

Learning in Sparsely Connected and Sparsely Coded Systems

James Anderson
Ersatz Brain Project Working Note
August 4, 2005

Abstract

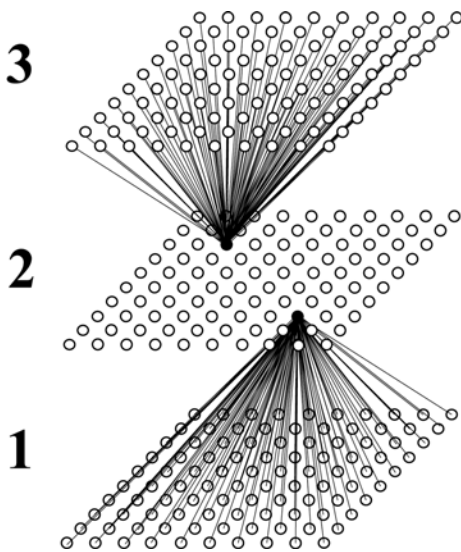
This paper considers the implications for learning and performance in sparsely connected models of the cerebral cortex that also use sparse data representations for events. Sparse connectivity combined with sparse data representation give rise to some interesting models. Sparse neural networks using scalar elements - most common neural network neuron approximations - are not very useful. However, when sparseness assumptions are combined with the Network of Networks modular architecture, where module activity can be described as vectors, more powerful systems emerge because pattern based activity allows enough path selectivity to make sparsely connected systems useful. We discuss some of these systems further.

There is a road, no simple highway
Between the dawn and the dark of night
And if you go no one may follow
That path is for your steps alone.

Ripple
Robert Hunter/Jerry Garcia

Sparse Connectivity and Sparse Coding

Full Connectivity. Most neural network learning models in the literature assume full or nearly full connectivity between layers of units. Units are most often arranged in layers because it is an arrangement that pays homage to the neuroanatomy of mammalian cerebral cortex where cortical regions project to other cortical regions over long-range projection pathways through the white matter. It also has an agreeable input-output anatomy that agrees with intuitions about pattern recognition and signal processing. Full connectivity means that a unit in one layer projects to, and it projected to, by all the units in layers above



and below it.

Fully connected systems can often be analyzed easily using standard mathematical techniques. They can perform a number of powerful information processing operations, and, combined with simple local learning algorithms such as the Hebb rule, they can be used to build adaptive systems with a number of useful applications in both engineering and science.

The number of connections in fully connected systems grows rapidly, order n^2 , where n is the number of units. The Figure, a simple three layer fully connected feed-forward neural net, suggests how dense interconnections can be even in a small system. Only connections to a single unit are shown.

Sparse Connectivity. Although these models use full or high connectivity, the actual cerebral cortex is **sparsely connected**, that is, each neuron projects to relatively few others given the potential number they could connect to, even in projections from one cortical region to another. Although it is often stated that cortical pyramidal cell might have as many as 50,000 connections, cortex has at least 10,000,000,000 cells, suggesting exceedingly sparse overall cortical connectivity, on the order of 0.001 percent. There is also evidence that many of the connections are inactive during operation. Most cortical neurons seem to have many "silent" connections that can be unmasked if the environment changes. Studies of cortical plasticity have indicated that in the somatosensory system, for example, loss of a digit allows the rewiring of the selective cells formerly connecting to that digit so they now respond to new inputs, usually the adjacent digits.

Learning models in layered systems almost all incorporate some version of the Hebb synapse. In order to attain satisfactory performance in a fully connected network, it is necessary to develop elaborate error-correcting systems because different stored pattern associations otherwise will interact strongly. The best known such error correcting system is backward error propagation (**back propagation**) which assumes error information moves backwards from output layer to input layer. The backward connections are assumed to have the same strengths as the forward connections. Back propagation is acknowledged to be highly unbiological though it has many useful

engineering applications. Among its problems are long learning times, a tendency to overfit training data, and catastrophic unlearning, where learning new material destroys memory for older stored information. However, back propagation is a very general algorithm and is capable of associating many pairs of input and output patterns with high accuracy.

Sparse Coding for Data Representation. Besides sparse connectivity, there is reasonably strong experimental evidence that **sparse coding** is used for data representation in the cortex, that is, information is represented by the activities of relatively few units. A review by Olshausen and Field (2004) comments, "In recent years a combination of experimental, computational, and theoretical studies have pointed to the existence of a common underlying principle involved in sensory information processing, namely that information is represented by a relatively small number of simultaneously active neurons out of a large population, commonly referred to as 'sparse coding.'" (p. 481). Many of these ideas first emerged in a classic paper by Barlow (1972).

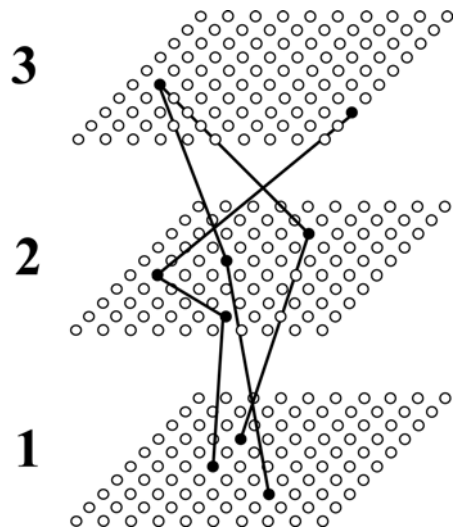
There are numerous advantages to sparse coding. Olshausen and Field mention that sparse coding provides increased storage capacity in associative memories and is easy to work with computationally, properties we will make use of. Among other virtues, sparse coding also "makes structure in natural signals explicit" (p. 481) and is energy efficient.

