In: Toward Replacement Parts for the Brain, MIT Press, pp.277-294

# Brain circuit implementation: High-precision computation from low-precision components

Richard Granger
Brain Engineering Laboratory
CS Dept, University of California
Irvine, CA 92697
granger@uci.edu
www.BrainEngineering.com

## Introduction

Attempts to understand, let alone augment or supplant, the operation of brain circuitry rely not only on our knowledge of isolated neuron, circuit, and brain slice behavior but also on the impressive computations achieved by assemblies of these components. The variability of the constituent elements of these circuits suggests that they are arranged to operate in ways not obvious from standard engineering points of view. The quite nonstandard (and in many cases, apparently substandard) unit components of these circuit designs (synapses and cells that are probabilistic, relatively low precision, very sparsely connected, and orders of magnitude slower than the typical elements of most engineering devices) provide constraints on their possible contributions to the overall computation of neural circuits. Since the resultant brain circuits outperform extant engineering devices in many realms of crucial application ranging from recognition of complex visual or auditory signals to motoric traversal of complex terrain, our ability to imitate them can lead to two related but distinct classes of scientific advance: new and unanticipated types of hardware devices based on the unveiled engineering principles, and enabling technologies for the integration of extrinsic devices with intrinsic brain circuitry. This latter capability implies two directions of improved communication: i) a heightened ability to "listen to" and interpret brain activity, and ii) a burgeoning faculty for "talking back" to the brain, which may ameliorate impaired brain function or enhance normal function.

## Background and Approach

The computations performed by human (or even rat) brains have not yet been matched by engineering approaches despite tremendous amounts of money spent on the attempts. It is not known why painstakingly developed artificial approaches continue to fall short of human performance. Even approaches that attempt to simulate complex human behaviors may well be based on descriptions of those behaviors that are incorrectly or incompletely specified. A prerequisite to replicating (let alone exceeding) human abilities may turn out to be a clear computational understanding of brain mechanisms. Among the vast body of data on brain mechanisms are a few key facts:

- The evolution of mammalian forebrain (telencephalon) gave rise to the emergence of new circuits that do not appear in other classes (e.g., reptiles, birds). Arguably the most advanced of these, thalamocortical circuits, are also by far the must numerous, accounting for allometrically disproportionate space in mammalian brains, including the vast majority of circuits in human brain.
- Study of the anatomical design and physiological operation of these circuits has led to the derivation of specific algorithms that these circuits may be carrying out.
- Derivation suggests that distinct brain regions carry out distinct algorithms, each contributing a different set of computations, with further composite algorithms arising from interactions among multiple brain regions.
- Key features of most telencephalic circuits include: sparse connectivity among neurons ($p \leq 0.001$); low precision synaptic connections ($\leq 4$ bits); simple processors (add, multiply); simple "learning" rules (fixed size increment or decrement); slow operation (milliseconds per operation vs. nanoseconds for typical computer hardware); variable, probabilistic responses (vs. the fixed, deterministic responses of engineering devices).

These characteristics should be a severe liability for brain circuits, begging the question of how they can achieve the advanced behavioral performance exhibited by organisms. Indeed, it is sometimes instead assumed that these substandard assessments of brain components are incorrect, driving an ongoing search for hidden precision in

brain components, including carefully timed synchronies, as opposed to the variable synchronies underlying EEGs, precisely timed sequences of neuronal firing (spike trains), as opposed to sequences driven probabilistically (such as Poisson sequences), added topography in brain wiring, as opposed to the coexistence of topographic and highly nontopographic circuits described by much quantitative anatomical research, complex high-precision synapses carefully arrayed on dendrites, as opposed to relatively low-precision synapses probabilistically arrayed. Typical artificial "neural networks" (ANNs) make use of some or all of these "improvements" in neural machinery, enabling higher precision computation. However, due to the costs of the higher precision machinery used, such networks sometimes have relatively high space and time complexity costs, resulting in systems that do not readily scale to problems of large size.

An alternative is that brain circuits do use sparse, probabilistic, slow, low-precision components to somehow perform the rapid, high-precision computations that apparently underlie our advanced sensory and motor capabilities, via algorithms that combine these components in such a way as to enable the emergence of precision computation. Work in our laboratory in recent years has shown how these biological components, although impoverished from the viewpoint of standard engineering, nonetheless may give rise to specific useful high-precision computational methods. This research has resulted in the derivation of a range of algorithms, including hierarchical clustering (Ambros-Ingerson et al., 1990; Kilborn et al., 1996), sequence prediction (Granger et al., 1994; Aleksandrovsky et al., 1997), time dilation (Granger et al., 1995), Bayes classification (Coultrip & Granger, 1994), high capacity storage and retrieval (Aleksandrovsky et al., 1996; Whitson, 1998; Rodriguez et al., 2003), reinforcement learning (Brucher 2000), novelty detection, data compression and hash coding (Granger et al., 1994; Rodriguez et al., 2003).

If these impoverished components can carry out these advanced algorithms, then the low precision of the components becomes an advantage rather than a liability: it enables relatively cheap (sparse, slow, low-precision) components to do a job that otherwise would require much more expensive apparatus. In particular, this suggests the utility of direct hardware implementations in which low-precision components carry out these algorithms. ANNs typically require computing units with 8 or more bits of precision, with relatively dense connectivity among these units, and their computations degrade in performance with reductions in either precision or connection density. A system reliant on these characteristics cannot use lower-precision components to carry out the same computations, and cannot readily scale to large sizes.

**Thalamocortical Circuits**

Figure 1 illustrates key anatomical architectural characteristics of thalamocortical circuits (Rodriguez et al., 2003). Neurons are vertically organized into "pyramidal cell modules" (White and Peters, 1993; Peters et al., 1994), roughly 30-35 μm across, consisting of distinct groups of layer V and layer II-III pyramidal cells whose apical dendrites are commingled. Functional "columns" which are physiologically defined in terms of receptive field properties, rather than anatomical boundaries (e.g., Mountcastle, 1957; 1978), are often described as 400-500 μm or more in extent, thus comprising perhaps 200 pyramidal cell modules apiece.

