A Brain-Like Computer for Cognitive Software Applications: The Ersatz Brain Project

James A. Anderson
Department of Cognitive and Linguistic Sciences
Brown University
Providence, RI 02912
James_Anderson@brown.edu

Abstract

We want to design a suitable computer for the efficient execution of the software now being developed that will display human-like cognitive abilities. Examples of these potential software applications include natural language understanding, text processing, conceptually based internet search, natural human-computer interfaces, cognitively based data mining, sensor fusion, and image understanding. Requirements of the proposed software are primary in shaping our hardware design. The hardware architecture design is based on a few ideas taken from the anatomy of mammalian neo-cortex. In common with other such attempts it is a massively parallel, two-dimensional array of CPUs and their associated memory. However, the design used in this project (1) uses an approximation to cortical computation called the ***network of networks*** *which holds that the basic computing unit in the cortex is not a single neuron but small groups of them working together in attractor networks; and (2) assumes connections in cortex are very sparse. The resulting architecture depends largely on local data movement.*

1. Goal of the Ersatz Brain Project

We want to design a suitable computer for the efficient execution of the software now being developed that will display human-like cognitive abilities. We base our fundamental architecture on a few ideas taken from the design of the mammalian neocortex, therefore we have named our project the Ersatz Brain Project. We gave it this name because, given our current state of knowledge, our goal can only be to build a shoddy second-rate brain. Even so, such a computer may be a starting point for projects that realize systems with the power, flexibility, and subtlety of the actual brain. We suggest that a "cortex-power" massively parallel computer is now technically feasible, requiring on the order of a million simple CPUs and a terabyte of memory for connections between CPUs.

Many groups over many years have proposed the construction of brain-like computers. To orient a reader to a complex history, see [3, 6, 7].

The human brain is composed of on the order of 10^{10} neurons, connected together with at least 10^{14} neural connections. These numbers are likely to be underestimates. Biological neurons and their connections are extremely complex electrochemical structures that require substantial computer power to model even in poor approximations. The more realistic the neuron approximation, the smaller is the network that can be modeled. Worse, there is strong evidence that **a bigger brain is a better brain**, thereby increasing greatly computational demands if biology is followed closely. **We need good approximations to build a useful brain-like computer**.

2. The Ersatz cortical computing unit The Network of Networks

Received wisdom has it that neurons are the basic computational units of the brain. However the Ersatz Brain Project is based on a different assumption. We will use the **Network of Networks [NofN]** approximation to structure the hardware and to reduce the number of connections required [4,8].

We assume that the basic neural computing units are not neurons, but small (perhaps 10^3 -10^4 neurons) attractor networks, that is, non-linear networks (modules) whose behavior is dominated by their

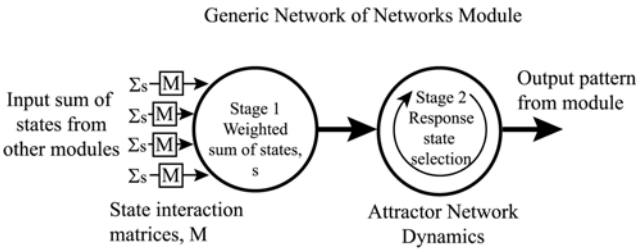

Figure 1. Generic Network of Networks module.

attractor states that may be built in or acquired through learning. There are many types of such networks. (See [2, 13] among others.) Basing computation on module attractor states and not directly on neural discharges reduces the dimensionality of the system, allows a degree of intrinsic noise immunity, and allows interactions between networks to be approximated as interactions between attractor states. Interactions between modules are similar to the generic neural net unit except scalar connection strengths are replaced by **state interaction matrices**. (Figure 1) The state interaction matrix gives the effect of an attractor state in one module upon attractor states in a module connected to it.

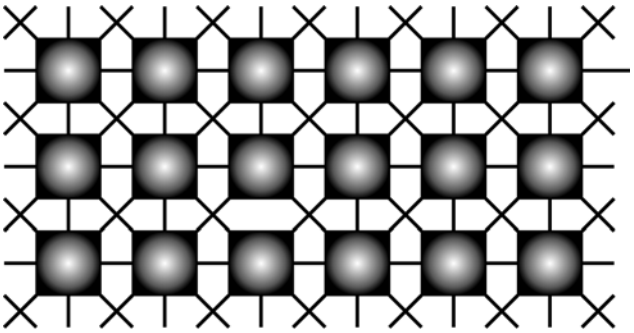

Figure 2. Network of Networks 2-D modular architecture.

Because attractors are derived from neuron responses, it is potentially possible to merge neuron-based preprocessing with attractor dynamics. The basic Network of Networks system is composed of very many of these basic modules arranged in a two-dimensional array. (Figure 2).

