NASA Technical Memorandum 102239

Ground-State Coding in Partially Connected Neural Networks

Yoram Baram

October 1989

(NASA-TM-102239) GROUND-STATE CODING IN PARTIALLY CONNECTED NEURAL NETWORKS (NASA) 21 p CSCL 09B

N90-12227

Unclas
G3/63 0243003

National Aeronautics and
Space Administration

NASA Technical Memorandum 102239

Ground-State Coding in Partially Connected Neural Networks

Yoram Baram, Ames Research Center, Moffett Field, California

October 1989

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

Ground-State Coding in Partially Connected Neural Networks

Yoram Baram*

Abstract

Patterns over $\{-1, 0, 1\}$ define, by their outer products, partially connected neural networks, consisting of internally strongly connected, externally weakly connected subnetworks. The connectivity patterns may have highly organized structures, such as lattices and fractal trees or nests. Subpatterns over $\{-1, 1\}$ define the subcodes stored in the subnetworks. The network code is defined as the set of permutations of the subcode words, one from each subnetwork, that agree in their common bits. It is first shown that the code words are locally stable states of the network, provided that each of the subcodes consists of mutually orthogonal words or of, at most, two words. Then it is shown that if each of the subcodes consists of two orthogonal words, the code words are the unique ground states (absolute minima) of the Hamiltonian associated with the network. The regions of attraction associated with the code words are shown to grow with the number of subnetworks sharing each of the neurons. Depending on the particular network architecture, the code sizes of partially connected networks can be vastly greater than those of fully connected ones and their error correction capabilities can be significantly greater than those of the disconnected subnetworks. The codes associated with lattice-structured and hierarchical networks are discussed in some detail.

*Y. Baram is with the Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel. He is also associated with the NASA Ames Research Center, Moffett Field, CA 94035. This work was supported in part by the Technion V.P.R. Fund - Albert Einstein Research Fund and in part by the Director's Discretionary Fund, NASA Ames Research Center.

1. Introduction

Neural networks, defined by sums of outer products of binary vectors, were shown by Hopfield (ref. 1) to converge to a local minimum of the Hamiltonian associated with the network. The final state was shown to be, with high probability, one of the given vectors, provided that they are nearly mutually orthogonal and their number does not exceed a certain fraction of the number of neurons. While the possibility of disconnected neurons was mentioned by Hopfield in general terms, the network has been generally perceived as one in which every neuron is connected to all the others. When the stored patterns are selected so as to satisfy orthogonality, they may be viewed as a code, the size of which cannot exceed the number of neurons and must be considerably smaller if a substantial error correction capability is desired. Horn and Weyers (ref. 2) have derived conditions under which orthogonal patterns are unique ground states (absolute minima) of the Hamiltonian. These states may be reachable by such mechanisms as simulated annealing (see Kirkpatrick et al., ref. 3). However, the size of the code allowed by fully connected networks is disturbingly small.

In this paper we consider neural networks defined by outer products of vectors over $\{-1, 0, 1\}$, which will be called the stored patterns. The vector of nonzero bits, ordered according to their order of appearance in a pattern, will be called the subpattern associated with the pattern. Assigning a neuron to every bit position, each group of subpatterns corresponding to the same bits defines a subnetwork by the associated neurons and by the interneural connections obtained by the sum of their outer products. Since the sum of outer products of $\{\pm 1\}$ vectors is likely to produce a significant number of nonzero terms, the subnetworks are said to be internally strongly connected. (They are not necessarily fully connected; the exact connectivity pattern will be determined by the information in the subpatterns). On the other hand, the connectivity between the subnetworks is said to be weak, since they only partly overlap. The network code is defined as the set of all permutations of subpatterns, one from each subnetwork, that agree in their common bits. We first show that the code words are locally stable states of the network, provided that each of the subcodes consists of mutually orthogonal words or of, at most, two words. The regions of attraction associated with the code words are shown to grow with the number of subnetworks sharing each of the neurons. Then we show that if each of the subcodes consists of two orthogonal words, the code words are the unique ground states of the Hamiltonian associated with the network. The network structure need not generally be highly ordered. However, in order to construct specific codes, some organization must be imposed. Depending on the particular network architecture, partially connected neural networks can have considerably greater code sizes than fully connected ones. We consider as examples the codes associated with networks structured as lattices and fractal trees or nests.

2. Patterns, Subpatterns, and Subnetworks

Consider a set of N neurons and a set of M patterns, defined as vectors of dimension N, whose components, having a one-to-one correspondence with the neurons, take their values from

$\{-1, 0, 1\}$. Let the neurons corresponding to the ± 1 bits of each of the patterns define a subnetwork and let the subnetworks be indexed by $i = 1, \ldots, L$. In general, the subnetworks may be of different sizes, but we assume for convenience that the subnetwork size, K, is uniform throughout the network. The vector of ± 1 bits of a pattern, arranged in their order of appearance, will be called a subpattern. Suppose that there are M_i patterns whose subpatterns are associated with the i'th subnetwork and let them be denoted $w_{(i)}^{(l)}$, $l = 1, \ldots, M_i$. A matrix of synaptic parameters relating the neurons to each other is defined by

$$W = \sum_{i=1}^{L} \sum_{l=1}^{M_i} w_{(i)}^{(l)} w_{(i)}^{(l)T} \tag{1}$$

The state x_i of the i'th neuron is updated asynchronously according to

$$x_i = \text{sign}\{\sum_{j=1}^{N} W_{i,j}\} = \begin{cases} +1 & \text{if } \sum_{j=1}^{N} W_{i,j} x_j > 0 \\ x_i & \text{if } \sum_{j=1}^{N} W_{i,j} x_j = 0 \\ -1 & \text{if } \sum_{j=1}^{N} W_{i,j} x_j < 0 \end{cases} \tag{2}$$

The pattern structures, that is, the distribution of the $\{-1, 0, 1\}$ values in the bit positions, determine the connectivity structure of the network, as illustrated by the following examples.