The exact number of units active for an event is not known in general. Cortex clearly does not display "grandmother cell" selectivity, where a single high-level active cell represents the concept "grandmother". However, sometimes codings can be remarkably selective. A recent *Nature* paper describes several such units recorded from the medial temporal lobe of awake humans. One unit responded only to various images of Halle Berry and, in addition, the character string forming her name. Another responded only to Jennifer Anniston (Quiroga, Reddy, Kreiman, Koch, and Fried, 2005). It is likely, though not shown experimentally, that multiple cells were active to "represent" these actresses because it is unlikely that the single observed active unit happened to be the one where the experimenters had stimulus materials available. However, information seems not to be widely distributed either.

“Higher” regions (for example, inferotemporal cortex) seem to show a greater degree of sparse coding than lower ones (for example, V1). Cells in the higher levels of the visual system also have less spontaneous activity than lower regions. Cells in inferotemporal cortex are silent much of the time until they find a specific stimulus that piques their interest.

Sparse Coding combined with Sparse Representation. The Figure shows a cartoon version of a system that shows both **sparse coding** (three active units in input layer 1, two active units in output layer 3) and **sparse connectivity**.

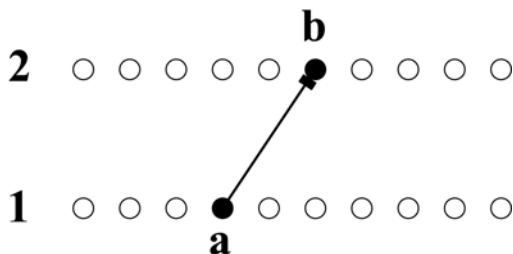
Instead of trying to derive very general pattern association systems like back propagation, using high connectivity, let us see if we can make a learning system that starts from the assumption of both **sparse connectivity** and **sparse coding**.



First, let us consider a trivially simple sparse system. For sparse connectivity, suppose we have n units in Layer 1 projecting to n units in Layer 2 with a single connection.

Second, the patterns to be associated consist of one active input unit and one active output unit.

There are then n possible input vectors and n possible output vectors. If we assume that connections are random, then the chance of being able to connect an arbitrary input unit to an arbitrary output unit is very small, $1/n$. This is not very useful in general, but let us assume we have the proper connection in place. (See Figure)



If the potential connection does exist, simple Hebb learning can learn it easily. Suppose the connection strength between unit **a** and unit **b** starts at zero. Simultaneous presentation of input and output patterns, that

is, with unit **a** active and unit **b** active, simple Hebb outer product learning rules will modify the strength proportional to the product of the two activities, that is,

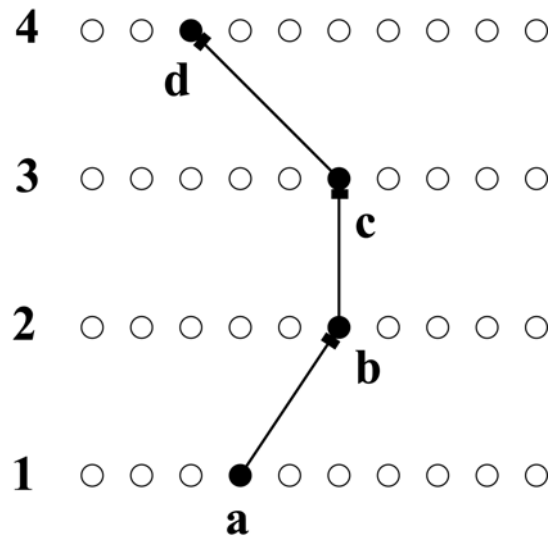
$$\Delta(\text{connection strength}) \sim (\text{activity of } \mathbf{a})(\text{activity of } \mathbf{b}).$$

The change in connection strength forms a link between **a** and **b**. By assumption, all the other units are inactive. Multiple presentations will strengthen the connection without limit, a common problem with simple Hebb learning.

Note that if we assume an input unit projects to every output unit, there is no loss of potential selectivity if we are using a sparse system. Since, by assumption, only one input unit and one output unit can be active at the same time, only that connection will change. Therefore, we have used the sparse data representation to build **selectivity** into system and we can then be **less selective** in the connection pattern.

Paths. In sparse systems, selectivity can come from other sources than a precise pattern of connection strengths. A critical notion in sparse systems is the idea of a **path**. A **path** connects a sparsely

coded input unit with a sparsely coded output unit. Paths have **strengths** just as connections do, but the strengths are based on the entire path, from input to output, which may involve intermediate connections. A path may have initial strength zero or start with a nonzero strength due to initial connectivity. The concept of a path provides a justification and suggests potential ways to structure initial wiring of the nervous

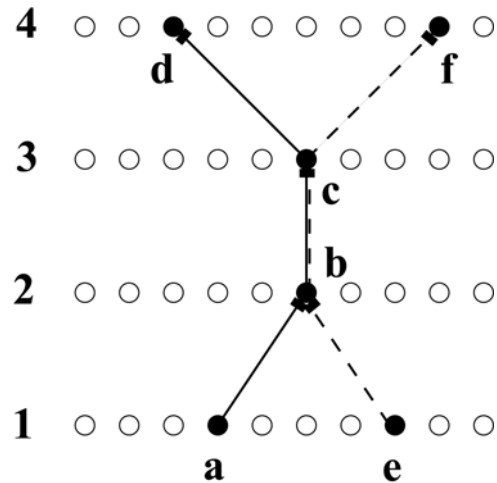


system. Connections that are likely to be useful can be sketched in initially. Learning can then tune and sharpen the pre-existing connections. This strategy seems to be found experimentally in many sensory systems, for some examples, see a discussion of many years of work on the development of different aspects of the visual system. (Cooper, Intrator, Blais, and Shouval, 2004)

For our very sparsely coded representations using a single active unit, even if there are multiple layers of units if a path exists between an input unit and an output unit, the path will be strengthened using simple Hebb learning.