2.1. Cortical columns

The most likely physiological candidate for the basic component of a modular network is the cortical column. Cortex is a large two-dimensional layered sheet, with a rather homogeneous cellular structure. One of its most prominent anatomical features is the presence of what are called columns, local groups of cells running perpendicular to the cortical surface. There are several types of columns present at different spatial scales. There was a recent special issue of the journal *Cerebral Cortex* devoted to cortical columns, their functions, and their connections. The introduction by Vernon Mountcastle [] provides a useful summary of the two types of columns that will most concern us.

"The basic unit of cortical operation is the minicolumn …. It contains of the order of 80-100 neurons, except in the primate striate cortex, where the number is more than doubled. The minicolumn measures of the order of 40-50 μm in transverse diameter, separated from adjacent minicolumns by vertical cell-sparse zones which vary in size in different cortical areas. Each minicolumn has all cortical phenotypes, and each has several output channels. … By the 26th gestational week the human neocortex is composed of a large number of minicolumns in parallel vertical arrays." (Mountcastle, 2003, p. 2)

Minicolumns form a biologically determined structure of stable size, form and universal occurrence in animals with reasonably complex cortices. What are often called "columns" in the literature are collections of minicolumns that seem to form functional units. Probably the best-known examples of functional columns are the orientation columns in V1, primary visual cortex. Vertical electrode penetrations in V1, that is, parallel to the axis of the column, found numerous cells that respond to oriented visual stimuli with the same orientation. Functional columns have since been found in most areas of cerebral cortex. The cells in a column are not identical in their properties and, outside of orientation, may vary widely in their responses to contrast, spatial frequency, etc. Clusters of minicolumns make up functional columns:

"Cortical columns are formed by the binding together of many minicolumns by common input and short range horizontal connections. The number of minicolumns per column varies probably because of variation in size of the cell sparse inter-minicolumnar zones; the number varies between 50 and 80. Long-range, intracortical projections link columns with similar functional properties. Columns vary between 300 and 500 μm in transverse diameter, and do not differ significantly in size between brains that may vary in size over three orders of magnitude … Cortical expansion in evolution is marked by increases in surface area with little change in thickness … ." [19, p. 3]

If we assume there are 100 neurons per minicolumn, and roughly 80 minicolumns per functional column, this suggests there are roughly 8,000 neurons in a column.

2.2. Connectivity

Besides modular structure, an important observation about the brain in general that strongly influences how it works is its very sparse connectivity. Although a given neuron in cortex may have on the order of 100,000 synapses, there are more than 10^{10} neurons in the brain. Therefore, the fractional connectivity is very low; for the previous numbers it is 0.001 per cent. Low connectivity has two implications: first, connections are expensive biologically since they take up space, use energy, and are hard to wire up correctly, and, second, the connections that are there are precious and their pattern of connection must be under tight control. This observation puts severe constraints on the structure of large-scale brain models. **One implication of expensive connections is that short local connections are relatively cheap compared to long range ones. The cortical approximation we will discuss makes extensive use of local connections for computation**.

2.3. Interactions between Modules

Let us discuss in a little more detail how to analyze interactions between small groups of modules. The attractor model we will use is the BSB network [2] because it is simple to analyze using the eigenvectors and eigenvalues of its local connections.

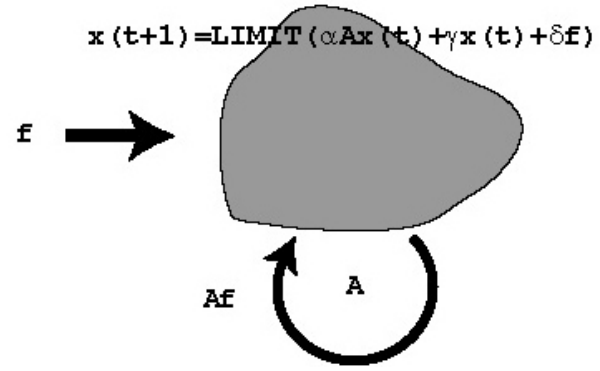

Figure 3. Basic BSB module

The BSB model (Figure 3) was proposed several years ago as a simple feedback nonlinear neural network. Its dynamics broke conveniently into a linear and a non-linear part. The analysis assumed it was a recurrent feedback network (See Figure 3). An input pattern, **f**, appears on an interconnected group of neurons, say from a sensory input. There is vector feedback through a connection matrix, **A, w**eighted by a constant, α, and an inhibitory decay constant, γ, with amount of inhibition a function of the amplitude of the activity. The state of the system is $\mathbf{x}(t)$. The system is linear up to the point where the LIMIT operation starts to operate. LIMIT($\mathbf{x}(t)$) is a hard limiter with an upper and lower threshold. Sometimes it is useful to maintain the outside input at some level; sometimes it is useful to remove the outside input. The constant δ performs this function. The basic algorithm for BSB is:

$$\mathbf{x}(t+1) = \text{LIMIT}\ (\alpha A\mathbf{x}(t) + \gamma\mathbf{x}(t) + \delta\mathbf{f}).$$

For our discussion of module interactions, let us assume f is an eigenvector of the connection matrix A with eigenvalue λ. For simplicity, let us also assume that the input, **f,** vanishes after $t=1$, that is, $\delta = 0$. Then in the linear region the behavior of the system is easy to compute. After t steps, starting from $t=1$, $\mathbf{x}(t+1) = (\alpha\lambda + \gamma)^t\,\mathbf{f}$. The temporal dynamics can show three kinds of qualitative behavior, based on this constant expression: (1) If $(\alpha\lambda + \gamma) > 1$ the amplitude of **f** grows without bound. (2) If $(\alpha\lambda + \gamma) = 1$ the amplitude is constant. (3) If $(\alpha\lambda + \gamma) < 1$ the amplitude approaches zero.