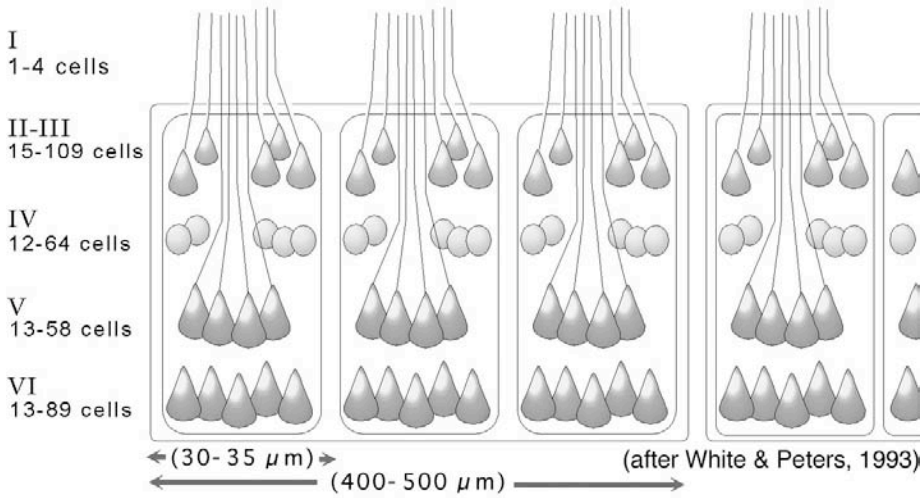


Figure 1. Pyramidal cell modules (White & Peters, 1993) are anatomically organized by grouping of apical dendrites of layer II, III and V cells, and have been proposed as candidate anatomical underpinnings of vertical organizations of cortex such as columns (Mountcastle, 1957) that have only been described physiologically. Shown are typical ranges of numbers of neurons within each module; the variance decreases if certain less typical regions such as primary sensory and motor areas are excluded. Modules are roughly 30-35 μm in size and may be constituents of larger columnar arrangements.

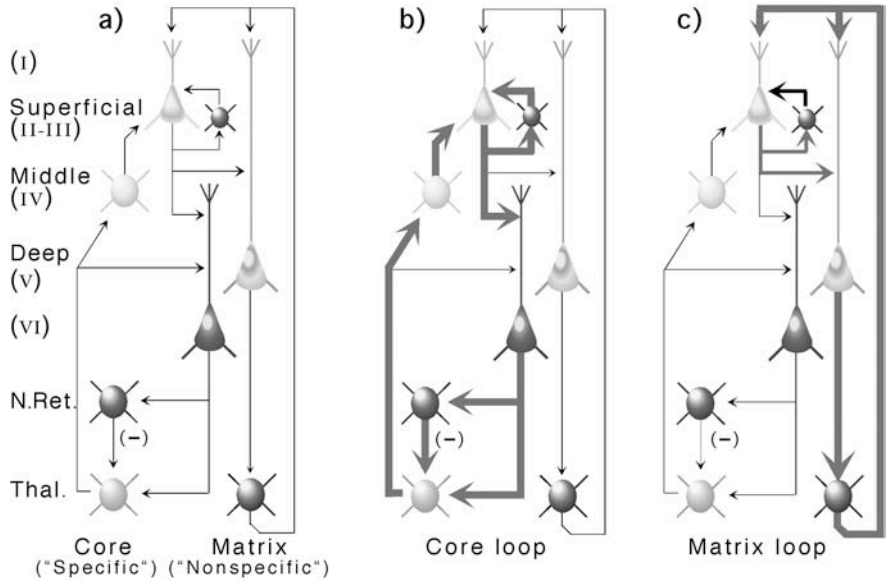


Figure 2. a) Components and organization of thalamocortical elements modeled. Included are features that recur throughout much of neocortex, especially in polysensory and association areas. Characteristics specific to primary sensory and motor areas are not modeled. b) One primary loop through thalamocortical circuitry: afferents from "core" thalamic nuclei (see text) project to layer IV and deep layer III; axons from superficial layer small pyramidal cells engage in local excitatory-inhibitory circuits as well as giving rise to collateral projections to deep layers, and to adjacent cortical regions (not shown); layer VI projects back to thalamic core as well as to overlying nucleus reticularis (N.Ret) neurons, which send inhibitory connections to thalamic core cells. Throughout the "core" loop, projections retain topographic relations. c) In the other primary thalamocortical loop, thalamic "matrix" nuclei project broadly and diffusely to layer I, contacting the apical dendrites of layer II, III and V cells. Layer V generates both descending projections (predominantly to motor systems) and to matrix thalamic nuclei (Bourassa et al., 1995; 1996).

Figure 2 illustrates the overall architecture of the thalamocortical system. Neurons throughout neocortex are organized into relatively stereotypical architectures. Although cortical studies describe some (subtle but potentially crucial) differences among various cortical regions (Galuske et al., 2000; Gazzaniga, 2000), there are sufficient shared characteristics to justify attempts to identify common basic functionality, which may be augmented by special purpose capabilities in some regions (Lorente de No, 1938; Szentagothai, 1975; Keller and White, 1989; Castro-Alamancos and Connors, 1997; Rockel et al., 1980; Braitenberg and Schuz, 1998; Valverde, 2002). Two parallel circuit types co-occur, involving the topographic projection of certain restricted thalamic populations and broad, diffuse projection of the remaining thalamic neurons. It has been found that these two populations of thalamic cells, distinguishable by their targets and topography, can also be identified by their differential immunoreactivity to two $Ca^{++}$ binding proteins: the former population, called thalamic "core" regions, exhibit immunoreactivity to parvalbumin, whereas the latter, termed thalamic "matrix" nuclei, are reactive to calbindin (Molinari et al., 1995; Jones, 2001). The topographically organized projections from thalamic core synapse largely on layer IV and deep layer III cells; the diffuse projections form synapses predominantly in layer I, on the apical dendrites of layer II, III and V cells. (Although the topographic afferents to middle cortical layers, e.g. LGN to primary visual cortex are often thought of as the primary input to sensory neocortex, these fibers actually comprise only about 6% of the synapses onto their primary targets (layer IV neurons), with the majority of the remaining afferents coming largely from lateral cortico-cortical connections (Freund et al., 1985; 1989; Peters and Payne, 1993; Peters et al., 1994; Ahmed et al., 1997).)