Example 2.1 The patterns

$$\begin{array}{ll} w_{(1)}^{(1)} = (+ + + + 00)^T & w_{(1)}^{(2)} = (+ - + - 00)^T \\ w_{(2)}^{(1)} = (00 + + + +)^T & w_{(2)}^{(2)} = (00 + - + -)^T \end{array}$$

where + and $-$ represent +1 and -1, respectively, define the six-neuron network, depicted in figure 1(a), which consists of two subnetworks of size four, sharing two neurons. The associated connectivity matrix

$$W = \begin{bmatrix} 2 & 0 & 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 & 0 \\ 2 & 0 & 4 & 0 & 2 & 0 \\ 0 & 2 & 0 & 4 & 0 & 2 \\ 0 & 0 & 2 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 & 0 & 2 \end{bmatrix}$$

defines the interneural connections represented in the figure by solid lines.

The above example shows that partial connectivity can be created not only by partial overlap between the stored subpatterns, but also by the information contents of the subpatterns themselves. This is further illustrated by the following example.

Example 2.2 The subpatterns

$$w_{(1)}^{(1)} = (+ + + +)^T \quad \text{and} \quad w_{(1)}^{(2)} = (+ + - -)^T$$

yield the matrix

$$W = \begin{bmatrix} 2 & 2 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \end{bmatrix}$$

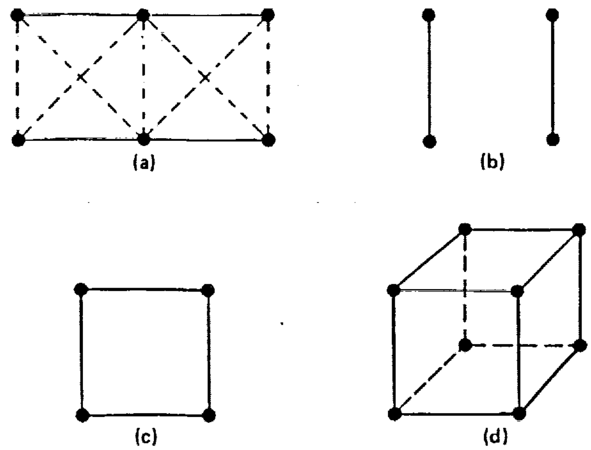


Figure 1: Partially connected subnetworks for examples 2.1 and 2.2.

which defines the four-neuron "subnetwork" depicted in figure 1(b). The subpatterns

$$w_{(1)}^{(1)} = (++\ ++)^T \quad \text{and} \quad w_{(1)}^{(2)} = (+-+-)^T$$

yield the matrix

$$W = \begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \end{bmatrix}$$

which defines the square subnetwork depicted in figure 1(c). Similarly, the subpatterns

$$w_{(1)}^{(1)} = (++\ ++\ ++\ ++)^T \quad \text{and} \quad w_{(1)}^{(2)} = (+-+-+-+-)^T$$

define the cubical subnetwork depicted in figure 1(d).

We call the set of subpatterns corresponding to a subnetwork a subcode. The code of the network is defined as the set of vectors consisting of permutations of subcode words, one corresponding to each of the subnetworks, that agree in their common bits and the converses (or negative versions) of these vectors. In example 2.1, the subcode stored in each of the subnetworks is $\{(++\ ++), (+-+-)\}$ and the network code is $\{(++\ ++\ ++), (+-+-+-), (------), (-+-+-+)\}$.

3. Local Stability and Attraction

3.1 Equilibrium States

It follows from equation (2) that the equilibrium states of the network are the states that satisfy the equation

$$x = \mathrm{sign}\{Wx\} = \mathrm{sign}\{\sum_{i=1}^{L}\sum_{l=1}^{M_i}(w_{(i)}^{(l)}, x)\, w_{(i)}^{(l)}\} \tag{3}$$

where $(x, y) = x^T y$, which implies that if x is an equilibrium point, so is its converse. Suppose that the network is probed by a code word $\bar{x}$ and let l_i denote the index of the subcode word in $\bar{x}$, corresponding to the i'th subnetwork. Suppose further that the k'th neuron is shared by n_k subnetworks. Then

$$[W\bar{x}]_k = K n_k \bar{x}_k + \sum_{i=1}^{L}\sum_{\substack{l=1\\ l\neq l_i}}^{M_i}(w_{(i)}^{(l)}, \bar{x})[w_{(i)}^{(l)}]_k \tag{4}$$

Denoting the indices of the subnetworks that share the k'th neuron $i_k = 1, \ldots, n_k$, it can be seen that a sufficient condition for a code word to be an equilibrium state of the network is

$$K n_k > \left| \left[\sum_{i_k=1}^{n_k}\sum_{\substack{l=1\\ l\neq l_{i_k}}}^{M_i}(w_{(i_k)}^{(l)}, \bar{x})\, w_{(i_k)}^{(l)}\right]_k \right| \tag{5}$$

If the subnetworks are disconnected, the condition becomes

$$K > \left| \left[\sum_{\substack{l=1\\ l\neq l_i}}^{M}(w_{(i)}^{(l)}, \bar{x})\, w_{(i)}^{(l)}\right]_k \right| \tag{6}$$

which is stricter than (5).

