most active (largest valued) input line being approximately 1, while the other dimensions are zero (equal to no neural activity). The Thalamus is an analogy for the biological anterior thalamus' function of cleaning up the action selected by the output of the basal ganglia, by enabling the corresponding thalamo-cortical circuits, thereby inverting the basal ganglia's inhibitory output to an excitatory output, and cleaning it up by thresholding its output.

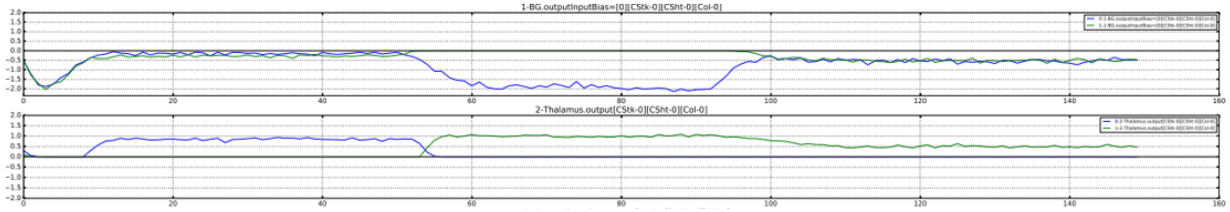

Figure 45: Signals from the Basal Ganglia and Thalamus in the NEG Granger model

In Nengo, the Thalamus network, forces the one 'winning' dimension from the BasalGanglia class to a strong 1, or high activity, while suppressing the other dimensions' outputs to zero, or no activity. Figure 45 shows the output of the basal ganglia circuit, which in the Granger model is fed the dot-products of each of the feature detectors across its dimensions, outputting a series of signals with the strongest signal (the highest valued one) representing the "winner" (most active) while the other signals are reduced to values below zero. The thalamus then takes this signal from the basal ganglia and outputs the "winner" with a value closest to one and the other signals are effectively suppressed to no neural activity, which shows as a value of zero. These two networks are used to implement a winner-take-all effect in the Granger model, simulating the effect of cortical inhibition and excitation, and the attractor dynamics of the layer I, II and III, of cortex.

For more information on the pre-built networks Nengo offers see: https://pythonhosted.org/nengo/networks.html.

5.7.8. Collecting Experimental Data: NEF Probes

class nengo.Probe (target, attr=None, sample_every=Default, synapse=Default, solver=Default, label=Default, seed=Default)

The NEF probe object can be used to monitor the value of almost all elements within a Nengo simulation such as spike data, represented values, and neuron voltages. The data is recorded within the probe class and then can be read out after the simulation is over. In our model, we graph the results and create PDF files of them, as well as store the results from the probes into CSV files for later analysis.

We have now seen how the NEF model's neural populations at an activity, representational, learning, and computational level, implementing both non-dynamic and dynamical systems using various neural response models, all in a biologically plausible way. For examples of NEF models of biological systems see (Eliasmith & Anderson, 2003).

6. The NEF Granger Core circuit

So far, we have covered;

- The anatomical background of the cortex and the thalamus
- Many theories of the functioning of the thalamus and cortex and their interconnected circuits
- A review of categorization algorithms and implementations of them
- The details of the Granger model of the categorization function of the thalamo-cortical circuits
- The NEF methods for modelling neural populations
- The Nengo classes that implement the NEF

We are now, finally, in a position to turn our attention to the of building an NEF implementation of the Granger Core circuit. We set out to extend the Granger model's core loop implementation in a more biological plausible way by adding more biologically realistic features through implementing it as a Nengo recurrent spiking neural network. For this we developed an NEF model of the Core circuit. This is a novel implementation as far as we know. There has not been a spiking neural model of the Granger model yet developed. We did not implement the Matrix circuit due to time constraints, however we present a design for it.

6.1. NEF Implementation of the Granger Core Circuit Flow

For an overview of the flow of the circuit showing the major NEF objects and their relationship in the processing order, see Figure 46.

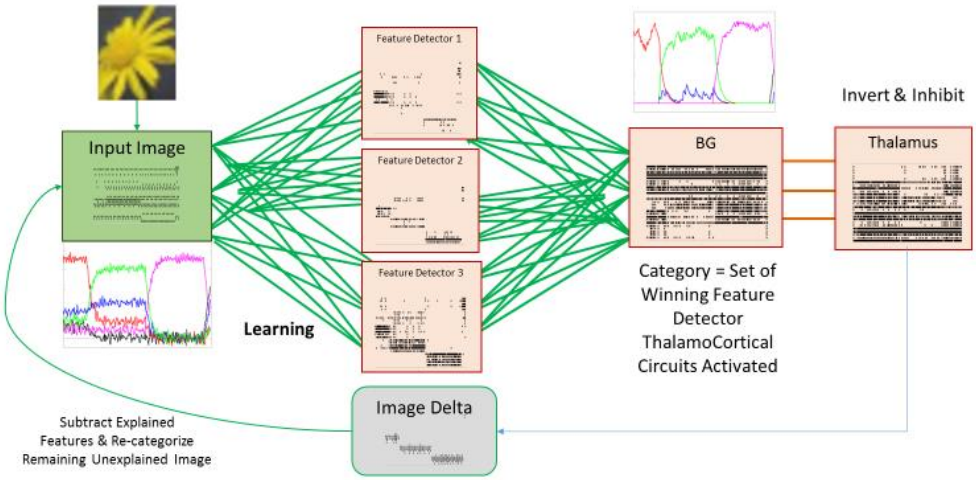

Figure 46: NEF Hierarchical Categorizer Core Circuit Conceptual Flow Orientation diagram. Green lines are neural connections, multi-coloured neural activity plots are shown next to the ensembles, dark neural spike-raster plots are shown within the ensemble boxes.

Before we attempt to describe the circuit flow in our model, we need to clarify the behaviour of recurrent circuits such as ours. While following the steps below, it is important to look carefully at the diagrams and take all of the step-wise breakdown of flow in these recurrent, dynamic neural models as a guide for how the circuit starts out. The activity will flow around the circuit such that the previous inputs influence activity that then merges with the next inputs, and changes both the next inputs as well as the activities in response to the next stimuli. When the circuit has many loops of varying latencies, this also means that later inputs affect the circuits' processing of previous inputs. A recurrent circuit will respond to the same input differently based on what happens before, and after that same input.

The major steps in our NEF implementation of the Granger model are as follows;

i. ***Stimulus:*** Images are read in and converted to vectors and input to the network.

ii. ***Response:*** The vectors are presented in series to the NEF implementation of the Granger circuit composed of a series of feature detectors (FDs), each having a learned feature pattern that it is uniquely specialized in responding to. These feature detectors are the analog for the cortical column cells in the brain and the categorizer circuit in the Granger model. The feature detectors react to the incoming stimulus by spiking: the closer the match, the higher the spiking rate.

iii. ***Competition & Categorization:*** The most active feature detector is the one whose learned pattern matches the input the best. This feature detector will be selected by a winner-take-all (WTA) process implemented in the cortex by lateral inhibition and attractor dynamics. This winning feature detector is the category or sub-category of the stimulus, depending on the level of recurrent processing we are currently at. (See below for the recurrence step that brings us back here eventually).

iv. ***Learning:*** The winning feature detector is allowed to learn, which moves its learned pattern closer to the inputs by a small amount controlled by a learning rate setting (see Appendix -**Simulation Parameters** for the learning rate settings).

v. ***Explained Stimulus:*** The portion of the input that was "recognized" by the winning feature detector plus the incremental learning shall be referred to as the "explained" stimulus. It is called that because it is now known to the Core circuit, given that there were neurons that coded for it, and their preferred direction vectors were updated to move closer to the current stimulus during the current categorization step.

vi. ***Inhibition:*** The feature detectors other than the one that won via the WTA process, are all inhibited from firing and learning so that losing feature detectors do not changes their PDVs.

vii. ***Subtraction:*** One of the Granger Core circuits' major mechanisms to implement sequential, hierarchical categorization is that the explained portion of the stimulus is subtracted by the inhibitory mechanisms in the reticular nuclei (explained in section 2.2.7 above), which were activated by the cortico-thalamic feedback originating from layer V of cortex.

viii. ***Sub-Categorize***: The feature detector that wins is the categorization result. So over time as the feature detectors compete, win, learn, then subtract the explained parts of the stimulus, and finally categorize the stimulus remaining, the un-explained components of the stimulus cause another feature detector to win. The entire sequence of winning feature detectors is the hierarchical categorization of the original stimulus. This process proceeds iteratively, until the inputs stimulus has been subtracted to nothing or there is another input which begins the process again. This is the major function of the Granger model, to produce a set of cascading, sequential categorizations of a stimulus by categorizing the stimulus, then subtracting what was identified and then re-categorizing the remaining stimulus.

ix. ***Sequential Inputs***: Note that the inputs appear in a stream. The sequential inputs are thus changing, and the recurrence in the circuit means that the feature detectors are always detecting in an environment of inputs that are part new stimuli, from the specific thalamic cells per our discussion above in section 2.3, and also part recurrent activity from the previous input. The next input could be the same image when the inputs are static, or a completely new image, or a moving image transiting across the visual field. Regardless, the recurrent circuits are always capturing some of this flow, even in the Granger Core circuit. The Granger Matrix circuit explicitly deals with the learning and recall of these sequences directly but as mentioned above that is outside the scope of our work.

6.1. MNIST Test Categorization Data

We used a modified version of the Lecun version of the MNIST data set (Lecun, Cortes, & Burges, 2017) of handwritten digits as our categorization dataset. We took the Lecun version, which is mass centered on a larger background than the original MNIST dataset.

Figure 47: Sample of MNIST digits (Meng, Appiah, Hunter, & Dickinson, 2011), reproduced with permission from the IEEE under open access for academic use license.

The MNIST dataset comprises 50,000 training images and 10,000 test images. Due to the processing time for such a large neural model as this, we only used small, randomly chosen sets of test images to run our model on. For reasons also of performance, the original MNIST images, which are twenty-eight (28) pixels square, are reduced to 10 pixels by 10 pixels in dimension to allow the NEF network to run in a reasonable time on the available PC with one GPU used for this work.

The full-size images can be run by increasing the number of neurons substantially (from 4,000 per image to upwards of 40,000 or more) used in representing the images and in the feature detectors, but the associated runtimes increase substantially so smaller images and less of them were used. The original goal of building up to processing the whole set of MNIST images at the original resolution was not possible given the time compute and research time available

Looking at the MNIST digits, it is apparent that there are many different features within each digit's images as well as between all digits. The MNIST images are fed into the categorization model and, by competitive interaction of the circuit (see following descriptions), like the other self-categorizing algorithms discussed in Chapter 3, the network determines the features that the data will be categorized on.

6.2. NEF Implementation Components of the Model

The overall NEF Core circuit that implements this flow is pictured in Figure 48. Starting on the left and working across the flow of the circuit, the nodes, neural ensembles, ensemble arrays and connections are as follows.

6.2.1. Image Input Node X

The reduced size MNIST images are taken and converted to a vector of either grayscale (values for each pixel from 0 for white to 1 for black), or pure black and white (either 0 or 1). The resulting vector is passed to a Nengo node object named ImageX of 100 (10 pixels x 10 pixels) dimensions as the input part of the network. This is analogous to raw sensory input. There is a major drawback of the Granger model, and this implementation here as well: the model attempts to understand the upper visual system's categorization operation, but uses raw direct image stimulus as input. It is a necessary but unfortunate simplification. For more discussion of this please see Section 6.3.

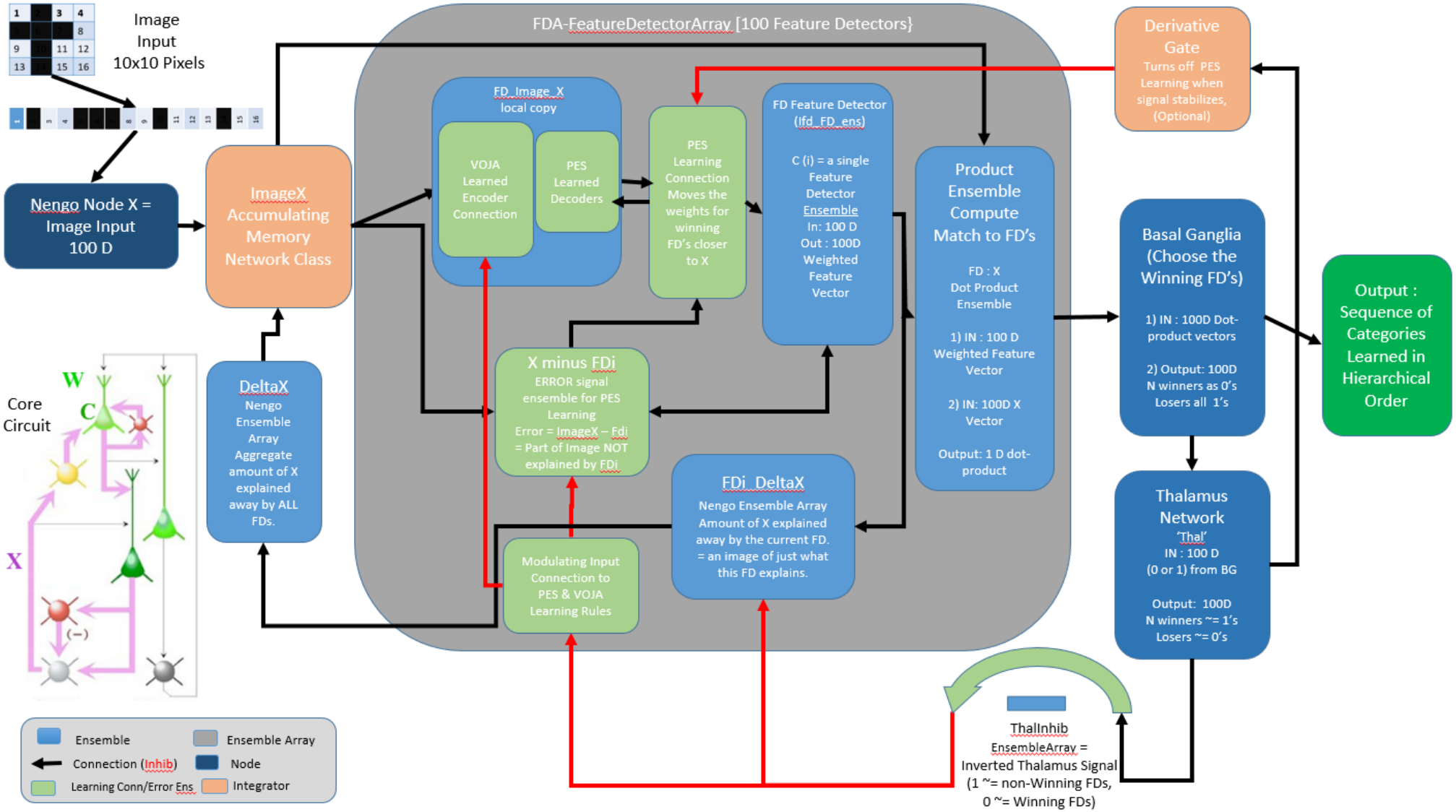

Figure 48: NEF model of the Granger Core circuit

6.2.2. The Choo Accumulating Memory Network (AMN)

We now need a way to represent the stimulus in neurons in way that can mimic the ability of the thalamic core cells to stimulate the cortex with the image but then, when the thalamo-cortical feedback comes back from layer V, remove the parts of the stimulus that are explained away by the cortex. For this we need to move the representation from an input vector into an NEF neural representation. For that we will use the Nengo network shown in Figure 49.

This network is a very slightly modified version of the one originally developed by Dr. Feng-Xuan Choo (Choo X. , 2016) specifically for this modelling work based on a small part of his extensive work in modelling working memory (Choo F.-X. , 2010a) and (Choo F.-X. , 2010b).

Refer to Figure 49 for the following discussion of the operation of the AMN network.

i. Inputs:

There are 3 inputs to the AMN;

a. ***The Thalamic Core Cell Stimulus:*** The original stimulus is constantly fed from the ImageX node into the output node of the AMN. The entire network will then create a subtraction signal which will be removed from this input to create the desired output of the AMN which is a progressive subtraction from the original image as the cortex explains away the input signal.

b. The Cortical Layer V Explained Stimulus: The input_s signal is the cortical layer V analogue, sending the output of the feature detectors from cortex back to the thalamic core cells.

c. A Reset Signal: An image reset signal is fed into the AMN to clear the integrators between images. This is required in our model due to our implementation, but it is unknown if there is a similar function in the biology, although some suggest that the burst mode (as discussed above) may be the signal of a new input and as such may cause a similar clearing effect.

ii. Input Change Detection:

The network needs to detect a change in the signal from the layer V cortical cells to know when to subtract the explained part of the stimulus. We simulate that here by taking the derivative of the incoming input_s signal and using that value to be the gate signal to the Accumulator sub-network of the AMN.

iii. Two Input Gated Memories:

There are two input-gated memories (IGMs) in the AMN. They are sequentially connected on the input side but recurrently connected on the gating signal. The dynamics of this sub-network can be summarized as follows:

a. The input from the layer V cortical analogue, input_s, flows to the second IGM where it is integrated in the first memory network of this IGM.
b. When the gating signal triggers, signalling a change in the input from the layer V cortical cells, it flows into the second memory component of the first IGM.
c. The second IGM network receives the output of the first IGM, essentially a copy of the explained stimulus signal from the cortical layer V cells.
d. This second IGM then accumulates these difference signals over time and feeds that difference signal to the output node to be subtracted from the original image signal, held in the thalamic core cells.

iv. Output:

This results in the output node holding at any time (the AMN's primary outputs) the as-yet-unexplained part of the original stimulus being fed into cortex all the time, thus simulating the output of the thalamic core cells to cortex layer 4. This entire circuit is used to model the effects of the thalamic core and RTN cells and their modulation by the cortical layer V cortico-thalamic feedback connections.

For more details on this type of architecture and working memory in general, please see Dr. Choo's work (Choo F.-X. , 2010a; Choo F.-X. , 2010b) .

v. Dynamics and Timing:

Figure 50 shows the graphs that result from Nengo probes on each of the input and output connections in a one-dimensional run of the AMN model. In our model, we will use the full 100 dimensions of our stimulus images. The AMN works with any number of dimensions. The upper graph shows the incoming signal rising and then being held constant. The explained part of the signal from layer V cortex feature detectors then arrives and once it stops changing the AMN then subtracts it from the original stimulus. Note the time delay on this due to the neural time constants. The output is then the result of this subtraction. The lower graph shows the operation of the change detection network, showing the change detection signal rising when the incoming cortical layer V signal changes and then falling back to no activity once it stops changing.