Peripheral inputs activate thalamic core cells, which in turn participate in topographic activation of middle cortical layers; e.g., ear > cochlea > auditory brainstem nuclei > ventral subdivision of medial geniculate nucleus (MGv) > A1; in contrast, matrix nuclei are most strongly driven by corticothalamic feedback (Bender 1983; Diamond et al. 1992a; 1992b), supporting a system in which peripheral afferents first activate core nuclei, which in turn activate cortex (via a stereotypical vertically organized pattern: middle layers > superficial layers > deep layers), which then activate both core and matrix nuclei via corticothalamic projections (Mountcastle 1957; Hubel and Wiesel 1977; Di et al. 1990; Kenan-Vaknin and Teyler 1994).

Three primary modes of activity have typically been reported for thalamic neurons: tonic, rhythmic and arrhythmic bursting. The latter appears predominantly during non-REM sleep whereas the first two appear during waking behavior (McCarley et al., 1983; Steriade and Llinas, 1988; McCormick and Feeser, 1990; Steriade et al., 1990; McCormick and Bal, 1994; Steriade and Contreras, 1995). There is strong evidence for ascending influences (e.g., basal forebrain) affecting the probability of response of excitatory cells during the peaks and troughs of such "clocked" inhibitory cycles. The most excitable cells will tend to fire in response even to slight afferent activity whereas less excitable neurons will only be added in response to stronger input; this excitability gradient selectively determines the order in which neurons will be recruited to respond to inputs of any given intensity.

Axons of inhibitory interneurons densely terminate preferentially on the bodies, initial axon segments, and proximal apical dendrites of excitatory pyramidal cells in cortex, and thus are well situated to exert powerful control over the activity of target excitatory neurons. When a field of excitatory neurons receives afferent stimulation, those that are most responsive will activate the local inhibitory cells in their neighborhood, which will in turn inhibit local excitatory cells. The typical time course of an excitatory (depolarizing) postsynaptic potential (PSP) at normal resting potential, in vivo, is brief (15-20 msec), whereas corresponding GABAergic inhibitory PSPs last roughly an order of magnitude longer (80-150 msec) (Castro-Alamancos and Connors, 1997). Thus excitation tends to be brief, sparse, and curtailed by longer and stronger feedback lateral inhibition (Coultrip et al., 1992).

Based on the biological regularities specified, a greatly simplified set of operations has been posited (Rodriguez et al., 2003). Distinct algorithms arise from simulation and analysis of core vs. matrix loops (see Figure 2).

*Thalamocortical "core" circuits*

In the core loop, simulated superficial cells that initially respond to a particular input pattern become increasingly responsive not only to that input but also to a range of similar inputs (inputs that share many active lines, e.g., small Hamming distances from each other), such that similar but distinguishable inputs will come to elicit identical patterns of output from layer II-III cells, even though these inputs would have given rise to slightly different output patterns before synaptic potentiation. These effects can be described in terms of the mathematical operation of clustering, in which sufficiently similar inputs are placed into a single category or cluster. This can yield useful generalization properties, but somewhat counterintuitively, it prevents the system from making fine distinctions among members of a cluster. For instance, four similar inputs may initially elicit four slightly different patterns of cell firing activity in layer II-III cells but after repeated learning / synaptic potentiation episodes, all four inputs may elicit identical activation patterns. Results of this kind have been obtained in a number of different models with related characteristics (von der Malsburg, 1973; Grossberg, 1976; Rumelhart, 1985; Coultrip et al., 1992).

Superficial layer responses activate deep layers. Output from layer VI initiates feedback activation of nucleus reticularis (NRt) (Liu and Jones 1999), which in turn inhibits the core thalamic nucleus (Ct). Since, as described, topography is preserved throughout this sequence of projections, the portions of Ct that become inhibited will correspond topographically to those portions of L.II-III that were active. On the next cycle of thalamocortical activity, the input will arrive at Ct against the background of the inhibitory feedback from NRt, which has been shown to last for hundreds of milliseconds (Huguenard and Prince, 1994; Cox et al., 1997; Zhang et al., 1997). Thus it is hypothesized that the predominant component of the next input to cortex is only the uninhibited remainder of the input, whereupon the same operations as before are performed. The result is that the second cortical response will consist of a quite distinct set of neurons from the initial response, since many of the input components giving rise to that first response are now inhibited relative to their neighbors. Analysis of the second (and ensuing) responses in computational models has shown successive sub-clustering of an input: the first cycle of response identifies the input's membership in a general category of similar objects (e.g., flowers), the next response (a fraction of a second later) identifies its membership in a particular subcluster (e.g., thin flowers; flowers missing a petal), then sub-sub-cluster, etc. Thus the system repetitively samples across time, differentially activating specific target neurons at a series of successive time points, to discriminate among inputs. An initial version of this derived algorithm arose from studies of feedforward excitation and feedback inhibition in the olfactory paleocortex and bulb, and was readily generalized to non-olfactory modalities (vision, audition) whose superficial layers are closely related to those of olfactory cortex, evolutionarily and structurally. The method can be cast in the form of an algorithm (see Table 1) whose costs compare favorably with those in the (extensive) literature on such methods (Ambros-Ingerson et al., 1990; Granger and Lynch, 1991; Gluck and Granger, 1993; Kilborn et al., 1996; Rodriguez et al., 2003). Elaboration of the algorithm has given rise to families of computational signal processing methods whose performance on complex signal classification tasks has consistently equaled or outperformed those of competing methods (Coultrip and Granger, 1994; Kowtha et al., 1994; Granger et al., 1997; Benvenuto et al., 2002).