These are functionally very different outcomes. In the nonlinear BSB network with growing activity, the state of the system will reach an attractor state based on the LIMIT function, usually the corner of a hypercube of limits. In practice, if **f** is an eigenvector the final BSB attractor state is close to the direction of **f**.

The transitions from growing activity to no activity and

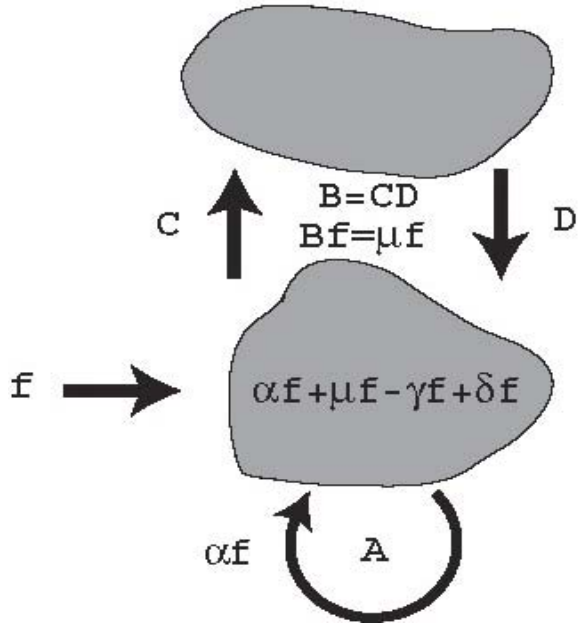

Figure 4. Two modules linked by an associative pattern feedback loop.

the opposite are under control of α, γ and λ, and, as we shall see, parameters derived from model interactions

with other modules. These relationships provide a way of controlling network behavior.

Let us consider the implications of connections from other structures. In particular, we know two relevant

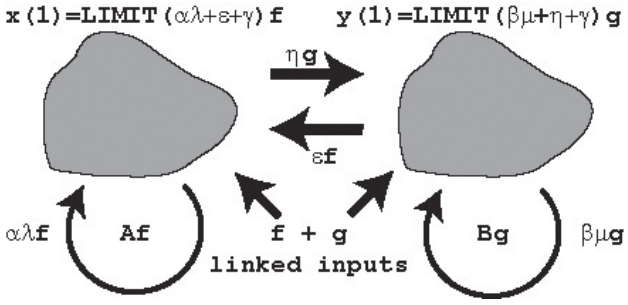

Figure 5. Two coupled modules receiving a pair of linked input patterns, **f** and **g**.

facts about cortex: (1) one cortical region projects to many others and, (2) there are back projections from the target regions to the first region that are at least of equal size as the upward projections. Therefore let us consider the anatomy shown in Figure 4. Note that the upward association can have any vector as an output association. The downward association has the eigenvector, **f**, as its output, perhaps mixed with other eigenvectors. Hebbian learning operating at the higher and lower ends of the loop will tend to construct **f** as an eigenvector because it is present at both input and output. These loops are reminiscent of the Bidirectional Associative Memory [BAM] of Kosko [14]. Analysis again is easy if **f** is an eigenvector. Let us assume δ is zero, as before. Analysis is the same as in the basic model, except that we now have a new term in the BSB equation corresponding to the contribution from the loop. Let the loop feedback eigenvalue be μ. Then we have a modified equation, after *t* steps, (assuming we have not reached the LIMIT) starting from $t=1$, $\mathbf{x}(t+1) = (\alpha\lambda+\mu+\gamma)^t \mathbf{f}$. (Figure 4.**)**

We can also propose a computational mechanism for binding together a multi-module input pattern using local connections. If two modules are driven by two simultaneously presented patterns, f and g, associative links between f and g can be formed, increasing the gain of the module and therefore the likelihood that later simultaneous presentation of the patterns will lead to module activity reaching a limit. Local pattern co-occurrence will form local pattern associative bonds, letting larger groupings act as a unit, that is, a unit that increases and decreases in activity together. Large-scale patterns will tend to bind many module activities together since they take place embedded in a larger informational structure. (Figure 5.)