A question of interest is whether code words can be guaranteed to be equilibrium points. It can readily be seen that when the subpatterns corresponding to each of the subnetworks are mutually orthogonal, each of the code words is an equilibrium point of the network, as in this case

$$(w_{(i)}^{(l)}, \bar{x}) = 0 \quad \text{for all } l \neq l_i$$

and condition (5) is satisfied. Suppose next that each of the subnetworks stores, at most, two subpatterns, associated with $w_{(i)}^{(1)}$ and $w_{(i)}^{(2)}$, $i = 1, \ldots, L$. Let a code word $\bar{x}$ consist of a permutation of these subpatterns, one per subnetwork, $w_{(i)}^{(l_i)}$, $i = 1, \ldots, L$, $l_i = 1$ or 2. Then

$$[W\bar{x}]_k = K n_k \bar{x}_k + \sum_{i_k=1}^{n_k}\sum_{\substack{i_k=1\\ l\neq l_{i_k}}}^{2}(w_{(i_k)}^{(l)}, \bar{x})[w_{(i_k)}^{(l)}]_k \tag{7}$$

Since

$$|\,(w_{(i_k)}^{(l)}, \bar{x})\,| \leq K$$

with equality if and only if $w_{(i_k)}^{(l)}$ is equal to $w_{(i)}^{(l_{i_k})}$ or to its converse, (5) is satisfied. It follows that $\bar{x}$ is an equilibrium point.

We have seen that the code words are equilibrium points of the network if the subcode words stored in each of the subnetworks are mutually orthogonal or if their number does not exceed two. The question arises whether in these cases the code words are the only equilibrium points. As the following example shows, this is not necessarily the case, even when each of the subcodes consists of two orthogonal words.

Example 3.1 Consider the network of example 2.1. It can readily be verified that the code words, $\bar{x}_1 = (+ + + + + +)^T$, $\bar{x}_2 = (+ - + - + -)^T$, $\bar{x}_3 = -\bar{x}_1$ and $\bar{x}_4 = -\bar{x}_2$ are equilibrium points of the network. It can also be verified, however, that the pattern $x = (+ + + - + -)^T$, which differs from $\bar{x}_2$ by a single bit, is also an equilibrium point, satisfying (3). In the following sections it will be shown that such equilibrium states are local, not absolute, minima of the associated Hamiltonian.

3.2 Local Attraction

We next examine the regions of attraction associated with the code words. Let $\bar{x}$ denote a code word, consisting of the subpatterns corresponding to $w_{(i)}^{(l_i)}$, $i = 1, \ldots, L$. A pattern x is in the region of attraction of $\bar{x}$ if

$$\operatorname{sign}\{Wx\} = \bar{x} \tag{8}$$

If the network's state is within the region of attraction of $\bar{x}$, then it will converge to the latter with probability 1, provided that the probability of each of the neurons being selected for update on each step is nonzero. Suppose first that each of the subnetworks stores two subpatterns, which may not be orthogonal, corresponding to $w_{(i)}^{(1)}$ and $w_{(i)}^{(2)}$, $i = 1, \ldots, L$. Indexing, as before, the subnetworks sharing the k'th neuron $i_k = 1, \ldots, n_k$, and writing

$$[Wx]_k = \sum_{i_k=1}^{n_k} (w_{(i_k)}^{(l_{i_k})}, x)[w_{(i_k)}^{(l_{i_k})}]_k + \sum_{\substack{i_k=1 \\ l \neq l_{i_k}}}^{n_k} (w_{(i_k)}^{(l)}, x)[w_{(i_k)}^{(l)}]_k \tag{9}$$

where $l_{i_k} = 1$ or 2, it can be seen that a sufficient condition for x to be in the attraction region of $\bar{x}$ is

$$\left| \sum_{i_k=1}^{n_k} (w_{(i_k)}^{(l_{i_k})}, x) \right| > \sum_{\substack{i_k=1 \\ l \neq l_{i_k}}}^{n_k} \left| (w_{(i_k)}^{(l)}, x) \right| \tag{10}$$

A sufficient condition for (10) is

$$\sum_{i_k=1}^{n_k} (w_{(i_k)}^{(l_{i_k})}, x) > \sum_{\substack{i_k=1 \\ l \neq l_{i_k}}}^{n_k} \left| (w_{(i_k)}^{(l)}, x) \right| \tag{11}$$

Next, suppose that the stored subcode words are mutually orthogonal and that their number in each of the subnetworks is uniformly M. Further, suppose that the subpatterns $w_{(i)}^{(l_i)}, i =$

$1, \ldots, q$, agree in their common bits, forming a code word. Denoting by $d[x, y]$ the Hamming distance between x and y, we have

$$[Wx]_k = \sum_{i=1}^{n_k} (K - 2\,d[w_{(i)}^{(l_i)}, x])[w_{(i)}^{(l_i)}]_k + \sum_{i=1}^{n_k} \sum_{\substack{l=1 \\ l \neq l_i}}^{M} (K - 2\,d[w_{(i)}^{(l)}, x])[w_{(i)}^{(l)}]_k \tag{12}$$

Since there is no conflict between the k'th bits of $w_{(i)}^{(l_i)}$, $i = 1, \ldots, n_k$, it follows from the neural update rule (2) that the distances $d[w_{(i)}^{(l_i)}, x]$, $i = 1, \ldots, n_k$ will decrease (if $x_k \neq [w_{(i)}^{(l_i)}]_k$) or remain the same (if $x_k = [w_{(i)}^{(l_i)}]_k$) if

$$\sum_{i=1}^{n_k} (K - 2\,d[w_{(i)}^{(l_i)}, x]) > \sum_{i=1}^{n_k} \sum_{\substack{l=1 \\ l \neq l_i}}^{M} |\, K - 2\,d[w_{(i)}^{(l)}, x] \,| \tag{13}$$

