The Initiative for Dynamic Link Neurocomputing

General Objectives and Organization of the Program

Prepared By

Richard B. Wells

University of Idaho

MRC Institute, Laboratory for Computational Neuroscience & Technology Research

Abstract

This white paper describes the general objectives and organization for foundational research in computer architecture derived from models of the dynamical organization of cortical and non-cortical bio-circuits in the central nervous system (CNS). Our principal General Objectives are: 1) to understand how this dynamical organization in brain structure can be viewed in a computational context; 2) to discover what implications dynamical organization holds for information processing; and 3) to develop a new class of computing devices organized and designed to take advantage of these findings. Within the framework of these General Objectives there are Discipline Objectives in support of these General Objectives. This research will be carried out by an interdisciplinary team composed of researchers drawn from the disciplines of microelectronics, computer engineering, computer science, neurobiology, and computational neuroscience. The discoveries and findings from this research are expected to make fundamental contributions to the art and science of computing technology as well as to our basic knowledge of brain organization and the biological substrates of intelligence. It is expected that the experimental and theoretical findings of this research will profoundly affect the future course of computer design and computer science, and will establish a new break-through paradigm leading to more truly intelligent and robust computing devices. It is not improbable that the foundations laid by this research will lead to an entirely new industry of biomimetic neurocomputer systems based on dynamic link architectures.

I. Introduction

Whatever one's philosophical position may be on the degree to which the brain can or should be regarded as some type of computer, there is no reasonable doubt that the brain and the computer can both be regarded as information processing systems. The modern digital computer can trace its origins back to the pioneering work of McCulloch and Pitts in 1943[1], and its development is largely credited to the work of John von Neumann, in whose hands the McCulloch-Pitts neuron model became the basic logic circuits used today in all digital computers[2]. In its most common forms, the digital computer is a device that carries out sequential logical and arithmetic operations on data operands. The

sequential nature of its operations limits the speed with which results can be obtained and is the mathematically fundamental limiting factor in its performance.

Numerous computer architectures have been proposed over the years with the aim of overcoming the fundamental speed limitations imposed by sequential processing, and for overcoming the fundamental limitations on machine reliability imposed by this same sequential information processing architecture. These architectures are generically called parallel computer architectures and fall into three general classes: single-instruction/multiple data (SIMD) architectures; multiple-instruction/multiple data (MIMD) architectures; and parallel associative arrays (PAA). Classic examples of these different classes include ILLIAC-IV (SIMD)[3], the Burroughs D825 (MIMD)[4], and the Goodyear Aerospace STARAN (PAA)[5]. There have been many decades of work carried out on the issues, factors, and problems associated with parallel computing[6], and this work is still on-going today.

From their beginning in the late 1950s and early 1960s, neural networks have been regarded as one possible approach to uncovering fundamental principles of parallel computing and to realizing more of the potential that has long been believed to lie within the framework of parallel computing. But even today it is argued by many that there are theoretical questions concerning the mathematical foundations of parallel computing still unanswered[7], and it is strongly argued by some that the current literature on neural networks and neurocomputing contains a wealth of experimental results but few theoretical results regarding the main mathematical issues of concern to parallel computing[8]. Artificial neural networks (ANNs) have long demonstrated the capacity to solve certain classes of problems that prove to be difficult to solve by ordinary computer methods. They have likewise long demonstrated an impressive degree of fault tolerance. This includes both being able to deal with problems similar to yet different than those for which they have been trained, and being able to tolerate damage to the ANN system itself. In addition, ANNs possess the remarkable capacity for self-optimization, a capacity that can be loosely characterized as "the ability to program themselves."

On the other hand, critics of artificial neural networks point out that these capacities have been demonstrated only at the level of relatively small-scale problems, and that there is significant evidence in support of the contention that the small-scale problems where neural networks have proven successful do not translate up to fit large-scale problems. Some of these concerns have been shown experimentally to be misplaced, but others remain unresolved. One of the most critical issues here is the question of whether or not the number of bits of accuracy required of neuronal weights rises linearly with the size of the support set of the problem to be solved. This is potentially a formidable problem because it implies a geometrically increasing computing precision is required as the scale of the problem increases. This "bit explosion" phenomenon was proven as a theorem (the Minsky-Papert stratification theorem) for one particular class of neural network[8], and it was conjectured that the same

result will be obtained for other classes of connectionist ANNs. Findings from recent research conducted at the University of Idaho have demonstrated that the computational precision required as the problem scales up <u>depends on the topology of the network</u>[9]. Although these findings do not necessarily refute the main result of the Minsky-Papert stratification theorem, they do indicate that computational precision requirements are affected by the structure of the neural network system, and that significant improvements in the required precision occur with relatively innocent-looking changes in network topology.

The point that Minsky and Papert raised with their theorem was not that artificial neural networks were incapable of solving complex parallel-processing problems, but rather that a stubborn adherence to a "one-size-fits-all" ANN architecture was short-sighted, and that the particular architecture most commonly used in connectionist neural networks was vulnerable to the same objections they raised against earlier perceptron networks. Our recent findings are not in disagreement with their position on this. Furthermore, it has long been known that biological neural networks do not possess the attribute
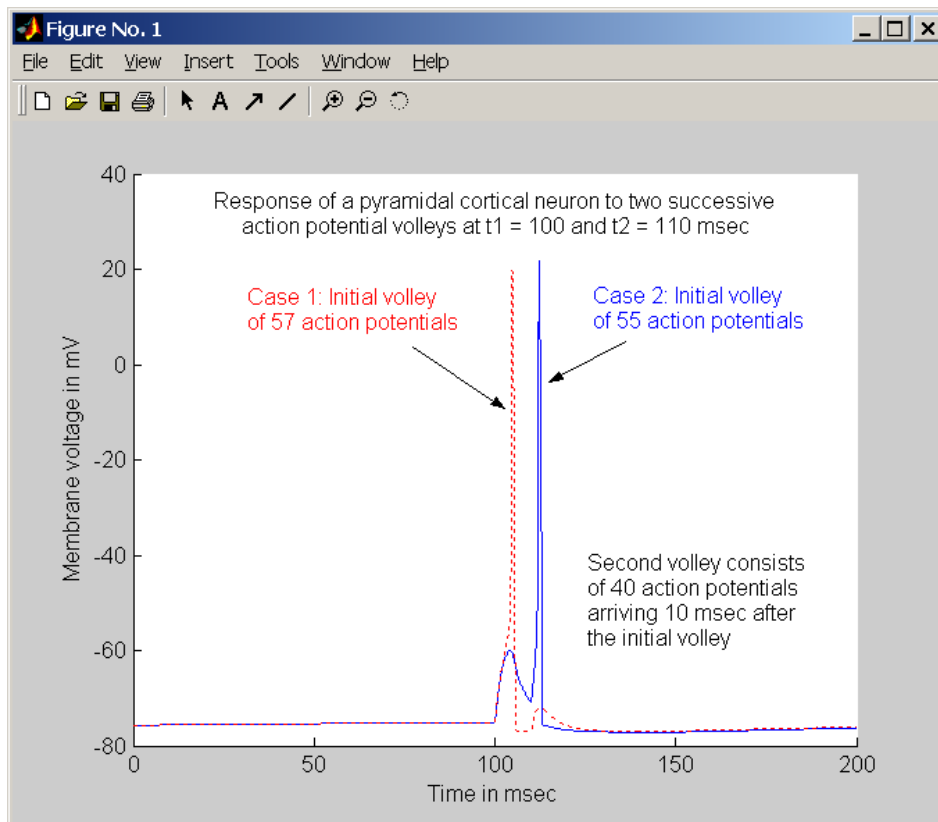

Figure 1: Simulated response of a pyramidal cortical neuron to two similar input stimuli. In case 1 the initial action potential volley arriving at t = 100 msec consists of 57 APs. In case 2 the initial volley consists of 55 APs. This is well within the typical range of variations in the CNS. The second volley in both cases is 40 APs, which in the absence of the first volley is insufficient to trigger a firing response in the neuron.

of having high computational precision in the biological mechanisms by which they carry out their signal processing and "computational" tasks[2]. Figure 1 illustrates the variability in the response of a typical pyramidal cortical neuron to two cases of stimulation that lie well within the normal range of statistical fluctuations observed in the neocortex. Although this figure shows a pronounced difference in the response of the neuron to the two cases, it is highly unlikely that these two cases could represent any significant difference in the information processing carried out in the cortex. Biological neural networks function quite normally in the face of variations such as that illustrated here, which implies that they are robust to lack of precision in the operational parameters of the network. Because such robustness is not ascribable to the neurons themselves, it must be viewed as a property of the neural network topological organization of the neocortex.

Another computationally significant issue is found when one considers the problem of learning rate in adaptive neural networks. Although no proof yet exists, there is strong reason to believe that Rivest-Shamir-Adleman public-key functions are trapdoor functions, and if they are then it follows as a theorem that training sets for linear-threshold neural networks (a class that describes most existing ANNs) are not polynomial-time learnable in terms of any polynomial-time-computable representation[10]. This is a significant and serious problem because it means that conventional ANNs that work for small-scale problems cannot simply be scaled up to solve large parallel computing problems. Although recently there has been a major advance made in machine learning for ANNs in terms of the scale of problems to which they can be applied[11], this issue of the "computational cost explosion" involved in scaling these networks for application to large problems remains a potentially fatal flaw in the connectionist paradigm. Again, however, when we turn to biological neural networks in living animals the learning process does not appear to be consistent with what would be expected on the basis of this theorem, even though the learning problems posed to living animals fall into the large-scale-problem category. This again implies that there is something in the topological/structural organization of the CNS networks that avoids the learning rate issue facing conventional connectionist networks.