Consider a path with several intermediate units, as shown in the Figure. If there is no initial transmission of activity through the path, there will be no learning. If, however, when the input unit, **a**, becomes active and can thereby give rise to a small activity at the output unit, **d**, learning can take place. Hebb learning will take the activity of **d** -- post-synaptic activity -- multiply it by the activity through the path from **a**, and increment the connection strength of unit **d**. The path becomes stronger. Continued learning will further strengthen that particular connection. If the path is strictly feedforward, as in the Figure, connections earlier in the path (i.e. on **b** or **c**) cannot be changed without additional assumptions, for example, some form of backward information flow. Selectivity in the first system becomes concentrated at the top layer. There is also the likely possibility that multiple paths may exist for the same single active element association.

Common Parts of Paths. Suppose there is a common portion of a path for two single active unit associations, that is, **a** with **d** (**a>b>c>d**) and **e** with **f** (**e>b>c>f**). We cannot weaken or strengthen the common part of the path (**b>c**) because it is used for multiple associations. Only if there are multiple paths that we can play against one another will it be possible to separate the pathways.



Initial Biases in Connectivity. For this simple system with **scalar** weights by far the best strategy would be to somehow construct **independent** paths for each single unit association.

If independent paths are desirable, a useful initial construction bias for such sparsely connected systems would be to make available as **many** potential paths as possible.

This bias differs significantly from back propagation, where there are almost always fewer units in the hidden layers than in the input and output layers, therefore **fewer** potential paths. In a fully connected system, adding more units than contained in the input and output layers would be redundant. This is not so in sparse systems. The biology suggests a huge expansion in number of units going from retina to thalamus to cortex. In V1, a million input fibers drive 200 million V1 neurons.

Therefore the wiring of the nervous system looks qualitatively more like a sparse, path model than a statistical fully connected back propagation model.

Of course, we know that real data representations do not connect a single input unit to a single output neuron. However, we conjecture the qualitative properties of this extreme system should carry over to the more biological sparse input representations. For example, if we had 1,000,000 units at input and output, two or three or even a dozen active units in the data representations should work about the same way statistically as a single active unit. Even a degree of distribution in a sparse system opens up a range of interesting possibilities, as we shall see in a later section.

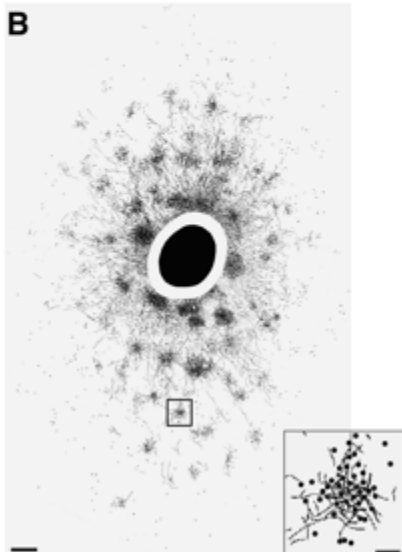
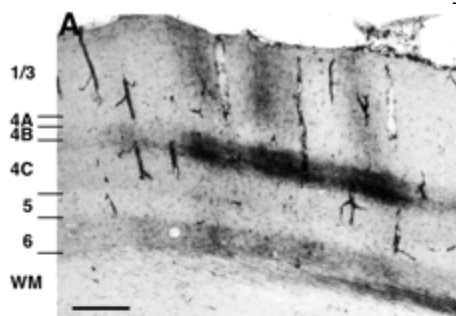
However, such extreme restrictions on connectivity and representation do not seem at first to form a very promising information processing system. I suspect this is why it has not been looked at as a nervous system model, even though it fits the qualitative structure of the cortex better than the high connectivity assumptions that underlie almost all common neural network models.

Sparseness in Systems Using the Network of Networks.

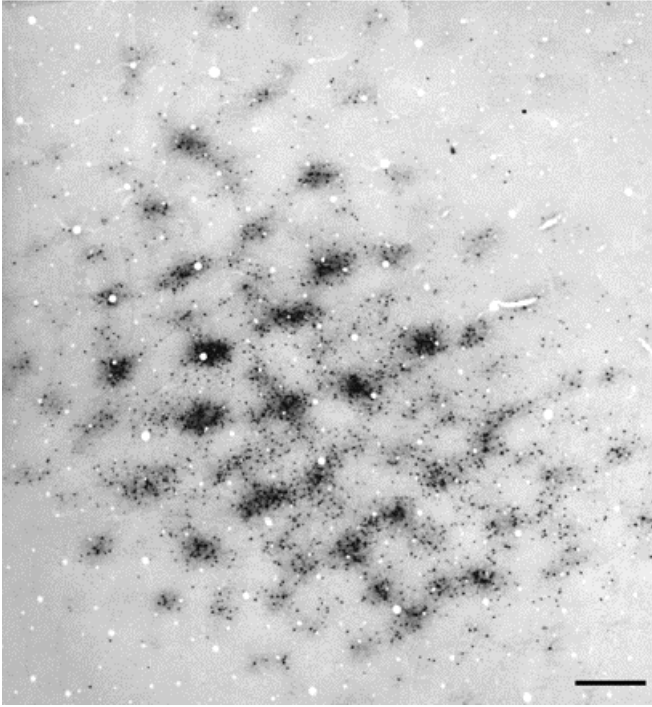
In the Ersatz Brain Project, we have been working with a computational approximation of neocortex we call the **Network of Networks**. Its primary assumption is that the basic unit of computation is not a single unit but a small group of units [a **module**] that act together as a nonlinear attractor network. The attractor network has several attractor states that dominate its behavior. (Anderson, 1993,1995; Hopfield, 1882)

Comments on Biology. Previous discussions of the Ersatz Brain Project have suggested that a likely physiological substrate for the modules is the cortical column, one of the key organizing features of mammalian neocortex.