Table 1

```
for input X
    for C ∈ win(X,W)
        W_j ⇐ W_j + k(X − C)
    end_for
X ⇐ X − mean(win(X,W))
end_for
```

where
X = input activity pattern (vector);
W = layer I synaptic weight matrix;
C = responding superficial layer cells (column vector);
k = learning rate parameter;
win(X,W) = column vector in W most responsive to X before lateral inhibition [e.g., $\forall_j$, max($X \cdot W_j$) ]

[see Rodriguez et al., 2004]

Analysis demonstrates good time and space costs for the derived algorithm. The three time costs for the processing of a given input X are: i) summation of inputs on dendrites; ii) computation of "winning" (responding) cells C; iii) synaptic weight modifications. For n learned inputs each of dimensionality N, in a serial processor, summation can be performed in O(nN) time, computation of winners take O(n) time, and modification of weights is O(N log n). When carried out with appropriate parallel hardware, these three times reduce to O(log N), O(log n), and constant times respectively, i.e., better than linear time. Space costs are similarly calculated: given a weight matrix W, in order to achieve complete separability of n cues, the bottom of the constructed hierarchy must contain at least n units, as the leaves of a tree consisting of log Bn hierarchical layers, if B is the average branching factor at each level. Thus the complete hierarchy will contain ~ n[B/(B-1)] units, and the space required to learn n cues of dimensionality N will be linear, or O(nN) time (Ambros-Ingerson et al., 1990; Kilborn et al., 1996; Rodriguez et al., 2003).

*Thalamocortical "matrix" circuits*

In contrast to the topography-preserving projections in the "core" loop between Ct and cortex, the diffuse projections from L.V to Mt and from Mt back to cortex in the "matrix" loop are modeled as sparsifying and orthogonalizing their inputs, such that any structural relationships that may obtain among inputs are not retained in the resulting projections. Thus input patterns in Mt or in L.V that are similar may result in very different output patterns, and vice versa. As has been shown in previously published studies, due to the nontopographic nature of layer V and Mt, synapses in L.V are very sparsely selected to potentiate, i.e., relatively few storage locations (synapses) are used per storage/learning event (Granger et al., 1994; Aleksandrovsky et al., 1996; Whitson 1998; Rodriguez et al., 2003). For purposes of analysis, synapses are assumed to be binary (i.e., assume the lowest possible precision, that synapses are either naïve or potentiated). A sequence of length L elicits a pattern of response according to the algorithm given above for superficial layer cells. Each activated superficial cell C in turn activates deep layer cells. Feedforward activity from the matrix thalamic nucleus Mt also activates L.V. Synapses on cells receiving activation from both sources (the intersection of the two inputs) become potentiated, and the activity pattern in layer V is fed back to Mt. The loop repeats for each of the L items in the sequence, with the input activity from each item interacting with the activity in Mt from the previous step (see Rodriguez et al., 2003).

The activation of layer V in rapid sequence via superficial layers (in response to an element of a sequence) and via Mt (corresponding to feedback from a previous element in a sequence) selects responding cells sparsely from the most activated cells in the layer (Coultrip et al., 1992) and selects synapses on those cells sparsely as a function of the sequential pattern of inputs arriving at the cells. Thus the synapses potentiated at a given

step in L.V correspond both to the input occurring at that time step together with orthogonalized feedback arising from the input just prior to that time step. The overall effect is "chaining" of elements in the input sequence via the "links" created due to coincident layer V activity corresponding to current and prior input elements. As in the operating rule described by (Granger et al., 1994) the sparse synaptic potentiation enables L.V cells to act as a novelty detector, selectively responding to those sequential strings that have previously been presented. The implicit data structures created by the operation of this system are trees in which initial sequence elements branch to their multiple possible continuations ("tries," Knuth, 1997). Sufficient information therefore exists in the stored memories to permit completion of arbitrarily long sequences from just prefixes that uniquely identify the sequence. Thus the sequence "Once upon a time" may elicit (or "prime") many possible continuations whereas "Four score and seven" elicits a specific continuation.

The resulting algorithm (see Table 2) can be characterized in terms of computational storage methods that are used when the number of actual items that occur are far fewer than those that in principle could occur. The number of possible eight-letter sequences in English is 268, yet the eight-letter words that actually occur in English number less than 10,000, i.e., less than one ten-millionth of the possible words. The method belongs to the family of widely-used and well-studied data storage techniques of "scatter storage" or "hash" functions, known for the ability to store large amounts of data with extreme efficiency. Both analytical results and empirical studies have found that the derived matrix loop method requires an average of less than two bits (e.g., just two low-precision synapses) per complex item of information stored. The method exhibits storage and successful retrieval of very large amounts of information at this rate of storage requirement, leading to extremely high estimates of the storage capacity of even small regions of cortex. Moreover, the space complexity of the algorithm is linear, or O(nN) for n input strings of dimensionality N; i.e., the required storage grows linearly with the number of strings to be stored (Granger et al., 1994; Aleksandrovsky et al., 1996; Whitson 1998; Rodriguez et al., 2003).

Table 2

```
for input sequence X(L)
        for C ∈ TopographicSuperficialResponse(X(L))
                for V(s) ∈ C ∩ NNtResponse(X(L-1))
                   Potentiate( V(s) )
                   NNt(L) ⇐ NontopographicDeepResponse(V)
                end_for
        end_for
end_for

(where
L = length of the input sequence;
C = columnar modules activated at step X(L);
V(s) = the synaptic vector of responding layer V cell,
NNt(L) = response of nonspecific thalamic nucleus to feedback from layer V.)
```

[see Rodriguez et al., 2004]

**Circuits of the Striatal Complex**

The basal ganglia, or striatal complex, is a collection of disparate but interacting structures including the caudate-putamen, globus pallidus, subthalamic nucleus, and substantia nigra pars compacta (SNc). It is the second largest telencephalic component after thalamocortical circuits, phylogenetically predating the mammals and operating as the primary brain engine for reptiles. In reptiles, and in mammals with relatively small brain / body size ratios, the primary anatomical efferents of striatal complex descend to brainstem nuclei, presumably driving complex sequential motor movements including species-specific behaviors (e.g., stalking, grooming). In all mammals, the striatal complex is tightly linked to anterior neocortex. As brain size grows allometrically larger, a number of fundamental relationships between neocortex and striatal complex are altered, predominantly due to disproportionate growth of the anterior neocortex:

1) Fascicular growth: the connection pathways between anterior and posterior neocortex grow, greatly increasing the relative size of the large axon bundles (fasciculi) connecting them.
2) Cortico-striatal loop growth: the dual efferent pathways from the striatal complex, one descending to brainstem nuclei and one ascending to anterior neocortex (via ventral thalamus), change in relative size: the cortical outputs grow far larger than the descending outputs.
3) Pyramidal tract growth: descending outputs from anterior cortex to motor systems grow disproportionately larger than the descending striatal motor outputs.