A third significant issue arising from conventional ANN topologies is that of stability in the firing rate patterns of the cortex. Large-scale cortical behavior predicted from statistical models of conventional ANNs indicates that neuronal activity in the neocortex is either bistable, with spontaneous firing rates of cortical neurons either zero or saturation at a high firing rate, or else it is monostable with the network having a stable steady state only at a zero firing rate[12,13,14]. For bistable configurations there is an unstable "ignition point" at relatively low firing rates. Instantaneous firing rates below the ignition point decay to zero firing; those above it go to saturation of the network. This is illustrated in figure 2. Again the theoretical model appears depart from actual cortical behavior as
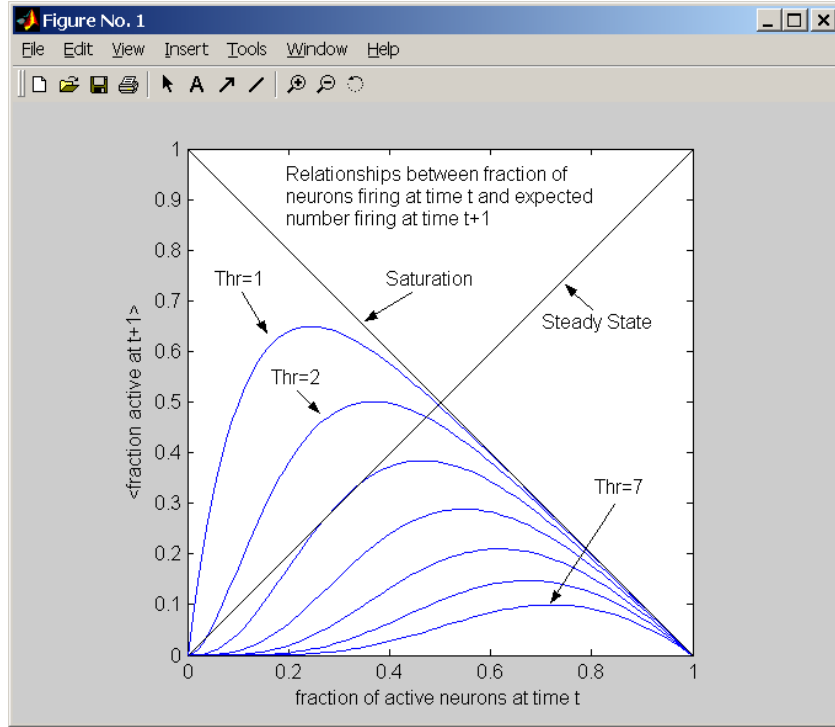
Figure 2: Expected statistical activity of randomly-connected perceptron networks. The figure illustrates the case for networks in which 20% of the neurons are inhibitory, which is approximately the case in the neocortex. The steady state behavior of the network is given by the intersections of the blue curves with the steady state line. Curves depict different firing threshold levels. At both very low and very high thresholds there is only one stable configuration at the zero firing rate. For intermediate levels there are two stable operating points, zero and a level near saturation of the entire cortex.

measured by PET or MRI scans. This is because firing thresholds in cortical neurons appear to be consistent with model conditions under which bistable operation should be expected, whereas in actual brain activity stable near-saturation firing levels are not observed. Neural network topology is again implicated here, since the theoretical results obtained from the topological model do not appear to be consistent with experimentally observed behavior.

II. Dynamic Link Architecture

ANNs are commonly said to be "biologically inspired," but it is usually the case that the accent has been on the "inspired" rather than on the "biologically". The basis of connectionist ANN models is drawn from what is known as the firing rate model of neuronal coding[15]. While it does appear that peripheral nerves do in fact encode information in the form of firing rate, the situation is different in the neocortex. Here the evidence, both theoretical[13,16] and empirical[17,18,19], is consistent with the hypothesis that information is encoded as synchronized firing by groups of cells.

The organization of the neocortex is consistent with this hypothesis. It has long been the prevailing view in neuroscience that the neocortex is organized along the lines of specialized "functional columns" dedicated to specific response processes. While the existence of functional columns appears

to be well established, mapping studies have also brought to light the interesting fact that in at least parts of the cortex, such as the visual cortex, there appear to be more functional columns than the cortex has space[20,21]. One hypothesis set down to explain this apparent paradox is to assume that functional columns are transitory, i.e. that the interconnects among neurons in a given region allows for dynamic re-organization of the cell group in response to stimuli and what can be termed "control signals" originating in other regions of the brain. It is known that some regions of the nervous system, e.g. the pyloric network in the stomatogastric ganglion which controls stomach muscles, are in fact capable of dynamically re-organizing their network connections[22].

In part to explain the apparent transitory nature of functional columns and in part to provide a theoretical foundation for dynamical laws of cortical phenomena, von der Malsburg has proposed the Dynamic Link Architecture (DLA) model[23]. DLA is an outgrowth of the correlation theory of brain function[24]. Mathematically the brain's data structure is regarded as having the form of a graph, and these graphs are dynamic, i.e. both the nodes and links depict activity variables and change on a rapid functional time scale. Mental objects are formed by binding together different nodes representing the different parts of the object. This interpretation of data structure is consistent with Damasio's model of how objects and events are represented in the brain[18,19].

In von der Malsburg's theory, synaptic weights are capable of relatively rapid elastic fluctuations relative to a long-term average value. A number of physiological mechanisms are known to exist that can provide the biological substrate for these modulatory actions[25]. It is known that the neocortex is extensively innervated by metabotropic afferents employing as neurotransmitters (NTXs) norepinephrine (NA), dopamine (DA), serotonin (5-HT) and acetylcholine (ACh). In addition, some inhibitory neurons in the neocortex colocalize various neuropeptides (CCK, NPY, VIP, SP, or somatostatin) with the inhibitory NTX GABA[26]. NA, DA, 5-HT, ACh, and the neuropeptides are all known to stimulate modulatory second-messenger metabotropic reactions capable of modulating the activity and/or sensitivity of neurons to their ionotropic inputs. From their known physiological effects in the pyloric network, it is clear that these modulators are capable of the type of dynamical "re-wiring" of neural connections necessary to account for the transitory functional column hypothesis.

What emerges from this theory is a picture of the cortex as a "network of networks". At the level of organization immediately above that of the individual neuron are dynamically-reconfigurable functional units (often referred to in the literature as "barrels", "columns", "bands", "slabs", and, less frequently, "blobs") of usually vertically-oriented groups of cell assemblies. The assemblies are dynamically reconfigurable by means of control signals within a limited range of overlapping cell groups. Local stability of functional cell assemblies is mediated through interconnections from local
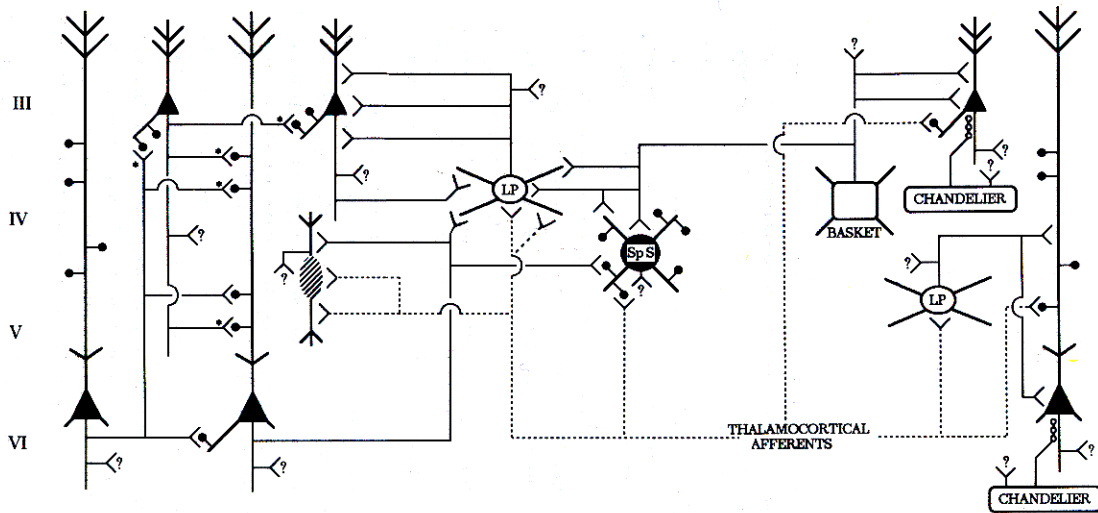
Figure 3: Summary illustration of synaptic connections in the neocortex that have been conclusively demonstrated in anatomical studies. Triangular-shaped symbols represent pyramidal neurons. SpS = spiny stellate cell. The cross-hatched cell is a non-spiny bipolar cell. LP=local plexus cells of unidentified types. Excitatory neurons are shown as solid black symbols, inhibitory interneurons are shown with unfilled (white) symbols. Roman numerals indicate the layer in the neocortex in which these cells are found. In the neocortex excitatory cells outnumber inhibitory ones by about 4:1.

inhibitory interneurons, whereas most excitatory stimulus to the assembly arrives via the white matter from other assemblies[13].

Units constitute first-order networks (FONs) in the network-of-networks architecture. It is known that the neocortex contains at least twelve distinct classes of neurons[27], and some of the circuit pathways have been established through anatomical studies[26], although no complete "wiring diagram" of neocortical circuits has yet been obtained. Figure 3[26] illustrates the synaptic cortical connections that have been conclusively demonstrated. In this figure it is obvious that not all classes of neocortical neurons are represented, and the principal objective of the neurobiological component of our proposed research is to supply more detail about the "wiring diagram" for these missing neuronal elements.

FONs receive numerous inputs from, and project numerous outputs to, other cell assemblies, including sub-cortical assemblies as well as other neocortical FONs. It is known that the great majority of ionotropic input afferents to the neocortex, as well as the output efferents from the pyramidal cells, are excitatory. Inhibitory interconnections appear to be strongly localized. This anatomical arrangement is consistent with the synfire chain model of activity transmission in the neocortex[13,28]. Synfire chains are feed-forward chains of cell assemblies characterized by synchronous firing behavior. A synfire node is an assembly of neurons that receive inputs from a common group of neurons in the previous link in the chain, and transmit outputs to a common group of neurons in the subsequent link in the chain. In our present context, synfire-chain-like structures constitute second-order networks (SONs). The synfire hypothesis was put forward by Abeles to explain two critical

attributes of cortical signal processing: reliability of transmission, and robustness of operation in the face of cell deaths. It has been shown from theoretical considerations that something like a synfire chain is necessary for reliable propagation of information through the neocortex[13,29]. This reliability is achieved through synchronous firing of the neurons in a synfire node, which is known to be a more effective mechanism for stimulating firing responses in neurons in the next link of the chain. In addition, it is known that neuronal cell death proceeds at a relatively constant rate throughout the adult lifetime, and synfire chains are robust in the face of loss of neuronal elements due to cell death that would seriously compromise most other proposed schemes of information transmission.