We are assuming columns connect to other columns. Some cortical biology suggests columnar identity is maintained in both forward and backward projections (Lund, 2003), as we are assuming. As Lund says, "The anatomical column acts as a functionally tuned unit and point of information collation from laterally offset regions and feedback pathways." (p. 12) and "... feedback projections from extra-striate



(A) Pia to white matter section of squirrel monkey area V1 showing label to the side of a large injection of HRP (not shown) through layers 1-6. Columns of heavy label are seen in Layers 1-3; heavy label in layers 4B/upper 4C α shows higher density under each column of label in layers 2/3; similar increased terminal densities occur in layers 5 and 6 in alignment with the regions of densest label in overlying layers. Scale bar: 100 μ m. Modified from Rockland and Lund, 1983. (B) Surface view of a 2D composite reconstruction of labeled lateral connections in layers 2/3 of macaque monkey area V1, showing a clustered pattern of anterogradely labeled neurons (inset) surrounding an injection of the tracer CTB (black oval). The patches of terminals and cells are the cross-sections of columns of label of the kind seen in (A). Small square: labeled patch shown in higher power in inset. Note that both retrogradely labeled cells and orthogradely labeled terminal axon processes are present in each patch, indicating reciprocity of connection with the injection site. Scale bar: 500 μ m. Inset: higher power drawing of patch in small square. Scale bar: 100 μ m. (Original Caption, Figure 3, Lund et al. (2003).



Bright-field photomontage of a tangential section through layer 4B of macaque area V1, showing CGB labeled cell bodies (cells of origin of feedforward projections to extra-striate cortex), resulting from an injection site in dorsal area V3. Note the patchy pattern of inter-areal connections. Scale bar: 500µm. (Original caption, Figure 6, Lund et al., 2003).

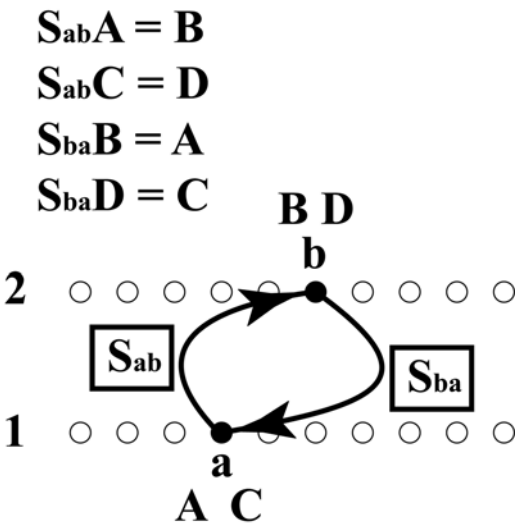
cortex target the clusters of neurons that provide feedforward projections to the same extra-striate site. ... The inter-areal connections, therefore, seem to be a component of the same V1 anatomical column system as the intrinsic connections." (p. 22). (See Figure)

Implications of Pattern Selectivity. A critical difference between a model neuron and a module in the Network of Networks is that activity in a model neuron is described as a **scalar** and activity in a module is a **vector** since many units are contained in a module and modules are connected to each

other through multiple connections. Modules send patterns of activity to each other.

Simple **scalar** activity is not selective. In a common path, activity arising from different sources cannot be told apart. However, there are other mechanisms that can produce path selectivity than keeping paths independent. If we use the Network of Networks approximations, activity along a path is not scalar but **vector**.

Our earlier discussion still holds. Now, however, the primitive objects are not simple unit activity but **module activity**. In the simplest associators, if activity in a



single module is connected to a single module, then we have a common pattern associator.

If a pattern **A** on module **a** is input to Layer 1, the activity, **B**, on **b** is given by $S_{ab}A$. We assume as part of the Network of Networks architecture that there are reciprocal connections between modules. Therefore, if pattern **B** is present on module **B** of layer 2, the activity on **A** is given by $S_{ba}B$. This loop between **a** and **b** is similar to Kosko's Bidirectional Associative Memory [BAM] model (Kosko, 1988)

If pattern **A** on **a** and pattern **B** on **b** are simultaneously present, the increment in strengths are given by standard Hebb learning, that is,

$$\Delta S_{ab} \sim BA^T.$$

$$\Delta S_{ba} \sim AB^T.$$

where η is a learning constant.

If no other pattern associations are stored in the connection matrices, after learning, if pattern **A** is present at module **a**, something like pattern **B**, multiplied by a constant, will appear at module **b**. The sparsely represented association (the only activity is on **a** and **b**) has been learned.

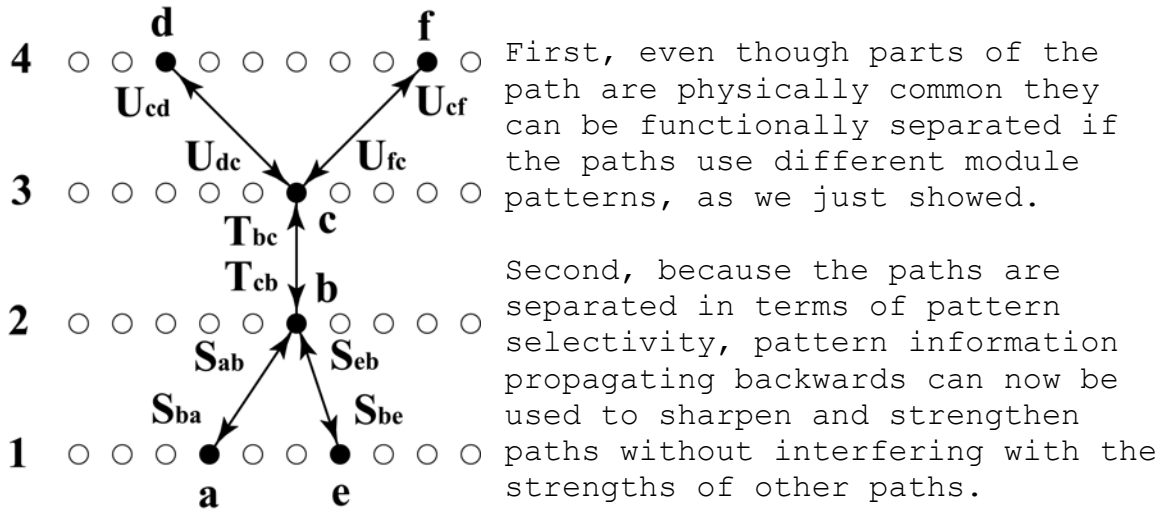
Multiple Patterns Can be Associated. Unlike scalar connections between units, a single connection between modules can learn multiple associations.