2.4. Interference Patterns and Traveling Waves

Because we have suggested many important connections are local, much information processing takes

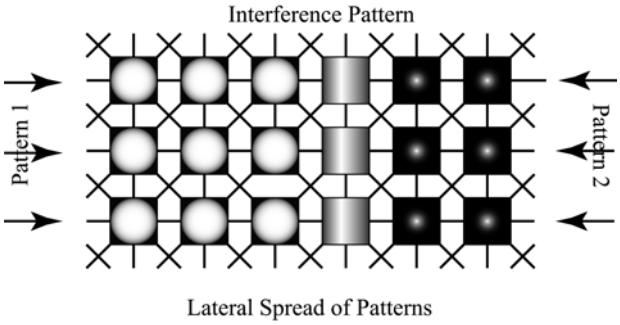

Figure 6. Two patterns move laterally across an array forming an interference pattern.

place by movement of information laterally from module to module as shown in Figure 6. This lateral information flow requires time and some important assumptions about the initial wiring of the modules. There is currently considerable experimental data supporting the idea of lateral information transfer in cerebral cortex over significant distances. The lateral information flow allows the potential for the formation of the feature combinations in the interference patterns, useful for pattern recognition. Since the individual modules are nonlinear learning networks, we have the potential for forming new attractor states when an interference pattern forms, that is, when two patterns arrive simultaneously at a module over different pathways. (Figure 6.)

The Network of Networks model is a general computational model. However, there have been a number of related computational models specifically designed for vision that have assumed that image processing involves lateral spread of information. An early example is Pitts and McCulloch [21] who suggested, "A square in the visual field, as it moved in and out in successive constrictions and dilations in Area 17, would trace out four spokes radiating from a common center upon the recipient mosaic. This four-spoked form, not at all like a square, would the be the size-invariant figure of a square (p. 55)." In the 1970's Harry Blum [10] proposed the "grassfire" model where visual contours ignited metaphorical "grassfires" and where the flame fronts intersected produced a somewhat size invariant representation of an object. The propagating waves are computing the **medial axis representation**, that is, the point on the axis lying halfway between contours.

There are now many examples of traveling waves in cortex. Bringuier et al. [11] observed long-range interactions in V1 with an inferred conduction velocity of approximately 0.1 m/sec. Lee, Mumford, Romero, and Lamme [18] discuss units in visual cortex that seem to respond to the medial axis. Particularly pertinent in this context is Lee [17] who discuses medial axis representations in the light of the organization of V1. In psychophysics, Kovacs and Julesz [15] and Kovacs, Feher, and Julesz [16] demonstrated threshold enhancement at the center of circle and at the foci of ellipses composed of oriented Gabor patches forming a closed contour. These models assume that an unspecified form of "activation" is being spread whereas the Network of Networks assumes that selective pattern information related to module attractor states is being propagated.

3. Ersatz Hardware: A Brief Sketch

How hard would it be to implement such a system in hardware? This section is a "back of the envelope" estimate of the numbers. We feel that there is a size, connectivity, and computational power "sweet spot" about the level of the parameters of the network of network model. If we equate an elementary attractor network with 10^4 actual neurons, that network might display perhaps 50 well-defined attractor states. Each elementary network might connect to 50 others through 50x50 state connection matrices. Therefore a brain-sized artificial system might consist of 10^6 elementary units with about 10^{11} (0.1 terabyte) total numbers involved in specifying the connections. Assume each elementary unit has roughly the processing power of a simple CPU. If we then assume 100 to 1000 CPU's can be placed on a chip there would be a total of 1000 to ten thousand chips in a brain sized system. These numbers are large but within the upper bounds of current technology.

Therefore, our basic architecture consists of a potentially huge (millions) number of simple CPUs connected to each other and arranged in a two dimensional array (Figure 7). The 2-D arrangement is is simple, cheap to implement, and corresponds to the actual 2-D anatomy of cerebral cortex. This intrinsic 2-D topography can make use of the spatial data representations commonly used in cortex for data representation for vision, audition, the skin senses and motor control.

3.1. Communications

The brain has extensive local and long-range communications. **The brain is unlike a traditional computer in that its program and its computation are determined primarily by strengths of its connections.** Details of these relatively sparse interconnections are critical to every aspect of brain function. And for our system, the details of CPU connectivity are equally critical.

3.1.1. Short-range connections. There is extensive local connectivity in cortex. An artificial system has many

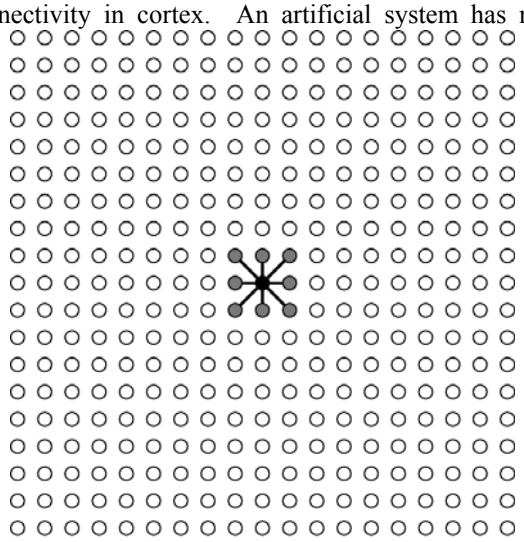

Figure 7. A 2-D array of modules with local connectivity. The basic Ersatz architecture

options. The simplest is NEWS connectivity. Our experience has been that expanding NEWS to include at least the diagonal CPUs works significantly better.