For many years it was assumed by most neuroscientists that information processing in the CNS was hierarchical. Under this assumption, successive layers of neurons performed successive transformations on incoming data building up toward object and event representation at an output layer in which one or more neurons indicated recognition of that object or event by their firing activity. Such an "output layer" cell has often been called a "grandmother cell." However, over the course of the past two decades this traditional model of neural network processing has been shown to be at odds with clinical findings from studies carried out on patients suffering from various forms of injury- or disease-induced brain damage. These studies clearly indicate that the "grandmother cell" hypothesis is completely inconsistent with the cognitive capabilities found to still exist in brain damaged patients. A summary of this evidence has been provided by Damasio[18]. The disagreement between the traditional model and empirical findings has led to the convergence zone hypothesis of cortical organization[19]. In this hypothesis, outputs from multiple functional units at one level in the neural structure converge on small assemblies of neurons, called convergence zones (CZs). CZs detect synchronized firing activities among diverse groups of functional nodes, and provide feedback signals returning to these layers that tends to maintain synchronized firing patterns from these groups.

Figure 4 illustrates the basic scheme of the convergence zone dynamical architecture as it applies to the sensory pathways. A similar structure is proposed for the motor pathways. Synchronous firing activities among various functional units in the early and intermediate sensory cortices converge on and are detected by CZ assemblies. The CZs provide feedback to the immediately previous layers of functional units, tending to promote and maintain their activities. The principal function of the CZ assembly is to produce binding codes for the representation of entities and events. Entities and events are represented by the suite of synchronous firing activities from the specific functional units involved. CZs also project forward to the next higher level of CZ assemblies. These, in turn, provide feedback projections to the immediately prior CZ layer.

Two types of binding codes are produced by CZs. Type I binding codes produce entity representations through synchronization of firing patterns in the functional units in sensory cortices.
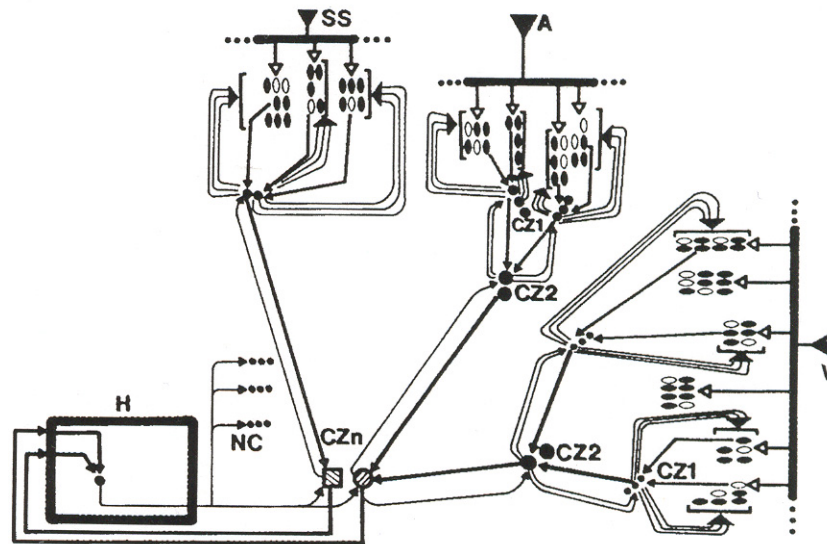
Figure 4: Simplified schematic of convergence zone structure for the perceptual pathways. A similar structure is also proposed for the motor pathways. V, SS, and A depict early and intermediate sensory cortices in visual, somatosensory, and auditory modalities. Separate functional units are depicted by the open and solid ovals. CZ = convergence zone. H = hippocampal system. NC = noncortical neural stations in the basal forebrain, brain stem, and neurotransmitter nuclei. Feedforward pathways (heavy lines) and feedback pathways (light lines) are not rigid channels but, rather, are facilitated pathways that become active when concurrent firing in early cortices or CZs takes place. Furthermore, they terminate over neuronal assemblies in a distributed fashion rather than at specific single neurons.

Type II binding codes produce representations of events unfolding in time by binding sequences of firing patterns, both in the sensory cortices and in the motor cortices. Separation of the pieces of an object, represented by specific cell groups, is called underlinesegmentation. Representation of an entity by putting these pieces back together is called superposition. Both are fundamental issues in the theory of neural networks and lead to one of the most difficult problems in current neural network theory: the binding problem. Von der Malsburg has given a very clear explanation of what the binding problem is all about[59], and dynamic link architecture is his, and Damasio's[18,19], proposal for solving it.

In addition to the direct feedback supplied by the CZ assemblies, late-stage CZs also project to non-cortical neurons in the hippocampal system, basal forebrain, brain stem, and neurotransmitter nuclei. These, in turn, send modulatory signals to various cortical circuits at a variety of levels. These modulatory signals facilitate attention, learning, and other higher cognitive functions. In the network-of-networks topology of DLA the interconnected assemblies at this level constitute third-order networks (TONs).

The features and topological organization of the DLA model of neural structure just described are consistent with the conjectures made by Minsky and Papert concerning what sort of neural network organization would likely be required to overcome the most serious mathematical issues raised by

their theoretical work on perceptron networks[8]. Indeed, it is noteworthy that Minsky's and Papert's conjectures predated the empirical findings that led to both the DLA model and the convergence zone hypothesis, both of which were proposed independently of the Minsky-Papert theory. Furthermore, our own research work has strongly suggested that dynamic links can improve the performance of even classical feedforward neural networks. This finding is based on our observation that complex non-linear optimization problems involving mappings from an input retina to an output vector do often produce one-to-many optimal mappings dependent on both spatial and temporal correlations in the input signal set. Furthermore, our research has shown that this multiplicity of optimal solutions truly does represent a one-to-many set of optimal solutions (the Vongkunghae set) and is not merely an artifact of so-called "local minima" in the performance surface of the network[30,31]. In a classical feedforward network Vongkunghae sets cannot be implemented, and a suboptimal re-mapping is necessary to obtain the one-to-one or many-to-one mapping functions that these classic ANNs are able to implement. DLA networks, on the other hand, are inherently capable of spatio-temporal modulations which retain the optimal one-to-many optimal solution sets.

III. The Research Initiative

General Objectives and Funding of the Proposed Work

The principal General Objectives for this project are: 1) to understand how the dynamical network-of-networks (DLA) organization in brain structure can be viewed in a computational context; 2) to discover what implications dynamical organization holds for information processing; and 3) to develop from this knowledge a new class of computing devices organized and designed to take advantage of these findings. These objectives constitute the top-level engineering outcomes of this research. These objectives can be met by the achievement of the following four general Scientific Objectives: 1) to formulate a canonical graph-theoretic description of dynamic link neurocomputing; 2) to find and implement solutions to the problems of segmentation, binding, and neural coding; 3) to derive from the graph-theoretic expression of neurocomputing a canonical description of data structures, formal grammars, and state-dependent neural coding to be used to connect the theory of dynamic link neurocomputing with the main body of computer theory; and 4) to demonstrate the feasibility of our theory by constructing prototype dynamic link neurocomputer networks in VLSI. How these four objectives come to be the principal scientific objectives is explained in the discussion of discipline-specific objectives described below.

This research is inherently interdisciplinary, involving contributions from neuroanatomy, computational neuroscience, computer science, computer engineering, and microelectronics teams. Team members from each of these disciplines have discipline-specific objectives, stated below, in

support of the three principal General Objectives. Most of these discipline-specific objectives cross-couple with the objectives of the other disciplines, requiring close research collaboration in order to achieve success in this program of work. Externally funded research contracts for multi-disciplinary Specific Research Topics are required in order to carry out this work in support of this Initiative, and the Specific Research Objectives these projects must serve the General or the Scientific objectives of the Initiative.

Because discoveries in each discipline have important implications for the others and vice versa, the work within each discipline is expected to contribute to the following additional program outcomes: A) identification from disciplinary findings of collateral research questions for other disciplines; and B) interdisciplinary collaboration in writing and submitting collateral research proposals to be submitted to National Science Foundation, the National Institutes of Health, or other appropriate external funding agencies. It is expected that these collateral topics for additional research will arise as an outgrowth of team efforts to accomplish the top-level objectives. As an example for how this is envisioned to take place, let us imagine the following situation: Discipline A develops an hypothesis or discovers a question such that specific knowledge from Discipline B must be obtained in order to test the hypothesis or address the question; Discipline B finds itself unable to supply this information from present knowledge, and proposes collateral work for obtaining this knowledge; Disciplines A and B jointly develop an interdisciplinary research proposal for obtaining this new Discipline B knowledge and for applying it in Discipline A. Situations such as this arising from collaborative work produce research synergy, i.e. "the whole is greater than the sum of its parts."

Given the size and scope of this project, an average of two new proposals per pair of disciplines per year (10 proposals per year), at an average per-proposal funding level of $500k over 3 years, is a necessary level of proposal productivity. Achievement of this outcome with a 15% to 30% success level in contract awards is necessary to maintain momentum in the Initiative. Critical mass for on-going Program funding required to maintain continuity in the Team and to make significant progress toward the General Objectives is about 6 funded contracts providing an annual Program budget of about $1 million per year.

Discipline-specific Objectives of the Proposed Work

A. Neuroanatomy Objectives. The specific objectives of the neuroanatomical component of this research are: NA1) to obtain from experimental studies additional information on synaptic interconnections in the neocortex; NA2) to integrate this additional knowledge with existing knowledge to propose possible cortical circuit models for functional column organization within the neocortex; and NA3) to communicate these findings to other members of the team and assist in the

development of cortical circuit and network models. It is expected that this knowledge will help "fill in more of the missing pieces" in our understanding of cortical circuits. This information and the models consistent with these findings are critical to an understanding of stability conditions in the DLA network, as well as for understanding the pathways for the modulatory mechanisms at work in the dynamical re-organization of functional units.

