

Cerebellar Models of Associative Memory: Three papers from IEEE COMPCON SPRING '89

Michael R. Raugh, Editor

Containing papers by:

*James Albus
Pentti Kanerva
Egon Loebner*

March 1989

*IN-63
DATE OVERRIDE
58204
p-24
ANALYTIC
(1+3)*

Research Institute for Advanced Computer Science
NASA Ames Research Center

RIACS Technical Report 89.11

NASA Cooperative Agreement Number NCC 2-408



Research Institute for Advanced Computer Science
An Institute of the Universities Space Research Association

(NASA-CR-188838) CEREBELLAR MODELS OF ASSOCIATIVE MEMORY: THREE PAPERS FROM IEEE COMPCON SPRING 1989 (Research Inst. for Advanced Computer Science) 24 p CACL 09B N92-14694 --THRU-- N92-14697 Unclas 63/63 0058204

Primary

CEREBELLAR MODELS OF ASSOCIATIVE MEMORY:
THREE PAPERS FROM
IEEE COMPCON SPRING '89

Michael R. Raugh, Editor

Containing papers by:
James Albus, Pentti Kanerva, and Egon Loebner

Research Institute for Advanced Computer Science
NASA Ames Research Center

RIACS Technical Report 89.11
March 1989

Abstract. Modeling a real-world phenomenon proceeds in two directions: by hypothesis from experimental data or by construction of a mathematical model from which results can be deduced. It is noteworthy when models derived from different directions are similar. A theory of human long-term memory, known as Kanerva's *sparse distributed memory* (SDM), arose independently, with slight variations, from both directions. Kanerva's approach was abstract. He sought a mathematical model that could account for (1) a massive storage capacity such that any two objects in memory could be closely associated, (2) an ability to retrieve data given only partial cues, and (3) recall of long temporal sequences. Kanerva was lead to a surprisingly simple architecture based on the geometry of hypercubes of very high dimensions: a generalized random-access memory that is easily analyzed and engineered. Kanerva only later noticed the similarity between SDM and the cerebellum. By contrast, two earlier and independently discovered models -- that of James Albus and that of David Marr -- were deliberate attempts to model the mammalian cerebellum. The three models are very similar. In the first paper Kanerva describes his model, sparse distributed memory. In the second paper, Albus describes his and Marr's two earlier models of cerebellar cortex. In the last paper Loebner discusses an ongoing effort to understand the complete cerebellum in finer detail and its position and role within the central nervous system. Loebner is leading a collaboration between Hewlett-Packard Laboratories, RIACS, and NASA, to understand the operations of the cerebellum from an engineering perspective, subject to constraints imposed by findings of neuroscience research. Loebner's work helps to explain the importance of the cerebellum for computer engineering: The cerebellum coordinates and calibrates interactions of a very large number of complex sub-systems, and its extraordinarily regular structure aids in the analysis of its architecture.

The three papers appearing here were presented in San Francisco at IEEE's COMPCON Spring '89. They are reprinted from the *Proceedings of the 34th IEEE Computer Society International Conference* with permission from the IEEE and from the Physiological Society, Oxford, England. Albus is Chief of the Robot Systems Division, National Institute for Standards and Technology. Kanerva is Principal Investigator of the RIACS Sparse Distributed Memory Project. Loebner is Counselor for Science and Technology at Hewlett-Packard Laboratories. The RIACS portion of the work reported here was supported in part by Cooperative Agreement NCC 2-408 between the National Aeronautics and Space Administration (NASA) and the Universities Space Research Association (USRA).

ORIGINAL PAGE IS
OF POOR QUALITY

CONTENTS

A Cerebellar-Model Associative Memory as a Generalized Random-Access Memory	1
Pentti Kanerva	
The Marr and Albus theories of the cerebellum: Two early models of associative memory	8
James Albus	
Intelligent Network Management and Functional Cerebellum Synthesis	14
Egon Loebner	
ABOUT THE AUTHORS	20

N92-14695

A CEREBELLAR-MODEL ASSOCIATIVE MEMORY AS A GENERALIZED RANDOM-ACCESS MEMORY

Pentti Kanerva

Research Institute for Advanced Computer Science
 NASA Ames Research Center, M/S 230-5
 Moffett Field, CA 94035

ABSTRACT

A versatile neural-net model is explained in terms familiar to computer scientists and engineers. It is called the sparse distributed memory, and it is a random-access memory for very long words (for patterns with thousands of bits). Its potential utility is the result of several factors: (1) A large pattern representing an object or a scene or a moment can encode a large amount of information about what it represents. (2) This information can serve as an address to the memory, and it can also serve as data. (3) The memory is noise tolerant--the information need not be exact. (4) The memory can be made arbitrarily large and hence an arbitrary amount of information can be stored in it. (5) The architecture is inherently parallel, allowing large memories to be fast. Such memories can become important components of future computers.

Introduction

This paper deals with neurally motivated associative memory, which is a basic component of neurocomputing. One specific cerebellar-model associative memory is discussed. It is called the sparse distributed memory or SDM [1], and it is described here by comparing it to the ordinary random-access memory (RAM) of a computer. Many of its properties are shared by most neural models, but some are specific to cerebellar models and to the sparse distributed memory in particular. The two cerebellar models that predate the sparse distributed memory and that resemble it the most were developed by David Marr [2] and by James Albus [3, 4].

Description of the MemoryOverview

An ordinary computer memory is a memory for short strings of bits, typically 8, 16, 32, or 64 bits. The bit strings are often thought of as binary numbers or "words," but, in general, they are just small patterns of bits. The memory stores them in addressable locations. The addresses to the memory also are short strings of bits. For example, 20 bits will address a memory with one million locations.

The sparse distributed memory is likewise a memory for strings of bits, except that the strings can be hundreds or thousands of bits long. Because the strings are so long, they are best thought of as large patterns. The addresses to the memory also are long strings of bits, or large patterns. In an important class of these memories, the address and data patterns are of equal size. In the examples in this paper the patterns are rather small; they have 256 bits.

Behavior

The behavior of an ordinary computer memory can be described as follows: If the word *W* has been written with address *A*, then *W* can be read back by addressing the memory with *A*, and we say that *A* points to *W*. The condition for this is, of course, that no other word has been written with address *A* in the meantime.

The sparse distributed memory has like behavior: If the pattern *W* has been written with pattern *A* as the address, then *W* can be read back by addressing the memory with *A*, and we say that *A* points to *W*. However, the conditions for this are more restrictive than they are with ordinary computer memories, namely, that no other pattern has been written before or since with address *A* or with an address that is similar to *A*.

The added restrictions pay off in noise tolerance in two ways: To read the pattern *W* from a sparse distributed memory, the address pattern need not be exactly *A* (in ordinary RAM, the exact address *A* must be used to read *W*). This means that the memory can tolerate a noisy reference address; it can respond to a partial or incomplete cue. Tolerance for noisy data shows up as follows: If many noisy versions of the same target pattern have been written into the memory, a (nearly) noiseless target pattern can be read back.

Figure 1 illustrates the memory's tolerance for noise. This memory works with 256-bit patterns. For ease of comparing patterns with each other, they are displayed on a 16 x 16 grid with 1-bits shown in black. The nine patterns in the upper part of the figure were gotten by taking a circular pattern and changing 20 percent of the bits at random. Each of the patterns was written into the memory with itself as the address. The noisy tenth pattern was then used as the address for reading from the memory, and the relatively

noise-free eleventh pattern was retrieved. When that pattern was used as the next read address, the final, nearly noise-free pattern was retrieved. Worth special notice is that the noise-free circular pattern was never used as the write address nor was it ever written into the memory (i.e., the memory had never "seen" the ideal pattern; it created it from the noisy versions it had seen).

The method of storage in which each pattern is written into memory with itself as the address, as illustrated in Figure 1, is called autoassociative. With autoassociative storage, the memory behaves like a content-addressable memory in the following sense: It allows a stored pattern to be retrieved if enough of its components are known.

A more general method of storage in which an address pattern and the associated data pattern are different is called hetero-associative. Figure 2 illustrates its use in storing a sequence of patterns. The sequence is stored as a pointer chain, with the first pattern pointing to the second, the second to the third, and so forth. Any pattern in the sequence can then be used to read out the rest of the sequence simply by following the pointer chain. Furthermore, the cue for retrieving the sequence can be noisy, as shown in Figure 3, in which a noisy third pattern retrieves a less noisy fourth, which in turn retrieves an almost noiseless fifth pattern, and the sixth pattern retrieved is perfect.

If the memory's address and data patterns are of different size, only heteroassociative storage is possible, although it is not possible to store pattern sequences as pointer chains.

The term 'associative memory' refers in neurocomputing to this very general property of linking one pattern to another, or forming an

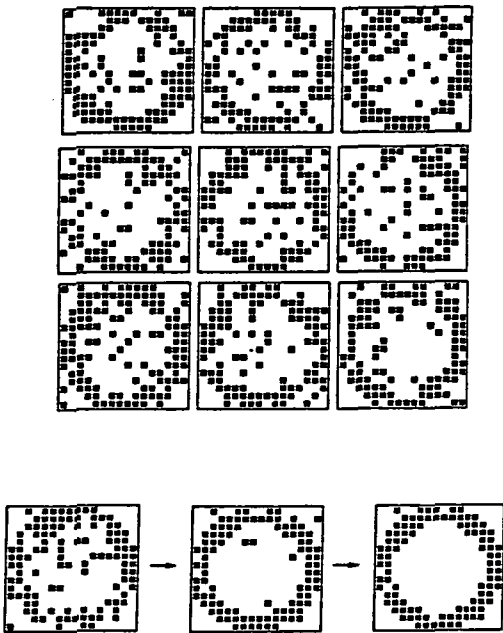


FIGURE 1. The sparse distributed memory's tolerance for noise.

association, the linkage being the association. In that broad sense, even the ordinary random-access memory is associative. However, the term is more specific in computer-engineering usage and is usually synonymous with 'content-addressable memory', which, in turn, is a tighter concept in computer engineering than in neurocomputing or psychology. As a neuro-computing term, associative memory implies also noise tolerance as illustrated in the examples above.