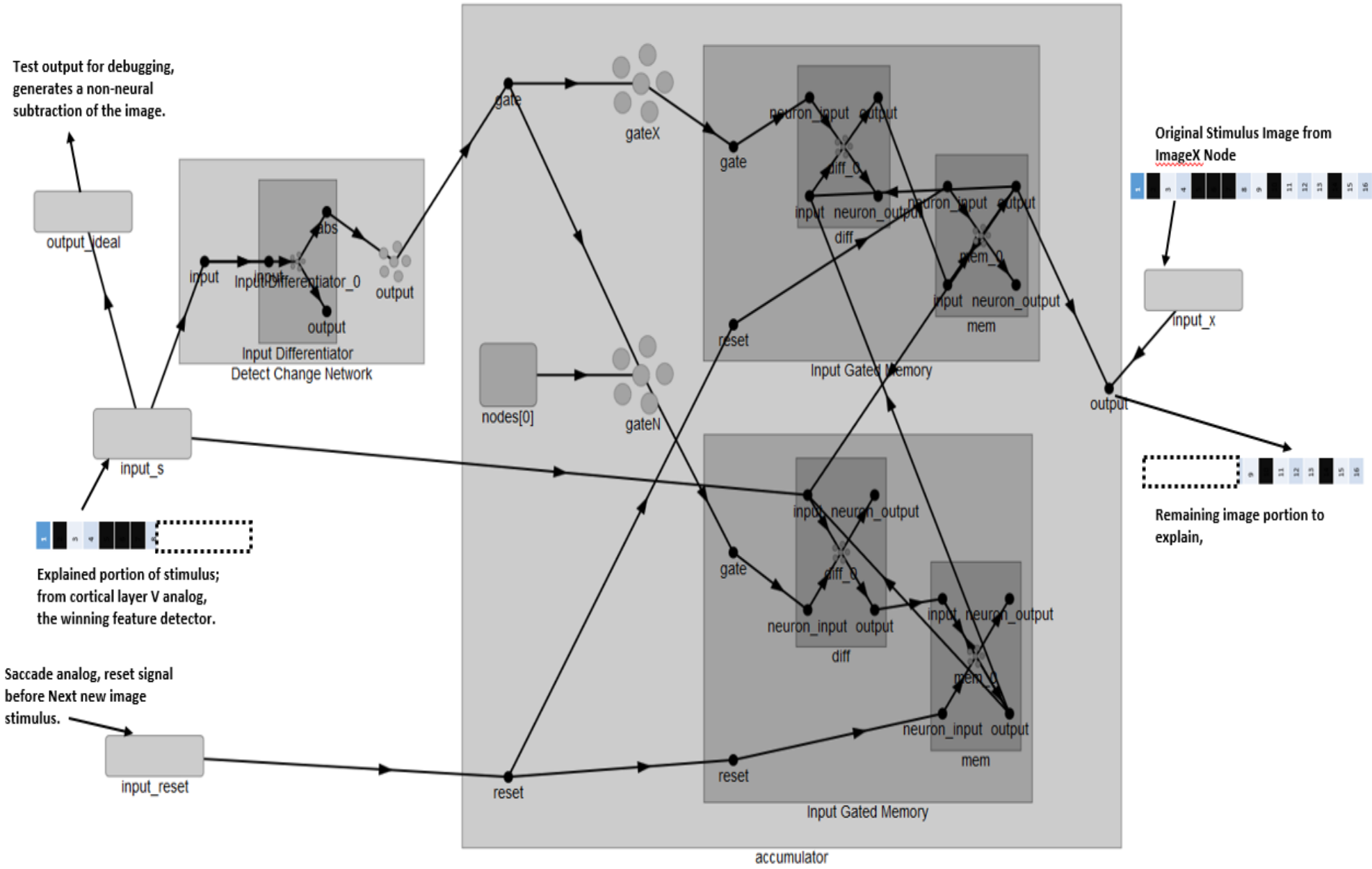

Figure 49: Nengo screenshot of Accumulating Memory circuit (Choo X. , 2016) adapted for use in simulating the subtraction of input by the RTN and thalamic core cells driven by layer V cortico-thalamic feedback.

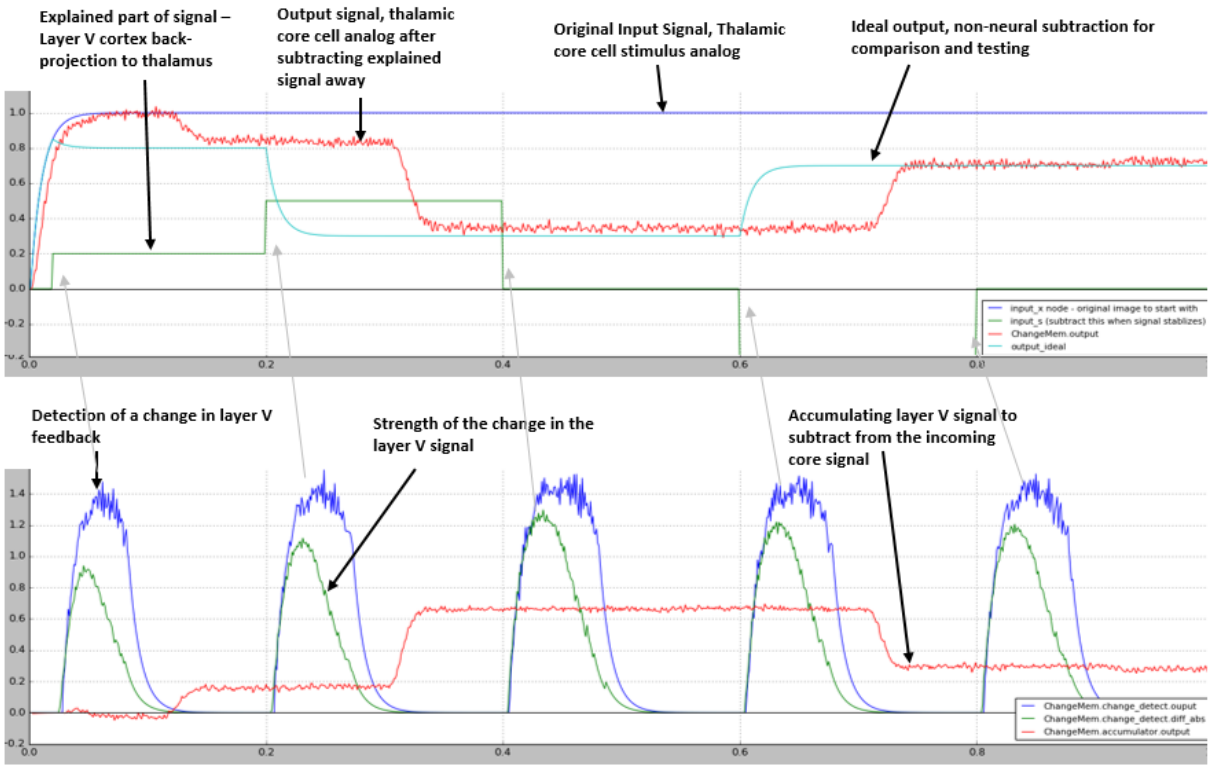

Figure 50: Dynamics of the incremental memory circuit.

6.2.3. CorticalSheet and CorticalColumn Classes

The CorticalSheet object holds a collection of objects and is described next. The sheet class is a container class for the cortical column objects. The cortical column class is a container class that is an analogue for a cortical macro-column, an instance of the set of layers of cortex that perform recognition in the Granger model on a specific feature set. This is an analogue for the upper layers attractor network theorized to be in layers I through IV. See Figure 51, which shows the nested, hierarchical relationship of the sheet, column and feature detector classes.

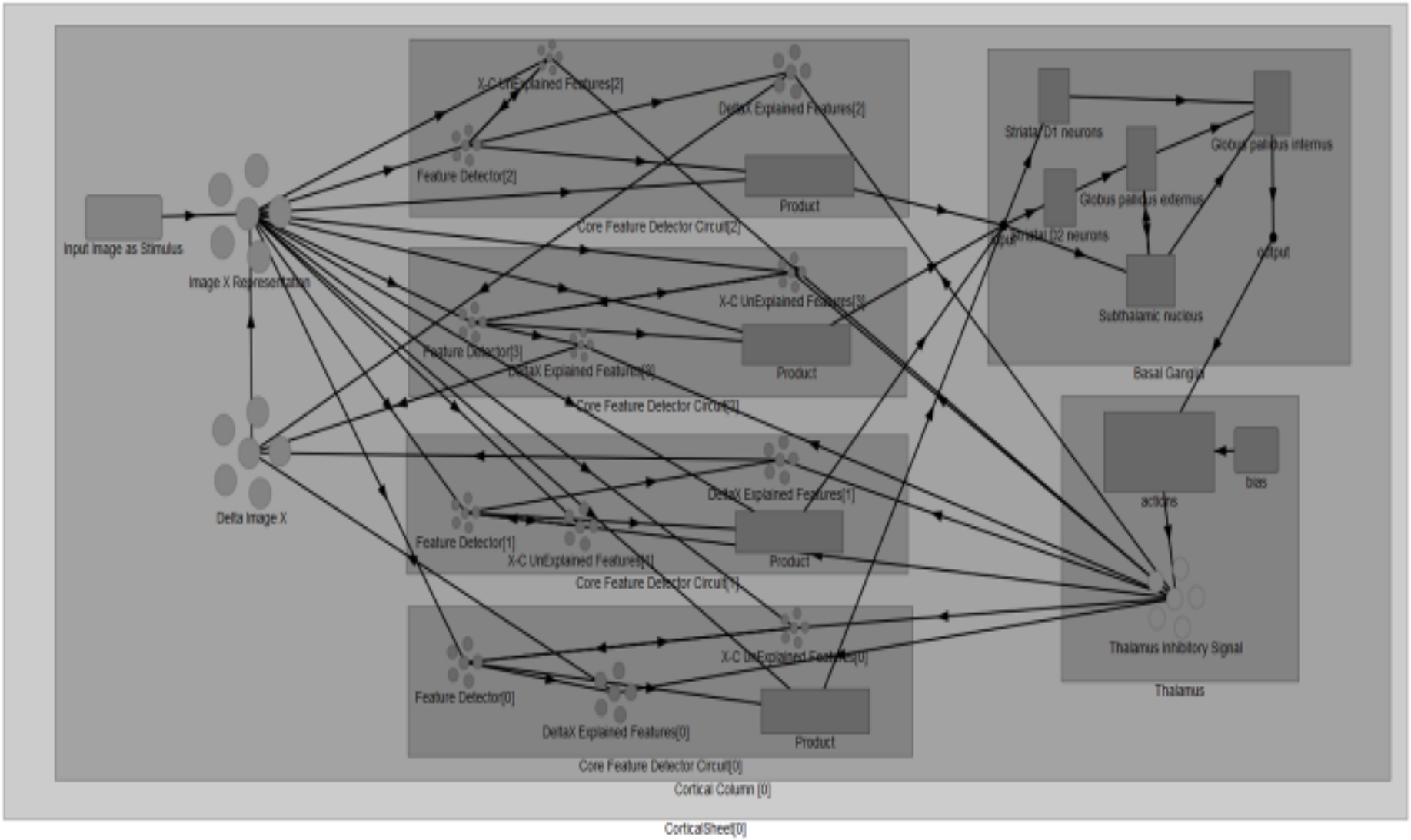

Figure 51: Nengo screenshot of NEF cortical sheet hierarchical categorizing Core circuit showing multiple feature detectors in a competitive array.

The actual categories that the feature detectors learn is determined by the competitive interaction between the various feature detector instances. The NEF neural ensembles learn the interactions between the dimensions of the stimulus images, just as cortical columns neurons learn the inter-relationships of their stimuli and at the same time that the columns compete for their own feature attributes to specialize in. This is how the structure of the inputs is learned.

6.2.4. Feature Detector Ensemble

Feature detectors are analogs to the cortical mini-column. The feature detector class represents a combination of the layer I, II, III and IV properties that implement this attractor network for a given feature in a specific orientation. The feature detectors are composed of inter-connected NEF ensembles of neurons. Each feature detector has a pattern that it learns to recognize, which starts initially from a random pattern determined by the assignment of random tuning curves to the neurons in the NEF ensemble at the core of the feature detector.

In section 6.1 we detailed the process flow for the whole model. Here we discuss the process flow within each feature detector itself. Refer to Figure 48 and Figure 52 in the discussion that follows.

i. ***Stimulus:*** The stimulus comes from the Image_X instance of the Choo Accumulating Memory Network detailed in Section 6.2.2.. Note that in the code, the stimulus is actually stored in a separate NEF ensemble within the feature detector class instance, as the NEF decoding of the stimulus is specific to each feature detector and the decoders are learned on each feature detector class instance. This emulates the individual cortical mini-columns' unique learning, and competitive preferred direction vector interactions. The main Image_X accumulating memory network stores the stimulus, as in the core cells of the thalamus. As such there is only one instance of this as it represents the stimulus as yet unexplained by all of cortex. As discussed in section 6.1, the explained part of this stimulus, as collectively learned by all feature detectors is subtracted from this stimulus progressively.

ii. ***Learning Features:*** The Image_X ensemble inside the feature detector is connected to the feature detector's feature ensemble (in the code this is the ifd_FD_ens ensemble). This ensemble stores the current learned stimulus features this feature detector has specialized to detect. Learning is accomplished with a combination of Voja and hPES learning. Learning in the model is therefore both supervised and unsupervised learning as detailed in (Bekolay, Kolbeck, & Eliasmith, 2013). The encoders on the local Image_X ensemble start as random encoders. The Voja learning connection then learns the encoders on the local copy of the Image_X vector, specializing this feature detector in recognizing those features that it has learned by virtue of its winning previous competitions, and the Voja learning rule being allowed to learn a bit more of the image being input. The

result is that, over time, each feature detector develops a unique set of specialized encoders. The local Image_X copy of the stimulus, or rather the special encoded version of the stimulus image for each feature detector, then drives a PES learning connection in each feature detector to this feature detector's reconstruction of the image from its learned features. This learned reconstruction is driven to attempt to replicate the original input stimulus image.

iii. Spatial Feature-Detector Learning: Cortex's many layers are organized in spatio-temporal maps. Within those maps the cortical mini-columns stimulate each other via their connections. The effect is to implement generally local mutual excitation and distal inhibition, all combined with strong network dynamic inhibition from the thalamus. The balance of these effects with layers of spatially arranged maps, which preserve the spatial order of the stimuli in lower levels and a representational ordering at higher levels (such as seen in the visual system's preferred direction vectors shown above in Figure 4), means that computation is very much spatial and temporal. Our model contains no direct implementation of the spatial nature at the intra-mini-column level. However, within each feature detector, local spatial correlations are learned by the hPES and VOJA connections over each feature detector's own preferred stimulus.

iv. ***Competing for Features:*** Feature detectors cannot learn unless they first 'win' the ability to learn. Both the PES and Voja learning rules are only allowed to learn in any given feature detector when the feature detector has won the competition among all the feature detectors by having the closest match to the input stimulus, measured by the dot-product between the stimulus image and the winning feature detector's PDV (stored in its ifd_FD_ens ensemble, see Figure 48). The error signal to the hPES learning rule is computed in the X_minus_FDi ensemble as the difference between the feature detector's PDV (in the ifd_FD_ens ensemble) and the original image. This winner-take-all (WTA) competition is implemented via the Nengo neural Product network, which computes the dot-product between the feature detector's PDV and the original Image_X stimulus. Once all the neural populations compute their feature detector's PDV and the product ensembles compute the dot-products, the WTA competition is implemented by an instance of the Nengo basal ganglia class which, as discussed above, only passes on the signal from the most active of its inputs.

Lateral inhibition is also supported in the model whereby the feature detectors all laterally inhibit each other such that the more active a feature detector network is, the more it exerts an inhibitory force on the activities of all other feature detectors.

In our model, the combination of the basal ganglia network and the cortical lateral feature detector inhibition implements the WTA mechanism theorized to be found in the attractor networks in layers I, II and III of cortex by the Granger model.

The Nengo Basal Ganglia network has the right functionality to model our cortical WTA analog functionality. Note, however that it is not being used in our model as a biological basal ganglia analog. The Nengo Thalamus network, smooths out the outputs of the basal ganglia network, and inverts it so that the winners are the most active feature detectors (the ones with the closest match of their PDV's to the image) and the 'losing' feature detectors show no activity. This 'winning' and 'losing' signal drives the recurrent inhibition to the Voja and hPES learning rules, ensuring that only the winning cortical mini-column(s) (our feature detectors) learn and specialize on the current input.

v. ***Implementing Explaining Away in our Model:*** As the network processes its inputs, there are two accumulators that are required to implement the desired behaviour. We require a neural working memory of the parts of the image that were explained away by the current winning feature detectors, and an accumulation of all current and previously explained parts of the image.

The former working memory is captured in the XminusFDi ensemble, while the later is accumulated in the FDi_DeltaX ensemble. The XminusFDi ensemble forms the error signal for the hPES learning connection from the image to the feature detectors while the FDi_DeltaX ensemble feeds the DeltaX accumulator which is subtracted from the accumulating memory in the ImageX instance discussed above.