B. Computational Neuroscience Objectives. The specific objectives of the computational neuroscience component of this research are: CNS1) to obtain from quantitative physiological modeling of small cortical networks a quantitative model of the types and distributions of modulatory signaling mechanisms necessary to realize dynamic links; CNS2) to develop new models of networks within the network-of-networks architecture that incorporate these modulatory mechanisms; CNS3) to develop computational models of the information processing implications of cortical circuit models proposed in NA2; CNS4) to communicate findings of issues and questions requiring additional knowledge of neural anatomy to the NA team for guidance of research direction; and CNS5) to analyze these model results to facilitate development of a mathematical theory of computation for parallel neurocomputing by means of DLA structures and communicate these findings to the CS, CE, and ME members of the team.

C. Computer Science Objectives. The specific objectives of the computer science component of this research are: CS1) to develop evolutionary computing algorithms, based on findings from NA1, NA2, CNS1, and CNS2, for the design of FONS, SONS, and TONS structures to serve as models for possible large-scale network configurations; CS2) to propose fitness functions that can serve as mathematical guides and propose questions for additional research work by the NA and CNS disciplines; CS3) to analyze evolved network structure results to facilitate development of a mathematical theory of computation for parallel neurocomputing by means of DLA structures; and CS4) to communicate these findings to the NA, CNS, and CE members of the team.

D. Computer Engineering Objectives. The specific objectives of the computer engineering component of this research are: CE1) on the basis of the functional unit paradigm of cortical organization, to propose possible interpretations of neuronal activity as data structures; CE2) to develop from the findings of CNS2 a mathematical language or other expression of data flow/information processing operations implied by these neuronal structures; CE3) to establish the mathematical properties of binding codes required to produce combinatorial and sequential entity- and event-oriented binding code representations; CE4) to communicate the findings of CE1-3 to the CNS and CS team members

for guidance and direction of their model development work and collaborate in the development of the mathematical theory of neural computation; CE5) to develop and pose parallel computation problems drawn from classical parallel computing tasks that can serve as exemplars and test cases for the mathematical theory developed in this research, and to communicate these test cases to the CS and CNS team members as guidelines for their research; and CE6) to propose a computer architecture for dynamic link neurocomputing seen as a parallel computer (e.g. in terms of SIMD, MIMD, PAA, or some combination thereof or some other form of computer organization as appropriate).

E. Microelectronics Objectives. The specific objectives of the microelectronics component of this research are: ME1) to design, fabricate, and test new VLSI circuit implementations of neural network units proposed by the CNS, CS, and CE members of the team; ME2) to carry out experiments posed by the CS and CE members of the team to determine if the outcomes of the proposed network structures are in agreement or disagreement with expectations based on the mathematical theory being developed; ME3) to communicate detailed findings and results to the other members of the team to provide them with experimental facts arising as consequences of the proposed models and paradigms; ME4) to analyze the feasibility of implementing proposed structures within the tolerance and parametric limitations imposed by existing VLSI technology; and ME5) to determine packaging, power dissipation, performance, and interconnect issues that must be addressed in practical implementations of neurocomputers based on the findings of this research.

Initial Program-Start-Up Methods for Accomplishing Proposed Objectives

Methods for Accomplishing the Neuroanatomical Objectives. At present the largest hole in our knowledge of the organization of the neocortex is in knowing the specific identity of elements postsynaptic to the intrinsic axonal projections of non-pyramidal neurons, especially projections to and among inhibitory interneurons. More knowledge of this organization is essential for determination of both the stability properties of functional columns and the information processing functions carried out within a functional column. It is noted here that obtaining a complete "wiring diagram" of cortical functional columns is an unreasonable expectation at the present state of the art in experimental neuro-anatomy, and such a complete picture is neither required nor expected. However, hypotheses based on experimental data can be used by the CS team members to evolve multiple putative cortical circuits consistent with anatomical findings, and these circuits can be theoretically examined by the CNS team members to determine possible different information-processing implications for the evolved circuits. It is at this NA-CS-CNS juncture where most of the cross-disciplinary collaboration can be expected

to originate. The theoretical models and putative circuits will pose additional anatomical questions, and research findings on these questions will lead to additional modeling work.

There are a number of technical reasons why fuller knowledge of cortical circuitry has been difficult to obtain, and these reasons have been reviewed in Chapter 3 of White[26]. It is widely agreed that the best approach to this problem is attempt to reconstruct the identity of elements through a series of thin sections in which both the pre- and post-synaptic neurons are labeled[32]. Labeling of postsynaptic cells might be accomplished using the well-known method of retrograde HRP (horseradish peroxidase) staining[33]. Presynaptic labeling might be accomplished by anterograde staining in which macromolecules are attached to HRP using immunohistochemical methods[34]. The primary technical difficulty in this method is obtaining sufficient relative isolation of labeled pairs from other labeled elements.

On the other hand, there is a large body of statistical information regarding the distribution of connections made by various non-pyramidal cells to various areas of the postsynaptic neuron[26], as well as on the expression of intracellular molecules in various electrophysiological classes of cortical neurons[27]. Various experimental findings on the expression of these molecular markers[35,36,37,38,39] combined with those of the known statistical distributions can be coupled with our staining experiments to propose <u>possible</u> circuit connectivities. These putative connectivities will then be subjected to analysis by both computational models (CNS team) as well as by evolutionary computing (EC) analysis (CS team) to determine what possible implications these connectivities have for the stability, responsiveness, and functionality of functional columns.

Our experimental studies can be carried out on male Sprague-Dawley rats. Experimental procedures will be done in accordance within the Institutional Animal Care and Use requirements. The whisker barrel cortex is suggested as the object of the study, both because these structures are a known example of functional column organization in the cortex and because the barrels are evident histologically in correspondence to their respective vibrissa. There is a good deal of literature published about the barrel cortex[40,41,42,43,44,45,46,47,48,49,50,51,52,53,54], and these findings are useful in analysis by other team members to determine the computational properties of the barrel. Thick sections (1-2 microns) will be cut from each tissue block using an ultramicrotome (LKB), mounted onto glass slides, stained, and examined with a light microscope to identify barrels in the cerebral cortex and orient to the layering in the cortex. The tissue block will be trimmed and/or re-oriented appropriately and thin sections (600-800 nm) will be cut from them and picked up onto copper grids (150-300 mesh). After cutting thin sections another group of thick sections will be prepared to serve as a guide when examining thin sections. Thin section analysis will be carried out using transmission electron microscopy. The barrel cortex from the right hemisphere will be used for studying the

synaptic inputs to neuronal cell bodies and horizontally-oriented structures. The barrel cortex from the left hemisphere will be used for studying synaptic inputs to apical dendrites and other vertically-oriented structures. The right hemisphere will be sectioned perpendicular to the surface so that each layer of neuronal soma is evident. The left hemisphere will be sectioned parallel/tangential to the surface so that vertically-oriented dendrites and axons will be cut in cross-section and therefore we shall be able to measure their diameters accurately.

Methods for Accomplishing the Computational Neuroscience Objectives. Computationally efficient Hodgkin-Huxley-like ionotropic signaling models for cortical neurons have been previously reported[55]. We will use the Wilson cortical neuron models will be used to simulate the behavior of the trigger zone and for small cortical neural netlets. We will use the MATLAB™ simulation language. The Wilson models will be augmented with new models of metabotropic second-messenger signal cascades for both monoamine and neuropeptide metabotropic mechanisms. Affects of metabotropic signaling mechanisms will initially be incorporated into these models using long-time-constant Rall alpha-function dynamics

$$\tau \frac{df_j}{dt} = -f_j + U[V_j - \Omega_j]$$
$$\tau \frac{dS}{dt} = -S + f_j$$

where f_j is the second-messenger activation variable for metabotropic synapse j, τ is the time constant for second-messenger action, U is the Heaviside step function, V_j is the membrane voltage of the presynaptic neuron, Ω_j is the presynaptic threshold for NTX release, and S is a conductance modification factor. This is the simplest possible dynamical model for second-messenger cascade effects, and it is anticipated that higher-order kinetic models must eventually replace the Rall model to obtain agreement between model results and physiological data. Diffusion models will be developed to propagate S in the near-vicinity of metabotropic synapses, and active transport models will be developed to propagate S to locations distal from the metabotropic inputs. Ionotropic channel conductance modification will initially be based on S according to $g = g_0 + S \cdot g_m$ where g_0 is the long-term synaptic weight value and g_m is the sensitization factor associated with effects of metabotropic second-messenger cascade reactions.

Simulations of small netlets under various signaling conditions will be used to develop simplified single- and multi-compartment neuron models composed of leaky-integrator-based feeding field dendritic inputs, leaky-integrator-based metabotropic linking field inputs, and a standard neuromime

firing compartment[56]. This form of model is computationally efficient enough to permit simulation of small networks of vertically-oriented cell groups in close proximity within the neocortex. These models will be available to the CS members of the team to support evolutionary computing work.

Simulations will proceed based on the hypothesis of dynamic re-organization of functional units. According to the correlation theory of brain function[24] projection (output) neurons in these assemblies fire synchronously. This hypothesis will be tested in short-chain synfire configurations and the degree of synchrony, degree of participation of synchronously-firing output projections, and phase delay relationships among these outputs will be characterized. This characterization permits the development of higher-level population models[57] of data flow signal processing with each node in the population model representing complex assemblies of neurons (nodes) for the network-of-network level representation of the system.

Analysis of the results of these models will be based on the hypothesis that the brain employs innate representations and binding codes that constitute the three Bourbaki "mother structures" (algebraic, order, and topological structures)[58]. Our analysis will attempt to identify the neuronal substrate of these structures, and positive identification of this substrate will lead directly to the identification of the computational operations carried out by the networks. Bourbaki structures have immediate implications for mathematical, learning, and computational properties of the system, and the findings of this analysis will be communicated to the CE and CS team members.