It follows that, since the stored subpatterns are mutually orthogonal, the maximum possible value of $|\, K - 2\,d[w_{(i)}^{(l)}, x] \,|$ is $2\,d[w_{(i)}^{(l_i)}, x]$ (it is obtained for either $d[w_{(i)}^{(l)}, x] = K/2 - d[w_{(i)}^{(l_i)}, x]$ or $d[w_{(i)}^{(l)}, x] = K/2 + d[w_{(i)}^{(l_i)}, x]$). This implies that convergence to the code word in question can be guaranteed if

$$\sum_{i=1}^{n_k} (K - 2\,d[w_{(i)}^{(l_i)}, x]) > \sum_{i=1}^{n_k} 2(M-1)\,d[w_{(i)}^{(l_i)}, x] \tag{14}$$

or

$$\sum_{i=1}^{n_k} d[w_{(i)}^{(l_i)}, x] < \frac{n_k K}{2M} \tag{15}$$

For disconnected subnetworks, a sufficient condition is

$$K - 2\,d[w_{(i)}^{(l_i)}, x] > 2(M-1)\,d[w_{(i)}^{(l_i)}, x]$$

yielding, for each of the subnetworks,

$$d[w_{(i)}^{(l_i)}, x] < \frac{K}{2M} \quad \text{for} \quad i = 1, \ldots, n_k \tag{16}$$

which is more restrictive than (15). The number of mutually orthogonal subpatterns that can be stored in a subnetwork is restricted, of course, by the subnetwork size. It can be seen from (15) and (16) that the maximal guaranteed regions of attraction are obtained for $M = 2$ (disregarding the trivial case $M = 1$).

We have seen that when the stored subpatterns corresponding to the subnetworks are mutually orthogonal, or if their number for each subnetwork does not exceed two, the code words are equilibrium points of the network, with regions of attraction that grow with the number of subnetworks sharing each of the neurons. The maximal guaranteed region of attraction is obtained for two orthogonal subpatterns corresponding to each of the subnetworks (excluding the trivial case of a single subpattern per subnetwork). It should be emphasized that even in this case there may be spurious local attractors, which are not code words.

4. Ground States

Suppose that the subpatterns stored in each of the subnetworks are mutually orthogonal and consider the Hamiltonian

$$H(x) = -x^T W x \tag{17}$$

which may be written as

$$H(x) = -\sum_{i=1}^{L}\sum_{l=1}^{M_i}(w_{(i)}^{(l)}, x)^2 \tag{18}$$

Noting that, for any code word $\bar{x}$

$$(w_{(i)}^{(l)}, \bar{x}) = K \tag{19}$$

we obtain

$$H(\bar{x}) = -K^2 L \tag{20}$$

Let

$$\hat{x}_{(i)}^{(l)} \equiv \frac{(w_{(i)}^{(l)}, x)}{||w_{(i)}^{(l)}||^2} w_{(i)}^{(l)} = \frac{1}{K}(w_{(i)}^{(l)}, x)\, w_{(i)}^{(l)} \tag{21}$$

where $||x||^2 = (x, x)$, denote the orthogonal projection of an arbitrary pattern x on $w_{(i)}^{(l)}$. We have

$$Wx = \sum_{i=1}^{L}\sum_{l=1}^{M_i}(w_{(i)}^{(l)}, x)\, w_{(i)}^{(l)} \tag{22}$$

Substituting (21) into (22) yields

$$Wx = K\sum_{i=1}^{L}\sum_{l=1}^{M_i}\hat{x}_{(i)}^{(l)} \tag{23}$$

Let us denote by $x_{(i)}$ the part of x corresponding to the i'th subnetwork and by S_i the subspace spaned by the subpatterns corresponding to this subnetwork, that is, $S_i = \text{span}\{w_{(i)}^{(l)},\ l = 1, \ldots, M_i\}$. The orthogonal projection of x on S_i is

$$\hat{x}_{(i)} \equiv \sum_{l=1}^{M_i}\hat{x}_{(i)}^{(l)} \tag{24}$$

It can be seen that

$$(x, \hat{x}_{(i)}) = (x_{(i)}, \hat{x}_{(i)}) \leq K \tag{25}$$

with equality if and only if $\hat{x}_{(i)} = x_{(i)}$, which is the case if and only if $x_{(i)}$ belongs to S_i. It follows that

$$H(x) = -K\sum_{i=1}^{L}(x_{(i)}, \hat{x}_{(i)}) \tag{26}$$

yielding

$$H(x) \geq -K^2 L \tag{27}$$

with equality if and only if $x_{(i)}$ belongs to S_i for all $i = 1, \ldots, L$. We have thus shown that an arbitrary state x has a minimal Hamiltonian value if and only if all its parts corresponding to the subnetworks, $x_{(i)}$, belong to the corresponding subspaces S_i, $i = 1, \ldots, L$.

Suppose that each of the subnetworks stores only two orthogonal words. Further suppose that x has minimal $H(x)$, hence that $x_{(i)}$ belongs to S_i for all i. We next show that this implies that x is a code word. Since the subcode words stored in each of the subnetworks agree in at least one bit and differ in at least one bit, and since $x_{(i)}$, whose components are ± 1, belongs to S_i, there exist two bits corresponding to the i'th subnetwork, say, the k'th and the m'th, and scalars c_1 and c_2 such that

$$[c_1 w_{(i)}^{(1)}]_k + [c_2 w_{(i)}^{(2)}]_k = \pm 1 \tag{28}$$

and

$$[c_1 w_{(i)}^{(1)}]_m + [c_2 w_{(i)}^{(2)}]_m = \pm 1 \tag{29}$$

yielding

$$c_1 + c_2 = \pm 1 \tag{30}$$

and

$$c_1 - c_2 = \pm 1 \tag{31}$$

However, taking the squares of both equations and subtracting, we obtain

$$c_1 c_2 = 0 \tag{32}$$

implying that either $c_1 = 0$ and $c_2 = 1$, or $c_2 = 0$ and $c_1 = 1$. But this means that $x_{(i)}$ is equal to either $w_{(i)}^{(1)}$ or $w_{(i)}^{(2)}$. Hence x is a code word. We have thus proven the following results.

Theorem Suppose that each of the subnetworks stores two orthogonal subpatterns. Then the only states of minimal H are the code words.