Construction

The ordinary computer memory is an array of addressable registers or memory locations. The locations are numbered sequentially, and the sequence number is the location's address. A memory with a thousand locations will therefore need ten-bit addresses. If the memory is built for eight-bit words, each location will have eight one-bit storage bins or flip-flops. This organization of the memory is shown in Figure 4. Each row in the figure is one memory location, with its address shown on the left and the storage bins on the right. In this figure, the memory's contents (the storage bins) have been set at random.

The sparse distributed memory also is an

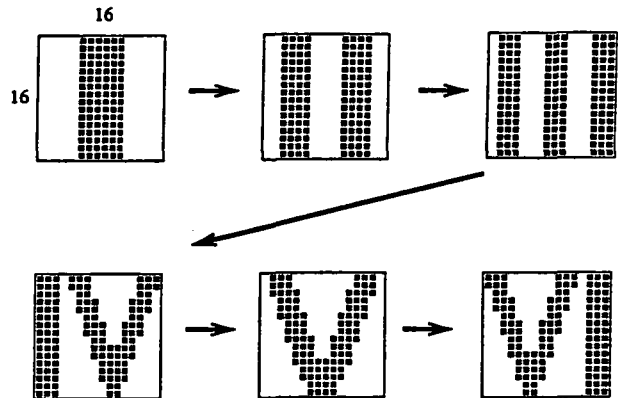


FIGURE 2. A sequence of patterns that is stored as a pointer chain.

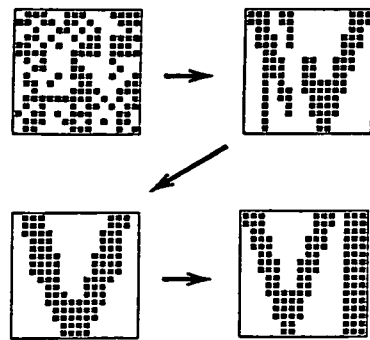


FIGURE 3. Iterated reading starting with a noisy third pattern.

array of addressable registers or memory locations. The addresses of the locations, however, are not sequence numbers but large bit patterns (256-bit addresses in the examples above). To store 256-bit data patterns, each location will have 256 storage bins for small up-down counters. This organization is shown in Figure 5, which is not that different from Figure 4. In Figure 5, the location addresses are random 256-bit patterns, and the memory's contents are shown after many patterns have been stored in the memory.

Because the memory addresses are large bit patterns, the number of addresses and hence of possible memory locations is astronomical. Only memories with a small subset of the possible locations can be built in practice, and that is why these memories are called sparse. Practical numbers range in the thousands to millions to billions of locations.

To move data into and out of the memory array, both kinds of memories have three special (input/output) registers: one for the memory address or cue, another for the word or

data pattern to be written into the memory, and the third for the word or data pattern being read out of the memory. In Figures 4 and 5 they are above and below the memory array.

In addition to the memory array and the input and output registers, Figures 4 and 5 show intermediate results of a memory operation. The numbers in the column or columns between the address matrix (on the left) and the contents matrix (on the right) indicate whether a memory location is selected for a given read or write operation. The selection depends on the contents of the address register, on the location's address, and on the selection criterion, as will be explained shortly. Figure 5 (of SDM) has, in addition, a row of sums as a way of getting from the contents of the memory locations to the final output pattern.

Operation

Reading and writing in ordinary computer memory is simple in concept. Both operations start with specifying a memory address in the address register. That selects one location from the memory array--the location with the matching address. The selection is indicated by the single 1 in the select column of Figure 4, the rest of the values in that column being zeros. If the memory operation is a write, the word being written is placed in the data-in register, and it will replace the word stored previously in the selected location; if it is a read, the contents of the selected location are copied into the data-out register.

Reading and writing in the sparse distributed memory likewise start with addressing the memory. However, when an address is specified in the memory-address register, the memory array will usually not have a location with that exact address. This is overcome by selecting many locations at once--and by modifying the rules for writing and reading accordingly.

The criterion for selecting or activating a location is similarity of address patterns: If the location's address is sufficiently similar to the address in the address register, the location is selected. Hamming distance between address patterns provides a simple measure of similarity, and it is used in Figure 5 and in subsequent examples. The column next to the address matrix in Figure 5 shows these Hamming distances, and the column next to it has ones where this distance does not exceed 112 bits. These than are the selected (nearby, active) locations, the unselected (distant, inactive) locations being indicated by zeros in the select column. As a rule of thumb, the selection criterion should be such that many locations are active at once, but their number should not exceed the square root of the total number of memory locations.

A (data) pattern is written from the data-in register into the memory by adding it into all selected locations (in an ordinary RAM, new data replace old in one location). It can be added simply by incrementing the counters under the ones of the data-in register and by decrementing the counters under the zeros.

Figure 6 shows the writing of two patterns into a very small memory that is initially

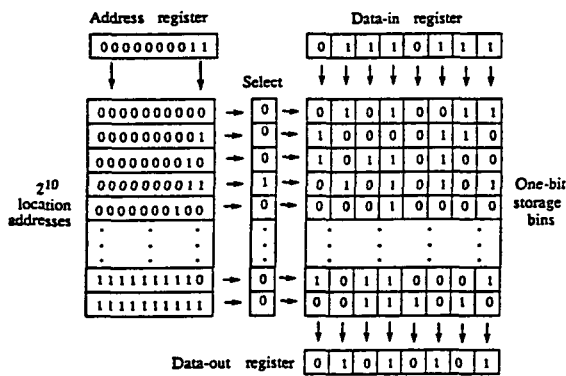


FIGURE 4. The organization of a random-access memory as an array of addressable locations.

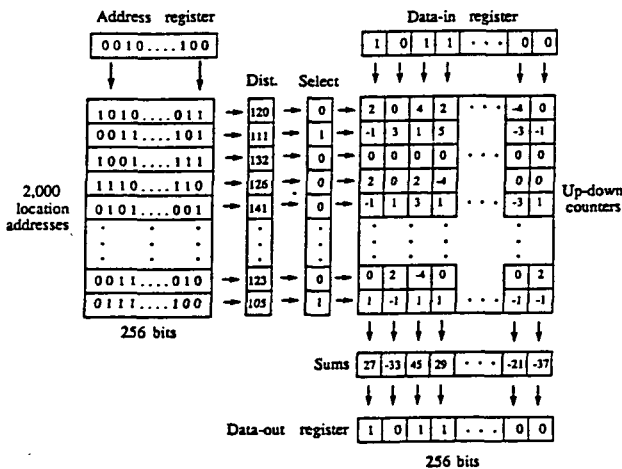


FIGURE 5. The organization of a sparse distributed memory as an array of addressable locations.

empty (all counters initially zeros). The selected locations are shown in white and the unselected in gray. As more and more data are written into the memory, individual counters can reach their capacity. When this happens, attempts to increment a counter past its maximum value or to decrement it past its minimum value are ignored.

A pattern is read out of the memory (from the selected locations) by computing an average over the contents of the selected locations. A simple average is gotten by adding the contents

(vector addition) and by thresholding the sums at zero, with a sum larger than zero yielding a 1 in the output pattern, and a sum smaller than or equal to zero yielding a 0. A bit of the output pattern will then be 1 if, and only if, the patterns written into the currently active locations have more ones than zeros in that bit position, constituting a bitwise majority rule. Figures 7a and 7b illustrate reading at and reading near the second write address, respectively. In both cases, the second written pattern is retrieved (cf. Fig. 6b).

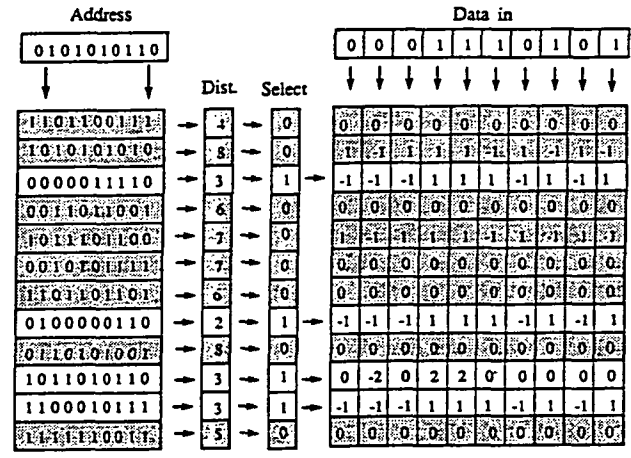
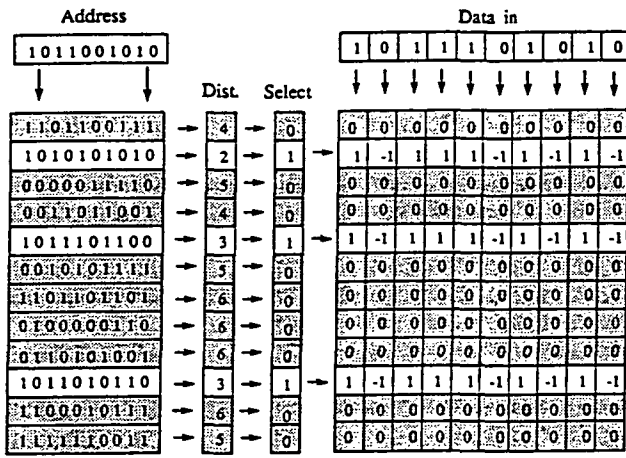


FIGURE 6. Writing two patterns into a tiny sparse distributed memory. First the pattern 101101010 is written at 1011001010 (a), and then the pattern 0001110101 at 0101010110 (b).