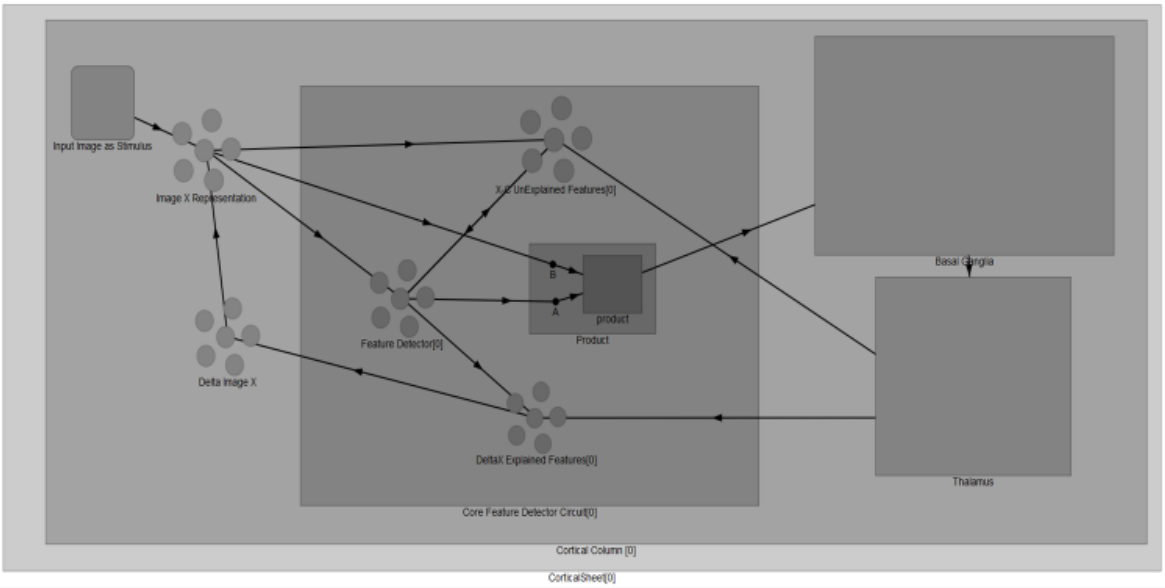

Figure 52: Nengo screenshot of one cortical feature detector.

vi. Sequential, Hierarchical Categorization Cortical States: The sequence of states that the network goes through with the dynamically formed winning feature

detector networks being activated and then deactivated, forms the progressive, hierarchical categorization over time of the input stimulus to the network. The learning occurring in the feature detector networks models the adaptive nature of the cortical circuits. In theory, in the cortex, the forward and backward connections that occur from and to the winning feature detector circuits potentiate the predictions of the next stimuli. These connections also constrain the interpretation of the current stimuli. In this way the progressive, sequential, and hierarchical categorization of this circuit could allow for network activities to represent the predictions of the future states of the general classes of the current stimuli. The connections could also constrain the recognition of the sequence of stimuli to resolve conflicting interpretations and aid in recognition and processing of stimuli. It is a compelling theory and one which Granger has written on subsequently to the original paper our modelling work was based on (Granger, Kilborn, Rodriguez, & Cabanne, 2017).

6.3. Shortcomings of the Model

There are many shortcomings of the model to be mentioned over and above those already discussed in the text preceding this section.

i. ***Contextual Flow of the Model:*** Biologically, the upper layers of the visual system, which the Granger model is working to understand the functioning of, take in highly processed representations. They are several layers into the visual system. The Granger model, and our model, jump right in to the middle of that circuit without having built up the rest of the visual system, and tries to study the plausibility of the functioning of the theorized categorization behaviour of that circuit. However, if the ultimate desire is to begin to understand the interaction of the thalamo-cortico-cortical circuits, then such models are a useful exercise, despite its limits of translation back to the biology, especially given the theorized importance of both categorization and of thalamo-cortical coordination to everything from perception to consciousness.

ii. ***Lack of Invariance:*** As mentioned above, the model is only a partial model of the image processing in the brain, while being fed images as test data, it tries to represent higher order circuit dynamics, which in the brain would be receiving already processed activities from the entire lower areas of the visual system, including both the upper and lower pathways. One of the things that the visual system does, before the Granger model circuits, is create invariance to the translation, rotation, deformation, and occlusion of the stimulus objects. This missing capability is a very serious limitation of our model's biological plausibility. It is however a limitation of the Granger model as well.

iii. Lack of Spatio-Temporal Learning: As discussed above, feature detectors model spatial learning reasonably well, but spatial characteristics are not learned

between mini-columns in the connectivity directed way that cortex is likely performs it. The temporal nature of the computation comes from the differing time constants across the interacting layers and columns. The difference between the strength, duration, and distance of inhibition versus excitation in cortex is a major design feature we have mentioned above. Our model does not account for this.

iv. ***Lack of Spreading Activation & Prediction:*** Our model also does not implement any form of spreading prediction or activation, nor do we model the effects of prediction, long-range inhibition, or the interaction of the multi-attractor layers, layers V and VI implementing a theorized top-down and bottom-up computation within the brain's thalamo-cortical circuits.

v. ***Failures of Biological Plausibility of the Model:*** No artificial neuron model is a replacement for the real biological circuits. That said, in generating our model, we tried to model biologically plausible circuits at the chosen level of abstraction. In some areas we achieve reasonable plausibility, in other areas the way in which the functions are achieved by the biology are unknown, or our tools do not easily allow us to model the biology, or simply our skills at modelling failed to generate a plausible model. This increases the risk to the biological plausibility of any and all observations made from such a highly recurrent, large and interconnected model such as ours.

   In two particular places, we have used network components that we know are not like the biological structures whose functionality we are attempting to model, but to complete the model we have tried to use these components to achieve the same functionality in order to study a functioning model.

   First, to simulate the winner-take-all (WTA) functionality amongst the thalamo-cortical circuits the Granger model uses, as discussed above, we have used the Nengo Basal Ganglia Network (as discussed in Section 5.7.7). This NEF network was designed to model the WTA mechanism for action selection in the biological basal ganglia. Here it performs a reasonable functional analog for the computational role we require to select the winning feature detector, but it is not biologically plausible as implemented.

   Second, as discussed above, in the Granger model, the core cells in the thalamus continue to provide stimulus after and during stimulus presentment to cortex. Portions of this information flow are then inhibited by the TRN cells. We have modelled this stimulus memory with subtraction using the Choo Memory Network, as documented above, but it does not directly model the TRN or the core cells in a biologically plausible way. It does however, hold the stimulus and subtract the stimulus in accordance with the feedback from the feature detectors, allowing our overall model to function.

7. Simulations and Results

The model was run on single MNIST images to validate that the feature detectors were learning, that the image was being progressively explained away and that the model produced a group of sequential categorizations from the input images. Batches of 1 to 1,000 MNIST images were run in the development of the model. The many parameters of the model were varied both together and in isolation to determine their effects on the model's performance.

Results for one image and 1,000 image runs are discussed below as examples of how the model performs. We discuss the variation in classifications seen from the result of the 1,000 image runs to assess to what extent a small dataset produces any learning of features, and in what sequences.

7.1. Single Image Simulation Example

7.1.5. Image Processing Overview Chart

The simulation code produces a series of time plots for each image processed, the main one being that shown in Figure 54 below. Note that these plots of a single image run are created on a run of the NEF Granger model using an artificially high learning rate to allow the plots to show conceptually what is happening quickly for the purposes of this explanation. There are a few adverse artefacts of this that we mention.

The top two graphs show the output of the basal ganglia and the thalamus (for explanation of the dynamics see Figure 45). The next two graphs (graphs #3 and #4) from the top show a cleaned up and an inverted copy of the thalamic output respectively. These signals are used to drive the learning switch and data collection functions of the model. Once a feature detector "wins", it is allowed to learn for a configurable time, typically 20 ms.

The bottom two plots, image_X and delta_X, show the dimensions of the image ensemble and the delta ensemble. The Image_X plot records the current amount of input still to be explained-away or categorized in the accumulating memory network. The delta_X plot shows the explained away features for the current hierarchical categorization level, resetting after each categorization level is completed. The transitions in the signals are driven primarily by the stabilizing of the output from the basal ganglia network (WTA analogue), which drives the model to subtract the accumulated explained features and begin categorizing again. This emulates the cortico-thalamic feedback pathway in the Granger model from layer V of cortex to the thalamic RTN cells, which restrict the thalamic core cells as explained previously.

The images on the bottom row of Figure 45 show, starting from the left, the end states of: (1) the node inputting the original modified MNIST image to the model, (2) the value of the accumulating memory left after the learning has explained away some of the image, (3) the value of the core of the feature detectors have learned so far, and lastly (4) a check image which is the result of simply adding the explained parts of the model to the remainder unexplained as yet. Note that it would take many runs to cause the feature detectors to learn all of the input stimuli due to the overall learning of features and the effects of the slow learning rates used, typically in the ranges of 1x 10^{-3} (for testing as herein) to 1 x 10^{-6} for simulation runs.

7.1.6.Decoded Circuit Component Time Series Plots

The probe data for the source and decoded images are also plotted on time series plots as shown in the following figures. These are all decoded values. The ensembles store the actual neural activities that we can apply the NEF decoders to and generate the representational equivalent of the activities, but with some noise.

7.1.6.1. Input Node

Figure 53 shows the input to the model from the node, which is a clean representation of the actual pixel values from the modified MNIST image.

Figure 53: Node input image stimulus reconstructed from node activity output.

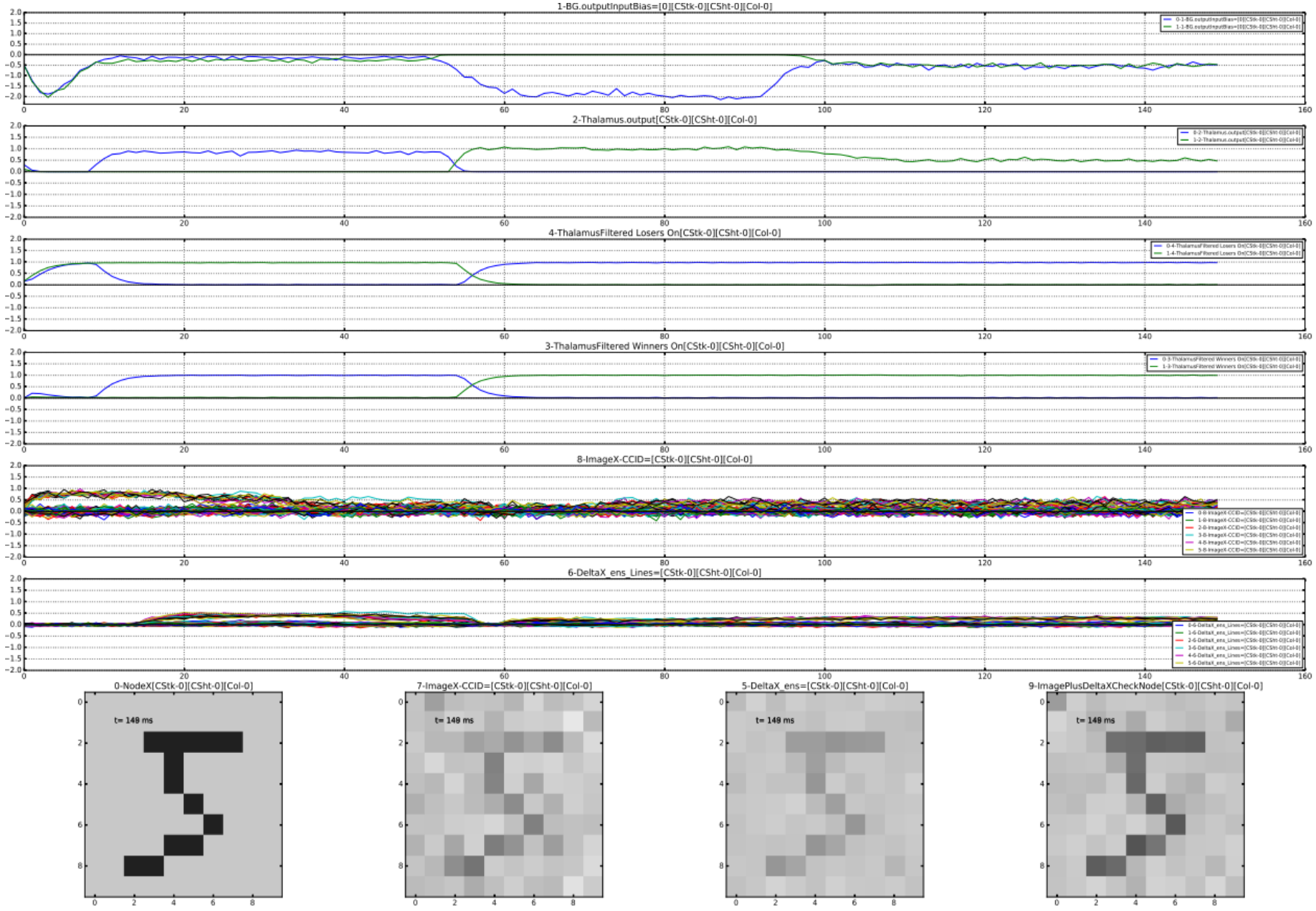

Figure 54: Summary image categorization neural activities plot.

7.1.6.2. Delta X

Figure 55 shows the accumulating explained components of the image, in total, across all feature detectors. It resembles the image quickly due the high learning rate in this demonstration run, and then is cleared when the next level of categorization begins, that is once the first level winning signal from the basal ganglia has stopped changing (as discussed above). The explained away portion then builds up again at the next level but in a weaker representation, as it is categorizing the remaining stimuli in the image on the second level categorization because the accumulating memory has subtracted away the previously explained stimuli.

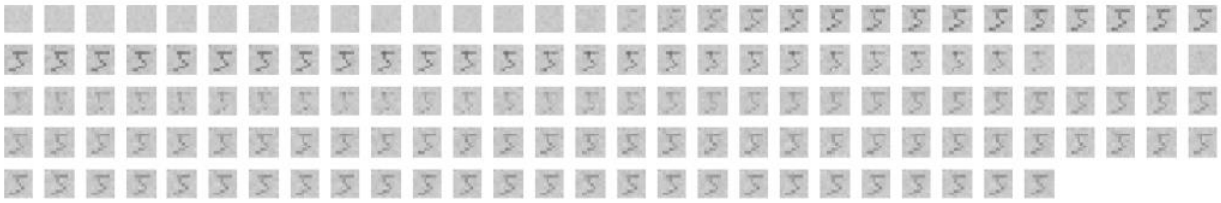

Figure 55: DeltaX.

7.1.6.3. Feature Detector Learned Features

The explained away portion is the sum of the feature detectors, learned features. Figure 56 shows the decoded representation of the winning feature detector over time. Again, due to the artificially high learning rate, the feature detector rapidly learns a good portion of the image, but then with only a few feature detectors in this test run, this detector wins on the next level pass over the inputs while the subtracting away is occurring. It then seems to unlearn its preferred stimulus, only to learn it solidly again at the end of the now stable presentation of the remaining stimulus to be categorized.

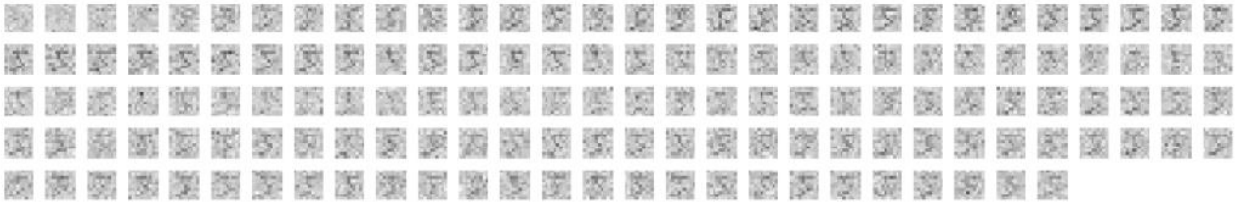

Figure 56: Feature detector learning over time.

7.1.6.4. Image Accumulating Memory

Figure 57 shows the decoded representation from the accumulating memory network which, as we have discussed, models the core thalamic stimulus cells and their interaction with the inhibitory overlaying RTN cells. As expected the image builds up initially under strong stimulus from the Image_X node, and then fades as portions of the image are explained away. But, the small test size error noted in Section 7.1.6.3 above causes the value of Delta_X to fall and then rise again. So, the progression of the explained features rises, falls and then rises again. This is an aberrant effect of the small one image test run.

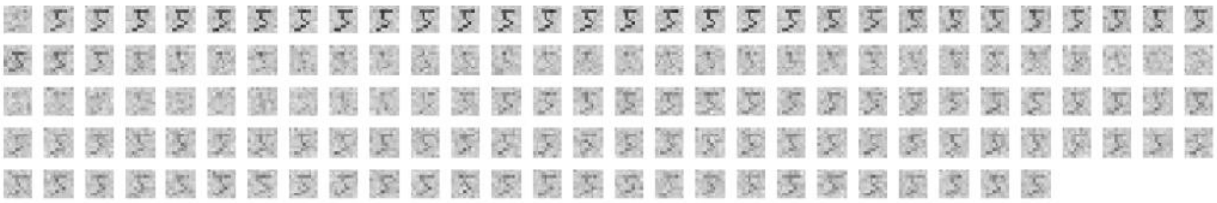

Figure 57 : Image explaining away over time in circuit. Reconstructed from neural activity in Layer II & III analogues.

7.1.6.5. Model Check Image: Unexplained + Explained Features Check

A check probe records the sum of the remaining to be categorized stimuli in the accumulating memory network (Image_X) and the already explained features in the Delta_X network. The result should always resemble the original input signal plus some noise. As seen in Figure 58, it largely does with slight variations in intensity.

This is evidence that our model is behaving reasonably. It does also shows that the way the model works, using a time duration of primary stimulus presentation without sensing when the primary stimulus is fully inhibited, causes incorrect responses as the primary stimulus continues to elicit feature detector selection and therefore learning that biases the feature detectors' learned patterns.

7.2. Feature Detector Evolution Over Time

Let us consider one feature detector's evolution from a random starting state to the end of the simulation, by which time it has learned a significant portion of the stimulus. Again, this is a high-learning-rate, single-image case we are using for illustration purposes. In Figure 59 plots 1 and 2, we see the winning versus losing feature detectors signals again. This feature detector is the winning one. The dot-product of this feature detector is the highest of all and therefore it is the winner in our simulated cortical-WTA process (until the explaining away process reduces the input image in one step after this winning feature detector is allowed to learn the stimulus at our artificially high demonstration rate).

Figure 58: Image plus Delta-X check

We see the effects of the noise inherent in the neural simulation, as well as the effects of the learning over time that occurs in the feature detector's internal representation in plot 7 of Figure 59. The check image at the bottom show the addition of now learned features and the residual stimulus signal which has fallen to a uniform nearly all white block as shown in the plot 4 of Figure 59.