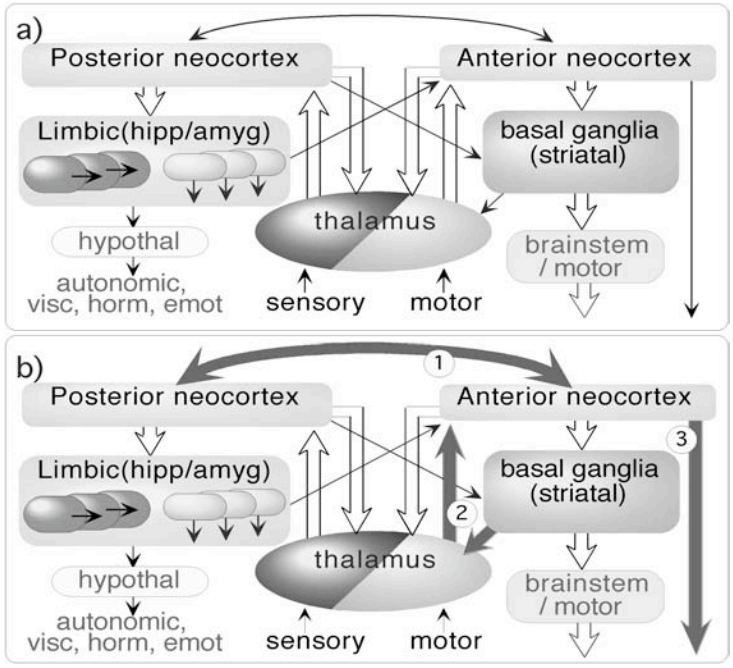


Figure 3. Primary constituents of mammalian telencephalon and subcortical connections. a) Posterior neocortex makes strong reciprocal connections with dorsal thalamic nuclei, as well as connections with the limbic system, which in turn provides descending projections to hypothalamic regulatory systems. Anterior cortex, with corresponding strong reciprocal connections to ventral thalamic nuclei, also projects to the basal ganglia (striatal complex), which in turn projects to brainstem nuclei. b) In mammals with larger brain-body size ratios, a number of allometric changes occur, of which three of the largest are highlighted: 1) growth of the fascicular connections between posterior and anterior cortex, 2) growth of efferent pathways from striatal complex to cortex via ventral thalamus, and concomitant reduction of the relative size of the descending projections from striatal complex, and 3) increase in size of the descending pyramidal tract projections from cortex to motor systems, as though in compensation for the reduced descending striatal pathway.

These changes in anatomical design are illustrated in Figure 3; they grow disproportionately with increase in brain-body size ratios, becoming most notable in humans. In relatively small-brained mammals such as mice, the primary motor area of neocortex is an adjunct to the striatally driven motor system. Whereas damage to motor cortex in mice causes subtle behavioral motor impairments, damage to motor cortex in humans causes paralysis. In this example of encephalization of function (Jackson, 1925; Ferrier, 1876; Karten, 1991, Aboitiz, 1993) motor operations are increasingly 'taken over' by cortex as the size of the pyramidal tract overtakes that of the descending striatal system. The role of the striatal complex in mammals with large brain-body ratios is presumably altered to reflect that its primary inputs and outputs are now anterior neocortex; in other words, it is now primarily a tool or "subroutine" available for query by anterior cortex. Its operations then are most profitably viewed in light of its dual utility either as organizer of complex motor sequences (in small brained mammals) or as informant to anterior cortex (in large brained mammals).

Figure 4 schematically illustrates the primary components of the striatal complex. Striking differences from the circuitry of the thalamocortical system include the very different, apparently specialized designs of the components: matrisomes (matrix), striosomes (patch), globus pallidus, pars interna and externa (pallidum), tonically active cholinergic neurons (TACs), and substantia nigra pars compacta (SNc). In contrast with the thalamocortical system, dominated by glutamatergic and local GABAergic neurotransmitter systems, the striatal complex depends on a broad variety of different neurotransmitter pathways including GABA, glutamate (Glu), dopamine (DA), acetylcholine (ACh), and Substance P (Sp) among others.

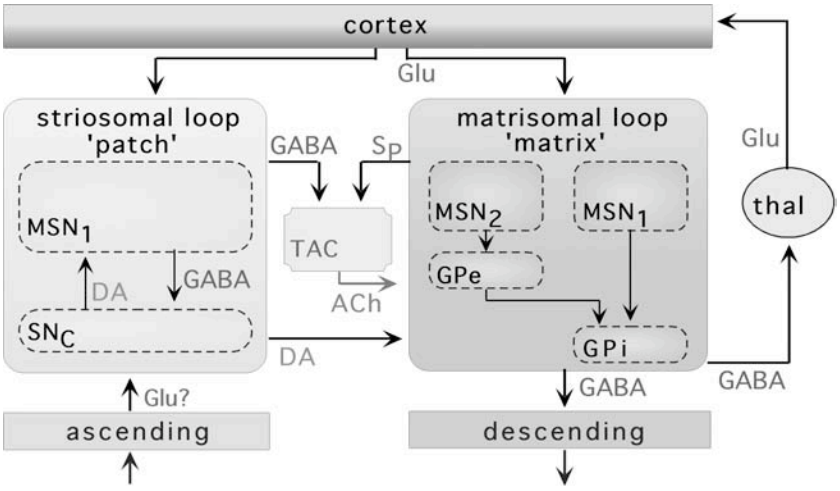


Figure 4. Schematic illustration of striatal complex (basal ganglia). Glutamatergic cortical afferents activate both matrisomal (matrix) and striosomal (patch) targets. Two GABAergic matrix pathways from medium spiny neurons (MSN) through pallidum project to brainstem motor systems, and back to cortical targets via thalamus. Patch projects GABAergically to SNc and to tonically active cholinergic neurons (TACs), which in turn make cholinergic projections to matrix. Both patch and matrix receive dopaminergic input from substantia nigra, pars compacta (SNc), which in turn receives ascending information conveying external "reward" and "punishment."