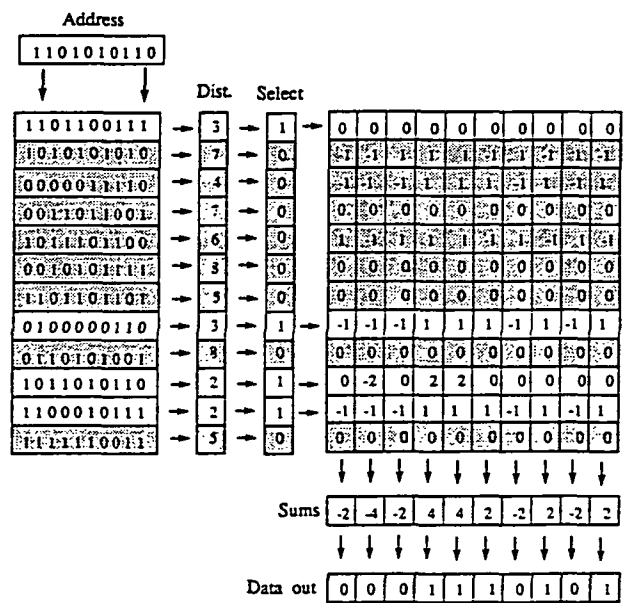
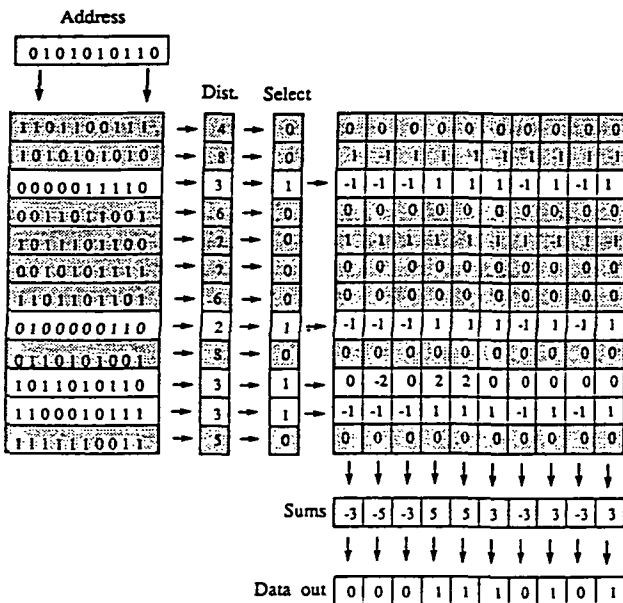


FIGURE 7. Reading at (a) and reading near (b) a previous write address.

Why Does the SDM Work?

A premier property of the sparse distributed memory is sensitivity to similarity, or noise tolerance. It is the result of distributing the data, that is, of writing into and reading from many locations at once, and it is explained mathematically by the amount of overlap, counted in active memory locations, when the memory is addressed with two different patterns. If two address patterns are very similar to each other, the sets of locations they activate have many locations in common; if they are dissimilar, the common locations are few or none. This can be seen in Figures 6 and 7: The second read address (Fig. 7b) differs from the second write address (Fig. 6b) by one bit only (the two addresses are very similar), and the number of locations selected by both--the overlap--is 3; it differs from the first write address (Fig. 6a) by five bits (dissimilar), and the overlap is 1 location. Thus, when we read near the second write address (Fig. 7b), the second written data pattern has a weight 3 and the first a weight 1 in the sums accumulated from the selected locations, allowing the second pattern to be recovered in thresholding.

The example illustrates that, in a sparse distributed memory, common address bits translate into common memory locations, and common memory locations translate into weights for stored patterns when reading from the memory. Thus, the memory is a means of realizing a weighting function that gives low weights to most of the patterns written into the memory and high weights only to a small number of "relevant" patterns, the relevance being judged by similarity of address.

The operation of the memory is statistical, and the actual output is affected not only by the construction of the memory but also by the structure of the data. The results discussed above are demonstrated most readily when the addresses of the locations and the data are a uniform random sample of their respective spaces of bit strings. There is the further condition that not too many patterns have been written into the memory. The memory works in the manner described if the number of stored patterns is no more than 1-5 percent of the number of memory locations.

Closely Related Architectures

Ordinary RAM as a Special Case of SDM

We can now demonstrate the close kinship of the two kinds of memories. Let us start with a random-access memory that has just over 16 million (2 to power 24 , 2^{24}) locations for 32-bit words. The memory address is then 24 bits long. This memory can be thought of as a sparse distributed memory with the following parameters: an array of 2^{24} memory locations, with 24-bit addresses and 32 one-bit up-down counters for holding the data. The address matrix would contain each of the 2^{24} possible addresses exactly once, and the Hamming distance for selecting a location would be zero. That would mean that each possible address would select exactly one location, and two different addresses would

always select two different locations.

Writing into this memory causes the old contents of the location to be lost to overflow and underflow, because the location's counters have only one bit each. Reading from it fetches the contents of one location--whatever was written there last--and thresholding will not change the bits. This example shows that the sparse distributed memory indeed is a generalized random-access memory; it yields the ordinary RAM as a special case. In the terminology of the preceding section, the data pattern associated with the read address has weight one and all other patterns have weight zero.

Extensions of the Basic Model

In the basic model of sparse distributed memory, the pattern components are binary. The model can be generalized to allow many-valued components, including continuous, and an important case is one in which the components are ternary. The most convenient three values are -1 , 0 , and 1 , and useful interpretations for them are 'off', 'don't know or don't care', and 'on', respectively. The activation of a location must be based on a measure that is more general than the Hamming distance, for example, on the inner (dot) product of the location's address with the address in the address register. Writing into the memory is by adding the input-data pattern into the active locations, much as before, and reading is by summing over the active locations, except that to get the final output pattern, we need two thresholds instead of one. If this model is restricted to the values -1 and 1 , and the two thresholds are both equal to 0 , it is equivalent to the basic model with binary components.

Other variations of the model are gotten by adjusting it to the data being stored. The more the data deviate from the "ideal," that is, from being a uniform random sample of the underlying space, the more important the adjustments are. Real-world data are never ideal in that sense, and so the adjustments are essential in systems for real-world applications. The adjustments include: choosing the addresses of the memory locations based on the addresses in the data; activating a fixed number of closest locations in any given read or write operation instead of all locations within a certain distance; having individual selection distances for individual locations; adding correction vectors into the memory instead of, or in addition to, data-pattern vectors; weighting active locations in a read operation according to their contents; and adjusting the thresholds that determine the final output.

Some variations of the basic model would take it outside the realm of cerebellar models. Adjusting the addresses of the memory locations as a part of "training" the memory for a given data set is the most important of such variations. In the cerebellar models, the address of a location, once defined, stays fixed, setting them apart from more general models, such as multilayer back-propagation nets [5], which resemble the cerebellar models in many other respects. Another characteristic

of the cerebellar models, as compared with most other models, is that any given read or write operation activates many locations but leaves most locations inactive: a location is either on or off, as indicated in the select column of Figure 5. These constraints of the cerebellar models simplify the construction of memories based on the models, making it possible to build very large memories that can be trained reasonably fast.

In the taxonomy of adaptive networks or artificial neural nets, the sparse distributed memory is a 'fully connected three-layer feed-forward' net. The address register (see Fig. 5) corresponds to the input layer of such a net, the location-address matrix holds the input weights of the hidden layer (each memory location--a row--is one hidden unit), the select vector is the output of the hidden layer, the contents matrix (the up-down counters) are the weights of the output layer (each column is one output unit), and the data-out register has the outputs of the output layer. 'Fully connected' means that each bit of the input address is seen by each memory location and that each memory location can contribute to each output bit. 'Feed forward' means that the output of one layer goes to the next or subsequent layers only (no direct feedback to the layer itself or to its predecessors), which in turn means that the outputs of a layer are logically independent of each other. The term 'three-layer' is a misnomer, as is evident when several such nets are cascaded or pipelined. Cascading three of them will not result in a nine-layer net but in a seven-layer net, which suggests that the original net really is a two-layer net (and a cascade of three of which is a six-layer net). Thus, the network input (the address register) should not be counted as a separate layer.

Relation to the Cerebellum

The reason for calling the sparse distributed memory, and the models of Marr and of Albus, cerebellar models is largely historical. After developing these neural models of associative memory, the developers noticed and pointed out remarkable similarities in the wiring diagrams of their models and the wiring of the cortex of the cerebellum and, based on the similarities, suggested functions for several cell types of the cortex. The significance of the models is in giving us a mathematical way to look at a major part of the brain, in the perspective of the cerebellum as an associative memory with billions of locations, in motivating further research into the cerebellum, and in arming researchers with useful questions.

Why Associative Memories?

Nature has solved problems that appear to be beyond the capacity of even the most powerful computers. These problems include taking a complex signal from the world, such as the raw input to our visual, auditory, olfactory, and tactile systems, and producing from it over time a coherent model of the world--and of the self in it--that allows us to function in the world. In our ability to do so, we think of ourselves as intelligent and would call systems

with similar powers intelligent. How do intelligent systems work? We will consider this question only as it relates to associative memories for large patterns.

The perceptual task of identifying an incoming signal based on experience can be divided into sensory analysis and pattern matching. In sensory analysis, the senses extract features from the signal, and further processing of the signal is in terms of those features. If two scenes produce very similar patterns in terms of the extracted features, the two will be identified as the same by an associative memory (cf. Fig. 1). This is exactly what an intelligent system has to do in identifying objects from different views of it. However, it is important that the features are appropriate for the task. For a counter-example, the pixels of a bit map (a raw retinal image) are poor features for vision, because shifting the figure only slightly or viewing it from a different distance can change a large portion of the features. Human and animal perceptual systems have attentional mechanisms, including feedback from memory, that help the sensors to extract appropriate features.