3.1.2. Long-range connections. Many of the most important operations in brain computation involve pattern association where an input pattern is transformed to an associated output pattern that can be different from the input. A sensory stimulus connected to a motor response would fall into this category. Anatomically, one group of neurons connects to another group. The functional connections consists of information transmitted through physical links, that is, axons, structures that are thin cell processes a few microns across that can be over a meter long in humans.

3.2. CPU Functions

The CPUs have to handle two quite different sets of operations. First, is communications with other CPUs. Much of the time and effort in brain-based computation is in getting the data to where it needs to be. Second, when the data arrives, it is can be then used for numerical computation. A limited instruction set should be more than adequate for both functions.

4. Assemblies of Modules

The brain shows large differences in scale. Understanding how neurons work together in groups of larger and larger size is a key to understanding brain

function. We have already suggested one intermediate level of structure in the modules of the network of network and have discussed how small numbers of attractor modules might interact. We can take this idea a bit further where we suggest the possibility of the formation of stable "module assemblies" as the next step up in intermediate level structure.

4.1. Data from Inferotemporal Cortex

We reviewed some of the properties of cortical columns in Section 2. It is very hard to study detailed properties of columns. However, optical imaging of intrinsic cortical signals has allowed visualization of structures of the size of cortical columns. The spatial resolution of this difficult technique can be a factor of 50 or more better than fMRI and gets into the region where perhaps we can see some of the kinds of specialization for cortical computation that we want to see. A small body of intrinsic imaging work has been done on inferotemporal cortex [IT] in primates. As the highest visual area, IT contains the results of previous processing presumably in a form that congenial for further use. Several Japanese groups have studied the organizational properties of infereotemporal cortex (area TE) from a computational point of view (Tanaka and Tsunoda et al., [22,23]). They proposed a model for inferotemporal cortex function that is in rough harmony with the basic architecture we assume for the Ersatz Brain

Tanaka's early work on inferotemporal cortex used (1) computer simplified stimuli and (2) microelectrode recordings. Cells in area TE respond to few stimuli and have relatively little spontaneous activity. Once Tanaka found an image that drove a cell, the next step was to perform a series of abstractions of it until an optimal simplified image – a "critical feature" -- was found that adequately drove the cell that but any further simplifications of it did not. Cells in a small region of TE tended to have similar critical features. For example, a number of cells in a small region might respond to various "T-junctions." The T-junctions would be different from each other in detail, but seemed to be examples of the same 'T' structure. Regions of similar response seemed to have roughly the size (300 μm) and character of functional columns found elsewhere in cortex. However, nearby "columns" had completely different critical features. There was no sign of the continuous gradation of response found, for example, in orientation columns in V1.

Complex objects excited several columns. From four to about a dozen columns were excited by presentation of complex objects. Such sets of observations led Tanaka [22]and Tsunoda et al. [23] to propose a sparsely distributed, column based model of image recognition:

> "… objects are represented by combination of multiple columns in a sparsely distributed manner. The region activated by a single object image is only 3.3 ± 2.5% of the entire recording area (number of examined object images, 37). … These results are consistent with feature-based representation, in which an object is represented by the combined activation of columns each responding to a specific visual feature of the object." [23, p. 835]

"Feature based" does not mean choosing one from a small class of basic geometrical features, by analogy with distinctive features in speech or with the "geons" in the model for object recognition proposed by Biederman [9]. The "features" inferred from the data in these imaging papers are more complex, diverse and far less "universal" than a small set of geons

4.2. Module Assemblies

If only two modules are associatively linked, we have a situation similar to the Bidirectional Associative Memory [BAM]. But in situations involving multiple interacting modules, we have the potential for forming interesting structures through Hebbian learning

The **cell assembly** is an idea that was first proposed by Donald Hebb in his 1949 book, *Organization of Behavior* [12]. Classic **Hebbian learning** was originally proposed

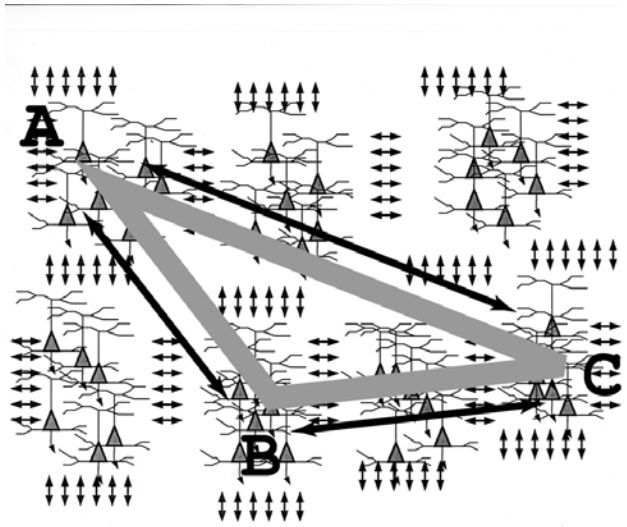

Figure 8. A linked "module assembly."