As in the case of fully connected networks (ref. 1), the Hamiltonian is nonincreasing along any trajectory in the state space, due to the symmetry of W and the asynchronous neural update rule. This, however, does not imply that a ground state will be reached by direct convergence from any initial state, as there can be local minima, corresponding to higher values of the Hamiltonian. A ground state will be attained by direct relaxation if the initial state falls within its region of attraction. This would be the case if the initial state represents a sufficiently close approximation of a code word. For larger errors, more complex mechanisms, such as simulated annealing (see Kirkpatrick et al., ref. 3) may be necessary for reaching the ground states.

5. Some Structural Codes

So far, no assumptions have been made on the network's connectivity structure, which is generally determined by the stored patterns. Highly structured networks allowing only certain connections between neurons may result from the storage of similarly structured patterns, or may be given prior to the storage of information. In the latter case, the stored information will impose a structure within the given structure. It should be emphasized that for all purposes of this paper, the network structure merely defines the connections between neurons and between subnetworks, which may be physically performed by fibers taking any geometric form or no uniformly prescribed form at all. Specific geometries would be meaningful, however, in the physical construction of neural networks.

5.1 Lattice Networks

We consider a network of N neurons, grouped into subnetworks of K neurons each, whose internal connections are to be determined by the stored information. The relationship between the subnetworks is defined by a lattice structure, in which the subnetworks form the building blocks or the fundamental units (see, e.g., Conway and Sloan, ref. 4). In order to characterize the code of a lattice neural network (not to be confused with "lattice codes," which are the centers of the fundamental units, ref. 4), we form a set of "chinese boxes" as follows. Select some subnetwork to be the innermost, or the first box. This subnetwork also defines the first shell. The subnetworks directly connected to it define the second shell. Together, the first box and the second shell define the second box. The subnetworks directly connected to the second box form the third shell and together they define the third box, etc.. Let us denote the subcode for the j'th subnetwork of the k'th shell c_j^k and the code corresponding to the k'th box C_k. Denoting by r_k a word of C_k, by $r_{k,k}$ its part in the $(k-1)$'th box and by r_k^k its part in the k'th shell, let $c_j^k(r_{k,k})$ be the code allowed by $r_{k,k}$ in the j'th subnetwork of the k'th shell. Denoting by L_k the number of subnetworks in the k'th shell, the set of permissible permutations of words in the subnetworks of the k'th shell associated with $r_{k,k}$ is given by the Cartesian product $\otimes_{j=1}^{L_k} c_j^k(r_{k,k})$. The box codes may be defined progressively by

$$C_{k+1} = \{r_{k+1} : r_{k+1,k+1} \in C_k,\ r_{k+1}^{k+1} \in \otimes_{j=1}^{L_k} c_j^k(r_{k,k})\} \tag{33}$$

with $C_1 = c_1^1$, as the innermost box consists of a single subnetwork. Denoting by M_j^k the number of subwords in $c_j^k(r_{k,k})$, and by Q the number of shells, the size of the network code is given by

$$M = \Pi_{k=1}^{Q} \Pi_{j=1}^{L_k} M_j^k \tag{34}$$

These expressions provide a general characterization of lattice neural codes. More concrete formulations will require specifying both the particular lattice structure and the subcodes stored in the subnetworks. The following examples, which involve networks structured as lattices in the plane with relatively small subnetworks, illustrate that certain lattice structures can yield larger codes than others.

Example 5.1 Consider the diamond (or "checkerboard") lattice network depicted in figure 2. It is not difficult to see that the addition of the Q'th shell to the $(Q-1)$'th box adds $4(Q-1)$ subnetworks, which, in turn, adds $8Q-4$ neurons $(3 \times 4 + 2(4(Q-1)-4))$. The size of a network of Q shells is readily obtained as $N = 4Q^2$. Suppose that a subcode consisting of the two subwords, $(++++), (+--+-)$, is stored in each of the subnetworks. It can be seen that the addition of a shell to the network multiplies the code size by 16 (there are two possible subwords for each of the four corner subnetworks; the words in the other subnetworks of the shell are determined by the previous box). The code size for a network of Q shells is, then, $M = 4 \times 16^{Q-1}$ (including converses). The ratio of code size to network size

$$\rho = \frac{16^{Q-1}}{Q^2}$$

is a monotone increasing function of Q. For the given three-shell network, the code consists of $4 \times 16^2 = 1024$ words, two of which are shown in the figure.

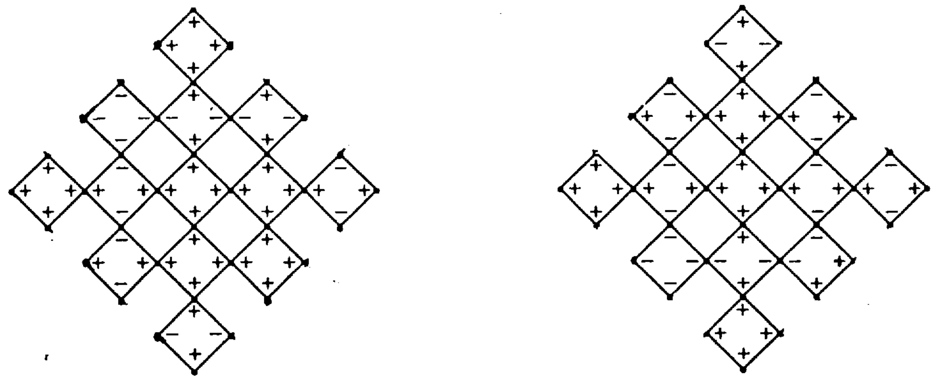


Figure 2: Two of the 1024 code words of a three-shell diamond lattice network storing the subcode $\{(+ + ++), (+ - + -)\}$ in each of the subnetworks.

Example 5.2 Now consider the circular lattice of figure 3. Suppose that the subcode to be stored in each of the subnetworks is $\{(+ + + + ++), (+ - + - + -)\}$. It is not difficult to see that the first two shells uniquely determine the neural state values for the rest of the network and that the network code size cannot be greater than eight (including converses), regardless of the network size. The reason for this finite code size is the tight packing of the lattice, which causes each subnetwork to share pairs of neighboring neurons with pairs of neighboring subnetworks. Half the code is shown in the figure (the converses of the shown words constitute the other half).