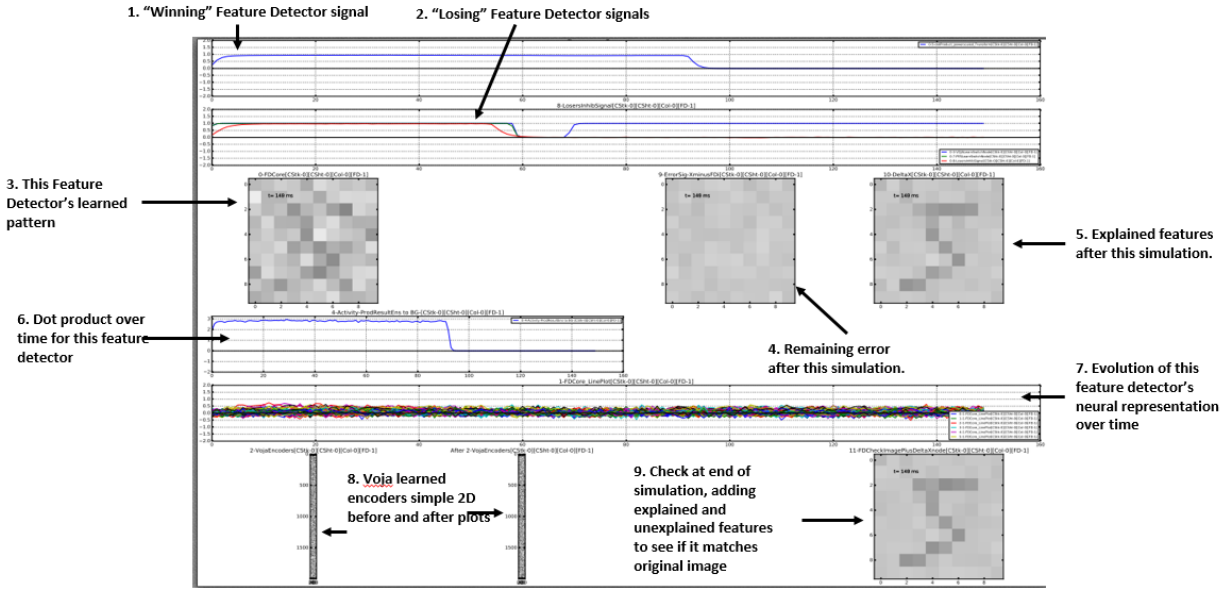

Figure 59: Feature Detector 1 Activity Plot.

7.3. Inhibitory Signals Controlling Feature Detector Learning

Figure 60 shows a plot of the inhibitory signals that control learning and the subtraction of explained features in the simulation. Many of these signals do not result from neural dynamics, rather they are control signals generated by the procedural logic that feeds the Granger cortical circuit components of our model. A broader more biologically plausible simulation would implement these signals directly using neural dynamics as we have done in the rest of the model. We will present them and discuss them here to make our review of the model complete.

In this example, the first thing that happens is that the image reset signal, generated by the image change logic, goes high. The new stimulus flows to the feature detectors and they compete as discussed above. The winning feature detector then receives a high signal on its corresponding basal ganglia dimension, this allows it to learn for a brief period of time and then flow its explained features signal into the accumulating memory. This happens as this signal causes the accumulating memory change-detect signal to go high. This allows the output of the feature detector to begin to accumulate as the explained-away features of the stimulus. After a delay, the learning inhibition then turns on again stopping this feature detector from learning more features.

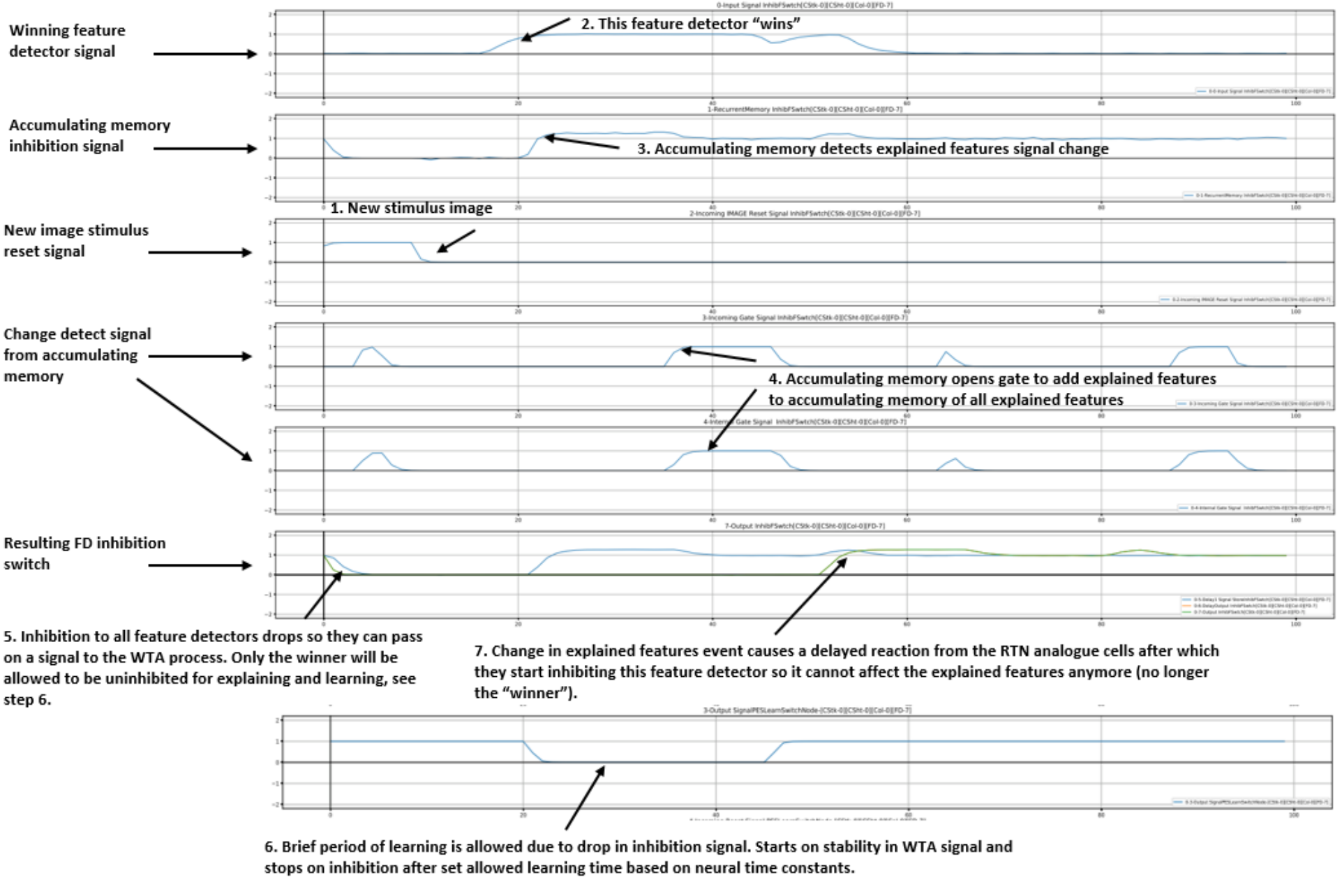

Figure 60 : Competitive explain-away and learning inhibition / disinhibition signals resulting from stabilizing dynamics of the simulated cortical Winner-Take-All circuit (implemented using the Nengo basal-ganglia network).

7.4. Activity Dependant Explaining Away

The learning and inhibition signals of the previous section are locked into a feedback loop with the evolution of the feature detectors' category representations. Figure 61 shows the multi-dimensional signal flows inside the accumulating memory circuit that sums the explained features as learned and then output by the winning feature detectors over time. The model then subtracts this summed explained-features representation from the remaining stimulus image to form an analogue for the behaviour of the thalamic core cells. In the biology, this subtraction is implemented by progressive inhibition of the thalamic core cells by the RTN cells under the direction of the feedback from layer V of cortex in the Granger model, as we discussed previously.

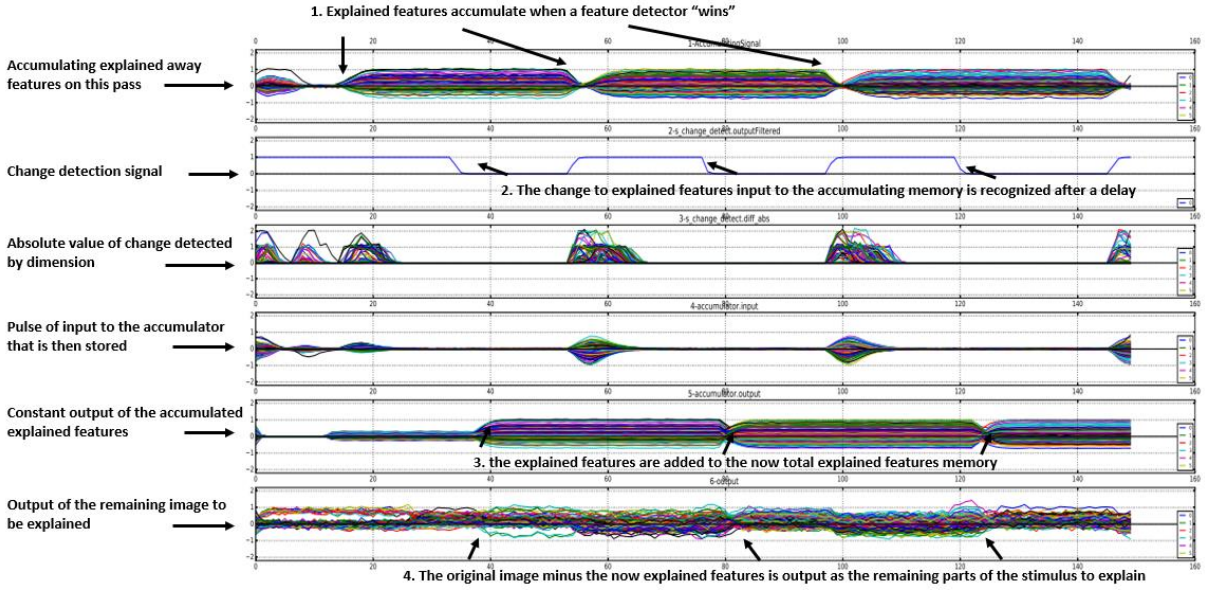

Figure 61: Accumulating explaining away circuit neural activity, simulating the effect of the Core circuit's thalamic reticular neurons inhibiting the explained portion of the inputs.

In this circuit, the winning feature detectors output the explained away features is shown in the top graph in Figure 61. This signal is monitored by the accumulating memory's change detect ensemble based on its derivative, that change detection triggers the accumulator gate to open, releasing its inhibitory grip on the accumulator. The second graph from the bottom shows the progressively growing signals from all the explained away features that are added by the accumulating memory. The last graph shows how the signal from the accumulating memory is then subtracted from the original image representation, that is the output of the accumulating memory.

7.5. Feature Detector Competition

As shown in chart A of Figure 62, the feature detector output is compared to the current remaining image stimulus, and the dot-product is computed neutrally. All the resulting dot

products are then fed as the basis of the WTA process to the basal ganglia model as in chart B.

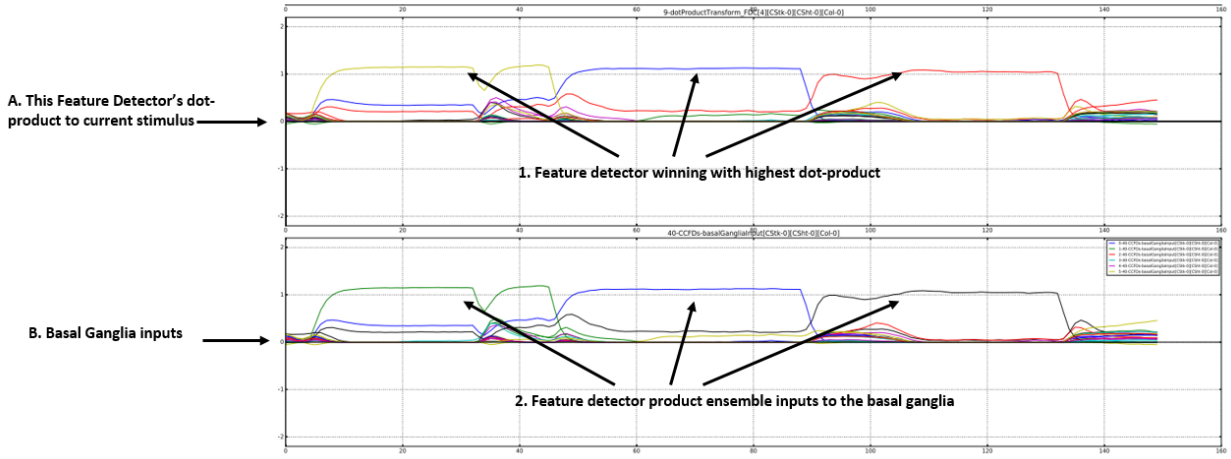

Figure 62 : Feature detector dot-product over time, simulating layer IV competitive WTA based on neural comparison of receptive fields to input stimulus from thalamus.

This sequence of the feature detectors that won for each image becomes the hierarchical categorisation for each image.

7.6. Multiple Image Simulations

Once the full model was run on single images, comprising approximately 500,000 to over 1,000,000 neurons, we ran the model on multiple images, testing the models' ability to produce feature detectors that learn complex features generated from multiple images. We found only partial success. Computational limits were the largest problem. The model's current design proved too cumbersome for current simulation resources. Each image took approximately four (4) minutes on the available GPU servers. This meant that a full MNIST run of 50,000 training images and 10,000 test images would have taken over 150 days to complete.

In addition, the model was limited by available memory to 30 feature detectors. In order to have a robust simulation we would have needed to many more feature detectors, perhaps hundreds. Given these limitations we opted to study smaller runs of 50 to 1,000 MNIST images. The results show some signs of robust feature detector competition, learning, and resulting explaining away, which are the core features of the Granger model. Given the low number of training samples possible on such small runs, to some extent it is surprising there is any consistency in the runs. Given the insufficient training runs that we were limited to, and the inconsistent results, we present these results for inspection only.

Figure 63 shows the result of a 50-image run. The images are shown along with the sequence of the identification numbers of the winning feature detectors for that image.

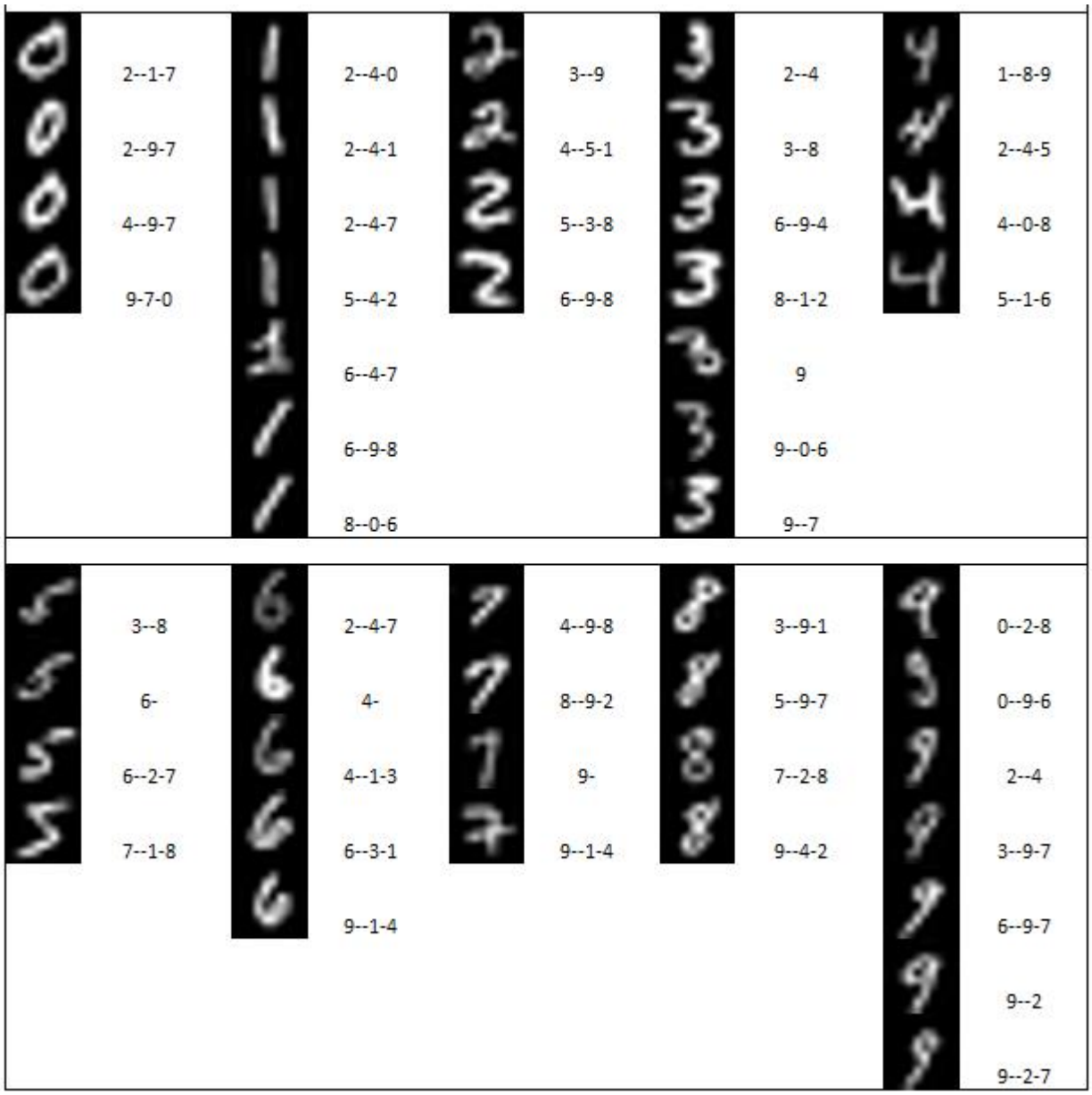

Figure 63: 50 MNIST image run results, showing the sequence of feature detectors that classify the image. The sequence of feature detectors is shown by the feature detector identification numbers in the string of digits separated by dashes beside each image. (Note: single or double dash seperators are equivalent).