Methods for Accomplishing the Computer Science Objectives. The principal theoretical problems of neural organization are those of segmentation, superposition, binding, and the neural code[59]. Traditional neural network theory has always dealt well with the problem of segmentation, but combining segmentation and superposition has always posed special difficulties, collectively called "the binding problem." These issues have been thoroughly reviewed by von der Malsburg[59]. Dynamic links and a graph-theoretic paradigm have been proposed as a biologically realistic method for solving the binding problem. An algorithm extending the basic concept of the Hopfield network has been demonstrated which integrate these concepts at the FON level[60]. This algorithm jointly optimizes a pair of Hamiltonian functions over a network of McCulloch-Pitts-like neural elements (perceptrons), and these Hamiltonian functions can serve as a starting point for a fitness function evaluating networks at the FON level. Evolved FONs are a likely point of origin for collaboration with the NA and CNS team members, and evaluation of FON information processing characteristics is a likely junction point for the CS, CE, and ME team members.

At the SON and TON levels of the system, additional complications arise, and do so in large part because of the as-yet-unresolved issue of neural coding. There is growing acceptance that firing rate

by itself is not a sufficient description of the neural code. Strong theoretical arguments have been put forth in support of the hypothesis that short-term synchronization of firing by groups of neurons is an important form of neural coding at the SON level of neural organization[61], although there is also evidence that single spike events also sometimes contain important information on the external stimulus[62] in animals lower on the phylogenetic tree. Detection of synchrony and coherence against a background of stochastic activity poses a difficult detection problem in neural networks as well as issues of time-scale requirements. Pawelzik has proposed a mathematically simple model which quantifies coherence in information-theoretic terms[63], and his model functions can serve as a starting point for fitness function development at the SON level of network organization. Because synfire chains constitute a type of data flow information processing network, they are a likely junction point for collaboration between the CS and CE team members.

At the TON level of network organization still more complications arise from the convergence of SON signals onto CZ assemblies, and from the requirement for temporal binding codes to represent objects as time-sequenced events (see figure 4). Chief among these complications is the question of how and when induction of long term potentiation (LTP) and long term depression (LTD) of synaptic plasticities takes place. In traditional neural network theory the LTP/LTD induction problem is approached statistically, and it is this statistical approach that contributes to the NP-complete issue of neural learning raised by Minsky and Papert. Reinforcement learning and the actor-critic architecture approach to network learning has more recently become widespread[64,65]. This method bears a striking analogy to the putative function of the hippocampal formation and non-cortical structures in figure 4. The critic element of an actor-critic architecture fills the role of a "reward" system, and this role is analogous to that played by elements of the limbic cortex, amygdala, hippocampus, ventral tegmental area, and nucleus accumbens in the central nervous system[66]. This system employs metabotropic signaling mechanisms and is believed to play an important role in learning.

The main difference between the usual actor-critic approach and the architecture of figure 4 lies in the location of the critic and the non-objective nature of its value function within the system. In somewhat loose terms, signaling from the brain's reward system has the character of impulsive and/or emotional drives rather than that of a direct and objective measure of system performance. Wells has argued that this non-objective character of the critic/value system is a logically necessary character of any system whose task is to stimulate concept formation in a neurocomputer[67]. Similarly, Picard has concluded that an element of "emotional intelligence" is necessary for even simple tasks in pattern recognition, machine learning, and machine decision-making[68] and has proposed simple models for objective functions of this emotional intelligence. Finally, there is psychological evidence that an

affective "value system" and affective "interests" complement the energizing character of affectivity in the development of human intelligence[69].

In this project we should confine our critic-value fitness function to primitive functions that respond to the presence or absence of stability in firing patterns converging on CZ networks. Psychologically, the emotion of "happiness" appears to be what some psychologists have called a "neutral gear," i.e. a point of stable system operation[70]. A measure of such operation is coherence detection carried out by elements of CZ assemblies coupled with identification of non-coherent upstream signaling assemblies. The critic would respond to coherence/incoherence through metabotropic signaling to induce LTP in network connections expressing coherence and LTD in network connections involved in incoherent signaling. Findings from this work are likely junction points for collaboration among the CS, CNS, and CE team members, and have implications for the ME team as well.

These classes of fitness functions can be used to evolve network structures at the FON, SON, and TON levels. The evolved networks can then be modeled as weighted digraphs, and their information processing capacities can be measured using the Frobenius theorem[71] with capacity indexing by the maximum eigenvalue of their graph spectrum[72]. These findings will likely be of direct and urgent interest to the CE team.

Methods for Accomplishing the Computer Engineering Objectives. Computer theory, with its myriad subdisciplines and specialties, is without doubt the most well-developed, thoroughly-researched instantiation of automata theory in existence. It contains a vast wealth of mathematical formalism expressing everything from computability theory (e.g. Turing machines) to formal grammars and languages to algorithm theory[73,74,75,76]. Now, when we look at brain theory (neuroscience) *as* a science, we have no other valid choice than to likewise regard the brain as an automaton regardless of whatever other philosophical or religious views we hold outside the proper realm of science. This is neither positivism nor scientific materialism (although it might be called Kantianism). It is merely a requirement of the scientific method. The scientific and mathematical idea of the brain-as-automaton goes back to von Neumann[2], and it is indeed possible to show a one-for-one correspondence between the fundamental elements of computer organization and the psychological epistemology of brain function[67]. It follows from this that a general theory of neurocomputing must be expressible with the same rigor and in the same mathematical formalism with which general computer theory (automata theory) is expressed.

Development of a computer paradigm for the neurocomputer system requires translation of the ideas inherent in neural network theory into more traditional set-theoretic forms of expression, e.g.

logic functions and operations. In DLA, data structures are represented in the form of digraphs describing network connections, and such a graphical representation has a well-known corollary to data structure representation as a linked list[77]. Here the nodes in the graph represent data cells and the arcs represent link pointers. Linked list structures and the graphs that describe them likewise play a fundamental role in the construction of formal grammars[78], which can here be regarded as the basis for algorithm expression in a neurocomputer.

It has long been known that neural networks are capable of expressing fundamental logic functions. It is likewise known that pulse-coded neural networks can express combinatorial logic functions[79]. Many gradient-descent-based algorithms for learning combinatorial logic expressions have been published over the years. Wells and Brennan have also demonstrated a Quine-McCluskey[80]-like adaptive algorithm for learning logic function expressions in the context of a content-addressable parallel processor[81]. Learning algorithms based on gradient descent using forward-backward methods have also been proposed for recurrent neural networks[82].

More difficult, however, is the problem of developing and expressing sequential algorithms. At the simplest level, sequence learning is related to learning in recurrent neural networks. However, algorithms for recurrent networks reported to date address only static learning and do not express temporal sequences in their networks. The problem of learning and representing temporal sequences in neural networks is tightly coupled to the binding problem, e.g. Type II binding codes in Damasio's model of neural function. Likewise, a single sentence in a formal grammar provides a basic entity object in an algorithmic representation, but it does not provide an algorithm because an algorithm expresses a sequence of operations and representations.

One way to approach the sequential representation problem is along the lines of Wood's model for reasoning agents[83]. Wood's loop is shown in figure 5. However, we require a less abstract representation of this process that will be compatible with the primitive and basic information processing operations present in the neurocomputer system. Recall that Type II binding codes are binding codes related to the motor cortex, i.e. to actions taken by the system. The "perception" of any action is a series of temporally-bound event-objects, each of which is represented by the integration of fragmentary representations from SON-level structures in figure 4. Type II binding codes can therefore be regarded as constituting a sliding activity window, which in turn is merely a state-dependent-coding[84] of object representations at any particular moment in time. The sequence binding problem therefore minimally involves two distinct but interacting learning operations: 1) formation of
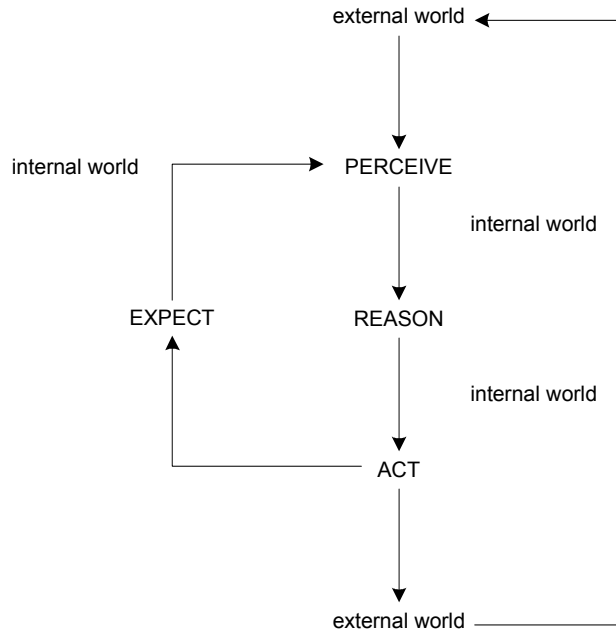
Figure 5: Woods' reasoning loop model of an intelligent agent. The agent constructs for itself a "world model" of the external world, which it uses in determining its actions and which it keeps updated in the face of changing circumstances or unexpected results.

the initial binding codes for object representations coexistent in time; and 2) anticipation of subsequent object representations based on a representation of the present state (Wood's EXPECT function).

The first of these learning operations is merely superposition of fragments of firing activities from some set of upstream networks at the SON level by a Type I binding code. By itself such a binding code does not produce within one representational state at one moment in time a representation of the distinct entity objects within the superimposed representation. Rather, such a representation must be regarded as a syncretic representation in which no sub-division of the object into a multiplicity of objects is immediately presented (not even the distinction between foreground and background objects). For purposes of terminology we will call this representation an intuition[85]. Psychological studies have shown that such syncretic perceptions are in fact the raw material from which intelligence is built up in young children[86,87]. Syncretism is universal in child development, and consequently we can make the hypothesis that the formation of syncretic Type I binding codes is a primitive learning operation in neurocomputing, and this hypothesis then requires a neural substrate. This substrate can be called an "attention network" and corresponds to the short-time-scale modulation of synaptic weights as described by von der Malsburg[59]. In graph-theoretic terms, this corresponds to dynamic weighting of the arcs connecting the participating SONs, but also requires the superposition of a second graph (the control graph), which forms a TON structure.