Consider our earlier system with connected modules and a common path.

In the simplest situation, suppose we look at modules **a** and **b** as before, but assume that **a** and **b** can display two orthogonal patterns, **A** and **C** on **a** and **B** and **D** on **b**. Subject to the usual limitations of simple pattern associators, (that is, it works best if **A** and **B**, and **C** and **D**, are orthogonal) such a network can easily learn to associate **A** with **B** and **C** with **D** using the same connections.

But this now means we have a way of using path selectivity to overcome some of the limitations of scalar systems.

Consider the common path situation again. We want to associate patterns on two paths, **a-b-c-d** and **e-b-c-f** with link **b-c** in common.



Backwards Projections. There is no reason to assume that the backward projections and the forward projections have the same strength. This assumption is perhaps the most biologically unrealistic assumption of back propagation. All that may be required initially for sparse systems is that modules be sufficiently well connected to produce activity along all points in the path.

First consider the two associative chains, up and down, **a>b>c>d** and **d>c>b>a**.

For simplicity, assume vectors are normalized with learning constant η and to a first approximation orthogonal (or at least independent). Since this is supervised learning, we assume that **a** and **d** are present at input and output.

We will first consider learning in the **linear** range of operation of the system. As the path gets stronger, we conjecture that modules in the path will develop internal states corresponding to path activation, that is, attractors. Our conjecture is that selective paths should develop as the system becomes more non-linear.

Consider the two intermediate modules on the path. Note that patterns **b** and **c** are the result of information moving upwards from the input layer and downwards from the output layer, that is, initially

Upward:

$$\begin{aligned}\text{Activity on } \mathbf{c} &= \mathbf{T}_{bc}\mathbf{b} + \mathbf{U}_{dc}\mathbf{d} \\ \text{Activity on } \mathbf{b} &= \mathbf{S}_{ab}\mathbf{a} + \mathbf{T}_{cb}\mathbf{c}\end{aligned}$$

If these activities are non-zero patterns of any type, we have the ability to use simple Hebb learning to change the strengths of coupling between modules using their connection matrices.

There are two active modules connected to \mathbf{b} , up from \mathbf{a} and down from \mathbf{c} . The change in upward coupling between \mathbf{a} and \mathbf{b} , through \mathbf{S}_{ab} is given by

$$\Delta(\text{coupling term } \mathbf{S}_{ab}) = \eta \mathbf{b}\mathbf{a}^T$$

The change in coupling between \mathbf{c} and \mathbf{b} is

$$\Delta(\text{coupling term } \mathbf{T}_{cb}) = \eta \mathbf{b}\mathbf{c}^T$$

Similarly, the two changes for module \mathbf{c} are:

$$\begin{aligned}\Delta(\text{coupling term } \mathbf{U}_{dc}) &= \eta \mathbf{c}\mathbf{d}^T \\ \Delta(\text{coupling term } \mathbf{T}_{bc}) &= \eta \mathbf{c}\mathbf{b}^T\end{aligned}$$

Next, we can compute coupling between \mathbf{a} and \mathbf{d} . We start by assuming that the initial coupling is small, but large enough to allow some patterns to emerge at \mathbf{b} and \mathbf{c} , that is, get through the path, weakly. The claim is that repeated learning would eventually erase the initial connectivity. (Something like this assumption is made in virtually all simulations of back propagation.)

If the modules are not connected, that is, not on the same sparse path, or if the activity of \mathbf{d} for that path is zero, path activities will be zero and there will be no learning.

Because we are initially assuming simple Hebb learning, we can compute the strength of the path after the first learning event.

Assume pattern \mathbf{a} is presented at layer 1. and \mathbf{d} is zero.

Then

$$\begin{aligned}\text{Pattern on } \mathbf{d} &= (\mathbf{U}_{cd}) (\mathbf{T}_{bc}) (\mathbf{S}_{ab}) \mathbf{a} \\ &= \eta^3 \mathbf{d}\mathbf{c}^T \mathbf{c}\mathbf{b}^T \mathbf{b}\mathbf{a}^T \mathbf{a}\end{aligned}$$

$$= (\text{constant}) \mathbf{d}$$

Pairs of intermediate terms become scalars since $\mathbf{a}^T \mathbf{a}$, $\mathbf{b}^T \mathbf{b}$ and $\mathbf{c}^T \mathbf{c}$ are scalar inner products. This result is well known for simple linear pattern associators.

As time goes on, Hebb learning will serve to increase the constant that couples the patterns on \mathbf{a} and \mathbf{d} and will increase the strength of this particular associative path.

Note that the somewhat arbitrary activities in \mathbf{b} and \mathbf{c} - the pattern sum of initial upward and downward connections - serves as a 'hidden layer' link between input and output patterns. This argument holds for more intermediate layers, as long as initial paths both up and down exist.