The actions of humans and animals are accomplished by the selective contraction and relaxation of large numbers of muscle fibers controlled by large numbers of motor neurons. The configuration of active and inactive motor neurons at any one time is a large pattern, and the state of no single neuron is critical for the performance of a given action. The activation patterns of motor neurons are therefore appropriate for an associative memory, as is the learning of actions as responses to sensory patterns--the actions being associated with the sensations. Actions can also be associated with internal states of the system that reflect the system's past in complicated ways, which means that a system based on an associative memory can learn complex, coordinated actions.

The relative merits of associative memories in these tasks derive from how information is packaged. Conventional computers work with small bit patterns (words) that represent a quantity, an index, or a small vector of features. Many such patterns are needed to describe a complex object or a moment of experience. However, at the top level, a single, short index describes or encodes it. The top-level description is precise, as two slightly different indexes can point to two entirely different objects, but it is also almost totally uninformative. To find out anything about the object, it is necessary to fetch from memory further indexes and associated data fields. This allows objects to be described in arbitrary detail, but it also tends to hinder fundamental operations such as the comparison of objects to see how they are related--it makes "seeing" objects in whole difficult; they are seen in tiny fragments.

In contrast, systems based on neurally motivated associative memories work with large patterns (e.g., 10,000 bits) as units. A single pattern can encode a large amount of information about an object--hence it is highly informative, yet it need not be precise. It can serve as the (top-level) description of an object, and it can also serve as an index.

These properties of the descriptions, together with the properties of the memory, are helpful with operations such as the comparing of objects. They also make it easy to describe events that occur over time: a moment (of experience) can be encoded by a single pattern, and an event by a sequence of patterns that is stored in the memory as a pointer chain. A single pattern can include sensory and motor components, plus components that encode the internal (subjective) state of the system, and hence a sequence of patterns can encode interactions of all of these components.

The memory's ability to store associations, and pattern sequences in particular, gives it the power to predict, and the failure of a prediction signals an occasion for learning. Learning is by training through a set of examples rather than by explicit programming. This is referred to as learning from experience. The term is particularly appropriate if the training patterns encode real-world phenomena.

Among traditional methods, multivariate statistical analysis resembles associative-memory-based methods, and there are important connections to coding theory and to adaptive filters. All of these exploit the richness of the geometry of very-high-dimensional spaces, something that conventional computer methods tend not to do.

Pattern Computing

Neurally motivated associative memories and, more generally, adaptive networks or artificial neural nets are computing architectures for very large patterns. They are therefore classified appropriately as pattern computers, as contrasted with conventional numeric and symbolic computers. This classification is based on practical considerations, as a computer in any one class can be used to emulate those in the other two, except that

the emulations tend to be too slow to be of practical interest. The speed of pattern computers in dealing with very large patterns is achieved by large numbers of relatively simple processors working in parallel.

Today's computers combine components for numeric and symbolic computing. We can expect future computers to add more and more pattern-computer components to them, as we learn to build and use pattern computers. That, in turn, will broaden the scope of computing and the usefulness of computers--it may well revolutionize computing.

Acknowledgements. This work was supported by the National Aeronautics and Space Administration (NASA) cooperative agreement NCC2-408 with the Universities Space Research Association (USRA). Michael Raugh fended off interruptions and made it possible to write the paper. The examples were run on an SDM simulator written for the SUN workstation by Robert Brown. Louis Jaeckel came up with the idea for Figure 1. James Keeler first drew the matrix pictures of the SDM as they appear in this paper, and he called the zeros in the trinary model don't-know-or-don't-care bits. David Rogers prepared the illustrations. The author's sincere thanks to all.

References

- [1] P. Kanerva, Sparse distributed memory, MIT Press, 1988.
- [2] D. Marr, "A theory of cerebellar cortex," J. Physiology, vol. 202, p. 437, 1969.
- [3] J. S. Albus, "A theory of cerebellar functions," Mathematical Biosciences, vol. 10(1/2), p. 25, 1971.
- [4] J. S. Albus, Brains, behavior, and robotics, Byte Books, 1981.
- [5] D. E. Rumelhart and J. L. McClelland, eds., Parallel distributed processing, vol. 1, MIT Press, 1986.

N92-14696

The Marr and Albus theories of the cerebellum
Two early models of associative memory

by

James S. Albus

Robot Systems Division
National Institute of Standards and Technology

Abstract

The Marr and Albus theories of the cerebellum are compared and contrasted. They are shown to be similar in their analysis of the function of the mossy fibers, granule cells, Golgi cells, and Purkinje cells. They both predict motor learning in the parallel fiber synapses on the Purkinje dendrites mediated by concurrent climbing fiber input. This prediction has been confirmed by experimental evidence. In contrast, Marr predicts these synapses would be facilitated by learning, while Albus predicts they would be weakened. Experimental evidence confirms synaptic weakening.

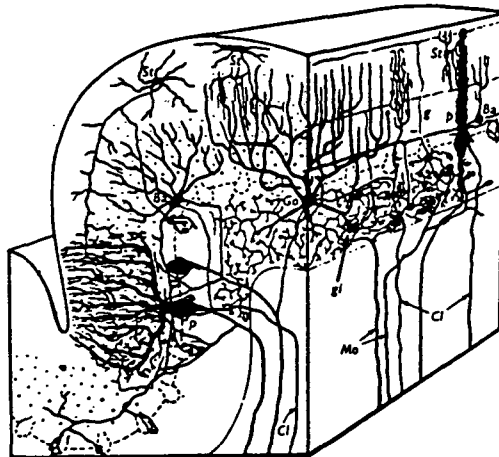
Introduction

Two papers published in 1969 and 1971 by David Marr and James Albus form the basis for what has become known as the Marr-Albus theory of the cerebellum. Both of these papers were inspired by, and draw most of their data from, a book by Eccles, Ito, and Szentagothai entitled *The Cerebellum as a Neuronal Machine* [Eccles67]

"A diagram of the general cerebellar cortical structure appears in Fig. 1. The cortex has two types of afferent fiber, the climbing fibers (Cl) and the mossy fibers (Mo). Each climbing fiber makes extensive synaptic contact with the dendritic tree of a single Purkinje cell (p), and its effect there is powerfully excitatory. The axons of the Purkinje cells leave the cortex (they form the only cortical output) and synapse with cells of the cerebellar nuclei.

"The second input, the mossy fibers, synapse in the cerebellar glomeruli (gl) with the granule cells. Each glomerulus contains one mossy fiber terminal (called a rosette), and dendrites (called claws) from many granule cells. The glomerulus thus achieves a considerable divergence, and each mossy fiber has many rosettes."

"The axons of the granule cells rise (g) and become the parallel fibers, which synapse in particular with the Purkinje cells whose dendritic trees they cross. Where the granule cell axons (i.e. the parallel fibers) make synapses, they are excitatory.



"Fig. 1. Diagram of cerebellar cortex (from Eccles et al. 1967, Fig. 1). The afferents are the climbing fibers (Cl) and the mossy fibers (Mo). Each climbing fiber synapses with one Purkinje cell (p), and sends weak collaterals to other cells of the cortex. The mossy fibers synapse in the cerebellar glomeruli (gl) with the granule cells whose axons (g) form the parallel fibers. The parallel fibers are excitatory and run longitudinally down the folium: they synapse with the Purkinje cells and with the various inhibitory interneurons, stellate (St), basket (Ba) and Golgi cells (Go). The stellate and basket cell axons synapse with the Purkinje cells, and the Golgi cell axons synapse in the glomeruli with the granule cells. As well as their ascending dendrites, the Golgi cells possess a system of descending dendrites, with which the mossy fibers synapse in the glomeruli. The Purkinje cell axons form the only output from the cortex, and give off many fine collaterals to the various inhibitory interneurons."

"The remaining cells of the cortex are inhibitory interneurons. The Golgi cells (Go) are large, and have two dendritic trees. The upper tree extends through the molecular layer, and is driven by the parallel fibers. The lower dendrites terminate in the glomeruli, and so are driven by the mossy fibers. The Golgi axon descends and ramifies profusely: it terminates in the glomeruli, thereby inhibiting the granule cells. Every glomerulus receives a Golgi axon, almost always from just one Golgi cell: and each Golgi cell sends an axon to all the glomeruli in its region of the cortex.

"The other inhibitory neurons are stellate cells, the basket (Ba) and outer stellate (St) cells. These have dendrites in the molecular layer, and are driven by the parallel fibers. Both types of cell synapse exclusively with Purkinje cells, and are powerfully inhibitory.

"Finally, the cortex contains various axon collaterals. The climbing fibers give off weak excitatory collaterals which make synapses with the inhibitory interneurons situated near the parent climbing fiber. The Purkinje cell axons give off collaterals which make weak inhibitory synapses with the cortical inhibitory interneurons, and perhaps also very weak inhibitory synapses with other Purkinje cells. These collaterals have a rather widespread ramification.

"Behind this general structure lie some relatively fixed numerical relations. These all appear in Eccles et al. (1967), but are dispersed therein. It is therefore convenient to set them down here.

"Each Purkinje cell has about 200,000 (spine) synapses with the parallel fibers crossing its dendritic tree, and almost every such parallel fiber makes a synaptic contact. The length of each parallel fiber is 2-3 mm (1 1/2 mm each way), and in 1 mm down a folium, a parallel fiber passes about 150 Purkinje cells. Eccles et al. (1967) are certain each fiber makes at least 300 (of the possible 450) synaptic contacts with Purkinje cells, and think the true number is nearer 450. There is one Golgi cell per 9 or 10 Purkinje cells, and its axon synapses (in glomeruli) with all the granule cells in that region, i.e. around 4500. There are many granule cells (2.4×10^6 per mm of granule cell layer), each with (usually) 3-5 dendrites (called claws): the average is 4.5 and the range 1-7. Each dendrite goes to one and only one glomerulus, where it meets one mossy fiber rosette. It is,

however, not alone: each glomerulus sees the termination of about 20 granule cell dendrites, possibly a Golgi cell descending dendrite, and certainly some Golgi axon terminals, all from the same Golgi cell. Within each folium, each mossy fiber forms 20-30 rosettes, giving a divergence of 1 mossy fiber to 400-600 granule cells within a folium. The mossy fiber often has branches running to other folia.