Example 5.3 In order to examine the error correction capability of a lattice network, we considered the portion of the network of figure 3, consisting of the seven inner circular subnetworks. The subcode $\{(+ + + + ++), (+ - + - + -)\}$ was stored in each of the subnetworks. Each of the eight code words was corrupted by errors, so that the probability of each of the bits having a reversed sign was p. Fifty corrupted versions of each of the eight code words were generated and presented to the network, which was allowed, for each such probe, to relax to a final state. The final error for each of the probes was calculated as the Hamming distance $(\bar{x} - x_f)^T(\bar{x} - x_f)/4$, where $\bar{x}$ is the corresponding uncorrupted code word and x_f is the corresponding final state. The average error was calculated for the entire set of 400 words. The experiment was then repeated for a network consisting of the disconnected subnetworks (seven subnetworks consisting of six neurons each), storing the same subcode and probed by the same code words as the lattice network. The results for several values of p are shown below, where $e_{lattice}$ denotes the average error for the lattice network and e_{dc} the average error for the network of disconnected subnetworks.

p :	0.0	0.1	0.2	0.3	0.4	0.5
$e_{lattice}$:	0.0000	0.2000	1.3400	4.3400	8.1600	11.5400
e_{dc} :	0.0000	1.3797	4.2602	9.0601	14.6398	21.4802

It can be seen that, although the lattice network consists of less neurons (30) than the network of disconnected subnetworks (42), it has a considerably higher error correction capability.

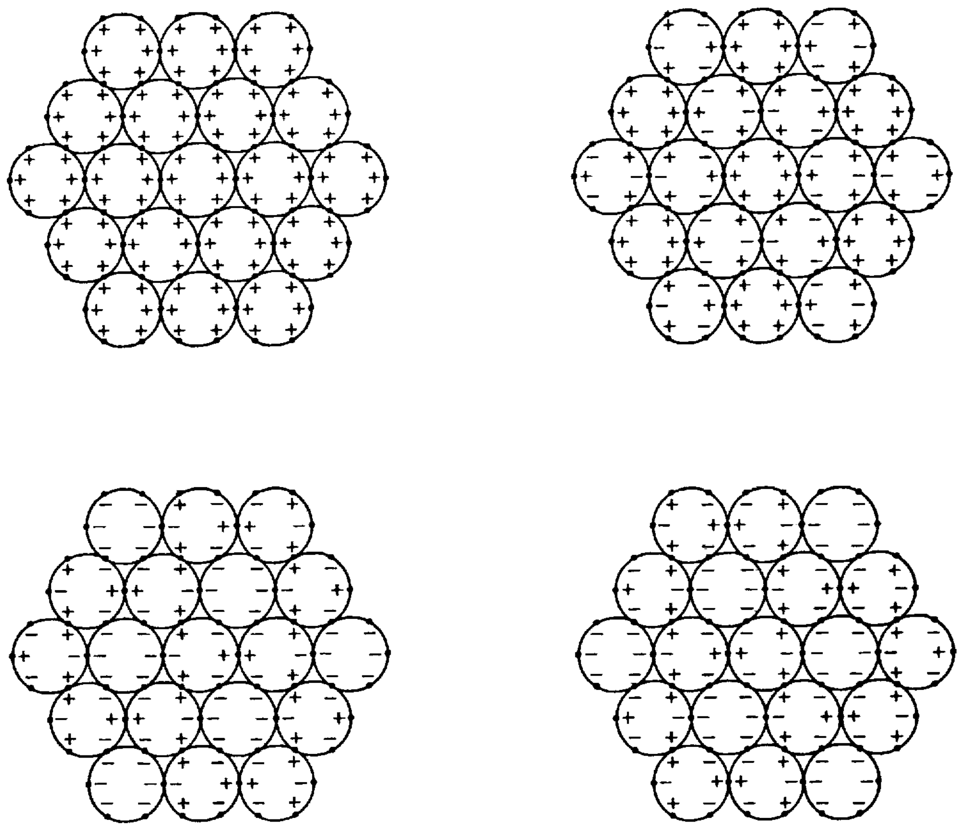


Figure 3: Half the code of a circular lattice network storing the subcode $\{(++++++), (+-+-+-)\}$.

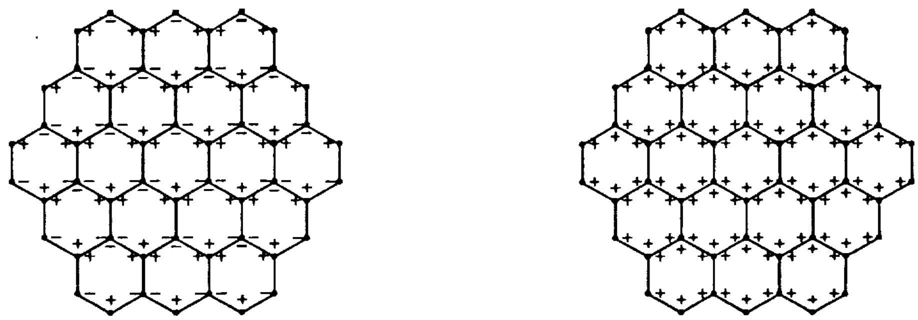


Figure 4: Half the code of the hexagonal lattice network storing the subcode $\{(++\,++\,++),(+-+-+-)\}$.

Example 5.4 Consider the hexagonal lattice network of figure 4. It can be seen that, if the subcode $\{(++\,++\,++),(+-+-+-)\}$ is stored in each of the subnetworks, there are only four code words, the first of which has $(++\,++\,++)$, the second $(+-+-+-)$ in all the subnetworks, and the other two are the converses of the first. Half the code is shown in the figure. This code size is a consequence of the fact that every two neighboring subnetworks share two neighboring neurons. Independently of the lattice geometry, every network having this property will have code size four.