The two pathways from cortex through the matrix components of the striatal complex involve different subpopulations of cells in matrisomes (matrix): i) MSN1 neurons, which express dopamine D1 receptors, project to globus pallidus pars interna (GPi), which in turn project to ventral thalamus and back to cortex; ii) MSN2 neurons, which express D2 receptors, project to globus pallidus pars externa (GPe), which in turn projects to GPi (and thence to thalamus and cortex). Unlike thalamocortical circuits, in which long axon projections are glutamatergic, the MSN and GP projections are GABAergic, inhibiting their targets. Thus cortical glutamatergic activation of MSN1 cells causes inhibition of GPi cells, which otherwise inhibit thalamic and brainstem targets; hence MSN1 cell activation disinhibits, or enhances, thalamic activation of cortex and

striatal activation of brainstem nuclei. In contrast, an extra GABAergic link is intercalated in the pathway from MSN2 neurons to the output stages of matrix; it can be seen that activation of MSN2 neurons decreases thalamic activation of cortex and striatal activation of brainstem nuclei. The two pathways from MSN1 and MSN2 neurons are thus termed "go" and "stop" pathways respectively, for their opposing effects on their ultimate motor targets. A complex combination of activated ("go") and withheld ("stop") muscle responses (e.g., to stand, walk, throw) can be created by coördinated operation over time of these pathways.

Two primary afferents to the striosomes are cortical and ascending inputs. The former are the same as the inputs to matrix (despite the schematized depiction in the figure, patch components are distributed throughout, and colocalized with, matrix). The ascending inputs to patch denote "reward" and "punishment" information and have been shown to up- and down-regulate dopamine responses from SNc (as well as other dopaminergic sites) in response to external stimuli carrying innate or learned valences (e.g., water to a thirsty organism). A cortically triggered action, followed by an ascending DA reward signal from SNc to patch, selectively enhances active cortical glutamatergic synapses onto both matrix and patch targets. Patch output back to SNc then inhibits DA response, so that increased cortical activation of patch (via enhanced synaptic contacts) will come to limit the DA input from SNc. On any given trial, then, the size of the DA signal from SNc comes to reflect the size of the actual ascending DA input (i.e., the reward signal) that occurred over previous trials. Thus with repeated experience, adaptive changes occur in both matrix and patch: initially random matrix responses to a cortical input become increasingly selected for responses that produce reward, and initial naïve striosomal responses will become increasingly good "predictors" of the size of the reward (or punishment) expected to ensue as a result of the action (Brucher, 2000).

Tonically active cholinergic neurons represent a small fraction (< 5%) of the number of cells in the striatal complex yet densely contact cells throughout matrix; thus they likely play a modulatory role rather than conveying specific information. The GABAergic inhibition of these cells by patch will come to increase for those patch responses that lead to reward, since in these instances the cortical drivers of these patch responses become synaptically enhanced. Thus in those circumstances where cortical inputs lead to expected reward, TAC cells will tend to have less excitatory effect on matrix. Since the TAC afferents to matrix are dense and nontopographic, they represent a random "background noise" input, which can increase variance in selected matrix responses to cortical inputs, making the striatally-selected motor response to a cortical input somewhat nondeterministic. The resulting behavior should appear "exploratory," involving a range of different responses to a given stimulus. With repeated exposure to the stimulus, as some responses differentially lead to reward, the corresponding synapses in both matrix and patch will be enhanced, leading to the increased probability of the selected responses (via matrix) and increasingly accurate "prediction" of reward size (via patch), as stated. An additional effect of synaptic increase in patch is that the afferent patch stimulation of TAC cells will increase, inhibiting TAC activity, and diminishing the breadth of the exploratory variability in the response just described. Thus as rewards occur, not only will reward-associated responses be increasingly selected by matrix, but the variability among those responses will decrease.

**Applications and Implementations**

The coordinated activity of simulated thalamocortical and cortico-striatal loops has yielded computational methods applicable to a variety of domains with surprising efficacy. Analysis of military signals showed these brain circuit derived methods specifically outperforming not only standard statistical approaches such as Bayesian nets but also methods based on typical "artificial neural network" approaches such as backpropagation (Kowtha et al., 1996). Analysis of electroencephalographic (EEG) information in populations of Alzheimer's patients and matched control populations have demonstrated that these brain circuit methods outperform even advanced statistical approaches (projection pursuit); the resulting analyses, applied to the task of classifying the subjects by diagnostic category (Alzheimer's vs. normal) proved significantly more effective than competing approaches (Benvenuto et al., 2002). These and related brain circuit algorithms are currently under development for a range of additional applications.

The simplifying constraints that led to derivation of these novel algorithms implies added efficiencies from direct hardware implementations. Figure 5 illustrates candidate designs for three constituent elements of thalamocortical circuitry: a) a "ladder" circuit that determines which cells in a thalamic nucleus will respond in terms of their order of excitability; b) a "winner take all" circuit that implements the effects of lateral inhibition in a typical cortical layer, selecting the most responsive excitatory unit(s) and suppressing responses from others; c) a sparse, random synaptic matrix connecting input axons (e.g., from thalamic matrix nuclei) to dendrites in neocortical layer I, via synaptic elements of various designs (e.g., floating gates; see Mead, 1989; Hasler et al., 1995; Shoemaker et al., 1991; 1996).