In looking through the probe data on the runs, and trying to trace what was causing this trial-to-trial instability, several significant limitations of the architecture of our model came to light. We believe the core problem lies with the computational limits of scaling the model to have a more biologically accurate number of columns and spatial arrangement of the feature detectors. This combined with the fact that the dot-product mechanism we are

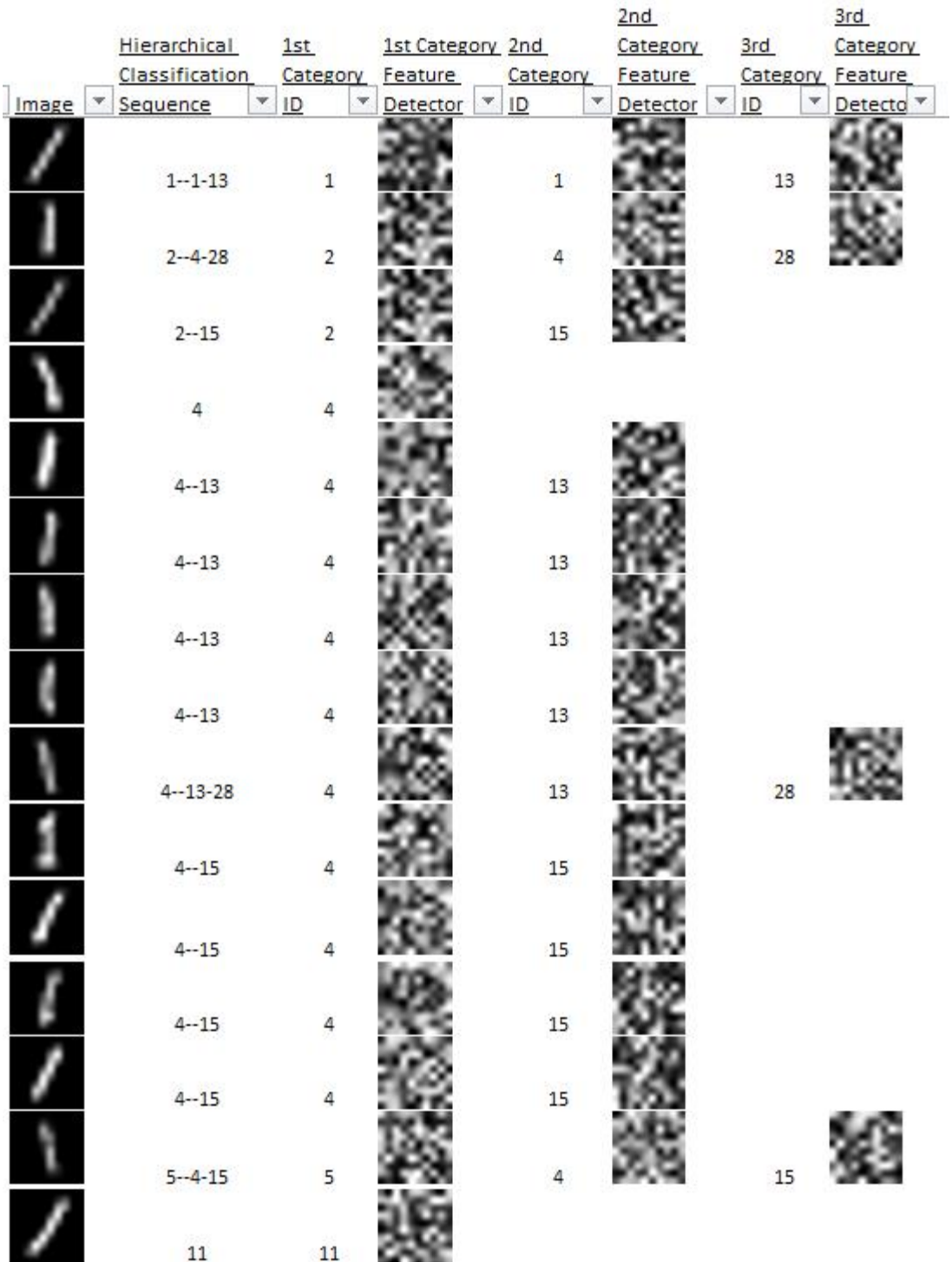

Image	Hierarchical Classification Sequence	1st Category ID	1st Category Feature Detector	2nd Category ID	2nd Category Feature Detector	3rd Category ID	3rd Category Feature Detecto
	1--1-13	1		1		13	
	2--4-28	2		4		28	
	2--15	2		15			
	4	4					
	4--13	4		13			
	4--13	4		13			
	4--13	4		13			
	4--13	4		13			
	4--13-28	4		13		28	
	4--15	4		15			
	4--15	4		15			
	4--15	4		15			
	4--15	4		15			
	5--4-15	5		4		15	
	11	11					

Figure 64: Results for images of 1's out of a 1,000 MNIST image training run, showing categorization of the images by matching feature detector.

using for comparison between the feature detectors when using 100-dimension vectors all with values in the 0 to 1 range and all computed across neurons is too noisy with too small variations to be a reliable measure to base the WTA mechanism on. The learning is then applied across multiple feature detectors for the same image. We conjecture that

the spatial competition arrangement, combined with lateral inhibition in real cortex allows multiple winning cortical pyramidal cells, all of which are in the same or related macro-columns. This would result in an over-complete, competitive representational space, which would make the system less reliant on exact repeatability against any one preferred direction vector. Our model has too few pyramidal neurons (aka feature detectors), and they are not redundant. The result of each image classification therefore rests solely on the exact winning sequence. In some ways, this effectively removes the population coding and redundant nature of biological neural networks that likely makes them so powerful. While we had hoped that learning over time would resolve this, either we did not have enough test runs to achieve that long run specialization through competition amongst the feature detectors, or our dot-product discrimination method's instability would not allow it.

Finally, although we were disappointed in the results of the 50 image runs, we wished to see if we could see any reasonable improvement due to learning when we ran 1,000 images (850 training and 150 test) over the course of 3 days of runtime. The learning rate for all learning was set to 1×10^{-6} and for the 1,000-image run this means that only about 100 images of 1's was trained on plus only 15 test images of 1's.

Figure 64 shows a sample of the classification of the 15 test images of 1's from the run. The column "Hierarchical Classification Sequence" shows the order of the feature detectors that matched the image in the order they were matched. The next columns show the state of the feature detector at the time that it matched the image, or the resulting residual image after the subtraction of the explained parts of the image. Due to the small number of runs, the feature detectors are all clearly early in their learning, not visibly resembling any one image strongly, but showing elements of what appear to be patterned learning. The number of 1's that are classified with a similar classification, appears to show the beginnings of a classification scheme, while other examples seem random or unrelated.

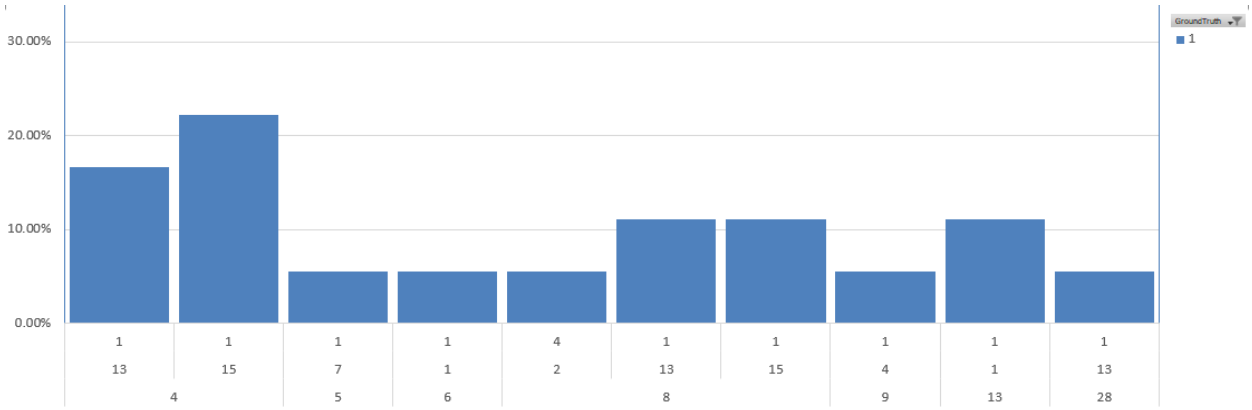

Figure 65: After 100 training images, hierarchical classification for MNIST images of 1's showing the percentage time each sequence of feature detector sequence matched to an image of a 1. To read the bottom axes, as an example, the right most category is feature detector 1 won, then 13 and finally 28; after which the stimulus was explained away to nothing. Negligible categories are not shown. This flat distribution across feature sequences suggests no hierarchical categorization of the shown images.

We then graphed the classification of all images across time for each classification hierarchy to see if in fact we can show that learning is occurring. Figure 65 shows the feature detector sequence classifications of the first 100 training images while Figure 66 shows the classification categories as performed on the test images after 750 training images.

In general, the 100-image run produces a more random distribution of feature detector sequences (Figure 65), the “categories”, than the tighter clustering of categories from the 750-image run (Figure 66). The images of the one’s are moving into a few more common clusters. Given the incomplete training, it is only circumstantial evidence of clustering working but is as would be expected given the short training process.

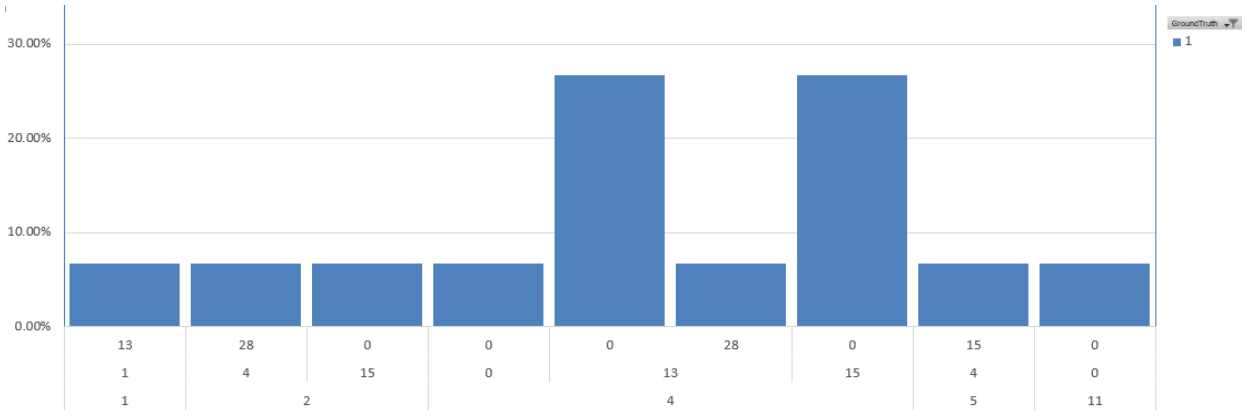

Figure 66: The classification sequences on the test images of 1’s; after 750 training images. Hierarchical classification for MNIST images of 1's showing the percentage time each sequence of feature detectors matched to an image of a 1. Negligible categories are not shown. The peaked distributions of classification sequences suggest improved clustering over that shown in Figure 65.

Turning to Figure 68, we see the entire dataset of classifications for a training run of 750 images. In Figure 67 we see the results after only 100 images were run. In comparing the two, we see the same progression to fewer clusters of greater explanatory frequency in the 750-image run than at the end of the 100-image run. In the 100-image run, the bars are roughly evenly distributed. In the 750-image run, the number of bars has reduced, as well as many of the remaining feature detectors frequency of “winning” bars increase in height. For each MNIST image the ground truth is clustering around an increasingly consistent sequence of feature detectors as also shown visually in Figure 64 for the run of the MNIST 1’s images from a 1,000-image training run.

There are also many examples of what appear to be random classifications or possibly unlearned feature associations. In all the data, while encouraging, is not conclusive in any empirically measurable way. Hopefully future studies continue to explore cortico-thalamo-cortical dynamic circuit modelling in a biologically plausible way to explain more of the categorization and classification behaviour that we have attempted to elucidate here.

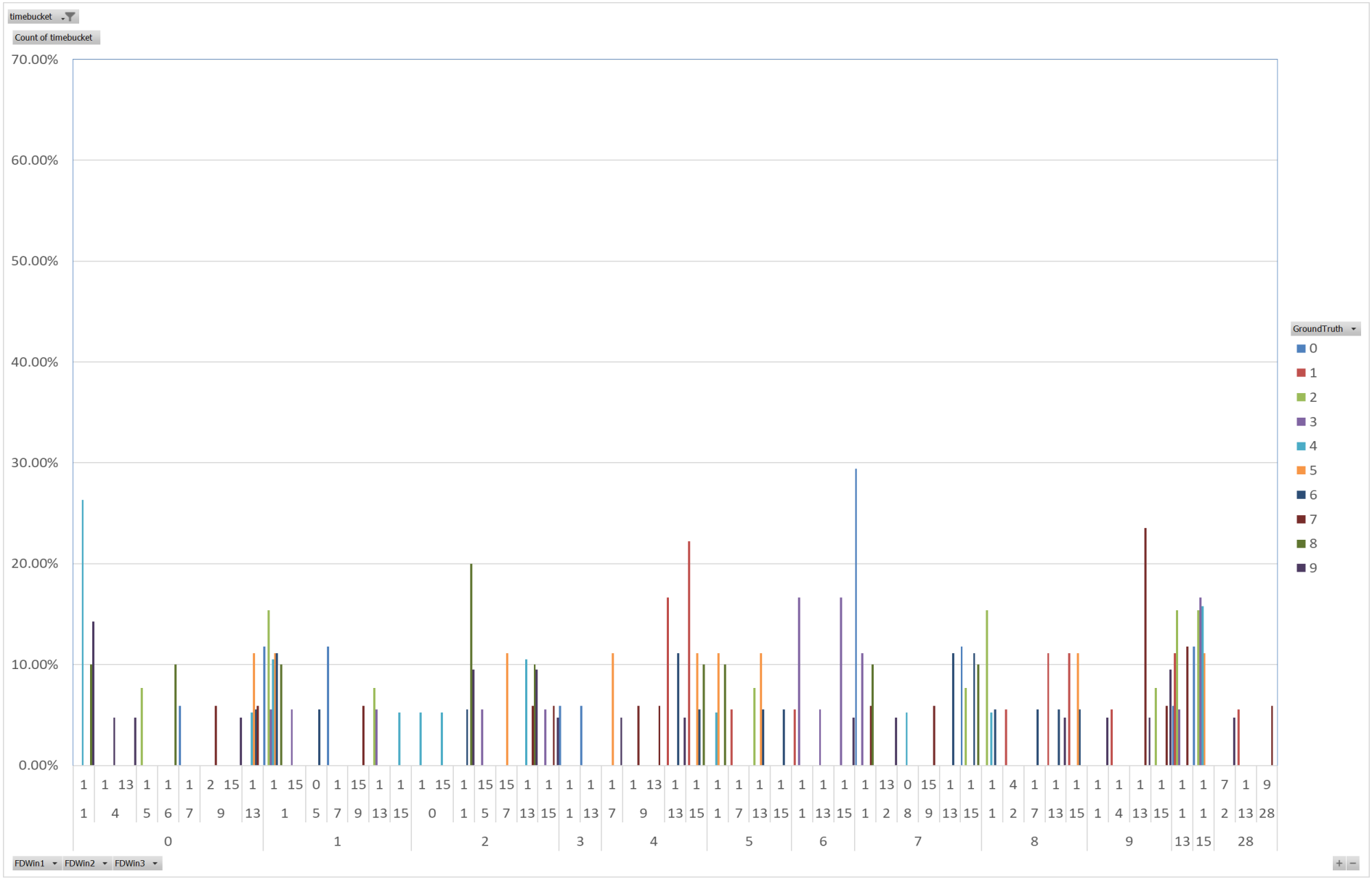

Figure 67: Feature detector categorization performance of MNIST images after 100 image training run.

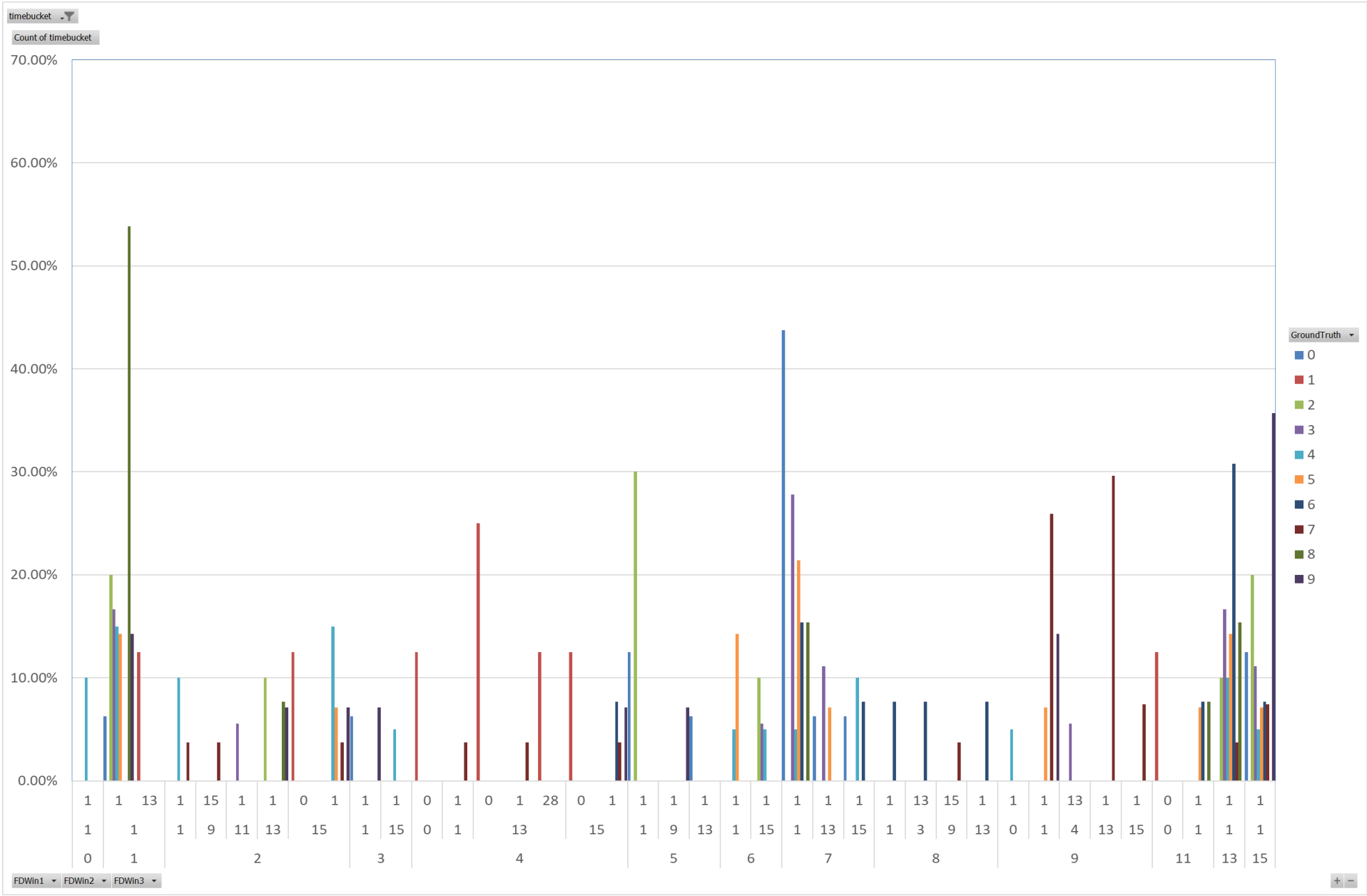

Figure 68: Feature detector categorization performance of MNIST images after 750 image training run.

8. Discussion and Future Work

8.1. Discussion

The main contribution of this thesis is that, the parts of our model that are neurally implemented have been shown to plausibly result in the sequential processing behaviour theorized in the Granger model. These parts include the competition between the feature detectors and the sequential explaining away of the input stimulus, and the dynamics of the neural representations of sensory inputs along the pathway from specific thalamus to cortical layers IV then I/II/III and then from layer V back to the thalamic RTN.