Therefore, upward and downward simple Hebbian learning has the potential to form selective, independent paths that could be used for associative learning in Network of Networks systems with sparse data representations and sparse connectivities. Both upward and downward paths must exist. However the downward path need not send error data, or any specific function of the output for that matter, as long as there is a connection.

Module Assemblies

Loops: Integration with Lateral Pattern Movement. We showed that simple Hebbian learning combined with sparse representations and connectivity can learn through selective strengthening and weakening of paths.

How can this system be integrated with our previous Network of Networks ideas?

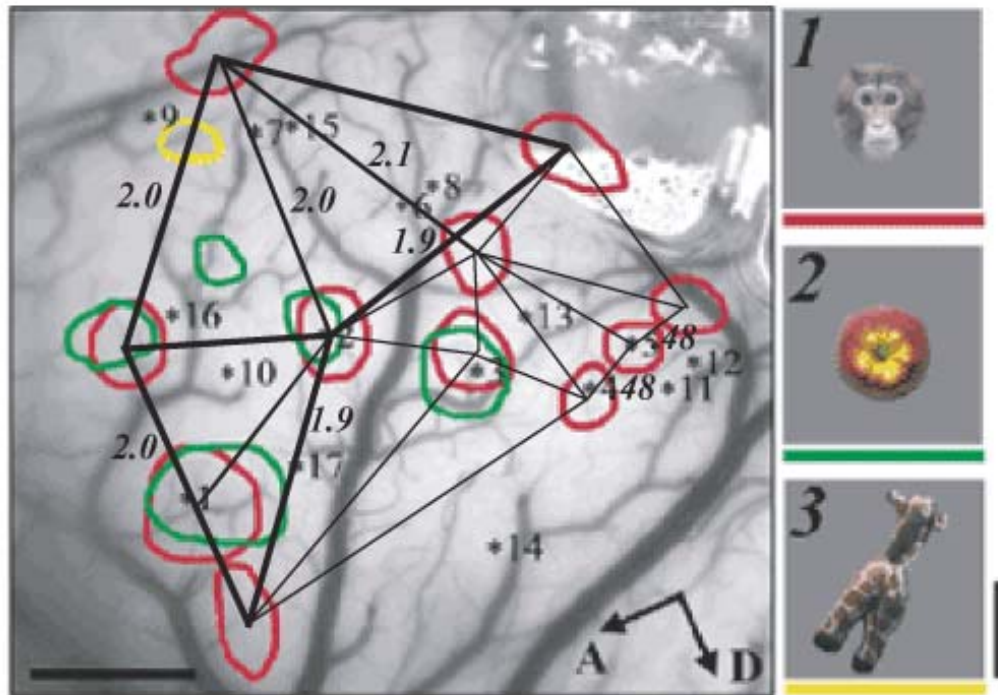
Our earlier discussion assumed lateral movement of pattern information between modules, and the potential formation of "interference patterns" (feature combinations) through non-linear learning of different patterns. For lateral movement to work well, local connections need to be relatively dense, certainly compared to the sparse connectivity we have been assuming for interlayer connections.

Biological Data on Local Connectivity. Let us assume that the intralayer connections are sufficiently dense so that active modules a little distance apart (not just nearest neighbors) can become associatively linked.

We can only make speculations about the degree of local density of connections in neocortex. Data from Lund (2003), presented earlier, suggest substantial connectivity over a region of a millimeter or so. Recurrent collaterals of cortical pyramidal cells form relatively dense projections around a pyramidal cell. The extent of lateral spread of recurrent collaterals in cortex seems to be over a circle of roughly 3 mm diameter. (Szentagothai, 1978). If we assume that a column is roughly a third of a mm, there are roughly 10 columns in a square mm. A 3 mm diameter circle has an area of roughly 10 square mm, leading to a suggestion that a column projects locally to about 100 other columns.

We have suggested before, inspired by the work of Tanaka (1996, 2003) and Tsunoda (2001) using intrinsic imaging on inferotemporal cortex, that we could use our learning mechanisms to form **module assemblies**, modeled on Hebbian **cell assemblies**. The experimental data presented in their papers suggested a dozen or so active columns [modules] represented a complex percept. The Figure shows the

response of inferotemporal cortex using intrinsic imaging to three images. Note that the monkey face (1) gives rise to activity in 11 columnar sized regions. The symmetric object has some overlap with the face and the unrelated figure (3) has no common response regions. The area of



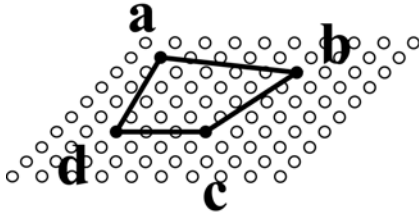
Relationship between intrinsic signals and spike activity in Area TE. ... Active spots elicited by three different stimuli (1, 2 and 3, right) ... The color of individual contours indicates the active spots elicited by the stimulus underlined with the same color ... A, anterior; D, dorsal. Horizontal scale bar (black) 1.0 mm; vertical scale bar 10° (applies to insets). Edited caption, Figure 2, Tsunoda (2001). Black lines added by Anderson to give distances (in mm) between active spots. Note the non-random distribution of distances.

cortical activation for the monkey face covers an area of about 4 mm on a side.

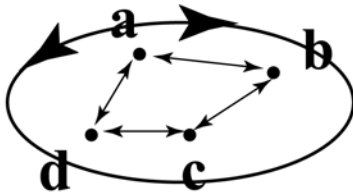
A conjecture (which Hebb also made about cell assemblies) is that these module assemblies correspond to a cognitive entity, that is, a concept. Hebb assumed, as we shall, that the assemblies were self-exciting, that is, information could travel around a loop of connected cells.