The second learning operation involves two temporal factors. First, the CZ assemblies that are eventually to produce the Type II binding codes that constitute the <u>memory</u> of a temporal sequence require a neural substrate in which Type I intuitive binding codes gradually "fade into" one another rather than switching absolutely abruptly. This can be regarded as constituting a run-length-limited coding sequence[84], and this dynamical process whereby active Type I binding sequences blend into one another can be termed <u>apprehension</u>[85]. In graph-theoretic terms such coding sequences can be formally analyzed using principal states and approximate eigenvector analysis[88], and the general theory of such sequences is describable in terms of sliding block code theory[89]. In terms of the neural substrate, the formation of a Type II binding code logically must involve detection of the "sliding code window" by a downstream (higher order) CZ assembly since the "codewords" that are "sliding" into one another are represented by the firing patterns of immediately upstream (lower order) CZ assemblies and not by early and intermediate perceptual networks (which represent no binding codes). The implementation of this temporal sequence recognition requires temporal binding to form what have been termed shifter circuits[59]. A content-addressable parallel processor[81] can perform such a task by using a sliding mask and its select-first-responder function, and a reasonable hypothesis is that a neural network equivalent of such a processor (TON level network) can implement the same temporal imprinting and detection algorithms.

Concurrently with the initial-imprinting/subsequent-detection process just discussed, a means of suppressing and/or ignoring other signaling activities, and for determining which activities are to be attended to by the process above, is required. This requires some local <u>attention mechanism</u> exercising local control of the "microtopic" being encoded. This approach is consistent with von der Malsburg's conjecture regarding the general characteristics for how temporal binding could first be achieved[59]. The Type II binding codes being formed by the above process produce "objective" representations; however, the task of attention is at root non-objective because it is a mechanism necessary for the possibility of representations of the objective type. The structures implicated for this task in Damasio's architecture correspond to the non-cortical elements in figure 4. There is a considerable body of psychological as well as neurological evidence that implicates at least some of these structures with the "emotional intelligence" aspect of the system, and so on this point there is a need for close collaboration of efforts with the critic-value system work described earlier under the CS method of approach. Based on the aforementioned psychological findings, Wells has proposed signal coherence/incoherence as the neurological substrate for critic-directed attention mechanisms[67,85]. This hypothesis is consistent with von der Malsburg's correlation theory of brain function[24] and implies the need for neural network control structures, namely coherence/incoherence detectors, either within the CZ assemblies or locally augmenting their function, which project to the non-cortical networks. These

non-cortical networks, in turn, would then have to provide control signals back to multiple CZ networks to "fix attention" on some signal pathways while suppressing it for others. A possible network scheme for the coherence/incoherence detection task is shown in figure 6. The scheme in effect amounts to a neural network analog of a correlator/phase-locked-loop function.

The second temporal factor concerns the neural substrate of Wood's EXPECT function. Once a temporal sequence has been encoded, detection of the possible beginning of its reoccurrence in some later sensory "experience" is ground for the system to <u>anticipate</u> that this temporal sequence is going to happen. Anticipation can be implemented by a sequence generator successively activating, in proper sequence and at the proper time, the binding code networks that have previously stored the binding code[59,19]. Anticipating CZ networks feeding back to the early and intermediate sensory networks condition these networks to fire. If the anticipated sensory data sequence occurs, then the anticipation is met; otherwise, there will be incoherence between the imaginary signaling sequence and the actual data path signaling sequence, and detection of this incoherence can be called a <u>disappointment</u> or a <u>frustration</u> in terms of the critic-value system network function. Occurrence of a disappointment or a
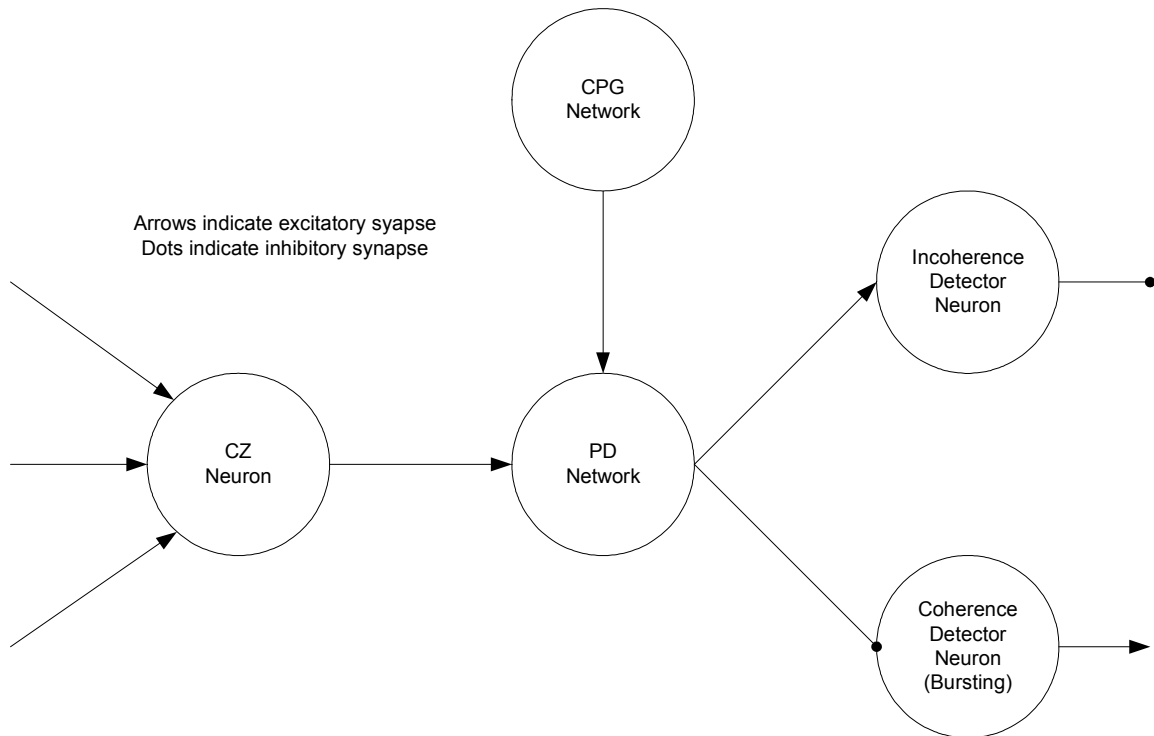

Figure 6: Possible coherence/incoherence detection scheme. CZ=convergence zone neuron or sub-assembly. PD=phase detector; CPG=central pattern generator. The general idea is that conditions of coherence and incoherence are characterized by the correlation found among signals. Synchronous time-locked signals converging on an integrate-and-fire CZ neuron will produce a periodic signal at its output at a firing rate that is an integer subharmonic of the incoming signal firing rates. This signal is compared (in frequency and phase) to a reference signal from a central pattern generator network. Phase/frequency errors produce output activity in the PD network which excites a neuron (or neurons) acting as the incoherence detector and simultaneously inhibits burst firing from the coherence detector neuron or neurons. These neurons project back to early and intermediate feature networks (a form of phase-locked loop structure) and to non-cortical critic-value networks.

frustration is an attention-signaling mechanism, but one acting across a much broader area of the system and at a much slower time scale than the attention mechanism discussed earlier[59]. Very compelling evidence from psychological experiments exists in support of the hypothesis that this mechanism is a fundamental characteristic of the way in which intelligence develops in human beings[90,91].

In terms of conventional computer theory, the mechanisms for binding discussed above are instances of state machine algorithms belonging to that class of state machine organization usually called the linked state machine architecture[92]. In a graph-theoretic context state machine sequences are commonly represented using trellis diagrams, a trellis diagram merely being a state diagram "unfolded" to make explicit the temporal sequence. A trellis constitutes a special case of order structure and provides a natural mode for mathematically expressing not only fixed state sequences but also time-varying state sequences and such parallel operation constructs as fork- and join-relations[93].

<u>Methods for Accomplishing the Microelectronic Objectives</u>. The ME team is responsible for implementing various neural network functions and structures in collaboration with the work of the other teams. The general schema for biomimic artificial neurons (BANs) has been previously established[25] and a number of implementations of elementary neuron circuits have already been reported[79,94,95,96,97]. Additional neuron designs will have been produced and tested, and standard cell implementations for most of the analog neuronal functions will have been developed, by Jan. 31, 2005 under the MRCI's Neurofuzzy Soft Computing Program. These designs provide the basic elements for the design of network implementations using the NSF MOSIS fabrication program. It is to be noted, however, that implementation of very large networks in a single VLSI chip is likely to require die-shrink re-design of these circuits in order to use higher-density (e.g. 0.5 micron) fabrication technologies.

Network structures developed by the CS and CE efforts will be translated into mixed-signal VLSI circuit designs using the established BAN and "forgetful logic" circuit designs as the starting point. It is noted that these designs will not necessarily be structured one-to-one with the block/signal-processing diagrams from these teams, owing to constraints in matching precision and achievable tolerances in standard VLSI technology. In such cases where warranted by these practical considerations, the ME team will work closely with the CS and CE members to come up with alternate but functionally equivalent network topologies capable of being implemented within existing technology constraints.

At the present state of the art, multichip implementations are probably unavoidable in constructing the complete neurocomputer. It is furthermore anticipated that most chip designs will be pinout-

limited unless appropriate partitioning design is carried out. The ME team will carry out the appropriate functional- and chip-partitioning to deal with these issues, and in those cases where pinout limitations are unavoidable the team will devise appropriate multiplexing methods for interchip communications[67,98,99]. The team will also carry out appropriate printed circuit board layout designs and mechanical packaging designs for the implementation of multichip breadboards and prototypes. In addition, the team will design an appropriate testing strategy so that the functionality of each integrated circuit can be verified in the laboratory.

For purposes of verifying the cognitive capabilities of the neurocomputer it will be necessary to provide the neurocomputer with a "body" – i.e. sensors, actuators, and a mechanical platform by which the neurocomputer can interact with the outside world. Test cases should include sensor arrays for auditory, and possibly simple visual, pixel arrays, motor drive circuits (e.g. for small wheeled or track vehicles to permit locomotion), and tactile arrays (e.g. "whiskers" for tactile exploration). The team will collaborate closely with the CS members to aid them in evolving appropriate sensory and motor control networks for processing the input data and output command flows. In addition, the design will incorporate battery sensors to detect low-battery conditions. This sensory information will flow to the non-cortical critic-value subsystem to sense "hunger" and trigger an appropriate "hunger response" (e.g. a behavioral analog to a baby's hunger vocalization reflex, which therefore requires vocalization transducer circuitry).