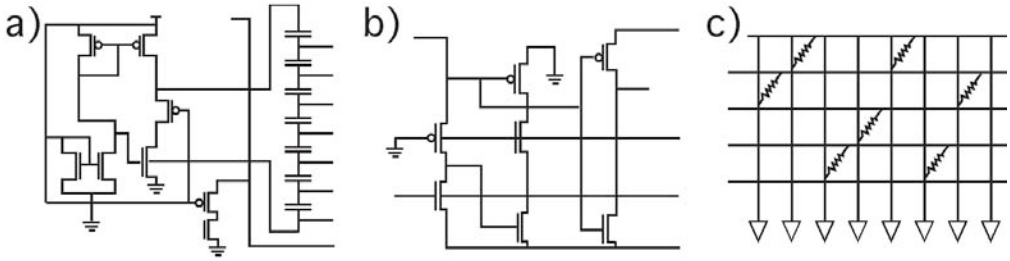


Figure 5. Designs for three constituents of thalamocortical circuitry. a) A "ladder" circuit determining order of cell response by order of excitability. b) A competitive or "winner take all" circuit implementing the effects of lateral inhibition in typical cortical local circuits (see text). c) sparse, random synaptic matrix connecting input axons (horizontal) with cortical dendrites (vertical) in nontopographic connectivity layers such as neocortical layer I.

Tasks that remain unsolved by current approaches range from complex visual processing (even simple static images, let alone complex movies) and auditory tasks (unconstrained voice recognition and speech processing) to broadened medical applications (increased diagnostic use as well as utility for amelioration and treatment of neurological impairments). Algorithms constrained by the simplicities and weaknesses of actual brain circuits and their constituents (Ambros-Ingerson et al., 1990; Anton et al., 1991; Coultrip et al., 1992; Coultrip and Granger, 1994; Granger et al., 1994; Aleksandrovsky et al., 1996; Kilborn et al., 1996; Shimono et al., 2000; Granger 2002; Benvenuto et al., 2002; Rodriguez et al., 2003) have been shown to have powerful computational properties and good space and time complexity costs, indicating their ability to scale to large size, and have given rise to novel systems for the analysis of complex time-varying signals in military, commercial, and medical applications domains. Current work is focused on the use of these findings for novel circuit designs, new approaches to medical signal processing, and enhancement of the two-way communication between brain systems and extrinsic systems. Although initial visual face and character recognition systems exist, these are kept from widespread use by their shortcomings such as high false-alarm rates. Similarly, current voice processing systems (e.g., automated telephone operators) can process only simple and very brief utterances, and remain limited by their inability to operate on extended speech streams. Development of these and related systems will not only advance our scientific understanding of mammalian telencephalic operation, and provide novel device designs applicable to a broad variety of commercial, military and medical task domains; they will also increasingly enable two-way communication between brain circuits and extrinsic systems, with concomitant medical and scientific benefits.

**REFERENCES**

Aboitiz F (1993) Further comments on the evolutionary origin of mammalian brain. Med Hypoth 41: 409-418.

Ahmed B, Anderson JC, Martin KAC, Nelson JC (1997) Map of the synapses onto layer 4 basket cells of the primary visual cortex of the cat. J Comp Neurol 380:230-242.

Aleksandrovsky B, Whitson J, Garzotto A, Lynch G, Granger R (1996) An algorithm derived from thalamocortical circuitry stores & retrieves temporal sequences. In: IEEE Int'l Conf Patt Rec, 550-554.

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. Biol Cybern 65:407-414.

Benvenuto J, Jin, Y., Casale, M., Lynch, G., Granger, R. (2002) Identification of diagnostic evoked response potential segments in Alzheimer's Disease. Exper Neurol 176:269-276.

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.

Braitenberg V, Schüz A (1998) Cortex: statistics and geometry of neuronal connectivity, NY: Springer.

Brucher F (2000) Unpublished dissertation, U. Calif.

Castro-Alamancos M, Connors B (1997) Thalamocortical synapses. Prog Neuro 51:581-606.

Coultrip R, Granger R (1994) LTP learning rules in sparse networks approximate Bayes classifiers via Parzen's method. Neural Networks 7:463-476.

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

Cox C, Huguenard J, Prince D (1997) Nucleus reticularis neurons mediate diverse inhibitory effects in thalamus. Proc Natl Acad Sci 94:8854-8859.

Deschenes M, Veinante P, Zhang Z (1998) The organization of corticothalamic projections: reciprocity v. parity. Brain Res Rev 28:286-308.

Di S, Baumgartner C, Barth DS (1990) Laminar analysis of extracellular field potentials in rat vibrissa/barrel cortex. J Neurophysiol 63:832-840.

Diamond M, Armstrong-James M, Ebner F (1992a) Somatic sensory responses in the rostral sector of the posterior group (POm) and in the ventral posterior medial nucleus (VPM) of the rat thalamus. J Comp Neurol 318:462-476.

Diamond M, Armstrong-James M, Budway M, Ebner F (1992b) Somatic sensory responses in the rostral sector of the posterior group (POm) and ventral posterior medial nucleus (VPM) of rat thalamus. J Comp Neurol 319:66-84.

Ferrier D (1876) The functions of the brain. London: Smith, Elder.

Freund TF, Martin KA, Whitteridge D (1985) Innervation of cat visual areas 17 and 18 by physiologically identified X- and Y-type thalamic afferents. I. Arborization patterns and quantitative distribution of postsynaptic elements. J Comp Neurol 242:263-274.

Freund TF, Martin KA, Soltesz I, Somogyi P, Whitteridge D (1989) Arborisation pattern and postsynaptic targets of physiologically identified thalamocortical afferents in striate cortex of the macaque monkey. J Comp Neurol 289:315-336.

Galuske RA, Schlote W, Bratzke H, Singer W (2000) Interhemispheric asymmetries of the modular structure in human temporal cortex. Science 289:1946-1949.

Gazzaniga MS (2000) Regional differences in cortical organization. Science 289:1887-1888.

Gluck M, Granger R (1993) Computational models of the neural bases of learning and memory. Ann Rev Neurosci 16:667-706.

Granger R, Lynch G (1991) Higher olfactory processes: perceptual learning and memory. Curr Opin Neurobiol 1:209-214.

Granger R, Whitson J, Larson J, Lynch G (1994) Non-Hebbian properties of long-term potentiation enable high-capacity encoding of temporal sequences. Proc Natl Acad Sci 91:10104-10108.

Granger R, Wiebe S, Taketani M, Ambros-Ingerson J, Lynch G (1997) Distinct memory circuits comprising the hippocampal region. Hippocampus 6:567-578.

Granger R (2002) Neural Computation: Olfactory cortex as a model for telencephalic processing. In: Learning & Memory (J.Byrne Ed) pp 445-450.