However, attempts to model the formation of cell assemblies using neurons were not successful. The major problem was that activity spread promiscuously as one cell assembly excited others.

Module Assemblies. We ran into a similar problem when tried to construct sparse paths using only scalar activities. However, we were able to incorporate what we conjecture are stable selective paths using the potential pattern selectivity available from intermodule connections.



Consider a set of interconnected modules. (See Figure).



Consider the set of four active modules, **a,b,c,d**, in the Figure. Assume they are densely connected locally. That will imply that **a** is connected to **b**, **c** to **d**, etc. Suppose that **a,b,c,d** are forced to hold a particular pattern by some outside mechanism, say another brain region or an outside supervising input.

Then the analysis we just performed on sparse paths with simple associators will hold, that is, as the path is traversed, each link becomes associated with its active connections.

However, this path closes on itself. If the modules **abcda** simultaneously active are learned, if **a** is present, then after traversing the linked path **a>b>c>d>a**, the pattern arriving at **a** in either direction will be a constant times the pattern on **a**. If the constant is positive and the starting pattern is still present, there is the potential for a positive feedback loop gain. The same analysis holds true for traversing the loop in the opposite direction, that is, **a>d>c>b>a**. Limitations on maximum activity in each module will stabilize activity as it becomes large so activity in the loop will not increase without bounds. All the modules in the loop are likely to reach attractors in synchrony after extensive learning.

Timing considerations are of importance. Adjusting travel time through the strength of associative connections is a way to control loop dynamics.

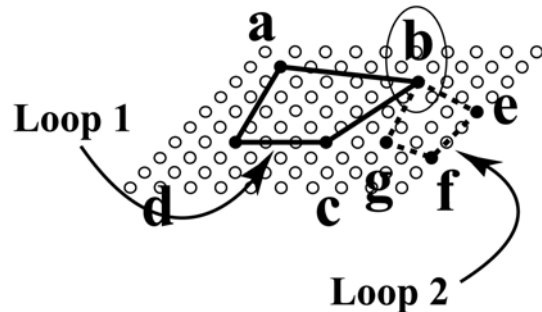
Interconnected Module Assemblies. The most difficult technical problem for realizing Hebb cell assemblies grew from the fact that if two loops contained common portions, activity propagated from one loop to another in what was often uncontrolled spread. It proved to be very difficult

to deal with this situation and attempts to do so introduced unnatural constraints. However, the greater selectivity possible with the patterned intermodule connections offer a way around this difficult.

Consider the situation in the Figure, where two loops contain a common module, that is, Loop 1 is **a-b-c-d** and Loop 2 is **b-e-f-g**. Module **b** is in common between Loop 1 and Loop 2.

There are two essential mechanisms to keep loops separate. The first is that modules can develop multiple attractor states. The second is that other members of the loop form a "context" that can be used for switching.

If the pattern on module **b** is different in the two loops, there is no problem. The intrinsic selectivity of the associative links will keep the loop activities distinct and activity from one loop will not spread easily into the others.



But suppose the pattern on **b** is initially **identical** in the two loops because, for example, the pattern on module **b** is arrives from elsewhere. This situation means on **b** is ambiguous and there is no *a priori* reason to activate Loop 1, Loop 2, or both.

Selectivity is still possible, though it requires additional assumptions to accomplish.

One possibility would be that a loop becomes active only if some critical number of nodes becomes active. This assumption uses context to do disambiguation.

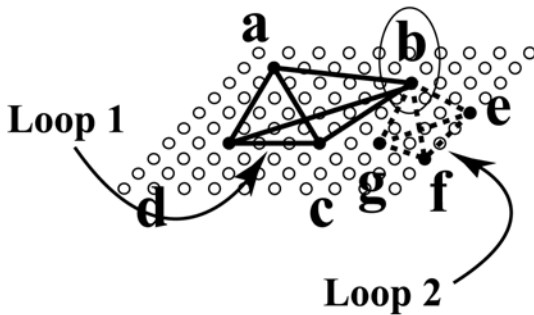
Another possibility is that a general inhibition process allows only a certain number of modules to be active at a time so it is not possible to have two separate loops active simultaneously. Some fine tuning of parameters might be necessary for this to work but it should be possible.

A third possibility is that there is reciprocal inhibition between the loops so Loop 1 inhibits loop 2 and vice versa.

This situation might arise because the active loop would see no activity in its local connections to other modules including those active in other loop and would learn that these activities should be zero when the first loop is active.

These possibilities are not mutually exclusive.

In the previous example, module **b** is only assumed to be

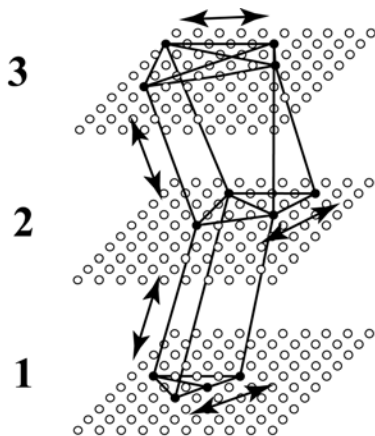


connected to two adjacent modules. More complex connection patterns are possible. One could easily assume richer interconnection patterns, where all allowable connections are learned so that **b** will receive input from **d** as well as **a** and **c**. A larger context would allow better loop disambiguation by increasing

the coupling strength of modules in one loop and perhaps inhibiting other loops more effectively.

Putting in All Together. Sparse interlayer connections and **dense intralayer** connections can work together.