To support the development of this platform we will sponsor one or more senior design projects. In addition, because it is likely that early neurocomputer prototypes will be too large and heavy to be incorporated into a small mechanical robotic platform, the ME team will provide a radio communication system linking the neurocomputer to the mechanical platform. The ME team will provide the systems-level specifications for this system, and we will implement it through a senior design project.

IV. Broader Technical Framework of this Research

The initial methods described in the previous section serve as a launching point for the Neurocomputing Initiative. In the long run it is necessary to have a vision of where all this research is leading and what we can expect the landscape to look like when we get there. All the work described in this document is based on the expected characteristics of a large-scale general-purpose neurocomputer as described in the Program's Vision Statement document[100]. The Vision Statement describes the general attributes and characteristics of the machine in terms of its functional organization. It describes the ways in which such a machine will be similar to von Neumann's model, and more importantly the ways in which it will differ. Input, output, arithmetic-logic, memory, and

control functions of a large-scale general-purpose neurocomputer must be regarded in a very different way than are these functions regarded in a standard digital computer. In addition to the biologically-inspired hardware substratum of the Machine, there is also an epistemological substratum that might be described as the idea of a "machine psyche." This substratum constitutes such cognitive functions as sensibility, apprehension, imagination, objective and subjective judgment, objective and practical reasoning, motoregulatory ("instinctive") action, and meanings, and how these cognitive traits are to be viewed in the organization of the Machine. The Vision Statement describes how to interpret the difficult ideas of intuition, concepts, affectivity, and ideas. As the Program progresses and the Machine takes shape, the research Team will be expanded to bring in contributors from the field of cognitive neuroscience in order to address important research questions contained within this epistemological substratum. Every co-PI in this Program is expected to become familiar with the broad framework expressed in the Vision Statement because this framework provides the unifying direction for all other work within this Program.

---

[1] McCulloch, W.S. and W. Pitts (1943), "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophysics*, 5: 115-133.

[2] Von Neumann, J., *The Computer and the Brain*, New Haven, CN: Yale University Press, 1958.

[3] Barnes, G.H., R.M. Brown, M. Kato, D.J. Kuck, D.L. Slotnick, and R.A. Stokes, "The ILLIAC-IV computer," *IEEE Trans., C-17,* vol. 8, pp. 746-757, Aug., 1968.

[4] Anderson, J.P., S.A. Hoffman, J. Shifman, and R.J. Williams, "D825 – a multiple-computer system for command and control," *AFIPS Proc. FJCC*, vol. 22, pp. 86-96, 1962.

[5] Comtre Corporation, *Multiprocessors & Parallel Processing*, P.H. Enslow, Jr. (ed.), NY: John Wiley, 1974.

[6] Rodrigue, G. (ed.), *Parallel Computations*, NY: Academic Press, 1982.

[7] Amari, S.-I., "Mathematical foundations of neurocomputing," *Proc. IEEE*, vol. 78, no. 9, 1990, pp. 1443-1463.

[8] Minsky, M.L. and S.A. Papert, *Perceptrons*, expanded ed., Cambridge, MA: The MIT Press, 1988.

[9] Hunter, D., *Parity-N Solutions Using Feedforward Neural Networks*, M.S. Thesis, University of Idaho, July, 2003.

[10] Natarajan, B.K., *Machine Learning*, San Mateo, CA: Morgan Kaufmann, 1991.

[11] Werbos, P., Plenary Address, 29th An. Conf. Ind. Electron. Soc. (IECON'03), Nov. 2-6, 2003, Roanoke VA.

[12] Beurle, R.L. (1956), "Properties of a mass of cells capable of regenerating pulses," *Philos. Trans. R. Soc. Lond.* [*Biol.*] 240: 55-94.

[13] Abeles, M., *Corticonics*, Cambridge, UK: Cambridge University Press, 1991.

[14] Anninos, P.A., B. Beek, T.J. Csermely, E.M. Harth, and G. Pertile, "Dynamics of neural structures," *J. Theoret. Biol.* (1970) **26**, 121-148.

[15] D.E. Rumelhart, J.L. McClelland, and the PDP Research Group (1986), *Parallel Distributed Processing* (in two volumes), Cambridge, MA: The MIT Press.

[16] Abeles, M., "Role of the cortical neuron: Integrator or coincidence detector?" *Israel J. Med. Sci.*, vol. 18, pp. 83-92, 1982.

[17] R. Eckhorn, H.J. Reitboeck, M. Arndt, and P. Dicke, "Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex," *Neural Comput.* **2**, 293-307 (1990).

[18] A.R. Damasio, "Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition," *Cognition*, **33** (1989), 25-62.

[19] A.R. Damasio, "The brain binds entities and events by multiregional activation from convergence zones," *Neural Comput.*, **1**: 123-132 (1989).

[20] Hubel, D.H. and T.N. Wiesel (1972), "Laminar and columnar distribution of geniculocortical fibers in the macaque monkey," *J. Comp. Neurol.* 146, 421-450.

[21] Livingstone, M.S. and D.H. Hubel (1987), "Psychophysical evidence for separate channels for the perception of form, color, movement, and depth," *J. Neurosci.* **7**: 3416-3468.

[22] Levitan, I.B. and L.K. Kaczmarek, *The Neuron: Cell and Molecular Biology*, Oxford, UK: Oxford University Press, 2002.

[23] Von der Malsburg, C., "Dynamic link architecture," in *Handbook of Brain Theory and Neural Networks*, 2nd ed., M.A. Arbib (ed.), Cambridge, MA: The MIT Press, 2003, pp. 365-368.

[24] Von der Malsburg, C. (1981), *The Correlation Theory of Brain Function*, MPI Biophysical Chemistry, Internal Report 81-2.

[25] Wells, R.B., "Modulation channels," *Proc. 28th An. Conf. Indus. Electron. Soc. (IECON'02)*, Nov. 5-8, Seville, Spain, pp. 3209-3214.

[26] White, E.L. and A. Keller, *Cortical Circuits*, Boston, MA: Birkhäuser, 1989.

[27] Toledo-Rodriguez, M., A. Gupta, Y. Wang, C.Z. Wu, and H. Markram, "Neocortex: Basic neuron types," in *Handbook of Brain Theory and Neural Networks*, 2nd ed., M.A. Arbib (ed.), Cambridge, MA: The MIT Press, 2003, pp. 719-725.

[28] Abeles, M., "Synfire chains," in *Handbook of Brain Theory and Neural Networks*, 2nd ed., M.A. Arbib (ed.), Cambridge, MA: The MIT Press, 2003, pp. 1143-1146.

[29] Nelkon, I. (1988), "Analysis of the activity of single neurons in stochastic settings," *Biol. Cybern.* 59: 201-215.

[30] Wells, R.B. and A. Vongkunghae, "A model-reference neural network scheme for obtaining laser printer modulation templates," *Proc. 28th Ann. Conf. Indus. Electron. Soc. IECON'02*, Seville, Spain, Nov. 2002, pp. 1872-1875.

[31] Vongkunghae, A. and R.B. Wells, "A method for obtaining robust modulation templates for laser printers," in *Intelligent Engineering Systems Through Artificial Neural Networks: Smart Engineering System Design: Neural Networks, Evolutionary Programming, Complex Systems and Artificial Life*, vol. 13, pp. 907-913, NY: ASME Press, 2003.

[32] Keller, A. and E.L. White (1989), "Synaptic organization of GABAergic neurons in the mouse Sml cortex," *J. Comp. Neurol.*

[33] Vanegas, H., H. Hollander, and H. Distel (1977), "Early stages of uptake and transport of horseradish peroxidase by cortical structures, and its use for the study of local neurons and their processes," *J. Comp. Neurol.* **177**, 193-212.

[34] Gerfen, C.R. and P.E. Sawachenko (1984), "An anterograde neuroanatomical tracing method that shows the detailed morphology of neurons, their axons and terminals: Immunohistochemical localization of an axonally transported plant lectin, *Phaseolus vulgaris* leucoagglutinin (PHA-L)," *Brain Res.* **290**: 120-125.

[35] Cauli, B., E. Audinat, B. Lambolez, M.C. Angulo, N. Ropert, K. Tsuzuki, S. Hestrin, and J. Rossier (1997), "Molecular and physiological diversity of cortical nonpyramidal cells," *J. Neurosci.*, **17**: 3894-3906.

[36] DeFilipe, J. (1993), "Neocortical neuronal diversity: Chemical heterogeneity revealed by colocalization studies of classical neurotransmitters, neuropeptides, calcium binding proteins, and cell surface molecules," *Cereb. Cortex*, **3**: 273-279.

[37] Defilipe, J. (1997), "Types of neurons, synaptic connections, and chemical characteristics of cells immunoreactive for calbinden-D28K, parvalbumin, and calretinin in the neocortex," *J. Chem. Neuroanat.*, **14**: 1-19.

[38] Kawaguchi, Y. and Y. Kubota (1997), GABAergic cell subtypes and their synaptic connections in rat frontal cortex," *Cereb. Cortex*, 7: 476-486.

[39] Wang, Y., A. Gupta, M. Toledo-Rodriguez, C.Z. Wu, and H. Markram (2002), "Anatomical, physiological, molecular, and circuit properties of nest basket cells in the developing somatosensory cortex," *Cereb. Cortex*, **12**: 395-410.

[40] Zucker, E. and W.I. Welker (1969), "Coding of somatic sensory input by vibrissae neurons in the rat's trigeminal ganglion," *Brain Res.* **12:** 138-156.

[41] Woolsey, T.A., C. Welker, and R.H. Schwartz (1975), "Comparative anatomical studies of the SmI facr cortex with special reference to the occurrence of 'barrels' in layer IV," *J. Comp. Neurol.* **164:** 79-94.

[42] Bennett-Clarke, C.A., N.L. Chiaia, & R.W. Rhoades (1997), "Contributions of raphe-cortical and thalamocortical axons to the transient somatotopic pattern of serotonin immunoreactivity in rat cortex," *Somatosens. Mot. Res.* **14:** 27-33.

[43] Pinto, D.J., J.C. Brumberg, D.J. Simons, G.B. Ermentrout (1996), "A quantitative population model of whisker barrels: Re-examining the Wilson-Cowan equations," *J. Comput. Neurosci.*, **3:** 247-264.