As the above examples indicate, the code size is largely affected by inter-subnetwork relations within the lattice structure. While the last two examples present cases of limited code sizes, the first example presents a case in which the code size increases with the network size. An increase of the subnetwork size, and thereby of the error correction capability, can be achieved in this case by extending the lattice structure to higher dimensions (see, e.g., Conway and Sloane, ref. 4). A three-dimensional extension of the diamondal lattice is the cubical lattice depicted in figure 5. It consists of eight neurons in each subnetwork and its code size for Q shells is $4 \times 256^{Q-1}$.

5.2 Hierarchical Structures

Information structures or physical development processes may give rise to hierarchical network architectures. Figure 6(a) depicts a network having a "fractal" (see Mandelbrot, ref. 5) tree structure. In this network, which consists of layers of subnetworks of the same number of neurons, each subnetwork in layers above the lowest consists of neurons of a lower layer, one from each subnetwork. Subsequently, each neuron in these layers is a member of two subnetworks belonging to two neighboring layers. The subnetworks at each layer may also be connected into

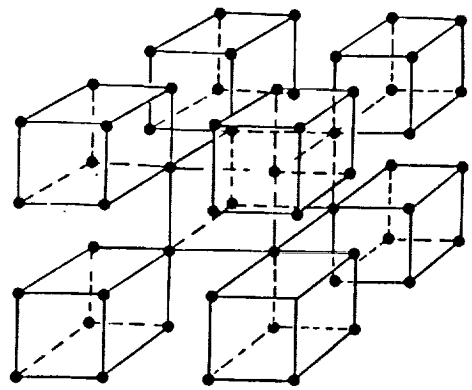


Figure 5: A three-dimensional checkerboard cubical structure.

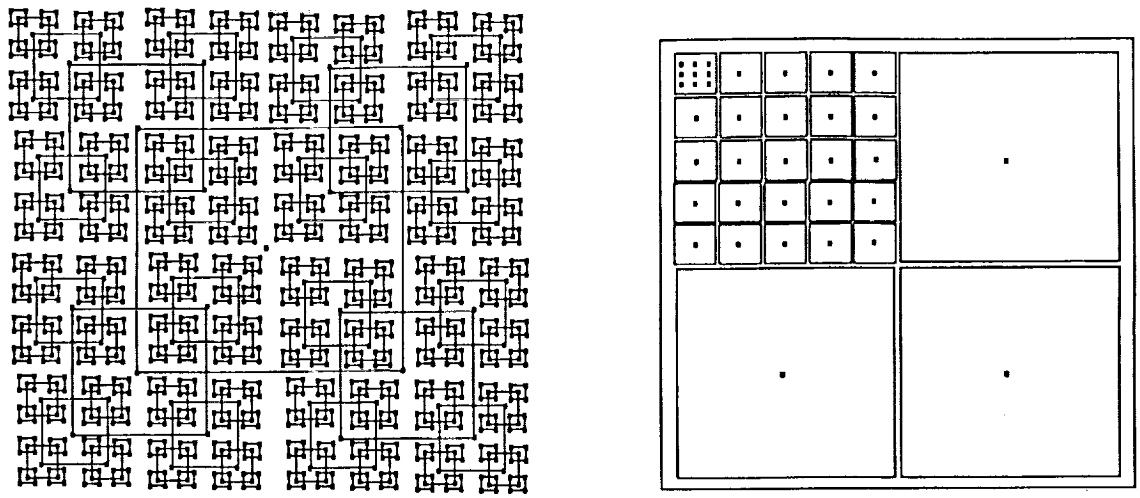
a lattice structure. In the network depicted in figure 6(b), where only the neurons of a single subnetwork in each layer are shown, each neuron at a given layer is also a member of all the layers beneath it. Hence, the entire network consists of the neurons of the lowest layer and the structure is defined solely by the interneural connections. Since higher layers are nested in lower ones, we call the structure a nested network. The following will apply to both types of structures.

In order to characterize the code of a hierarchical network, we index the layers from top to bottom and define the k'th pyramid as the structure consisting of the k top layers. The code construction is completely analogous to that for lattice networks, when analogies are drawn between shells and layers and between boxes and pyramids. Let us denote the subcode for the subnetworks of the k'th layer c_k and the code corresponding to the k'th pyramid C_k. We shall assume that the symbols in each word of c_k and of C_k are arranged in some order, forming vectors of appropriate dimensions. In the hierarchical structure, one of the bit positions of any subcode word position in the $k+1$'st layer is also shared by a subcode word position of the k'th layer. We shall assume for simplicity that this bit position, which will be called the "common" position, is the same for all the subcode word positions of a given layer. Since the words in the $k+1$'th layer must agree with those in the k'th layer in the common bit position, only certain permutations of the subcode words of the different layers are permissible. Denoting by r_k a word of C_k, by $r_{k,k}$ its part in the $k-1$'th pyramid, and by r_k^k its part in the k'th layer, let $c_j^k(r_{k,k})$ be the code allowed by r_k^k in the j'th word position of the k'th layer. The set of permissible permutations of words in the word positions of the k'th layer associated with $r_{k,k}$ is given by the Cartesian product $\otimes_{j=1}^{L_k} c_j^k(r_{k,k})$, where L_k is the number of subnetworks in the k'th layer. The pyramidal codes may be obtained recursively by the operation

$$C_{k+1} \equiv \{r_{k+1} : r_{k+1,k+1} \in C_k,\ r_{k+1}^{k+1} \in \otimes_{j=1}^{L_k} c_j^k(r_{k,k})\} \tag{35}$$

with $r_1 = c_1$. Denoting by M_j^k the number of words in c_j^k, the size of the code is given by

$$M = \Pi_{k=1}^{K} \Pi_{j=1}^{L_k} M_j^k \tag{36}$$



(a) (b)

Figure 6: Tree (a) and nested (b) network architectures.

where K denotes the number of layers in the network. Both (35) and (36) have the same forms as the corresponding expressions for the lattice networks. It should be noted that when the subnetworks consist of odd numbers of neurons, the associated subpatterns can only be made nearly, not strictly, orthogonal (their inner products would produce 1 instead of 0). For large subnetworks, this slight diversion from the orthogonality assuption is not expected to result in a significantly different network performance. Relatively small subnetworks may be restricted to even sizes in order to maintain the orthogonality requirement.