in the same book specifically to allow formation of cell assemblies. A cell assembly is a closed chain of mutually self-exciting neurons. Hebb viewed the assembly as the link between cognition and neuroscience. When an assembly was active, it corresponded to a cognitive entity, for example, a concept or a word. Although the basic idea is an appealing one, it is hard to make it work in practice because it is difficult to form stable assemblies. Two common pathological cases are (a) no activity in the network and, (b) even more distressing,

spread of activity over the entire network. Model neurons in a realistic system will participate in multiple assemblies and therefore activity spread will widely. It is possible to control this behavior by making strong assumptions about inhibition, but the resulting systems are not robust.

More modern versions of cell assemblies are structures like **synfire chains**, where a more complex assembly evolves in both space and time, giving rise to precise timing relations among a widely distributed set of active elements. Although there has been considerable interest in such models [1] the experimental evidence for them is sparse and controversial.

The activity of single neurons in most parts of cortex seems to be characterized by their average spike rate. Because average firing rate is a scalar it is hard to control the spread of activity from a particular cell assembly to other assemblies in a strongly interconnected network. The key assumption of the Network of Networks model is to have the basic computing element be interacting groups of neurons. This means module activity is not a scalar but a pattern of activity, that is, a high dimensional vector. Connections between modules are in the form of interactions between patterns. There is an intrinsic degree of selectivity and patterns may be less likely to spread promiscuously.

Because of this increased selectivity it should be possible to show that several modules can become stably linked together through Hebbian learning. We showed in Section 2.3 that associatively connecting modules together could increase the feedback coefficients in both modules. As we would expect from a positive feedback system with limits, BAM's and their numerous variants are extremely stable and robust in their behavior, usually heading to learned pairs of attractor states very quickly over a very wide range of system parameters. Even though their behavior is delightfully predictable, it has been hard to know what to do with them from an information processing point of view.

Let us see if we can look at the multimodule interactions that would occur in the Network of Networks. Consider the behavior of an associatively linked three module geometry. (Figure 8.) We can use exactly the same arguments as used in BAM's to predict that stable states can be formed in this system. Specifically, we have formed a loop with activity traveling back and forth in the loop.

Suppose we want to form a module assembly. First, all three modules should be driven simultaneously – say by the same object representation – so that Hebbian learning can form associative links between the modules.

Second, the strength of associative links must be sufficient so that all three modules, **A**, **B**, and **C**, become and stay active, exciting each other. The connections must be strong enough to ensure that the "loop gain" is effectively greater than one. It is now easier to see why the greater selectivity of module pattern association means that activity in a module assembly can easily be localized to the loop and will not spread easily to other modules. The criteria for effective spread require specific pattern learning sufficient to produce high loop gain.

There is no reason to stop with three modules. A number of interesting conjectures arise in more complex geometries. With more than three modules, multiple associative paths and loops become available for associative learning to operate. If strict timing relations, play a part in possible associative learning and feedback enhancement,, then the number of possible spatial-temporal solutions may not be very large. This possibility might provide an explanation for the relatively sparse number of modules coding complex objects seen in intrinsic images of IT.

The high level assemblies respond to what other levels tell them. A complex object presumably corresponds to a complex pattern of features and feature combinations at "lower" levels of the system. (Since there are both upward and downward connections, the distinction between lower and higher levels is not clean.) The Ersatz Brain architecture with its extensive lateral local processing allows feature combinations to be formed from local information flow. These feature combinations, (a) tend to be spatially localized and (b) can correspond to abstractions and generalizations of frequently occurring (but somewhat noisy) events.

Therefore higher layers are receiving a projection from lower layer (or layers) showing a topographically dispersed set of diverse selective feature combinations. Each module would then receive a simultaneous but distinct set of input patterns from lower levels, and, of course, will report its activity back to the lower level through reciprocal pathways. Note that all these proposed processing techniques develop representations that are somewhat sparse and topographically localized. This observation seems to correspond roughly to what is seen in cortex.

5. Topographic Computation

The cortex, and, in deliberate imitation, our computational model is a 2-D sheet of modules. Connections between modules, slow and expensive in the brain, as we have noted, perform the computation. Therefore topographic relationships between modules and their timing relationships become of critical

importance. We suggest that it is possible to use the topographic structure of the array to perform some interesting computations, in fact, **topographic computation** may be one major mechanism to perform a computation, and, therefore, a major tool to program the array.

In a neural network, one useful form of such a topographic representation is called a **bar code.** The value of the parameter represented depends on the location of a group of active units on a linear array of units. The price paid for this useful data representation is low precision and inefficiency in use of units. If a single parameter is represented only as a location on a surface, then many units are required to represent a value that a single unit could represent by an activity level. Topographic representations of sensory information are ubiquitous in cortex. Such a representation technique, and its variants are most naturally implemented in a system that is composed of many relatively inexpensive units, performing a function that is only secondarily concerned with high accuracy. Nanocomponent based computing devices may fall into this category, as does the nervous system.