While we did not show the repeatability of the Granger model's hierarchical classification in our model, we believe that our work does lend credence that dynamics stemming from neural connectivity and neural time-constants are a core component of cortical computation. We also believe that this is an area that needs much more research both to understand how brains work, but also as a key thrust in pushing the current state of artificial intelligence forward past the largely feedforward, non-dynamic approach of deep learning and other machine learning models.

Secondly, we have discussed an array of cortical models, focusing on those that are envisioned as dynamic ones, and shown in detail a path for transferring those anatomical and dynamic models to NEF spiking models both theoretically and practically. We hope that this serves as an end to end example of how this can be thought of in concert with the large body of published research on the NEF and Nengo neural computational models.

This modelling work provides some more credence to the hypothesis that the dynamics of networks in the brain are likely a major source of the function of higher order cognitive capabilities. We also believe that this lends support to the theories that underlie the Neural Engineering Framework (Eliasmith & Anderson, 2003), and the advanced dynamic brain model Spaun (Eliasmith, 2013) built with it. In Spaun, the dynamics integrate many different types of neural algorithms, including pattern separation and recognition ones like support vector machines (SVMs).

8.2. Future Work

We set out on a functionally broad challenge in this paper, holding up the functioning of thalamo-cortico-cortical loops as the explanatory challenge of our work. We now describe the many short-comings of our model compared to the Granger model, and compared to the biology. We have listed many already in Section 6.3, but we wish to mention a few more here. The following are suggestions for future work for anyone wishing to continue

an NEF model of the thalamo-cortical circuits using the Granger model as a starting point. They could:

a. Improve the model to operate consistently on a trial to trial basis, perhaps using much larger layers of feature detectors with spatially arranged lateral inhibition mimicking the spatial competition found in cortex. It would then be possible to extend the model into multiple sheets of many feature detectors per column to test hierarchical winner-take-all among the cortical columns, as opposed to narrower individual feature detector competition. This will require much larger computational resources.

b. Implement the Granger model's sequence circuit to learn sequences, and use those sequence predictions to constrain the competition amongst the feature detectors as cortex is theorized to do.

c. Implement a more realistic attractor-based cortical model within the layers of feature detectors inspired by Rolls' model, described above, and using Eliasmith's methods for building spiking attractors in the NEF (Eliasmith, 2005).

d. Investigate the Granger hierarchical classification model as a method of equating stimuli with neural symbolic representations, such as, the Semantic Pointer Architecture (SPA) to investigate the resulting model as a candidate for a biologically-plausible model of feature binding and symbol grounding. This could be done by connecting the layer II and III representations that the Granger model constructs (the feature detectors in our implementation) with SPA pointers in a learned way, and then evaluating what learning mechanisms would enable reliable learning over time. The SPA pointers could then also reverse the process, forming predictions by expanding the pointers to the learned stimuli that created them, to investigate Eliasmith's concepts of pointer expansion in a symbol grounded way.

8.3. Conclusion

This project was undertaken to investigate the biological plausibility of the Granger model, when modelled using spiking neurons. The results of the modelling work suggest that serial, hierarchical progressive categorization could be implemented biologically via the interaction of the excitation and inhibition between thalamo-cortical loops. Further study, including larger and improved models should be undertaken to evaluate the plausibility of dynamic, thalamo-cortico-cortical networks, to further our understanding of the biology, but also to search for improved learning algorithms in general. The cortex leaves much to be discovered.

Bibliography

Abeles, M. (1982). *Local Cortical Circuits: An Electrophysiological study.* Berlin: Springer.

Abeles, M. (1991). *Corticonics.* New York: Cambridge University Press.

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science, 9*(1).

Ambros-Ingerson, J., Granger, R., & Lynch, G. (1990). Simulation of paleocortex performs hierarchical clustering. *Science, 247*, 1344-1348.

Anton, P., Lynch, G., & Granger, R. (1991). Computation of frequency-to-spatial transform by olfactory bulb glomeruli. *Biological Cybernetics, 65*, 407-414.

Antzoulatos, E., & Miller, E. (2011). Differences between neural activity in prefrontal cortex and striatum during learning of novel, abstract categories. *Neuron, 71*(2), 243-249.

Applied Brain Research Inc. (2017, 10 19). *Associative Memory learning example*. Retrieved from Nengo core 2.6: https://www.nengo.ai/nengo/examples/learn_associations.html

Ballard, D. H. (2015). *Brain Computation as Hierarchical Abstraction.* Cambridge, MA: MIT Press.

Ballard, D. H. (2015). *Brain Computation As Hierarchical Abstraction.* (T. J. Sejnowski, & T. A. Poggio, Eds.) Cambridge, Massachusetts, USA: MIT Press.

Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communications*.

Bekolay, T., Bergstra, J., Hunsberger, E., DeWolf, T., Stewart, T. C., Rasmussen, D., . . . Eliasmith, C. (2013). Nengo: A python tool for building large-scale fnctional brain models. *Frontiers in neuroinformatics, 7*.

Bekolay, T., Bergstra, J., Hunsberger, E., DeWolf, T., Stewart, T. C., Rasmussen, D., . . . Eliasmith, C. (2013). The Nengo Neural Simulator. *Frontiers in Neuroinformatics*. Retrieved from Nengo: http://nengo.ca/

Bekolay, T., Kolbeck, C., & Eliasmith, C. (2013). Simultaneous unsupervised and supervised learning of cognitive functions in biologically plausible spiking neural networks. *In Proceedings of the 35th Annual Conference of the Cognitive Science Society*, 169-174.

Beul, S. F., & Hilgetag, C. C. (2015). *Towards a "canonical" agranular cortical microcircuit.* (P. Krieger, Ed.) Frontiers in Neuroanatomy.

Bienenstock, E. L., Cooper, L. N., & Munro, P. (1982). Theory for the development of neuron selectivity: orientation. *Journal of Neuroscience, 2*(1), 32.

Borji, A., & Dundar, A. (2017). A new look at clustering through the lens of deep convolutional neural networks. *arXiv*.

Bosman, C. A., & Aboitiz, F. (2015). Functional constraints in the evolution of brain circuits. *Frontiers of Neuroscience: Hypothesis and Theory*. doi:10.3389/fnins.2015.00303

Bosman, C. A., Lansink, C. S., & Pennartz, C. M. (2014). Functions of gamma-band synchronization in cognition: from single circuits to functional diversity across cortical and subcortical systems. *European Journal of Neuroscience*, 1982-1999. doi:10.1111/ejn.12606

Bourassa, J., & Deschenes, M. (1995). Corticothalamic projections from the primary visual cortex in rats: a single fiber study using biocytin as an anterograde tracer. *Neuroscience*(66), 253-263.

Branan, N. (2010). Are Our Brains Wired for Categorization? *Scientific American*.

Briggs, F., & Usrey, W. M. (2008). Emerging views of corticothalamic function. *Current Opinion in Neurobiology, 18*(4), 403-407.

Brodmann, K. (1909). Localization in the Cerebral Cortex: The Principles of Comparative Localisation in the Cerebral Cortex Based on Cytoarchitectonics. (J. A. Barth, Ed.) 292.

Burlet, S., Tyler, C. J., & Leonard, C. S. (2002). Direct and indirect excitation of laterodorsal tegmental neurons by hypocretin/orexin peptides: Implications for wakefulness and narcolepsy. *Journal of Neuroscience, 22*(7), 2862–2872.

Buschman, T. J., & Miller, E. K. (2010). Shifting the spotlight of attention: evidence for discrete computations in cognition. *Frontiers in Human Neuroscience*, 9.

Buzsaki, G., & Draguhn, A. (2004). Neuronal Oscillations in Cortical Networks. *Science, 304*(5679), 1926-9.

Byrne, F. (2015). *Symphony from Synapses: Neocortex as a Universal Dynamical Systems Modeller using Hierarchical Temporal Memory.* Dublin, Ireland: arXiv. Retrieved from https://www.youtube.com/watch?v=rrBhHDzmgUA

Cajal, R. y. (1899). Comparative study of the sensory areas of the human cortex. 314, 361, 363.

Cajal, R. y. (1909). *Histology of the Nervous System.* (N. Swanson, & L. W. Swanson, Trans.) New York, USA: Oxford Press.

Cajal, R. y. (1937). *Recollections of My Life.* Philadelphia, PA: American Philosophical Society.

Calvin, W. H. (1995). Cortical Columns, Modules, and Hebbian Cell Assemblies. In M. A. Arbib, *The Handbook of Brain Theory and Neural Networks* (pp. 269-272).

Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Science, 14*(1), 506-515.

Carlo, C. N., & Stevens, C. F. (2012, January 22). Structural uniformity of neocortex, revisited. *Proceedings of the National Academy of Sciences, 110*(4), 1488-1493.

Chandrashekar, A., & Granger, R. (2012). Derivation of a novel efficient supervised learning algorithm from cortical-subcortical loops. *Frontiers in Computational Nueroscience*.

Charand, K. (2005). *Action Potentials*. Retrieved from Georgia State University: http://hyperphysics.phy-astr.gsu.edu/hbase/biology/actpot.html

Choo, F.-X. (2010a). *The Ordinal Serial Encoding Model: Serial Memory in Spiking Neurons. Feng-Xuan Choo's Masters Thesis.* University of Waterloo.

Choo, F.-X. (2010b). A Spiking Neuron Model of Serial-Order Recall. *32nd Annual Conference of the Cognitive Science Society. 32.* Cognitive Science Society.

Choo, X. (2016). *Accumulating Memory Network.*

Churchland, P. S. (1989). *A Neurocomputational perspective: The Nature of Mind and the Structure of Science.* Cambridge, MA, USA: MIT Press.

Churchland, P. S. (2006). A neurophilosophical slant on consciousness research. In V. A. Casagrande, R. W. Guillery, & S. M. Sherman (Eds.), *Cortical Function: A View from the Thalamus.* Amsterdam, The Netherlands: Elsevier.

Clark, D. L., & Boutros, N. N. (2010). *The Brain and Behaviour: An Introduction to Behavioural Neuroanatomy.* (3rd, Ed.) Cambridge, UK: Cambridge University Press.

Cohen, H., & Lefebvre, C. (2005). *Handbook of categorization in cognitive science.* Elsevier.

Conley, M., & Diamond, I. (1994). Organization of the Visual Sector of the Thalamic Reticular Nucleus in Galago. *Europena Journal of Neuroscience, 7*, 463-476.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Costa, N. M., & Martin, K. A. (2010). Whose cortical column would that be? *Frontiers in NeuroAnatomy*. Retrieved from http://journal.frontiersin.org/article/10.3389/fnana.2010.00016/full#B29

Coultrip, R., & Granger, R. (1992). A cortical model of winner-take-all competition via lateral inhibition. *Neural Networks, 5*, 47-54.

Crick, F. (1984). Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Sciences, 81*, 4586-4590.

Crick, F. C., & Koch, C. (2005). What is the function of the claustrum? *Philosophical Transactions in the Biological Sciences of the Royal Society's London Branch, 360*(1458), 1271-1279.

D'Souza, R. D., & Burkhalter, A. (2017). A Laminar Organization for Selective Cortico-Cortical Communication. *Frontiers in Neuroanatomy*.

Dann, B., Michaels, J. A., & Schaffelhofer, S. (2016). Uniting functional network topology and oscillations in the fronto-parietal single unit network of behaving primates. *eLife*. doi:10.7554/eLife.15719

Das, D. R. (2017). Thalamus. *Thalamus*. Retrieved 04 12, 2017, from https://www.slideshare.net/RanadhiDas1/thalamusanatomyphysiologyapplied-aspects

Deschenes, M., Timofeeva, E., Lavallee, P., & Dufresne, C. (2005). The vibrissal system as a model of thalamic operations. *Progress in Brain Research, 149*.

Deschenes, M., Veinante, P., & Zhang, Z. (1998). The organization of corticothalamic projections: reciprocity versus parity. *Brain Research Review*(28), 286-308.

Diogo, A. C., Soares, J. G., Albright, T. D., & and Gattass, R. (2002). Two-dimensional map of direction selectivity in cortical visual area MT of Cebus monkey. *Anais da Academia Brasileira de Ciências, 74*, 463–476.

Douglas, J. R., & Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annual Review of Neuroscience*(27), 419-451.

Douglas, R. J., & Martin, K. A. (1991). A canonical microcircuit for neocortex. *Neural Computation*, 480-488.

Douglas, R. J., & Martin, K. A. (1991). A functional microcircuit for cat visual cortex. *Journal of Physiology 440*, 735-769.

Douglas, R. J., & Martin, K. A. (2007). Mapping the Matrix: The Ways of Neocortex. *Neuron, 56*.

Eliasmith, C. (2003). Neural Engineering: Unraveling The Complexities Of Neural Systems. *IEEE Canadian Review - Spring*.

Eliasmith, C. (2005). A unified approach to building and controlling spiking attractor networks. *Neural Computation*.

Eliasmith, C. (2007). Computational Neuroscience. In D. M. Gabbey, & P. Thagard (Ed.), *Philosophy of Psychology and Cognitive Science. Handbook of Philosophy of Science (Vol. 4).* Amsterdam: Elsevier.

Eliasmith, C. (2013). How to build a brain: A neural architecture for biological cognition.

Eliasmith, C. (2013). Neural Engineering: Unraveling The Complexities Of Neural Systems. *IEEE Canadian Review*.

Eliasmith, C., & Anderson, C. H. (2003). Neural engineering: Computation, Representation, and Dynamics in Neurobiological Systems. 350.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2013). A large-scale model of the functioning brain. *Science*, 1202-1205. doi:10.1126/science.1225266

Favorov, O. V., & Kelly, D. G. (1994). Minicolumnar organization within somatosensory cortical regions. *Cerebral Cortex, 4*, 408-427.

Felleman, D. J., & Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex, 1*(1), 1047-3211.

Fiston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication throuhg neuronal coherence. *Trends in Cognitive Science*, 414-480. doi:10.1016/j.tics.2005.08.011

Fukushima, K. (1980). Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics, 36*, 193–202.

George, D., & Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. *PLoS Computational Biology, 5*(10).

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., . . . Essen, D. C. (2016, July 20). A multi-modal parcellation of human cerebral cortex. *Nature*(536), 171-178.

Granger, R. (2006). Engines of the brain: The computational instruction set of human cognition. *AI Magazine, 27*, 15-32.

Granger, R. (2011). How Brains Are Built: Principles of Computational Neuroscience. *Cerebrum*.

Granger, R., & Lynch, G. (1991). Higher olfactory processes: perceptual learning and memory. *Current Opinion in Neurobiology, 1*, 209-214.

Granger, R., Kilborn, K., Rodriguez, A., & Cabanne, C. (2017). Evidence that categorization and recognition are performed by a single thalamocortical mechanism. *In submission*.

Greenstein, B., & Greenstein, A. (2000). *Color Atlas of Neuroscience.* Stuttgart: Georg Thieme Verlag.

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*.

Gross, C. G. (2000). Coding for visual categories in the human brain. *Nature Neuroscience*.

Grossberg, S. (2008). Spikes, synchrony, and attentive learning by laminar thalamocortical circuits. *Brain Research, 1218*, 278-312.

Grossberg, S. (2014). *Adaptive Resonance Theory*. Retrieved from Scholarpedia: http://www.scholarpedia.org/article/Adaptive_resonance_theory

Guillery, R. W. (2006). Anatomical pathways that link perception and action. In V. A. Casagrande, R. W. Guillery, & S. M. Sherman (Eds.), *Cortical Function: A View from the Thalamus.* Amsterdam, The Netherlands: Elsevier.

Guillery., R. W., & Sherman, S. M. (2011). Branched thalamic afferents: what are the messages that they relay to the cortex? *Brain Research Review*(66), 205-219.

Hawkins, J., & Blakeslee, S. (2004). *On Intelligence.* New York: Times Books.

Hawkins, J., Ahmad, S., & Cui, Y. (2017). *Why Does the Neocortex Have Layers and Columns, A Theory of Learning the 3D Structure of the World.* Redwood City, CA, USA: Numenta Inc.

Hebb, D. O. (1949). *The Organization of Behavior.* New York: Wiley and Sons.

Hermann, P. A., Lundqvist, M., & Lansner, A. (2013). *Nested theta to gamma oscillations and precise spatiotemporal firing during memory retrieval in a simulated attractor network.* Retrieved from www.sciencedirect.com.

Hinton, G. E., McLleland, J. L., & Rumelhart, D. E. (1986). Distributed representations. (D. E. Rumelhart, & J. L. McClelland, Eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, 1*, 77-109.

Horton, J. C., & Adams, D. L. (2005). The cortical column: a structure without a function. *Philosophical Transactions of the Royal Society Biological Sciences*. Retrieved from http://rstb.royalsocietypublishing.org/content/360/1456/837.long

Hubel, D. H., & Weisel, T. N. (1972). Laminar and Columnar Distribution of Geniculo-cortical Fibers in the Macaque Monkey. *Journal of Comparative Neurology, 146*, 421-450.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and fundamental architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106-154.

Hubel, D., & and Wiesel, T. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *Journal of Comparative Neurology*(158), 267-293.

Hum, B., Steriade, M., & Deschenes, M. (1989). The effects of brainstem peribrachial stimulation on reticular thalamic neurons: the blockage of spindle waves. *Neuroscience, 31*, 1-12.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron*.

Hwang, K. (2016). The human thalamus is an integrative hub for functional brain networks. *bioRxiv*.

Isik, L. (2017, 08 16). *Using neural decoding to study object and action recognition in the human brain*. Retrieved from Center for Brains, Minds and Machines (CBMM): https://cbmm.mit.edu/

Jandel, M. (2009). Thalamic Bursts Mediate Pattern Recognition. *Proceedings of the 4th International IEEE EMBS Conference on Neural Engineering*.

Jones, E. G. (1985). *The Thalamus* (I ed.). New York: Springer Science + Business Media.

Jones, E. G. (1986). Neurotransmitters in the cerebral cortex. *J. of Neurosurgery, 65*, 135-153.

Jones, E. G. (1998). The Core and Matrix of Thalamic Organization. *Neuroscience, 85*(2), 331-345.

Jones, E. G. (2001). The thalamic matrix and thalamocortical synchrony. *TRENDS in Neurosciences*, 595-600.

Jones, E. G. (2002). Thalamic circuitry and thalamocortical synchrony. *Philsophical Transactions of the Royal Society*.