Example 5.5 Consider the network depicted in figure 7. It consists of two layers of subnetworks of six neurons each, placed at the corners and the centers of pentagons, with the center neurons at the lower layer being shared with the subnetwork of the higher layer. For graphical clarity, only some of the interneural connections within the subnetworks are shown. Each of the subnetworks stores the subcode

$$\begin{bmatrix} + + + + + + \\ + - + - + - \end{bmatrix}$$

in which the first bit corresponds to center (common) neurons. It can be seen that the only permissible subcode word in the upper layer is $(+ + + + + +)$ and that the code size is $2^6 = 64$. Ten corrupted versions of each of the code words were generated, with an error probability p per bit. The network was probed by each of these 640 words and allowed to relax to a final state. The average error was calculated, as in example 5.3, for several values of p. The experiment was then repeated for a network consisting of the disconnected subnetworks of the lower layer. The results are given below, where e_{nest} and e_{dc} denote the average errors for the nested network and

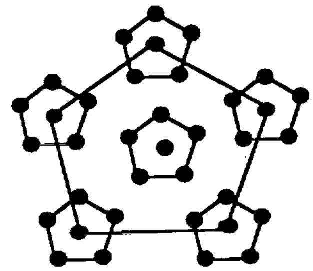


Figure 7: A two-layer nested network for example 5.5.

for the network of disconnected subnetworks, respectively.

p :	0.0	0.1	0.2	0.3	0.4	0.5
e_{nest} :	0.0000	0.8378	2.8004	5.0196	6.6997	8.2418
e_{dc} :	0.0000	0.6300	3.6900	9.1800	14.2200	17.8200

It can be seen that, for substantial errors, the nested network has a significantly higher error-correction capability than the network of disconnected subnetworks.

6. Conclusion

Outer products of vectors over $\{-1, 0, 1\}$ define partially connected neural networks, consisting of subnetworks corresponding to the nonzero bits. When each of the subnetworks stores, at most, two mutually orthogonal subpatterns, the code words, defined as the permutations of the ± 1 subpatterns that agree in their common bits, are the unique ground states of the associated Hamiltonian. These states can be reached by direct relaxation, if the initial state falls within their regions of attraction, or, otherwise, by such mechanisms as simulated annealing. Specific codes may be constructed by choice of the network structure and the subcodes corresponding to the subnetworks.

References

1. Hopfield, J. J.: Neural Networks and Physical Systems with Emergent Computational Abilities. Proc. Nat. Acad. Sci. USA, vol. 79, Apr. 1982, pp. 2554-2558.

2. Horn, D.; and Weyers, J.: Hypercube Structures in Orthogonal Hopfield Models. Physical Review A, vol. 36, no. 10, Nov. 1987, pp. 4968-4974.

3. Kirkpatrick, S.; Gelatt, C. D., Jr.; and Vecchi, M. P.: Optimization by Simulated Annealing. Science, vol. 220, no. 4598, May 1983, pp. 671-680.

4. Conway, J. H.; and Sloan, N. J. A.: Sphere Packings, Lattices and Groups. Springer Verlag, New York, 1988.

5. Mandelbrot, B.: The Fractal Geometry of Nature. W. H. Freeman and Co., New York, 1983.

Report Documentation Page

National Aeronautics and Space Administration

1. Report No. NASA TM-102239	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle

Ground-State Coding in Partially Connected Neural Networks

5. Report Date

October 1989

6. Performing Organization Code

7. Author(s)

Yoram Baram*

8. Performing Organization Report No.

A-89256

10. Work Unit No.

505-67-21

9. Performing Organization Name and Address

Ames Research Center
Moffett Field, CA 94035

11. Contract or Grant No.

13. Type of Report and Period Covered

Technical Memorandum

12. Sponsoring Agency Name and Address

National Aeronautics and Space Administration
Washington, DC 20546-0001

14. Sponsoring Agency Code

15. Supplementary Notes

Point of Contact: Dr. Leonard Tobias, Ames Research Center, MS 210-9, Moffett Field, CA 94035
(415) 694-5430 or FTS 464-5430

*Permanent Address: Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel.

16. Abstract

Patterns over $\{-1,0,1\}$ define, by their outer products, partially connected neural networks, consisting of internally strongly connected, externally weakly connected subnetworks. The connectivity patterns may have highly organized structures, such as lattices and fractal trees or nests. Subpatterns over $\{-1,1\}$ define the subcodes stored in the subnetworks. The network code is defined as the set of permutations of the subcode words, one from each subnetwork, that agree in their common bits. It is first shown that the code words are locally stable states of the network, provided that each of the subcodes consists of mutually orthogonal words or of, at most, two words. Then it is shown that if each of the subcodes consists of two orthogonal words, the code words are the unique ground states (absolute minima) of the Hamiltonian associated with the network. The regions of attraction associated with the code words are shown to grow with the number of subnetworks sharing each of the neurons. Depending on the particular network architecture, the code sizes of partially connected networks can be vastly greater than those of fully connected ones and their error correction capabilities can be significantly greater than those of the disconnected subnetworks. The codes associated with lattice-structured and hierarchical networks are discussed in some detail.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Neural networks Coding Associative memory	Unclassified-Unlimited Subject Category - 63

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	18	A02

NASA FORM 1626 OCT 86

For sale by the National Technical Information Service, Springfield, Virginia 22161