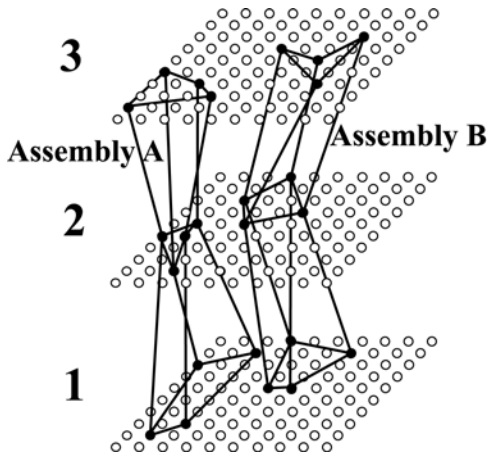
Once a coupled module assembly is formed, it could serve as a basic entity that can be linked to by other layers. Weak sparse paths coming from other layers will be strengthened by the same kinds of local Hebbian learning that we discussed earlier.



All connections bi-directional.

So now we can see the outlines of a dynamic, adaptive computational architecture that becomes both feasible and interesting. The basic notion is that local module assemblies form, perhaps at each layer. The sparse connections between each layer learn to couple the assemblies through learning. The individual sparse paths learn to work together to form complex multilevel

entities. We might end up with something like that shown in the Figure, tightly connected module assemblies within a layer, sparsely linked together. This rings (loops in layers) and strings (sparse connections between layers)

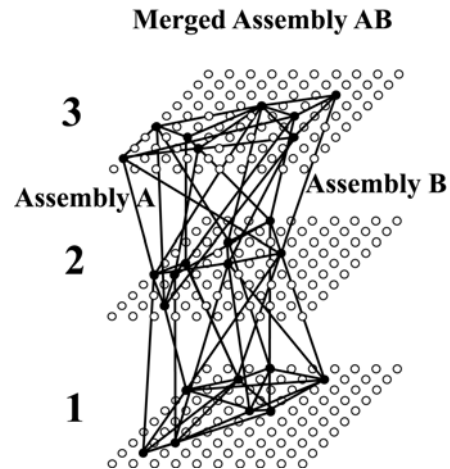


architecture becomes a way to get multi-area activation patterns bound together through common dynamics.

Computation and learning from this point of view is based on the formation of sparsely interconnected between layers, less sparsely interconnected at a layer, module assemblies working with sparsely represented data.

Combinations. A natural way for learning to progress with this approach is to build combinations of module assemblies. Early learning, we conjecture, might form small, but well separated module assemblies. The Figure shows two such module assemblies that are bound together by past learning.

As learning progresses, groups of module assemblies will bind together through the mechanisms we have discussed, if they co-occur. The small early assemblies can act as the "sub-symbolic" substrate of cognition and the larger assemblies, symbols and concepts. Because of the sparseness issues, smaller components will be largely independent of each other and will not interfere with each other. Cognitive learning would have something of the aspect of an erector set, where the parts start being independent and then get bound together to form a sturdy higher level spatially extended assembly. Note that there are more associative connections possible in the bound system than in the parts because there are many more possible paths. An interesting conjecture is that larger assemblies (words? concepts?) are more stable than their component parts.



Note this process looks very much like what is called compositionality. (Geman, 2005). The virtues of compositionality are well known. It is a powerful and flexible way to build information processing systems because complex mental and cognitive objects can be built from previously constructed, statistically well-designed pieces. What we are suggesting here is a possible model for the dynamics and learning of a compositional system. Note however, that this system is built based on constraints derived from connectivity, learning, and dynamics and **not** as a way to do optimal information processing. Perhaps compositionality as we see it manifested in cognitive systems is more like a splendid bug fix than a well chosen computational strategy.

Whether these very preliminary ideas can be made to work to build computational models of cognitive function remains to be determined.

References

- JA Anderson (1993). The BSB network. Pp. 77-103 in MH Hassoun (Ed.), *Associative Neural Networks*. New York, NY: Oxford University Press.
- JA Anderson (1995). *An Introduction to Neural Networks*. Cambridge, MA: MIT Press.
- JA Anderson and JP Sutton (1997). If we compute faster, do we understand better? *Behavior Research Methods, Instruments, and Computers*, **29**, 67-77.
- HB Barlow (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, **1**, 371-394.
- LN Cooper, N Intrator, BS Blais, and HZ Shoval (2004). *Theory of Cortical Plasticity*. Singapore: World Scientific Publishing.
- S Geman (in preparation). Invariance and selectivity in the ventral visual pathway. Division of Applied Mathematics, Brown University, Providence, RI.
- DO Hebb (1949). *The Organization of Behavior*. New York, NY: Wiley.

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

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

JS Lund, A Angelucci, PC Bressloff (2003). Anatomical substrates for functional columns in Macaque monkey primary visual cortex. *Cerebral Cortex*, **12**, 15-24.

BA Olshausen and DJ Field (2004). Sparse coding of sensor inputs. *Current Opinions in Neurobiology*, **14**, 481-487.

RQ Quiroga, L Reddy, G Kreiman, C Koch and I Fried (2005), Invariant visual representations by single neurons in the human brain. *Nature*, **435**, 1102-1107.

N Rochester, JH Holland, LH Haibt, and WL Duda (1956), Tests on a cell assembly theory of the action of the brain using a large digital computer. *IRE Transactions on Information Theory*. **IT-2**, 80-93.

DL Sheinberg and NK Logothetis (2002). Perceptual learning and the development of complex visual representations in temporal cortical neurons. In M Fahle and T Poggio (Eds.), *Perceptual Learning*, 95-124. Cambridge, MA: MIT Press.

J Szantagothai (1978). Specificity versus (quasi-) randomness in cortical connectivity. In MAB Brazier and H Petsche (Eds,) *Architectonics of the Cerebral Cortex*. New York, NY: Raven.

K Tanaka (1996). Inferotemporal cortex and object vision. W.M. Cowan, E.M. Shooter, C.F. Stevens, and R.F Thompson (Eds.) *Annual Review of Neuroscience*, **19**, 109-139.

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.

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.