"Just below the Purkinje cells are the Golgi cell bodies, and just above them are the basket cell bodies. There are 10-12% more basket cells than Purkinje cells, and about the same number of outer stellate cells. Each basket cell axon runs for about 1 mm transversely, which is about the distance of 10 Purkinje cells. The basket axon is liable to form baskets round cells up to three away from its principal axis, so its influence is confined to a sort of box of Purkinje cells about 10 long and 7 across. The distribution of the outer stellate axons is similar except that it has a box about 9 x 7, since its axon only travels about 0.9 mm transfolially. The outer stellates inhabit the outer half of the molecular layer, and the basket cells the inner third. There are intermediate forms in the missing sixth. None of these cells has a dendritic tree as magnificent as that of the Purkinje cell, and Eccles et al. (1967) do not venture any comparative figures. Some outer stellates are small, with a local axonal distribution. A lot of the synapses of parallel fibers with this last group of cells are directly axodendrite, but all other parallel fiber synapses are via spines, though these are of different shapes on the different sorts of cell. Calculations based on slightly tenuous assumptions suggest that each Purkinje cell receives connections from about 7000 mossy fibres." [From Marr 1969]

Both Marr and Albus agree on the nature and function of the mossy fibers, granule cells, and Golgi cells, i.e. that they recode input patterns of mossy fiber firing rates into patterns of parallel fiber activity.

Marr expresses the recoding in terms of codons.

"The synaptic arrangement of the mossy fibers and the granule cells may be regarded as a device to represent activity in a collection of mossy fibers by elements each of which corresponds to a small subset of active mossy fibers. It is convenient to introduce the following terms: a codon is a subset of a collection of active mossy fibers. The

representation of a mossy fiber input by a sample of such subsets is called the codon representation n of that input: and a codon cell is a cell which is fired by a codon. The granule cells will be identified as codon cells, so these two terms will to some extent be interchangeable. The size of codon that can fire a given granule cell depends upon the threshold of that cell, and may vary: and the mossy fibers which synapse with the granule cell determine the codons which may fire that cell.

"There are exactly $\binom{L}{R} = \frac{L!}{R!(L-R)!}$ codons of size R associated with a collection of L active mossy fibers. If two mossy fiber inputs each involve activity in L fibers of which W were common to the two, the two inputs are said to overlap by W elements; and they may be expected to have some codons in common. In fact the

number they share is precisely $\binom{W}{R}$. The ratio X of the number of shared codons to the number of codons each possesses is given by
$$X = \frac{\binom{W}{R} / \binom{L}{R}}{R} = \frac{W(W-1)\dots(W-R+1)}{L(L-1)\dots(L-R+1)}$$

which tends to (W/L) as W increases. The limiting values of X for relevant values of R appear in Table 1. It will be observed that the effect of the subset coding is to separate patterns, because similar inputs have markedly less similar codons.

TABLE 1. Overlap Table, i.e. values of $(W/L)^*$

(W/L)	R = 2	3	4	5
0.5	0.25	0.12	0.06	0.03
0.6	0.36	0.22	0.13	0.08
0.7	0.49	0.34	0.24	0.17
0.8	0.64	0.51	0.41	0.33
0.9	0.81	0.73	0.66	0.60

"The mossy fiber granule cell relay effectively takes a sample of the codon distribution of an input: the sample is small enough to be manageable, but large enough for the input event to be recoverable from it with high probability." [From Marr 1969]

Marr's concept of codons derives from Brindley [Bri69], and is elaborated in later papers by Marr. From analysis of codon theory, Marr predicts that the number of responses that can be stored by each Purkinje cell is less than 500, and probably around 200.

Albus expresses the recoding in terms of Perceptron theory [Ros61].

"Assume a decoder, or rather a recoder, that codes N input fibers (mossy fibers) onto $100N$ association cells (granule

cells). Such a recoding scheme provides such redundancy that severe restrictions can be applied to the $100N$ association cells without loss of information capacity. For example, it is possible to require that of the $100N$ association cells, only 1% (or less) of them are allowed to be active for any input pattern. That such a recoding is possible without loss of information capacity is easily proven, for 2^N is much smaller than 100^N things taken N at a time.

"That such a recoding increases the pattern-recognition capabilities of a Perceptron is certain, since the dimensions of the decision hyperspace have been expanded 100 times. The amount of this increase under conditions likely to exist in the nervous system is not easy to determine, but it may be enormous. It can be shown that $100N$ things taken N at a time is greater than 100^N . Thus 2^N possible input patterns can be mapped very sparsely onto 100^N possible association cell patterns. If this is done randomly, the association cell patterns are likely to be highly dissimilar and thus easily recognizable. The ratio $100^N/2^N = 50^N$ rapidly increases as N becomes large.

"The restriction that only 1% of the association cells are allowed to be active for any input pattern means that any association cell participates in only 1% of all classifications. Thus its weight needs adjusting very seldom and there is a fairly good probability that its first adjustment is at least in the proper direction. This leads to rapid learning." [From Albus 1971]

From analysis of Perceptron theory, Albus predicts that the number of patterns that can be recognized by each Purkinje cell is on the order of 200,000.

The large difference between Marr and Albus in predicting Purkinje discrimination capacity is due to differences in the hypothesized mechanism of learning. Marr suggests that learning takes place only by facilitation of positive synaptic weights between parallel fibers and Purkinje dendrites. Albus suggests a mechanism by which synaptic influence can effectively be adjusted in both positive and negative directions. This is accomplished through modification of parallel fiber synapses not only on Purkinje dendrites, but on Basket and Stellate b cells as well.

Marr and Albus agree in suggesting that climbing fibers control cerebellar learning by modification of synaptic weights between parallel fibers and Purkinje dendrites. There is, however, a

significant difference between Albus and Marr regarding the character of the climbing fiber influence. Marr uses only data from Eccles et al indicating that climbing fibers are powerfully excitatory. On this basis, Marr postulates that climbing fibers affect learning through strengthening of parallel fiber synapses on Purkinje dendrites.

In contrast, Albus includes additional data from other sources indicating that climbing fiber effects are much more complex.

"Each Purkinje cell is contacted by a single climbing fiber. In a conscious animal the climbing fibers fire in short bursts of one or more spikes at a rate of about 2 bursts/sec [5, 18]. Each climbing fiber burst causes a single spike on the Purkinje axon followed by a complex burst of spike-like activity in the Purkinje dendritic tree and intense depolarization of the Purkinje cell. The single axon spike is followed by a pause in the spontaneous Purkinje axon spike activity for 15-30 msec. This pause, accompanied by intense depolarization, was first observed by Granit and Phillips [8] and was termed the inactivation response to distinguish it from a normal pause in activity resulting from hyperpolarization. After the 15 to 30 msec inactivation response, the cell gradually recovers its spontaneous firing rate over a period of 100-300 msec [3]. As it approaches normal, the cell becomes once again responsive to parallel fiber input activity." [From Albus 1971]

On the basis of this data, Albus suggests that the primary effect of climbing fiber input is to cause the Purkinje to pause, i.e. the net results is inhibitory, despite the initial excitatory spike. He further hypothesizes that climbing fibers effect learning through weakening parallel fiber synapses, not only on Purkinje dendrites, but on nearby Basket and Stellate cells as well.

This is a counterintuitive idea which not only disagrees with Marr's theory of synaptic facilitation, but with virtually the entire tradition of neurophysiological and psychological learning theory. Almost without exception, previous theories had been influenced by the Pavlov, Hebb, Skinner presumption that learning occurs by facilitation of synapses due to their association with behavior leading to successful results; not by synapses being weakened by contributing to unsuccessful behavioral results. In fact, the entire branch of psychology founded by Skinner has generalized this notion to the point of opposing the principle of teaching by punishing incorrect behavior.

The notion of learning from error correction (i.e. weakening synaptic weights that contribute to undesirable results) comes from engineering. It is the fundamental principle of servomechanisms (i.e. negative feedback of an error signal). It was put into a neurological context by the Perceptron and its derivatives such as the Adeline[Wid85], the Cerebellar Model Articulation Controller (CMAC)[Alb75], and neural nets." [Hop82, Gro75].

Albus suggests as a possible mechanism for synaptic weakening that there exists a critical interval near the end of the inactivation response after the effect of the climbing fiber burst has worn off sufficiently so that the cell can be fired by parallel fiber input but before the dendritic membrane has returned completely to normal. If the Purkinje cell fires in this interval, this firing is an error signal that signals every active parallel fiber synapse to be weakened.

The amount of weakening of each synapse is proportional to how strongly that synapse is exciting the Purkinje cell at the time of error signal. The effect of this mechanism would be to train the Purkinje cell to pause at the proper times, that is, at climbing fiber burst times. After learning is complete, the Purkinje knows when to pause because it recognizes the mossy-parallel fiber pattern that occurred previously at the same time as the climbing fiber burst. Later, since each parallel fiber active synapse was weakened by the error signal, if the same mossy-parallel fiber pattern occurs again, the Purkinje will pause even without the climbing fiber burst. Thus, the Purkinje is forced to perform in a certain way by the climbing fiber teacher. After learning is complete, it behaves in that same way, under the same mossy fiber conditions, even in the teacher's absence. [Alb71]

Albus goes on to hypothesize that synaptic weakening also occurs at the parallel fiber synapses on Basket and Stellate b dendrites. This effectively provides both positive and negative training adjustments. Positive adjustments occur by weakening excitatory synapses on inhibitory interneurons, and negative adjustments by weakening excitatory synapses on the Purkinje output cells.

Albus argues that synaptic weakening is necessary as a learning mechanism for precise motor learning, because otherwise synapses quickly become saturated.