Grossberg S (1976) Adaptive pattern classification and universal recoding. Biol Cybern 23:121-134.

Hasler P, Diori C, Minch B, Mead C (1995) Single transistor learning synapses. Adv Neural Inf Proc Sys, 7: 817-824.

Hubel D, Wiesel T (1977) Functional architecture of macaque monkey visual cortex. Proc R Soc Lond B Biol Sci 198:1-59.

Huguenard JR, Prince DA (1994) Clonazepam suppresses GABAB-mediated inhibition in thalamic relay neurons through effects in nucleus reticularis. J Neurophysiol 71:2576-2581.

Jackson J H (1925) Neurological fragments. London: Oxford University Press.

Jones E (1981) Functional subdivision and synaptic organization of mammalian thalamus. Int Rev Phys 25:173-245.

Jones E (1998) A new view of specific and nonspecific thalamocortical connections. Adv Neurol 77:49-71.

Jones E (2001) The thalamic matrix and thalamocortical synchrony. Tr Neurosci 24:595-601.

Jones E, Hendry S (1989) Differential Calcium Binding Protein Immunoreactivity Distinguishes Classes of Relay Neurons in Monkey Thalamic Nuclei. Eur J Neurosci 1:222-246.

Karten H (1991) Homology and evolutionary origins of 'neocortex'. Brain Beh Evol, 38: 264-272.

Keller A, White E (1989) Triads: a synaptic network component in cerebral cortex. Brain Res 496:105-112.

Kenan-Vaknin G, Teyler T (1994) Laminar pattern of synaptic activity in rat primary visual cortex: comparison of in vivo and in vitro studies employing current source density analysis. Brain Res 635:37-48.

Kilborn K, Granger, R, Lynch, G (1996) Effects of LTP on response selectivity of simulated cortical neurons. J Cog Neurosci 8:338-353.

Knuth D (1997) The art of computer programming. Reading, MA: Addison-Wesley.

Kowtha V, Satyanarayana, P., Granger, R., Stenger, D. (1994) Learning and classification in a noisy environment by a simulated cortical network. In: 3rd Ann Comp Neural Sys, 245-250. Boston: Kluwer

Liu X, Jones E (1999) Predominance of corticothalamic synaptic inputs to thalamic reticular nucleus neurons in rat. J Comp Neurol 414:67-79.

Lorente de No R (1938) Cerebral cortex: Architecture, intracortical connections, motor projections. In: Physiology of the nervous system (Fulton J, ed), pp 291-340. London: Oxford.

McCarley R, Winkelman J, Duffy F (1983) Human cerebral potentials associated with REM sleep rapid eye movements: links to PGO waves and waking potentials. Brain Res 274:359-364.

McCormick D, Feeser H (1990) Functional implications of burst firing and single spike activity in lateral geniculate relay neurons. Neuroscience 39:103-113.

McCormick D, Bal T (1994) Sensory gating mechanisms of the thalamus. Curr Opin Neuro 4:550-556.

Mead C (1989) Analog VLSI and Neural Systems. New York: Addison-Wesley.

Molinari M, Dell'Anna M, Rausell E, Leggio M, Hashikawa T, Jones E (1995) Auditory thalamocortical pathways defined in monkeys by calcium-binding protein immunoreactivity. J Comp Neur 362:171-194

Mountcastle V (1957) Modality and topographic properties of single neurons of cat's somatic sensory cortex. J Neurophysiol 20:408-434.

Mountcastle V (1978) Brain mechanisms for directed attention. J R Soc Med 71:14-28.

Peters A, Payne B (1993) Numerical relationships between geniculocortical afferents and pyramidal cell modules in cat primary visual cortex. Cerebral Cortex 3:69-78.

Peters A, Payne B, Budd J (1994) A numerical analysis of the geniculocortical input to striate cortex in the monkey. Cerebral Cortex 4:215-229.

Rockel A, Hiorns R, Powell T (1980) Basic uniformity in structure of the neocortex. Brain 103:221-244.

Rodriguez A, Whitson J, Granger R (2004) Derivation and analysis of basic computational operations of thalamocortical circuits. J. Cog Neurosci, 16: 856-877.

Rumelhart D, Zipser, D (1985) Feature discovery by competitive learning. Cognitive Sci 9:75-112.

Shimono K, Brucher F, Granger R, Lynch G, Taketani M (2000) Origins and distribution of cholinergically induced beta rhythms in hippocampal slices. J Neurosci 20:8462-8473.

Shoemaker P, Hutchens C, Patil S (1992) Hierarchical clustering network based on a model of olfactory processing. Analog Integr Circ & Signal Proc, 2: 297-311.

Shoemaker P, Hutchens C, Copper J, Lagnado I (1996) Floating-gate synapse in thinfilm SOS and its application to a learning neural network chip. Analog Integr Circ & Signal Proc, 5: 104-110.

Steriade M, Llinas R (1988) The functional states of thalamus and associated neuronal interplay. Phys Rev 68:649-742.

Steriade M, Contreras D (1995) Relations between cortical and thalamic cellular events during transition from sleep patterns to paroxysmal activity. J Neurosci 15:623-642.

Steriade M, Datta S, Pare D, Oakson G, Curro R (1990) Neuronal activities in brainstem cholinergic nuclei related to tonic activation processes in thalamocortical systems. J Neurosci 10:2541-2559.

Szentagothai J (1975) The 'module-concept' in cerebral cortex architecture. Brain Res 95:475-496.

Valverde F (2002) Structure of the cerebral cortex. Intrinsic Organization and comparative analysis of the neocortex. Rev Neurol 34:758-780.

von der Malsburg C (1973) Self-organization of orientation sensitive cells in striate cortex. Kybernetik 14:85-100.

White E, Peters A (1993) Cortical modules in the posteromedial barrel subfield (Sml) of the mouse. J Comp Neurol 334:86-96.

Whitson J (1998) Unpublished dissertation, U. Calif.

Zhang S, Huguenard J, Prince D (1997) GABAa receptor mediated Cl- currents in rat thalamic reticular and relay neurons. J Neurophysiol 78:2280-2286.