[44] Simons, D.J. and G.E. Carvell (1989), "Thalamocortical response to transformation in rat vibrissa/barrel system," *J. Neurophysiol.* **61:** 311-330.

[45] Simons, D.J. (1985), "Temporal and spatial integration in the rat SI vibrissa cortex," *J. Neurophysiol.* **54:** 615-635.

[46] Gonzalez, M.F. and F.R. Sharp (1985), "Vibrissae tactile stimulations: (14C)2-deoxyglucose uptake in rat brainstem, thalamus, and cortex," *J. Comp. Neurol.* **231:** 457-472.

[47] Ito, M. (1985), "Processing of vibrissa sensory information within the rat neocortex," *J. Neurophysiol.* **54:** 479-490.

[48] Durham, D. and T.A. Woolsey (1985), "Functional organization in cortical barrels of normal and vibrissae-damaged mice: A [3H]2-deoxyglucose study," *J. Comp. Neurol.* **235:** 97-110.

[49] Simons, D.J. and T.A. Woolsey (1984), "Morphology of Golgi-Cox-impregnated barrel neurons in rat SmI cortex," *J. Comp. Neurol.* **230:** 119-132.

[50] Simons, D.J. (1983), "Multi-whisker stimulation and its effects on vibrissa units in rat SmI barrel cortex," *Brain Res.* **276:** 178-182.

[51] Simons, D.J. and T.A. Woolsey (1979), "Functional organization in mouse barrel cortex," *Brain Res.* **165:** 327-332.

[52] Simons, D.J. (1978), "Properties of vibrissa units in rat SI somatosensory neocortex," *J. Neurophysiol.* **41:** 798-820.

[53] Durham, D. and T.A. Woolsey (1978), "Acute whisker removal reduces neuronal activity in barrel of mouse SmI cortex," *J. Comp. Neurol.* **178:** 629-644.

[54] Durham, D. and T.A. Woolsey (1977), "Barrels and columnar cortical organization: Evidence from 2-Deoxyglucose (2-DG) experiments," *Brain Res.* **137:** 169-174.

[55] H.R. Wilson (1999), "Simplified dynamics of human and mammalian neocortical neurons," *J. Theor. Biol.* **200,** 375-388.

[56] Eckhorn, R., H.J. Reitboeck, M. Arndt, and P. Dicke, "Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex," *Neural Computat.* **2**, 293-307, 1990.

[57] Pouget, A., K. Zhang, S. Deneve, and P. Lathan (1998), "Statistically efficient estimation using population codes," *Neural Computation*, **10**: 373-401.

[58] Piaget, J., *Genetic Epistemology*, NY: W.W. Norton, 1971.

[59] Von der Malsburg, C. (1999), "The what and why of binding: the modeler's perspective," *Neuron*, vol. 24, pp. 95-104.

[60] Von der Malbsurg, C. and E. Bienenstock (1987), "A neural network for the retrieval of superimposed connection patterns," *Biophys. Lett.*, **3**(11): 1243-1249.

[61] Abeles, M. (1994), "Firing rates and well-timed events in the cerebral cortex," in *Models of Neural Networks II*, E. Domany, J.L. van Hemmen, & K. Schulten (Eds.), NY: Springer-Verlag, pp. 121-140.

[62] de Ruyter van Stevenlinck, R.R. and W. Bialek (1988), "Real-time performance of a movement-sensitive neuron in the blowfly visual system: Coding and information transfer in short spike sequences," *Proc. R. Soc. London Ser.* **B 234:** 379-414.

[63] Pawelzik, K. (1994), "Detecting coherence in neuronal data," in *Models of Neural Networks II*, E. Domany, J.L. van Hemmen, & K. Schulten (Eds.), NY: Springer-Verlag, pp. 253-285.

[64] Barto, A.G., "Reinforcement learning," in *Handbook of Brain Theory and Neural Networks*, 2nd ed., M.A. Arbib (ed.), Cambridge, MA: The MIT Press, 2003, pp. 963-968.

[65] Barto, A.G., "Reinforcement learning in motor control," in *Handbook of Brain Theory and Neural Networks*, 2nd ed., M.A. Arbib (ed.), Cambridge, MA: The MIT Press, 2003, pp. 968-972.

[66] Nestler, E.J. and R.C. Malenka, "The addicted brain," *Sci. Am.* vol. 290, no. 3, Mar. 2004, pp. 78-85.

[67] Wells, R.B., "Preliminary discussion of the design of a large-scale general-purpose neurocomputer," Nov. 14, 2003, http://www.mrc.uidaho.edu/~rwells/techdocs.

[68] Picard, R., *Affective Computing*, Cambridge, MA: The MIT Press, 1997.

[69] Piaget, J., *Intelligence and Affectivity: Their Relationship During Child Development*, Palo Alto, CA: Annual Reviews Inc., 1981.

[70] Carlson, J.G. and E. Hatfield, *Psychology of Emotion*, Fort Worth, TX: Harcourt Brace Jovanovich, 1992.

[71] Cvetković, D.M., M. Doob, and H. Sachs, *Spectra of Graphs*, NY: Academic Press, 1980.

[72] Akuzawa, K. and K. Ohnishi, "Design indices for information connection in decentralized systems," *Proc. 29th An. Conf. Indus. Electron. Soc. IECON'03*, Roanoke, VA, Nov. 2-6, 2003, pp. 2417-2422.

[73] Turing, A.M., "On computable numbers with an application to the Entscheidungsproblem," *Proc. London Math. Soc.*, Ser. 2-42 (1936), 230-265.

[74] Minsky, M.L., *Computation: Finite and Infinite Machines*, Englewood Cliffs, NJ: Prentice-Hall, 1967.

[75] Nelson, R.J., *Introduction to Automata*, NY: John Wiley, 1968.

[76] Kleene, S.C., *Introduction to Metamathematics*, Princeton, NJ: Van Nostrand, 1952.

[77] Preparata, F.P. and R.T. Yeh, *Introduction to Discrete Structures*, Menlo Park, CA: Addison-Wesley, 1974.

[78] Hopcroft, J.E. and J.D. Ullman, *Formal Languages and Their Relation to Automata*, Reading, MA: Addison-Wesley, 1969.

[79] Liu, B. and J.F. Frenzel, "A CMOS neuron for VLSI circuit implementation of pulsed neural networks," *Proc. 28th An. Conf. Indus. Electron. Soc. (IECON'02)*, Nov. 5-8, Seville, Spain, pp. 3186-3190.

[80] McCluskey, E.J., *Introduction to the Theory of Switching Circuits*, NY: McGraw-Hill, 1965.

[81] Brennan, A., *Binary Connectionist Networks*, M.S. Thesis, University of Idaho, Aug., 1998.

[82] Doya, K., "Recurrent networks: learning algorithms," in *Handbook of Brain Theory and Neural Networks*, 2nd ed., M.A. Arbib (ed.), Cambridge, MA: The MIT Press, 2003, pp. 955-960.

[83] W.A. Woods, Important issues in knowledge representation, *Proc. IEEE*, vol. 74, no. 10, 1986, pp. 1322-1334.

[84] Wells, R.B., *Applied Coding and Information Theory for Engineers*, Upper Saddle River, NJ: Prentice-Hall, 1999.

[85] Wells, R.B., *The Critical Philosophy and the Phenomenon of Mind*, unpublished manuscript.

[86] Piaget, J., *The Language and Thought of the Child*, 2nd ed., London: Routledge & Kegan Paul, 1932.

[87] Piaget, J., *Judgment and Reasoning in the Child*, Totowa, NJ: Littlefield, Adams & Co., 1966.

[88] Lempel, A. and M. Cohn, "Look-ahead coding for input-restricted channels," *IEEE Trans. Inform. Th.*, vol. IT-28, pp. 933-937, Nov., 1982.

[89] Adler, R., D. Coppersmith, and M. Hassner, "Algorithms for sliding block codes," *IEEE Trans. Inform. Th.*, vol. IT-29, no. 1, pp. 5-22, Jan., 1983.

[90] Piaget, J., *The Origins of Intelligence in Children*, Madison, CN: International Universities Press, 1952.

[91] Piaget, J., *The Construction of Reality in the Child*, NY: Basic Books, 1954.

[92] Clair, C.R., *Designing Logic Systems Using State Machines*, NY: McGraw-Hill, 1973.

[93] Davey, B.A. and H.A. Priestley, *Introduction to Lattices and Order*, 2nd ed., Cambridge, UK: Cambridge University Press, 2002.

[94] Wells, R.B. and B. Barnes, "Capacitor-free leaky integrator for biomimic artificial neurons," *IEE Electron. Lett.*, vol. 38, no. 17, pp. 974-976, Aug. 2002.

[95] Wells, R.B. and B.C. Barnes, "Delay-resistor implementation of integrators in biomimic artificial neurons," *Proc. 28th An. Conf. Indus. Electron. Soc. (IECON'02)*, Seville, Spain, Nov.5-8, 2002, pp. 3186-3190.

[96] Barnes, B.C., R.B. Wells, and J.F. Frenzel, "PWM characteristics of a capacitor-free integrate-and-fire neuron," *IEE Electron. Lett.*, vol. 39, no. 16, pp. 1191-1193, Aug., 2003.

[97] Barnes, B.C. and R.B. Wells, "A versatile pulse-mode biomimic artificial neuron using a capacitor-free integrate-and-fire technique," *Proc. 29th An. Conf. Indus. Electron. Soc. (IECON'03)*, Roanoke, VA, Nov. 2003, pp. 2968-2972.

[98] A.F. Murray, D. Del Corso, and L. Tarassenko, Pulse-stream VLSI neural networks mixing analog and digital techniques, *IEEE Trans. Neural Networks*, vol. 2, no. 2, 1991, pp. 193-204.

[99] A.F. Murray, Pulse arithmetic in VLSI neural networks, *IEEE Micro Mag.*, Dec., 1989, pp. 64-74.

[100] Wells, R.B., "Preliminary discussion of the design of a large-scale general-purpose neurocomputer," Nov. 14, 2003, http://www.mrc.uidaho.edu/~rwells/techdocs.