If a synaptic weight is increased each time it correctly fires, repeated learning will eventually cause it to saturate. This means that continued training in motor skills will produce degraded performance.

"Yet, it is an obvious fact that continued training in motor skills improves performance. Extended practice improves dexterity and the ability to make fine discriminations and subtle movements. This fact strongly indicates that learning has no appreciable tendency to saturate with overlearning. Rather, learning appears to asymptotically approach some ideal value. This asymptotic property of learning implies that the amount of change that takes place in the nervous system is proportional to the difference between actual performance and desired performance. A difference function in turn implies error correction, which requires a decrease in excitation upon conditions of incorrect firings." [Alb71]

Conclusions

Recent experimental data confirms the basic Marr-Albus hypothesis in three important respects:

- 1) motor learning does indeed occur in the cerebellum,
- 2) parallel fiber synapses on the Purkinje dendrites are modified, and
- 3) the modification is produced by concurrent activity of climbing fibers. [Ito84].

It has also been shown experimentally that cerebellar learning is accomplished through weakening of variable synapses, as predicted by Albus alone [Ito84]. Observations of negative as well as positive changes in synaptic strength have also been observed in the visual cortex [Rui69, Ros72]

Thus, the Marr and Albus theories have become two of the best working hypotheses currently available to cerebellar researchers.

Both the Marr and Albus theories make a number of additional predictions about neuronal function in the cerebellum, as well as the relationship between the cerebellum and other centers of motor control. These have not yet been either confirmed or disproven by experimental evidence. For example, there is as yet no evidence that the responsiveness of a basket cell to mossy fiber inputs is modified following conjunctive activation of the mossy fibers with climbing fibers. [Ito82]

In other areas, the CMAC model based on the Albus cerebellar theory is now being used to perform dynamic computations for fine motor control of robot arms [Alb75, Mil87]. A control system architecture based on CMAC principals has been used for the control of automated manufacturing facilities [Alb81], for controlling Multiple Autonomous Undersea Vehicles [Alb88], and will be implemented on the Flight Telerobotic Servicer [Alb87] being built for the NASA Space Station.

References

- Albus, J. "A theory of cerebellar function", *Math. Biosciences*, 10(1971), 25-61
- Albus, J. "A new approach to manipulator control: the Cerebellar Model Articulation Controller (CMAC)" and "Data storage in the CMAC", *Trans. ASME J. Dyn. Sys. Meas. and Control*, Sept 1975, 220-235
- Albus, J. Brains, Behavior, and Robotics, Byte/McGraw-Hill, Petersborough, NH, 1981
- Albus, J., H. McCain, and R. Lumia, NASA/NBS Standard Reference Model for Telerobot Control System Architecture (NASREM), NBS Tech Note 1235, National Bureau of Standards, Gaithersburg, MD. 1987
- Albus, J. System Description and Design Architecture for Multiple Autonomous Undersea Vehicles, NIST Tech Note 1251, National Institute of Standards and Technology, Gaithersburg, MD., 1988
- Brindley, G. "Nerve net models of plausible size that will perform many of very many simple learning tasks." *Proc. Roy. Soc. B* 1970
- Eccles, J., Ito, M., and Szentagothai, J. The Cerebellum as a Neuronal Machine, Springer, Berlin, 1967.
- Grossberg, S., Classical and Instrumental Learning by Neural Networks, Academic Press, New York, 1974
- Hopfield, J. "Neural networks and physical systems with emergent collective computational abilities" *Proc. Nat. Acad. Sci. USA*, 79, 2554-2558, April, 1982
- Ito, M., The Cerebellum and Neural Control, Chapter 10, Raven Press, NY, 1984
- Marr, D. "A theory of cerebellar cortex", *J. Physiol.*, 202(1969), 437-470

Miller, W., "Sensor-based control of robotic manipulators using a general learning algorithm", IEEE J. Rob. and Auto. RA-3, No. 2, 1987, 157-165

Rosenblatt, F., Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms, Spartan Books, Washington, D.C., 1965

Rosenzweig, M., K. Mollgaard, M. Diamond, and E Bennett, "Negative as well as positive synaptic changes may store memory", Psychological Review, 1972, Vol. 79, 93-96

Ruiz-Marcos, A. and F. Valverde, "Temporal evolution of the distribution of dendritic spines in the visual cortex of normal and dark-raised mice", Exptl. Brain Res. 8 (1969) 284-294

Widrow, B., and S. Stearns, Adaptive Signal Processing, Prentice Hall, NJ, 1985

INTELLIGENT NETWORK MANAGEMENT AND FUNCTIONAL CEREBELLUM SYNTHESIS

Egon E. Loebner

Hewlett-Packard Laboratories
1501 Page Mill Road, Palo Alto, CA 94304

ABSTRACT

Transdisciplinary modelling of the cerebellum across histology, physiology and network engineering provides preliminary results at three organization levels: I/O links to central nervous system networks, links between the six neuron populations in the cerebellum and computation among the neurons of the populations. Older models probably underestimated the importance and role of climbing fiber input which seems to supply write as well as read signals, not just to Purkinje but also to basket and stellate neurons. The well-known mossy fiber-granule cell-Golgi cell system should also respond to inputs originating from climbing fibers. Corticonuclear microcomplexing might be aided by stellate and basket computation and associative processing. Technological and scientific implications of the proposed cerebellum model are discussed.

INTRODUCTION

James Clerk Maxwell was a strong proponent of the "cross-fertilization of Sciences". In his Rede Lecture on "The Telephone", he honored Alexander Bell for not being a specialist who "builds up particular sciences", but for being one "who opens such communications between the different groups of builders as will facilitate a healthy interaction between them" [1]. Maxwell had exploited what he called "that partial similarity between the laws of one science and those of another which makes each of them illustrate the other" as a tool to build a unified theory of electromagnetism using mechanical analogies. Michael Idvorsky Pupin later adapted the very same tool to transform acoustical into electrical machinery [2]. The need to accelerate reciprocal transdisciplinary crossings between neuroscience and computer science was highlighted recently [3]. Some neuroscientists recognize the benefits to be expected from infusion of engineering and other ideas into their field [4] and anticipate a symbiotic relationship between modelling and experimental research [5].

This paper is a preliminary report of research recently undertaken under the auspices of RIACS at NASA's Ames Research Center. The work described here is carried out jointly with Jim Keeler (now at MCC), with Coe Miles-Schlichting of RECOM, and David Rogers of RIACS, both at NASA Ames Research Center. The goal is to develop a mathematical model of a mammalian cerebellum and to construct a functioning hardware implementation of one of its portions. We are attempting to preserve as many of its salient network topology and information processing features as is reasonably possible. In this we hope to follow the design philosophy of RCA's 1960-61 functional optoelectronic model of the frog retina [6] which culminated in the 1963 construction of the largest and most complex functional and parallel processing neural networks in existence at that time [7].

Thus far we critically sifted through books, bibliographies, abstracts and articles of a vast literature and selected those few that we expect to rely upon. As new experimental techniques produce more accurate findings, older theories and models get challenged and sometimes discarded. In order to synthesize the truest-to-life cerebellar functions, we have attempted to reconcile contradictions in reported facts and proposed interpretations. For example, we think that a modified functionality should be assigned to neurons targeted by climbing fiber collaterals since this seems to better fit recent physiological results.

During our attempts to classify the information and rank it, we tried to resist the "all too convenient" temptation to overlook inconvenient facts in order to simplify the model and to follow Einstein's dictum that everything should be made as simple as possible, but not any simpler.

CURRENT FRONTIERS OF NETWORK MANAGEMENT

Churchland and Sejnowski point to two recent reviews by Goldman-Rakic and Mountcastle (see their reference 8) which suggest a "democratic" organization of

processing in webs of strongly interacting networks in the association areas and the prefrontal cortex [5]. They think that this points to a distributed control instead of the more generally assumed single control center. They believe that "coming to grips with systems having distributed control will require both new experimental techniques and new conceptual advances". We agree. However, in our opinion their suggestion to study "models of interacting networks of neurons" needs to be paired and crossfertilized with research on networks of closely and loosely coupled state-of-the-art computers. The latter type research is exemplified by the pioneering work of Amnon Barak and coworkers who have been experimenting with a general-purpose, time-sharing operating system that induces a cluster of loosely connected independent homogeneous computers to act as a single-machine UNIX system [8,9]. We suspect that some of the principles employed by the Barak and other groups may aid in the study of interactivity between different parts of the central nervous system, and that some of the work suggested by Churchland and Sejnowski could in turn provide ideas and insights for future designs of intelligent distributed management within the rapidly growing networks of computing machines.

AN ENGINEERING VIEW OF THE BRAIN

Sir Charles Scott Sherrington, the corecipient of the 1932 Nobel Prize, had observed that the increase of brain complexity during vertebrate evolution correlates both with a greater functional unification of organisms (a closer functional welding of parts) and with greater dominance over their environment (richer and more manifold commerce with the environment). He stressed that connecting originally unconnected structures to act jointly, results in more than a simple sum of the activities of the separate component parts.

It has been pointed out that technological evolution follows principles closely analogous to biological evolution and that wholesale knowledge transfer from biology to technology is possible [10]. Maxwellian exploitation of their mutual similarities can and does provide technologically based inspiration and guidance for theory builders in biosciences. This is especially true for neuroscience and computer technology. The evolution of computer technology has already produced a greater functional unification within large and complex human organizations, as well as greater dominance over their environments.

In this paper we adopt a distributed computer network point of view of the brain. Because computers and their nets are still at a very early stage of their evolution, extreme caution is necessary in setting up the brain/computer analogy. It is well to remember that many brain functions are yet to be duplicated by engineers. Nevertheless, the recent revival of neural network modelling and building offers promise for overcoming conceptual barriers which impede transdisciplinary crossfertilization between technology and biology in general and between neuroscience and computer science in particular.