Jones, E., & Burton, H. (1976). Areal differences in the laminar distribution of thalamic afferents in cortical fields of the insular, parietal and temporal regions of primates. *Journal of Computational Neurology, 168*(2), 197-247.

Kachergis, G., Wyatte, D., O'Reilly, R. C., Kleijn, R. d., & Hommel, B. (2014). A continuous-time neural model for sequential action. *The Royal Society Philosophical Transactions in Biology, 369*(1655).

Kalisman, N., Silberberg, G., & Markram, H. (2005). The neocortical microcircuit as a tabula rasa. *Proceedings of the National Academy of Science*.

Kandel, E. R., Schwartz, J. H., & Jessell, T. H. (2000). *Principles of Neural Science* (4th ed.). New York: McGraw-Hill.

Kasthuri, N., Hayworth, K. J., Berger, D. R., Schalek, R. L., Conchello, J. A., Knowles-Barley, S., . . . Morgan, J. L. (2015). Saturated Reconstruction of a Volume of Neocortex. *Cell, 162*, 648-661.

Kerkoerlea, T. v., Selfa, M. W., Dagninoa, B., Gariel-Mathisa, M.-A., & Poorta, J. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proceedings of the National Academy of Sciences, 111*(40), 14332–14341.

Kohonen, T. (1982). Self-organizing formation of topologically correct feature maps. *Biological Cybernetics, 43*, 59-69.

Kriegeskorte, N. (2015). Deep neural networks: A new framework for modeling biological vision. *Annual Review of Vision Science, 1*, 417–446.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105.

Krueger, J. M., Rector, D. M., Roy, S., Van Dongen, H. P., Belenky, G., & Panksepp, J. (2008). Sleep as a fundamental property of neuronal assemblies. *Nature Reviews Neuroscience, 9*, 910–919.

Kwan, K. Y., Šestan, N., & Anton, E. S. (2012). Transcriptional co-regulation of neuronal migration and laminar identity in the neocortex. *Development, 139*, 1535-1546.

L.Feldman, M., & Peters, A. (1974). A study of barrels and pyramidal dendritic clusters in the cerebral cortex. *Brain Research, 77*(1).

Lamb, S., & Contributors. (2017). *LangBrain*. (R. University, Producer) Retrieved 06 15, 2017, from Language and Brain: Neurocognitive Linguistics: http://www.ruf.rice.edu/~lngbrain/main.htm

Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Science, 10*(11), 494-501.

Lecun, Y., Cortes, C., & Burges, C. J. (2017). Retrieved from THE MNIST DATABASE of handwritten digits: http://yann.lecun.com/exdb/mnist/

Lewis, L. D., Voigts, J., Flores, F. J., Schmitt, L. I., Wilson, M. A., & Halassa, M. M. (2015). Thalamic reticular nucleus induces fast and local modulation of arousal state. *eLife*. doi:10.7554/eLife.08760.001

Linsker, R. (1990). Perceptual neural organisation: some approaches based on network models and information theory. *Annual Review of Neuroscience, 13*.

Llinas, R. R., Leznik, E., & Urbano, F. J. (2002). Temporal binding via cortical coincidence detection of specific and nonspecific thalamocortical inputs: A voltage-dependent dye-imaging study in mouse brain slices. *Proceedings of the National Academy of Sciences, 99*(1), 449-454.

Lorente de Nó, R. (1934). Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. *Journal of Psychology and Neurology, 46*, 113–177.

Lorente de No, R. (1938). The Cerebral cortex: architecture, intracortical connections, moto projections. (J. Fulton, Ed.) *Physiology of the Nervous Systems*, 274-301.

Löwel, S., & Singer, W. (1992). Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science, 255*.

Lynch, G., & Granger, R. R. (2008). *Big Brain : The Origins and Future of Human Intelligence.* New York: Palgrave MacMillan.

Maass, W., & Markram, H. (2004). Theory of the Computational Function of Microcircuit Dynamics. In S. Grillner, & A. M. Graybiel (Ed.), *Report of the 93rd Dahlem Workshop on Microcircuits: The Interface between Neurons and Global Brain Function Berlin, April 25–30, 2004* (p. 470). Berlin: The MIT Press.

Mahmoud, D. M. (2017, 05 02). *Anatomy of the diencephalon*. Retrieved from Slideshare: https://www.slideshare.net/DrMohammadMahmoud/anatomy-of-diencephalon

Malekmohammadi, M., Elias, W. J., & Pouratian, N. (2015). Human Thalamus Regulates Cortical Activity via Spatially Specific and Structurally Constrained Phase-Amplitude Coupling. *Cerebral Cortex, 25*, 1618-1628.

Markram, H. (2008). Fixing the location and dimensions of functional neocortical columns. *Frontiers in Life Science*, 132-135.

Martellus, G. H. (1489). *Martellus World Map.* Florence.

Meng, H., Appiah, K., Hunter, A., & Dickinson, P. (2011). FPGA implementation of Naive Bayes classifier for visual object recognition. *ResearchGate*.

Miller, R. (2002). Wheels within Wheels: Circuits for Integration of Neural Assemblies on Small and Large Scales. In A. Schüz, & R. Miller (Eds.), *Conceptual Advances in Brain Research Series* (Cortical Areas: Unity and Diversity ed., Vol. 5, p. 481). Dunedin, New Zealand: Taylor and Francis.

Moruzzi, G., & Magoun, H. W. (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology, 1*(4), 455-73.

Mounstcastle, V. (1997). The columnar organization of the neocortex . *Brain, 120*(4), 701-722.

Mountcastle, V. B. (1957). Modularity and topographic properties of single neuron's of cat somatic sensory cortex. *Journal of Neurophysiology, 20*.

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics, 66*, 241-251.

Neumann, J. V. (1958). *The Computer and the Brain.* Yale University Press.

O'Reilly, R. C., Wyatte, D., & Rohrlich, J. (2014). *Learning Through Time in the Thalamocortical Loops.* arXiv.

Oja, E. (1982). Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology, 15*(2), 267-273.

Olshausen, B. (2013). *20 Years of Computational Neuroscience.*

O'Reilly, R. (2016). *Biologically-based Error Driven Learning in Thalamocortical Circuits. Presentation.*

O'Reilly, R. C. (1996). *The leabra model of neural interactions and learning in the neocortex. Ph D. Thesis.* Pittsburgh, PA: Carnegie Mellon University.

P.Tinuper, Montagna, P., Medori, R., Cortelli, P., Zucconi, M., Baruzzi, A., & Lugaresi, E. (1989). The thalamus participates in the regulation of the sleep-waking cycle. A clinico-pathological study in fatal familial thalamic degeneration. *Electroencephalography and Clinical Neurophysiology, 73*(2), 117-123.

Peters, A., & Yilmaz, E. (1993). Neuronal Organization in Area 17 of Cat Visual Cortex. In *Cerebral Cortex.* Oxford University Press.

Peters, A., & Yilmaz, E. (1993). Neuronal organization in area 17 of cat visual cortex. *Cerebral Cortex, 3*, 49-68.

Phillips, W. A. (2012). Self-Organized Complexity and Coherent Infomax from the Viewpoint of Jaynes's Probability Theory. *Informaton, 3*, 1-15. doi:10.3390/info3010001

Piantoni, G., Halgren, E., & Cash, S. S. (2016). The Contribution of Thalamocortical Core and Matrix Pathways to Sleep Spindles. *Neural Plasticity, 2016*, 10. doi:doi:10.1155/2016/3024342

Pinault, D. (2004). The thalamic reticular nucleus: structure, function and concept. *Brain Research Reviews, 46*(1).

Plebe, A. (2011). Self-Organization of Object Categories in a Cortical Artificial Model. In J. I. Mwasiagi (Ed.), *Self Organizing Maps - Applications and Novel Algorithm Design* (p. 714). InTech.

Plebe, A., & De La Cruz, V. M. (2016). *Neurosemantics: Neural Processes and the Construction of Linguistic Meaning* (Studies in Brain and Mind ed., Vol. 10). (G. Piccinini, Ed.) Switzerland: Springer.

Prattl, J., Dawson, N., Morris, B. J., Grent-t-Jong, T., Roux, F., & Uhlhaas, P. J. (2016). Thalamo-cortical communication, glutamatergic neurotransmission and neural oscillations: A unique window into the origins of ScZ? *Schizophrenia Research*. doi:10.1016/j.schres.2016.05.013

Quax, S., Jensen, O., & Tiesinga, P. (2017). Top-down control of cortical gamma-band communication via pulvinar induced phase shifts in the alpha rhythm. *PLOS Computational Biology*.

Rakic, P. (1988). Specification ofCerebral Cortical Areas. *Science*, 170-176.

Rakic, P. (1995). A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends in Neuroscience, 18*(9).

Rinkus, R. (1995). TEMECOR: An Associative, Spatiotemporal Pattern Memory for Complex State Sequences. *Proceedings of the 1995 World Congress on Neural Networks.*, 442-448.

Rodriguez, A., & Granger, R. (2016). The grammar of mammalian brain capacity.

Rodriguez, A., Whitson, J., & Granger, R. (2004). Derivation and analysis of basic computational operations of thalamocortical circuits. *Journal of Cognitive Neuroscience, 16*(5), 856-877.

Rolls, E. T. (2016). *Cerebral Cortex: Principles of Operation.* Oxford: Oxford University Press.

Rosch, E., & Lloyd, B. L. (1978). *Principles of Categorization.* Hillsdale, NJ: John Wiles & Sons.

Rouiller, E., & Welker, E. (1991). Morphology of corticothalamic terminals arising from auditory cortex of rat: a Phaseolus vulgaris-leucoagglutinin (PHA-L) tracing study. *Hear Res, 56*, 179-190.

Roy, A. (2017). The Theory of Localist Representation and of a Purely Abstract Cognitive System: The Evidence from Cortical Columns, Category Cells, and Multisensory Neurons. *Frontiers in Psychology*. doi:doi: 10.3389/fpsyg.2017.00186

Rummelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(9), 533-536.

Sanger, T. D. (1989). Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network . *Neural Networks*, 459-473.

Schmitt, L. I., Wimmer, R. D., Nakajima, M., Happ, M., Mofakham, S., & Halassa, M. M. (2017). Thalamic amplification of cortical connectivity sustains attentional control. *Nature*.

Scholarpedia. (2017, 06 30). *http://www.scholarpedia.org/article/Synfire_chains*. Retrieved from Scholarpedia.

Sejnowski, T., & Delbruck, T. (2012, october). The Language of the Brain. *Scientific American*, pp. 54-59.

Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal, 27*, 379-423. 623-656.

Sherman, S. M. (2005). Thalamic Relays and Cortical Functioning. In S. M. Sherman, *Progress in Brain Research* (Vol. 149). Elsevier.

Sherman, S. M. (2016). Thalamus plays a central role in ongoing cortical functioning. *Nature Neuroscience*.

Sherman, S. M., & Guillery, R. W. (1998). On the actions that one nerve cell can have on another: Distinguishing ‘‘drivers’’ from ‘‘modulators’’. *Proceedings of the National Academy of Sciences, 95*, 7121-7126.

Sherman, S. M., & Guillery, R. W. (2001). *Exploring the Thalamus.* San Diego, CA, USA: Academic Press.

Sherman, S. M., & Guillery, R. W. (2006). Exploring the Thalamus and Its Role in Cortical Function. MIT Press.

Sirosh, J., & Mikkulainen, R. (1997). Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Computation*(9), 577-594.

Solari, S. (2009). *PhD Thesis: A unified anatomical theory and computational model of cognitive information processing in the mammalian brain and the introduction of DNA reco codes.* San Diego, CA: University of California at San Diego.

Solari, S. V., & Stoner, R. (2011). Cognitive consilience: primate non-primary neuroanatomical circuits underlying cognition. *Futures in Neuroanatomy, 5*(65), 149-171.

Song, S., Sjostrom, P. J., Reigl, M., Nelson, S., & Chklovskii, D. B. (2005). Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits. *PLoS Biology, 3*(3).

Steriade, M., Dossi, R., Pare, D., & Oakson, G. (1996). Fast oscillations (20 – 40 Hz) in thalamocortical systems and their potentiation by mesopontine cholinergic nuclei in the cat. *Proceedings of the National Academy of Sciences of the USA, 88*, 2533-2538.

Stewart, T. C., Bekolay, T., & Eliasmith, C. (2011). Neural representations of compositional structures: Representing and manipulating vector spaces with spiking neurons. *Connection Science, 2*, 145-153.

Strack, B. (2013). Multi-column, multi-layer computational model of neocortex. (K. J. Cios, Ed.) *PhD Dissertation*, 134.

Swaminathan, S. K., & Freedman, D. J. (2011). Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. *Nature Neuroscience*, 315-320.

Swenson, R. S. (2006). *Revirew of Clinical And Functional Neuroscience*. Retrieved 07 01, 2017, from Dartmouth Medical School: https://www.dartmouth.edu/~rswenson/NeuroSci/chapter_10.html

Thomson, A. M. (2010, March 31). Neocortical layer 6, a review. (J. DeFelipe, Ed.) *Frontiers in NeuroAnatomy*, 14.

Ulinski, P. S., Jones, E. G., & Peters, A. (1999). *Cerebral Cortex.* New York: Springer Science and Business Media LLC.

Valiant, L. G. (2014). What Must a Global Theory of Cortex Explain? *Current Opinion in Neurobiology*, 15-19.

Van Essen, D. C. (2006). Corticocortical and thalamocortical information flow in the primate visual system. In V. A. Casagrande, R. W. Guillery, & S. M. Sherman (Eds.), *Cortical Function: A View from the Thalamus* (p. 302). Amsterdam, The Netherlands: Elsevier.

Van Essen, D. C., Donahue, C., Dierker, D. L., & Glasser, M. F. (2016). Parcellations and Connectivity Patterns in Human and Macaque Cerebral Cortex. In H. Kennedy, D. C. Essen, & Y. Christen (Eds.), *Micro-, Meso- and Macro-Connectomics of the Brain* (pp. 89-106). New York: Springer. doi:10.1007/978-3-319-27777-6

Voelker, A. R., Crawford, E., & Eliasmith, C. (2014). Learning large-scale heteroassociative memories in spiking neurons. *Unconventional Computation and Natural Computation.* Waterloo: Centre for Theoretical Neuroscience.

Wang, Y., Brzozowska-Prechtl, A., & Karten, H. J. (2010). Laminar and columnar auditory cortex in avian brain. *Proceedings of the National Academy of Science, 107*, 12676–12681.

Ward, L. (2011). The thalamic dynamic core theory of conscious experience. *Concious Cognition, 20*, 464-486.

Wen, H., Shi, J., Chen, W., & Liu, Z. (2017). Deep Residual Network Reveals a Nested Hierarchy of Distributed Cortical Representation for Visual Categorization . *bioRxiv*.

Willis, A. M., Slater, B. J., Gribkova, E. D., & Llano, D. A. (2015). Open-loop organization of thalamic reticular nucleus and dorsal thalamus: a computaitonal model. *Journal of Neurophysiology, 114*, 2353-2367. doi:10.1152/jn.00926.2014

Wimmer, V. C., Bruno, R. M., Kock, C. P., Kuner, T., & Sakmann, B. (2010, 05 07). Dimensions of a Projection Column and Architecture of VPM and POm Axons in Rat. *Cerebral Cortex, 20*(10), 2265-2276.

Yamins, D., Hong, H., Cadieu, C., Solomon, E., Seibert, D., & DiCarlo, J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 8619–8624.

Zikopoulos, B., & Barbas, H. (2006). Prefrontal Projections to the Thalamic Reticular Nucleus form a Unique Circuit for Attentional Mechanisms. *The Journal of Neuroscience*, 7348 –7361.

Zikopoulos, B., & Barbas, H. (2006). Prefrontal Projections to the Thalamic Reticular Nucleus form a Unique Circuit for Attentional Mechanisms. *The Journal of Neuroscience*, 7348-7361.

Zylberberg, A., Slezak, D. F., Roelfsema, P. R., Dehaene, S., & Sigman, M. (2010). The Brain's Router: A Cortical Network Model of Serial Processing in the Primate Brain. *PLOS Computational Biology, 6*(4).