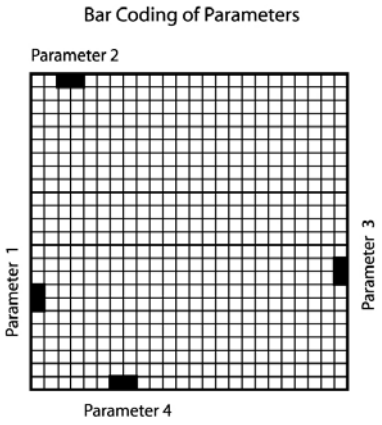

Figure 9. Four bar codes.

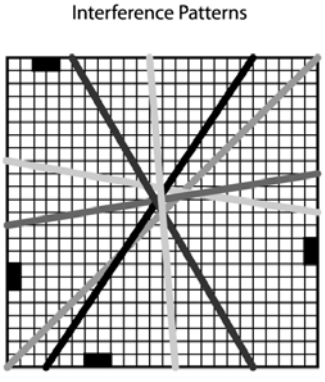

Figure 10. Formation of interference patterns.

5.1. Sensor Fusion

One potential application of our Ersatz Brain architecture is to **sensor fusion.** Sensor fusion involves integration of information from different types of sensor into a unified interpretation.

Suppose we have a set of numerical sensor readings, say four. We can do quite a bit with these values alone, for example, using them as a set of features for object recognition. However, this is not really sensor fusion since the sensor data is not integrated but used as information to determine a previously learned classification. This may not be possible if there is no previous classification, that is, in unsupervised learning.

Spatializing the data, that is letting it find a natural topographic organization that reflects the relationships between multiple data values, is a technique of great potential power, though an unfamiliar one. It is, however, a natural and effective way to compute for the two dimensional cortex and, of course, our two dimensional Ersatz Brain architecture. It also provides a way of "programming" a parallel computer. (See Anderson [4]).

Assume we have four parameters that we want to represent in the activity pattern that describes single entity (Figure 9). A radar pulse from a single emitter could be characterized by frequency, angle of arrival, pulse width, and intensity. (Some of these coding ideas grew out of a project to "make sense" of a complex radar environment. See Anderson et al. [5].)

Our architecture assumes local linear transmission of patterns from module to module according to the Network of Networks model assumptions. Modules have multiple stable attractor states. Patterns are transmitted laterally from module to module. The modules are non-linear for large signals but we assume, linear for small ones. When two different patterns arrive at a module simultaneously, there is the possibility for a new pattern to be generated in the module, now representing the coincidence as a new part of the module's repertoire of attractor states. Standard learning rules plus the non-linear network can perform this operation. There may be many two parameter interference patterns, the straight lines in Figure 10. Each pair of input patterns gives rise to an interference pattern, a line perpendicular to the midpoint of the line between the pair of input locations.

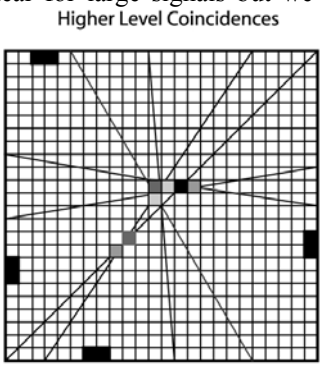

Figure 11. Formation of higher-level feature combinations.

5.1.1. Higher Level Features. Besides the interference patterns representing coincidences between pairs of patterns, in addition, there are often places where three or even four features coincide at a module. (Figure 11) Determining when or whether these coincidences instruct modules as to higher-level combinations of features is a

matter for further work. The possibility of their formation is clearly demonstrated however. The higher-level combinations represent partial or whole aspects of the entire input pattern, that is, they respond to the Gestalt of the pattern. In this sense they have **fused** a number of aspects of the input pattern and represented it as a new activity pattern at a specific location.

5.1.2. Formation of Hierarchical "Concepts" Perhaps the most intriguing aspect of this coding technique is the way it allows for the formation of what look a little like hierarchical concept representations. Suppose we have a set of "objects". For a simple demonstration, assume we have three parameter values that are fixed for each object and one value that varies widely from example to example. After a number of examples are seen, the system develops two different spatial codes (Figure 12).

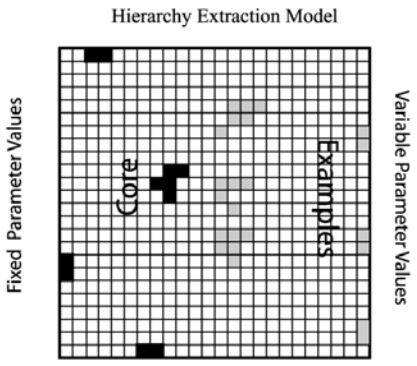

Figure 12. Mechanism for formation of hierarchical structure.