SYSTEM INTEGRATION OF THE CEREBELLUM

The cerebellum is a major part of the brain. The brain and the spinal cord constitute the central nervous system (CNS). A simplified brain taxonomy breaks up the brain into five parts: the end brain, the interbrain, the midbrain, the afterbrain, and the hindbrain. The end brain and interbrain constitute the forebrain, while the remaining three parts constitute the brain stem. The two major subsystems of the brain are the cerebral hemispheres, which are part of the end brain and the cerebellum which is part of the afterbrain. The other parts of the afterbrain, the pons and cerebellar peduncles, connect the cerebellum to other portions of the CNS. While the physical size of the cerebellum is smaller than that of the cerebral hemispheres, they contain similar numbers of neurons; i.e., between ten billion and one hundred billion.

The cerebellum subdivides into the cerebellar cortex and four pairs of deep cerebellar nuclei (DCN). The neuronal networks of the cerebellar cortex are compactly arranged within a folded three-dimensional matrix whose central layer comprises a regular two-dimensional lattice of flat Purkinje (P) neurons whose bodies define the Purkinje cell layer (PL). The layer above, toward the cortical surface, is the molecular layer (ML) and the layer below, toward the DCN is the granular layer (GL). Many rows of stacked P neurons combine into folia, which further combine into a hierarchically organized structure of sublobules, lobules and lobes. Many columns of P-cells are aggregated into separate zones which are associated with different axonic projections onto different DCN target neurons. This coordinate system allows a high degree of experimental reproducibility and permits the generation of "demographic" maps of sensory and motor projections onto relatively small populations of neurons spell within the cerebellar network [11].

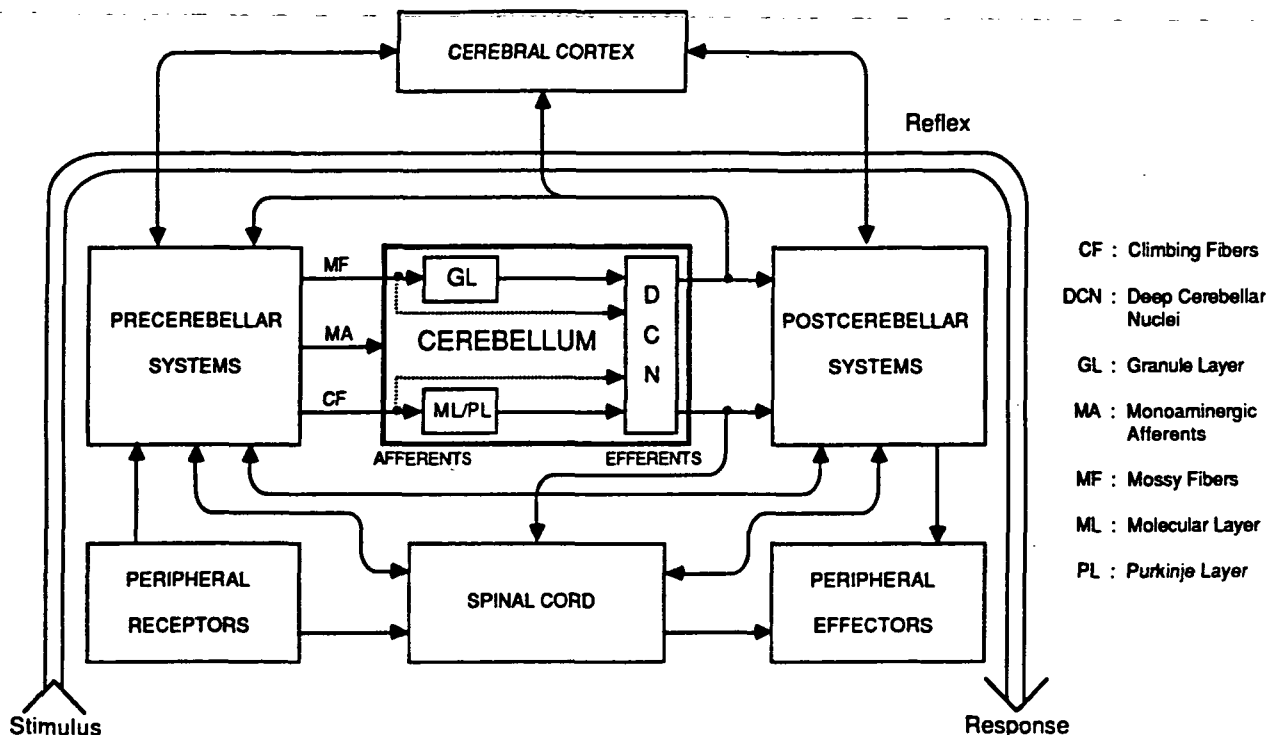


Figure 1. Cerebellar I/O Network

From a network and functional point of view the cerebellum is situated at the midpoint of a great multitude of reflex arcs, which are paths followed by nerve impulses that are responsible for many hundreds of different reflex actions. This we have indicated on Figure 1 which depicts a highly schematized flow diagram of impulse transmission from a sensory receptor source near the point of stimulation via afferent neurons to one or more reflex centers in the spinal cord or brain, and back from these centers through efferent neurons to a motor effector sink near a point of response. Our diagram lumps this great multitude of reflex centers and/or afferent and efferent relay stations into four generalized brain locations: the pre- and post-cerebellar systems, the cerebellum and the cerebral cortex. Neglecting the presence of a great variety of reflex and relay centers in each of these generalized locations, one can still deduce from the network topology of the diagram that there are at least thirty different general paths through this network which connect sensory sources to motor sinks. If we estimate the number of different paths through the large variety of individual reflex centers and relay centers, we arrive at many thousands of reflex arcs, a great fraction of which involve at least one passage through the cerebellum.

This should not be very surprising since the literature contains observations on many kinds of reflexes. A cursory examination revealed over 120. Ito's book lists at least 27 reflexes that involve the cerebellum.

The cerebellum receives three kinds of inputs and produces four kinds of outputs. It receives a high rate of pulses via mossy fibers (MF) which can originate from a very great multitude of precerebellar systems, the spinal cord or the cerebral cortex. It receives a much lower rate of pulses via the climbing fibers (CF) which originate in the inferior olive, a precerebellar system in the hindbrain, that receives inputs from over twenty other centers. These two kinds of inputs have quite different termination topologies. MFs terminate solely in the GL, while CFs terminate in all three layers. The MFs supply constant monitoring of sensory input data [12] while the CFs seem to be dedicated to inputting attention generating sensory data that signals time-uncertain or unanticipated events [13-14]. The third input to the cerebellum are monoaminergic afferent fibers (MA). There are at least two types. The noradrenergic type originates in the locus coeruleus and the serotonergic in the raphe complex. Their function remains obscure. Cerebellar output is produced

in DCNs. In rhesus monkey and cat, the ratio of input GL neurons to output DCN neurons is about a hundred thousand. If we allow for a twenty-five-fold increase in pulse rate from Purkinje to DCN cells, we estimate 4,000 input pulses per cerebellar output pulse. This ratio provides a measure of cerebellar processing power, i.e., its data rate reduction capability.

It should be noted that Figure 1 shows a direct connection of MFs and CFs to the DCNs, bypassing the cerebellar cortex. This supports the fact that absence of the cerebellar cortex does not result in loss of sensation or intelligence. It does result in ataxia, proprioceptive misperception, poor muscular coordination and inability to adapt to changing environmental conditions. Such behavior can be compared to an orchestra that lacks a conductor. The music score is followed but there are difficulties with coordination and synchronization of the players and any to-be-remembered changes in their performance.

CEREBELLUM AS A PROCESSOR OF INFORMATION

In synthesizing a functional model of the cerebellar processing architecture we try to adhere to the principle that reliable and up-to-date experimental biological knowledge should constrain inventive modelling. We desire to preserve relative numbers of various classes of neurons that form the "circuitry" and logic of the processor network. Their connectivities, as represented by their respective fan-outs and fan-ins should also be approximated. This can best be visualized with the aid of Figure 2 which has been constructed using our best estimates of numbers and topologies found in the massive but incomplete literature on the subject. It seems appropriate to remark at this point that this state of affairs has hardly changed since the days when Sherrington observed that exact knowledge regarding CNS anatomy and physiology is extremely inadequate although there exists a vast body of detailed fact. Since the numbers of the various kinds of