Appendix A: Code & Parameters

Source Code for Thesis Work

All the software for the models in this work can be obtained from:

https://github.com/ctn-waterloo/nengo_cortical

Simulation Parameters

gMNISTContinueWithNextImage = True	if true and running MNIST files, the next MNIST image after the last one processed will be used, so the run will continue, this is used with the useSameWeights and CustomWeights solver to restart runs in case they crash
gProbeSampleEvery = 0.010	default=0.001; number of ms to sample all probe data, simultion runs with dt=0.001 seconds usually, but no need to sample that often for plotting results
gMNIST_ImageScaletoNegOne2One = False	default=False; Controls whether the MNIST images (which are encoded as 0:1 pixel values (white=0, black=1) are changed to go from -1:1 (-1=white, 1=black) scaling to better match the receptive range of the image and feature detector ensembles
gImageStorageRadius = 1.2	default=2 in various places such as AccumulatingMemory and FeatureDetectors; controls the radius of the ensembles storing the images. images are very high dimensional ensembles and unless we use really large neuron counts (liek 40,000 neurons to store an image of 10x10 pixels) the image data is largely -1's and 1's (when gMNIST_ImageScaletoNegOne2One is True, 0's and 1's if false) for which saturates the ensembles at either end of the activation curve, the result is that with a radius of 1 (the nengo default) you end up with values closer to 0.7 and -0.7 (for -1:1 encoding) to compensate, making the radius =2 gives enough encoding room before saturation to allow the decoded values to be reasonably close to -1 and 1 even for ensembles with only 4,000 neurons (10% of the ideal).

gUseDerivativeAccumMemory = True	default=True; see CorticalCoumn.py ; controls whether to use the derivative method (deriv ~= 0) for determing when FDDeltaX is stable enough to consider this level of FD wins as succesful and ended if not derivative method then a PATH average value method is used. Note derivative method is far more biologically plausible.
gInputGatedMemoryDifference Gain = 5	default=15, in AccumulatingMemory.py; controls the gain from the accumulating memory to the memory for changes (note Xuan had recommended 15 as the value here)
gUseRandomTransformWeights OnFDs = True	default=True : controls whether FDs are initialized with a random transform vector on the connection from the FD to the Product ensemble note: in Nengo weights = encoders * transforms * decoders; so setting the transforms to be random initially means the feature detectors will start with random weights from the image to the feature detector ensemble this is what we want as we want each feature detector to start as a random weighted filter on the incoming stimuli image representing the random initial state of the cortical connections. The competitive process then selects a winner and that winner learns thier weights a little closer to the incomign stimuli, again as in cortex Flow: if gUseRandomTransformWeightsOnFDs=True then random weights will be used; if gUseSameWeights = True then the same ones from last one are used, else new ones (see CorticalFeatureDetector); causes the weights to be used as the transform, guseSameWeights setting determines whether the same weights are used, this switch determines whether they are used (same or new)
gFDRecurrence = False	Whether or not to use recurrent connection in the FD's to cause them to hold thier value over time
gFDReccurenceStrength = 1	the strength of the recurrent connection
gUseSameWeights = False	these are the weights on the transform on the connection from the image to the feature detectors, works with gUseRandomTransformWeightsOnFDs: if gUseRandomTransformWeightsOnFDs is true then transform weights are used, if

	gUseSameWeights is true then they are the same ones from last run. These serve to increase the FD variation at the start over which the features are learned.
gPESLearningOn = True	default=True; if False feature detectors do not learn via hPES weights learning
gPESInputSynapse = 0.005	default = 0.005; in CorticalFeatureDetector; synapse time constant on the inputs to the product network
gVOJALearningOn = True	default=True; if False feature detectors do not learn via VOJA encoders learning
gDefaultLearningRate = 5e-5	default = 1e-3; controls the rate of learning on the hPES connections. Smaller means FD's learn slower. Too large (greater than 1e-3 or so) and the circuit never stablizes long enough for the accumulating memory to see a stable FDDeltaX and so never causes it to subtract the explained parts of the image; and the circuit never progresses just oscillates in the FD competition process too small and it takes a long time to declare a good FD winner
FDEncodersAsImageTest = False	default=False; will set Feature Detector encoders = parts of the image (if 3 then in 3 parts) for testing, chops up image into FDs pieces and sets FD's encoders to those pieces
EncodersToMatch = []	index of which encoder to match directly to the image for debugging only
gReloadPreviousEncoders = True	default=True; if True encoders from last run are reloaded, if file not found then the gRandomEncodersType encoders are generated Used with gMNISTContinueWithNextImage to restart a run usually because a large run has crashed
gUseCustomEncoders = True	default=true; Whether to use custom encoders (chosen from list below) on featuredecoders objects or let nengo solve for them to start with, see CorticalFeatureDetectorsCollection.py This must be set to True for the gRandomEncodersType setting

	to be used, otherwise Nengo creates the encoders as default
gRandomEncodersType = "TrainingImageSamples"	default= NegOneOne. Which form of encoder initial value to use; "ZeroOrOne" = 0 or 1 integers; "ZeroToOne" = 0...1 floats; "NegOneOne" = -1..1 floats "NegOneOne" = -1 to 1 range encoders; note -1 on an encoder means that dimension doesn't matter "ZeroToOneRandom" = 0 to 1 range random encoders; "UsePrevious" = use the ones saved from the last run; "Ones" = all ones; "TrainingImageSamples" = bootstraps the encoders on the feature detectors by setting them equal to a different starter image to speed training, note careful with the choice of solver, a random solver will negate much of the effects of this possibly; "Hypershpere" or anything else = default setting of nengo.dists.UniformHypersphere(surface=True).sample(n=self.p_NeuronsPerFD, d=self.imageDim). Note: If using VOJA learning start with Hypersphere
gDefaultEncoderType = "ZeroToOneRandom"	default = "ZeroToOneRandom"; Only supported one currently. the type of encoder to use if the first type has no more options, such as when TrainingImageSamples is used but there are only 10 samples but we need to build 30 featuredetectors. Similar in theory to waht some say is the frist step in neural learning is to mirror the inputs, after which neural learning diverges to what the the network learns. This is intentional as the first classifications benefit from starting with known samples then the second and subsequent category encoders are best to start random (if using one of the random encoder types above for the default)
gUseSolverType = "RandomRange"	in CorticalFeatureDetector.py ;default="None"; one of : default="LstSq" = Nengo to use the Nengo default which is lstsq2 (least squares) in corticalFD.py; Zeros : start all feature detectors with all zeros; "Random" : random values in the range from -1 to 1; "RandomRange" : start with random weights in the range specified in the call to the constructor of the Solver; note weights are very small normally and should be from a negative to a

	positive value. Some real weights from a real feature detector with 1000 neurons are min=-0.00021212544 and the max=0.0001728448 so good defaults would be min=-0.001 and the max=0.001; "LoadCustomWeightsSolver" = loads the previous run's weights or uses RandomRange if the files are not found ; "RandomizedSVD" solverObj=nengo.solvers.LstsqL2(solver=nengo.utils.least_squares_solvers.RandomizedSVD()) ; faster for large ensembles
gFDInhibitionOn = True	default = True; controls whether FD's are be turned off from the thalamus output signal by inhibition; see CorticalFeatureDetector.py
gFDInhibitionSignalCutoff = 0.8	default - 0.6, in corticalfeaturedetector.py; dotProdOut function below this level the inhib signal is considered off, above this the FD is inhibited
gUseDelayedFDThalInhibGate = False	default=False; see FeatureDetector.py ; delays propagation of error signal in FDiDeltaX until the output from the Thalamus is stable. Note this could have a negative effect in that it can allow learning to go on until ALL FD's have learned all the features and thus are all the same can stop the competition of the FDs in this way that the model needs to allow FDs to specialize.
gInhibitLosersOnDeltaXBuild = True	default = True; if True inhibits the output of the FDs to the FDDeltaX build process; so the losers output does not propagate to DeltaX otherwise Losers would be shown as explained output. False setting is only used for debugging to see the full flow of FDiDeltaX output reflected in DeltaX.
gWinningFDSignalDuration = 40	default = 20 ; in InhibitFlipSwitch.py; number of simulation dt's (typically 1 ms each) before declaring a winning FD, the time for the network to settle.
gLateralInhibitionOn = True	In CorticalFeatureDetectorsCollection.py; controls whether or not the pyramidal layer IV neurons exhibit lateral inhibition to each other Lateral inhibition is the creation of a series of inhibitory connections between the FD's
gLateralInhibStrength = -0.5	default= -0.20 = -20 % * square of activity (0:1 range) lateral inhibition max; in CorticalFeatureDetectorCollection.py; the strength

	of the lateral inhibitory connections in the range of -100% to 0%; normally around -0.10. Biologically this inhibition is not the stellate cells inhibition in layer IV, this is an analogue for the network competition for the resulting final activated neural network selected by network competition to represent the inputs. The incoming signal which is either neural or numpy dot-product or the scaled activity sum is a value in the range -1 to 2, typically 0.5 to 1.2 it is squared and then scaled by gLaterIInhibStrength factor to yield an input to inhibit the neurons of the dot-product and therefore lower that FD's score in the competition to the basal ganglia. Transform Formula = function=lambda x: [[x**2 * gLaterIInhibStrength]]*self.neuronsPerEnsembleArray)
gLateralInhibStrengthTimeConstant = 0.020	default = 0.020; in CorticalFeatureDetectorsCollection.py; controls the TC for the lateral inhibition time constant
gPyramidalCompetitionType = "Activity"	default = "Activity", "NumpyDotProduct", "Neural" Can be "Neural" = in CorticalFeatureDetectors.py; controls whether the dot-product input to the basal ganglia is computed using the neural dot-product or whether a numpy.dot() function in a node is used "NumpyDotProduct". Or "Activity" based which is the sum total of the activity of the neurons
gActivityScaleFactor = 100	default = 100; the factor by which the summed activity of all the neurons in an ensemble is scaled-down by before being input to the Basal Ganglia using this formula: 1/(gActivityScaleFactor*self.p_NeuronsPerFD)
gAccumulateOnDerivativeStableMemoryOutputOnlyPositive = False	Default=False; in AccumulateOnDerivativeStableMemory object, sets the output to only respond to positive signal, means that if the feature detectors over subtract the explained features, neural activity will go to zero, this is biologically plausible and is what the Granger model says the need for this comes from the instablity in the BG outputs due to the changing signal levels from the input dot-products which don't stay within the range of the BG model (0:1.2) and bounce around up and down, this causes the BG

	then thalamus signal to oscillate and that causes the accumulating memory to subtract the explained features sometimes more than once. this is a very hard problem to solve and due the variability in the input images has been left as a future improvement and this setting caps the explained-away features at zero. So once the image has been explained-away to zero, the stimulus stops (ie: the image is gone all white, the background color) and all feature detection stops. This issue with the implementation of the dot-product, attempting to model the lateral inhibition found in cortex, is perhaps the major short-coming of the model and the cause of much of its trial to trial instability. After many attempts to resolve it and after discussing with the creator of the basal ganglia model, it remains outside the scope of the BG model to auto-scale the inputs (as perhaps it should be, being better solved on the inputs) but also it has proven a large project to try to auto-scale the inputs. There are several other possible alternatives but they are all either separate masters level problems in their own right and so this filter has been applied instead to allow this model to proceed to write-up. If this provokes interest, please contact me for further details and discussion. A BG model with auto-scaled dynamics, and perhaps many nengo inputs with auto-gain control would be extremely useful and a great Masters project and one which from my reading is very biolotgically plausible, I think auto-gain control is practically a core-feature of neurons and neural circuits in addition to selective gain control (which many think is behind attention).
gBGAutoInputBias = False	default = False ;; in CorticalColumn.py; whether or not to compute automatically an input bias for the basal ganglia, this is an amount to shift up (+) or down (-) the input to the basal ganglia to move the inputs into the range from 0:1
gBGInputBiasShiftFactor = 0	default = 0; in CorticalColumn.py; we need to find a factor that can shift the inputs to the basal ganglia downward to the range 0:1 for best results the BG is fed by the dot-products output of the feature-detectors compared to the incoming-image; these

	vary by the dimensions of the image and the values in the images in test runs on a 0:1 encoded (white:black) image with 5x5 pixels, the formula used was the max difference in the pixels (computed but normally ~= 1 of course) when dot-producted tests show this to be in the range of 0:7 typically. So the formula used is to compute -maxDiffernce * NoDimensionsSquare * gBGAutoInputBiasScaleFactor this has been giving an inputBias of around -5 or so, this causes the dot-product inputs to shift into the range of 0:1.5 approximately and that gives reasonable results for the output of the basalganglia. This scale factor will have to be re-determined for each image size and encoding method. Also note, when you set this shift factor, it has to be set in concert with the gproductRangeScaleFactor below. If the inputs are already scaled to be in the 0:1 range then do not shift the range again. If however the inputs are scaled right (all in a tight range of approx = 1 from min to max values) then use this to shift them down to end in the 0:1 absolute value range
gproductRangeScaleFactor = 1	DmensionsSquareRoot in the input image; see CorticalColumn.py; controls the scaling of the dot-product value computed by the product ensembles; the values from productEnsembles can grow as the sum of the number of dimensions of the input; need to scale it to range 0 to 1 for best results as input to the basal ganglia
gDOTProductInputMagnitude = 1	default=1; in CorticalFeatureDetector; this is the expected radius of the inputs to the product network. the coding of the image as 0:1, white:black means the radius should be 1
gDOTProductLookbackWindow = 20	default = 100 ; in CorticalFeatureDetector, number of ms to average the dot product signal over to stablize it. note this should be less or same as the reset window length so it clear before the reset is over
gBasalGangliaInputPower = 1	default=1; see FeatureDetector.py ; dot-product input to the basal ganglia is raised to this power for

	the neural method only; seperates out the inputs better. Inputs should be in the range 0 to 1 to start
gThalamus_Mutual_Inhibition = 0.8	default = 1; in CorticalColumn.py, sets the strength with which the Thalamus forces the inputs to inhibit each other, 1 means that only 1 winner will be chosen; less than 1 means more than 1 winner likely
gFDdotProdWinLooseCutOff = 0.2	default = 0.2; In CorticalColumn; this is the scaled dot-product value that determines the cut-off to determine if a FD "wins" based on the output of the thalamus ie: wins the competition to explain the inputs. Higher values mean more win and lower values less win. typically somewhere between 0.2 and 0.6, but changes by the model or other factors will make this set point move around See output graphs of the dotProduct and the Thalamus.output to determine the range of inputs to this cutoff
gDotProdtoBGSynapseTC = 0.010	default = 0.010; in CorticalFeatureDetector.py ; the time constant for integrating the dot-product signal over before feeding into the BasalGanglia
gBGtoThalamusSynapseTC = 0.010	default = 0.020; in CorticalFeatureDetector.py ; the time constant for integrating the BG signal over before feeding into the thalamus
gThalamus_threshold = 0.25	default = 0.25; the lowest input signal to consider, range from 0. to 1.
gThalFiltWinSignalValue = 1.0	default = 1; this is the value of the signal out of the cleaned up thalamus sent if a FD wins (or looses in case of the losers signal); normally 1 but when the signal bounces around 1 it can cause instability in the error and FD DeltaX signals which cause problems for stablity of the AccumMemory signal a value of 1.1 or 1.2 may stabilize the signal so it oscillates gently above 0 and therefore not inhibiting the learning & FD Delta ensemble
gThalFiltLosersSignalValue = 1.0	default = 1; Same use as above, except on the losing signals. As well note, the losing signal is transformed by -1 and used to inhibit the VOJA learning, this signal cannot go lower than -1 or the VOJA ensemble will start forgetting (negative learning) so only use values very close to 1 here.

gThalFiltOutputSynapse = 0.020	default = 0.020; in CorticalColumn.py : synapse value to be put on the Output added to filter and smooth the Thalamus output; 0.005 is the nengo default, 0.050 would be the maximum value; 0.010 seems to work well
gStartLearningAfter=0.100	default=0.100; how many seconds to wait before turning on learning from the start of the simulation
gStopLearningAfter = 1	default = 0.80; percentage of the run length of images to use as training; after this percentage then the images are processed without learning
gFlipLearningEvery=0.200	default=None; whether to turn learning off after it starts after some number of seconds of learning, reduces over-fitting note; this setting and gInhibitFlipSwitchLookBackWindow are related; the learning interval must be shorter (by at least 50 ms perhaps) than the gInhibitFlipSwitchLookBackWindow as during training an FD needs to win then be allowed to learn, then the signal has to stabilize long enough for the derivative to stay stable so that the imgae subtraction will occur in AccumulatingMemory
gInhibitFlipSwitchLookBackWindow = 300	default=gFlipLearningEvery+100 in ms though; in FeatureDetector.py in number of ms (not seconds!) for the feature detector change signal to propagate through before the winning feature detector is inhibited so the thalamus outputs the selected winning FD (the value of the thalamus output on that dimension goes to 1); then this is fed into the IhibitFlipSwitch for that FD; this then is delayed and after gInhibitFlipSwitchLookBackWindow number of ms the signal is allowed out (it is rounded to 1 or 0 but not averaged, just delayed) this signal then drives the inhibition of the activity in the FD product ensemble which feeds the basal ganglia, in effect turning off this FD from competing from then on, until the reset signal on the change of image is received. During the period of the inhibition signals delay, the learning can occur (and must be shorter than the delay period)
gRunSimgleImageMode = False	default=True, in granger.py, controls if true images are run one at a time and the simulator does a

	nengo.network.reset() between each, otherwise images are run one after another using the same state of the simulator
gDisableResetSignal = False	gDefaultLearningRate=True; in saccades.py, if true the reset between images is off, this is used only when running in single image mode
gInhibResetSignalMinLength = 0.100	default=0.100; number of seconds to hold reset signal high at end of image transitions, allows time for signal to propagate through net
gStartResetLeadTime = 0.050	default=0.050; number of sec to signal reset before it happens to start the process of shutting learning off so no learning happens during the reset
rInhibResetSignalThreshold = 0.90	default=0.90; in InhibiFlipSwitch; above this the reset signal passes through
gThalamusVOJATickLearningHurdleRate = 0.5	the signal level that is the cut off for learning being turned on
gVOJALearningRate = 1e-3	default 1e-3: the rate at which the encoders are learned
gThalamusPESTickLearningHurdleRate = 0.5	the signal level that is the cut off for learning being turned on
gVOJALearnInhibSignalValue = 1	default 1; the absolute value used to send to the hPES learning connection (after transformed to a negative to shut off learning)
gPESLearnInhibSignalValue = 1	default 1; the absolute value used to send to the VOJA learning connection (after transformed to a negative to shut off learning). Note: VOJA implements "forgetting" on the inhibit signal, so the value should not be less than 0
gLearninghInhibInputSignalHurdleValue = 0.2	default=0.3; in LearningSwitch.py; controls what is the level of output from the ThalamusFiltered.LosersOnOutput[self.pFDid] signal which goes to ~1 for a given dimension of it output, each corresponding to a feature detector, when that FD has LOST the competition the value from the thalamus, despite some filtering, does move around 1 a bit so a number approx. = 1 works
gAverageLearningSignalWindowLength = 10	default=1 in LearningSwitch.py; length of window in ms for signal averaging for learning switch determination

gLearnInhibResetThreshold = 0.90	default = 0.90; in LearningSwitch.py; when the reset signal goes above this value the learning signal inhibition is reset; so the feature detector being inhibitied is released
gLearningSignalResetTime = 0.100	default = 100, in LearningSwitch.py; number of seconds to hold the reset line high once triggered between image changes, will inhibit learning in this period
gUseVOJAIntercepts = False	default intercepts on an ensemble are set to nengo.dists.Uniform(-1.0, 1.0) by nengo, for VOJA learning the encoders move to learn the keys but must space themselves out according to the max dot-product of the pairs of keys being used; if VOJA learning is ON and this is ON then the setting below is used as the min-intercept = max dot-product between keys
gVOJAIntercept = 7	the VOJA encoder ensemble's neurons need to be set to not respond below the maximum dot-product between the images see https://pythonhosted.org/nengo/examples/learn_associations.html
gLearnSignalThreshold = 0.7	if the learning inhibition signal is over this level, round up to 1 or floor to 0 below it
gResetLearningSignalThreshold = 0.05	default=0.05; in DetectDerivativeChange.py; if the s_change_detect signal (DetectDerivativeChangeGate.s_change_detect.output) goes below this level, the LearningSwitch.reset() function is called this is to be called when the feature detectors subtract away the portion of the incoming signal that has been "explained" away
gTurnLearningOnThreshold = 0.8	default=0.80 ; in learningSwitch.py; when the learning signal goes above this level, learning is turned on
gAreaIntercepts = True	default=True ; in CorticalFeatureDetector.py; controls whether or not the AreaIntercepts distribution is used on the image ensemble in CorticalFeatureDetectors and the featureDetector ensembles as well to improve performance by allowing for the use of less neurons but with the same accuracy. Note: see also the setting for the

	solver above; the RandomizedSVD is a sub-solver to the standard LstSqrs one which is faster than LstSqrs, but if using one of the other solvers that just assigns the weights then this likely isn't a help. Also using gAreaIntercepts is only useful is solving for decoders and not using Random weights