In the first, a number of high order feature combinations are fixed since their three input "core" patterns never change. In the second, based on the additional spatial relations generated by the widely different examples, there is a varying set of feature combinations corresponding to the details of each specific example of the object. If the resulting spatial coding is looked at by an associative system, then two kinds of pattern can potentially be learned.

The first learned pattern corresponds to the unchanging **core** and might correspond to an abstraction of the commonalities of many examples. The second learned set of patterns corresponds to the core, plus the **examples** -- patterns associated with each specific learned parameter set. All the specific examples are related by their common core pattern. This spatial dichotomy has considerable similarity to the subordinate-superordinate relationships characterizing in a hierarchical semantic network [20].

We have **spatialized** the logical structure of the hierarchy. Because the coincidences due to the "core" (three values) and to the "examples" (all four values) are spatially separated, we have the possibility of using the "core" as an abstraction of the examples and using it by itself as the descriptor of the entire set of examples. It then acts like the higher level in a hierarchy, that is, all examples contain the core. The many-to-one relationship here – many low level examples, fewer high level examples -- is typical of hierarchical semantic networks. Many of the control mechanisms found in the cortex, most notably attention, are often held to work by exciting or inhibiting regions of the cortical array. The model we are proposing is well suited to such attention-like control structures.

6. Conclusions

We have presented brief summary of the features of the Ersatz Brain Project. It is neither neuroscience based or abstraction based but falls at an intermediate level of organization where a transition between largely continuous perceptual processes and largely discrete cognitive processes might occur. From this starting point we have discussed: (1) Some physiological evidence for our basic approximation; (2) A brief discussion of the basic formal structure; and (3) Provided an example of the potential utility of a topographic modular system for performing a useful computation: information fusion and resultant formation of hierarchical structures.

References

[1] M. Abeles (1982). *Local Cortical Circuits: An Electrophysiological Study*. Berlin: Springer-Verlag

[2] JA Anderson (1993). The BSB network. Pp. 77-103 in MH Hassoun (Ed.), *Associative Neural Networks*. New York, NY: Oxford University Press.

[3] JA Anderson (1995). *An Introduction to Neural Networks*. Cambridge, MA: MIT Press.

[4] JA Anderson (2003). Arithmetic on a parallel computer: Perception Versus Logic. *Brain and Mind*, **4**, 169-188.

[5] JA Anderson, MT Gately, PA Penz, and DR Collins (1990). Radar signal categorization using a neural network. *Proceedings of the IEEE*, **78**, 1646-1657.

[6] JA Anderson, A Pellionisz, and E Rosenfeld (Eds. 1990). *Neurocomputing 2*. Cambridge, MA: MIT Press.

[7] JA Anderson and E Rosenfeld (Eds. 1988). *Neurocomputing*. Cambridge, MA: MIT Press.

[8] JA Anderson and JP Sutton (1997). If we compute faster, do we understand better? *Behavior Research Methods, Instruments, and Computers*, **29**, 67-77.

[9] I Biederman (1987). Recognition by components: A Theory of Human Image Understanding. *Psychological Review*, **94**, 115-147.

[10] HJ Blum (1973). Biological shape and visual science (Part I). *Journal of Theoretical Biology*, **38** 205-87.

[11] V Bringuier, F Chavane, L Glaeser, and Y Fregnac (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science*, **283,** 695-699.

[12] DO Hebb (1949). *The Organization of Behavior*. New York, NY: Wiley.

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

[14] B Kosko (1988). Bidirectional associative memory. *IEEE Transactions on Systems, Man, and Cybernetics.* **18,** 49-60.

[15] I Kovacs and B Julesz (1994). Perceptual sensitivity maps within globally defined shapes. *Nature*, **370**, 644-6.

[16] I Kovacs, A Feher and B Julesz (1998). Medial point description of shape: a representation for action coding and its psychophysical correlates. *Vision Research*, **38**, 2323-2333.

[17] T-S Lee (2002). Analysis and synthesis of visual images in the brain. In P Olver and A Tannenbaum (Eds), *Image Analysis and the Brain*. Berlin: Springer.

[18] TS Lee, D Mumford, R Romero, and VAF Lamme (1998). The role of primary visual cortex in higher level vision. *Vision Research*, **38,** 2429-2454.

[19] VB Mountcastle (2003). Introduction to special issue on cortical columns. *Cerebral Cortex*, **13**, 2-4.

[20] GL Murphy (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

[21] W Pitts and WS McCulloch (1947/1965). How we know universals: The perception of auditory and visual forms. Reprinted in WS McCulloch (Ed., 1965), *Embodiments of Mind*. pp. 46-66 Cambridge, MA: MIT Press.

[22] K Tanaka (2003). Columns for complex visual object features in inferotemporal cortex: Clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex*. **13**, 90-99.

[23] K Tsunoda, Y Yamane, M. Nishizaki, and M. Tanifuji (2003). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*. **4,** 832-838.