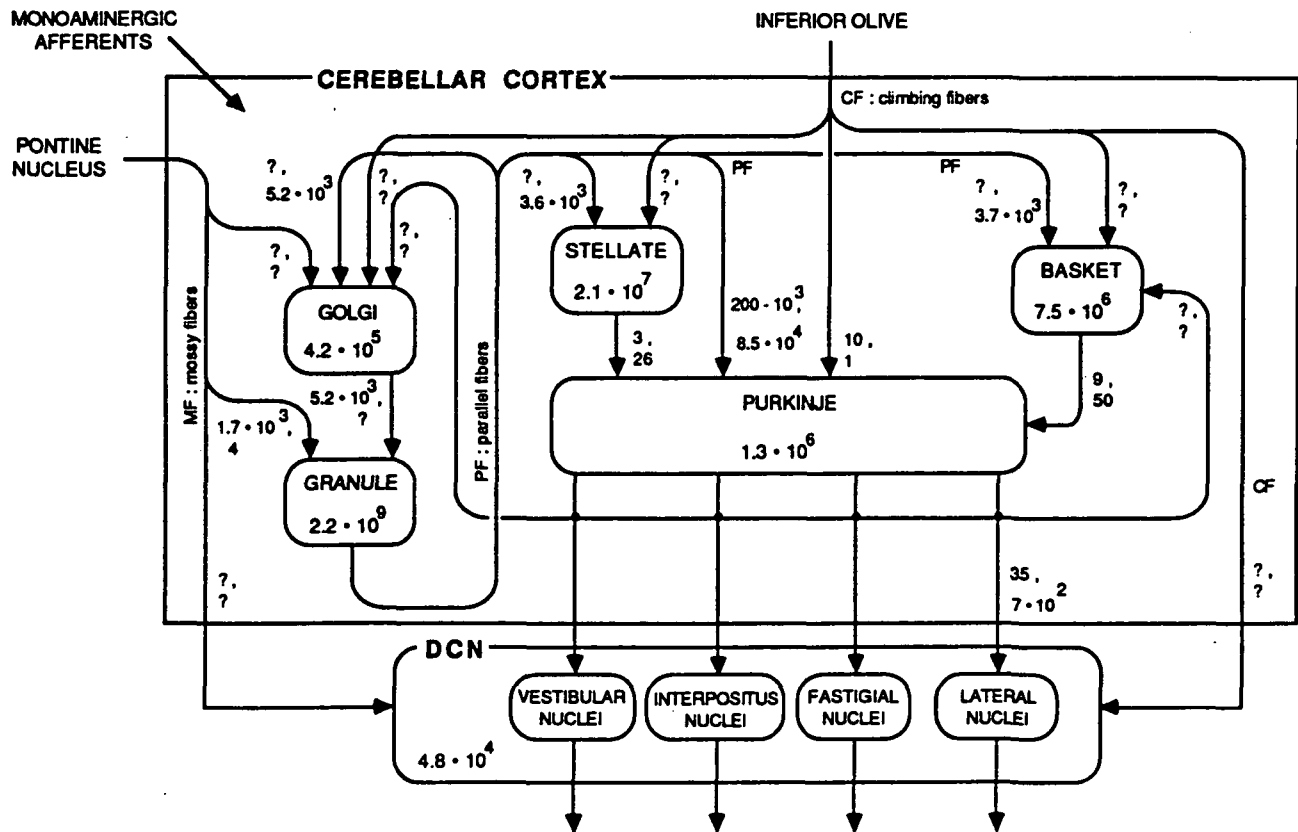


Figure 2. Cerebellar Interconnect Diagram

cerebellar neurons vary from specie to specie, we have standardized upon cat, whose facts are the most numerous and least inadequate.

Listed within their respective boxes are the population counts of the six kinds of neural cells found in the cerebellum. We discuss them in descending order. By far the most numerous are the very small granule cells. There estimated number is 2.2 billion in cat and 50 billion in man. They seem to be the most numerous neuron of the CNS in most species at the upper rungs of the evolutionary ladder. Then follow the two kinds of ML interneurons which in part interpose themselves in the major data processing path connecting the granule "input" cells to the Purkinje "output" cells. There are 20 million stellate cells and 7.5 million basket cells. The function of these interneurons has been thus far largely neglected by investigators of the cerebellum. The fourth kind of cell is the dominant Purkinje cell. It numbers 1.3 million in cat. This large and very regularly arrayed cell is also the most investigated one. Its false color photomicrograph adorns the cover of the special "Frontiers in Neuroscience" November 4, 1988 issue of SCIENCE. The photo belongs to a paper reporting microfluorometric imaging of intracellular calcium concentrations as a function of voltage-dependent electrical activity in cerebellar Purkinje cells [15]. The least numerous neurons of the cerebellar cortex are the Golgi cells. Their population count is less than half-a-million. They are among the most successfully modelled neurons of the cerebellum [16-18]. We agree with past modellers that the evidence is strong that Golgi cells regulate sensory data transmissions from the granule to the Purkinje cells via a negative feedback loop. However, in contradistinction to the presuppositions made in the above models [16-17] we think that Golgi cells receive inputs not only from MFs and granule cells, but that their activity is also subject to control by the second major cerebellar input, the CFs [19]. In comparison to the cell population counts in the cerebellar cortex, the population of DCN cells is truly diminutive. The largest DCN in the cat contains less than ten thousand cells while the sum total in all its DCNs is less than fifty thousand.

A concern of massive parallel processing design is fan-ins and fan-outs between successive processor stages. A major result of our preliminary investigation has been the establishment of histological facts about axonic connections projecting onto the six types of cerebellar

neurons. We show our findings in Figure 2. Where known, the directional interconnect gives two numbers. The upper number signifies the average number of target neurons which are reached by axons of a source neuron, while the lower number signifies the average number of source neurons that contribute inputs to target a neuron. On average, hundreds of Purkinje cells get an input from a granule cell while, about 85 thousand granule cells contact a Purkinje cell. The corresponding numbers for the stellate and basket to Purkinje connections are 3,16 and 9,50. These fan-ins of ML interneurons strongly suggest that they participate in logic processing, a role mostly overlooked by others. We believe that histologists need to fill-in numbers missing in our diagrams before their detailed functions can be clarified. Direct Purkinje cell to Purkinje cell links also need further attention. The large distributory role of CFs, supports Llinas' view that P-cells act in ensembles [12]. Fan-ins onto DCN targets give further credence to this view, especially when combined with an interpretation of the reported negative, as well as positive, changes in simple spike activities of P-cells [14]. The Marr model needs adjustment in light of the CF-Golgi connection and the CF read-out theory. In the absence of data we intend to simulate the above circuits. Our results increase the options for locating the thus far elusive, seat of memory in cerebellar network models.

I thank Coe Miles-Schlichting for help in preparing the above figures.

REFERENCES

- [1] J. C. Maxwell, "The Telephone", Nature XVIII, 159-63, 1878. [2]
- [2] E. E. Loebner, "From Maxwell's Art of Theorizing to Pupa's Art of Inventing, 1987 Annual Meeting of the Society for the History of Technology, unpublished.
- [3] E. E. Loebner, "Concurrency Assurance in Vertebrate Retinas", IEEE First International Conference on Neural Networks, Volume IV, 147-159, 1987.
- [4] S. P. Wise and R. Desimone, "Behavioral Neurophysiology: Insights into Seeing and Grasping", Science, 242, 736-741, 1988.
- [5] P. S. Churchland and T. J. Sejnowski, "Perspectives on Neuroscience", Science, 242, 741-745, 1988.

- [6] E. E. Loebner, "Image Processing and Functional Retina Synthesis", 1960 Bionics Symposium, USAF WADD Tech. Rept. 60-600, 309-338, 1960; republished as Chapter 39 in Human Factors in Technology, Eds. Bennet, Dugan and Spiegel, 492-518, McGraw Hill, NY, 1963; E. E. Loebner and T. Henneberger, "Image Processing in Synthetic Visual Organs", RCA Res. Tech. Rept. July 21, 1961.
- [7] M. B. Herscher and T. P. Kelley, IEEE Trans. Mil. Electr., MIL-7, 98-103, 1963.
- [8] A. Barak and A. Litman, "MOS: A Multicomputer Distributed Operating System, Software-Practice and Experience, 15, 725-737, 1985.
- [9] N. Alon, A. Barak and Manber U., "On Disseminating Information Reliably Without Broadcasting," Proc. 7th International Conference on Distributed Computing Systems, 74-81, Berlin, September, 1987.
- [10] E.E. Loebner, "Electronics and technoevolution," EDN 26 (20), 74-76, 1981.
- [11] Masao Ito "The Cerebellum and Neural Control", Raven Press, NY., 1984.
- [12] R. Llinas, "Functional Significance of the Basic Cerebellar Circuit in Motor Coordination", in Cerebellar Functions, J.R. Bloedel, J. Dickgens, and N. Precht, Editors, 170-185, 1985.
- [13] Jay-Shin Lou and J. R. Bloedel, "The responses of simultaneously recorded Purkinje cells to the perturbations of the step cycle in walking ferret: a study using a new analytical method - the real-time postsynaptic response (RTPR)", Brain Res., 365, 340-374 1986.
- [14] N. Mano, I. Kanazawa and K. Yamamoto, "Complex-Spike Activity of Cerebellar Purkinje Cells Related to Wrist Tracking Movement in Monkey," Journal Neurophysiol. 56, 137-158, 1986.
- [15] D.W. Tank, M. Sugimori, J.A. Connor, R.B. Llinas, "Spatially Resolved Calcium Dynamics of Mammalian Purkinje Cells in Cerebellar Slice", Science, 242, 773-777, 1988.
- [16] D. Marr, "A theory of cerebellar cortex," J. Physiol. 202, 437-470, 1969.
- [17] J. S. Albus, "A theory of cerebellar function," Math. Biosci. 10, 25-61, 1971.
- [18] Fujita, "Adaptive filter model of the cerebellum," Biol. Cybern. 45, 195-206, 1982.
- [19] J. Hamori and J. Szentagothai, "Identification under the Electron Microscope of Climbing Fibers and their Synaptic Contacts,: Exp. Brain Res. 1, 65-81, 1966. See also [12] above.

ABOUT THE AUTHORS

The authors are listed below in order of presentation of their papers at IEEE COMPCON Spring '89 and in the same order in which they have been reprinted in this report.

PENTTI KANERVA. Kanerva is Principal Investigator of the RIACS Sparse Distributed Memory (SDM) Project. Kanerva came to RIACS in 1985 from Stanford University to establish the project. While at Stanford, Kanerva developed a large terminal network, software and hardware for technical writing, typesetting software, algorithms for searching and sorting, new memory and new CPU architectures, and directed the technical development of a campus-wide network for text-handling. He was also a postdoctoral fellow and is presently a visiting scholar at Stanford. In 1965-1967 he organized the first computer center for the University of Tampere, Finland, and taught mathematics, statistics, and computer programming. Prior to working at Tempere, he worked in experimental design and statistical analysis at the Forest Research Institute of Finland and the Finnish State Computer Center. He has a PhD in Philosophy from Stanford (1984) and MS in Forestry from the University of Helsinki (1964). His current research is based on his dissertation on memory and is aimed at building practical computers with some of the desirable properties of human memory. Kanerva's book *Sparse Distributed Memory* (MIT Press, 1988) presents his dissertation with an additional chapter on the organization of an autonomous learning system.

JAMES S. ALBUS. Albus is Chief of the Robot Systems Division at the Center for Manufacturing Engineering, National Institute of Standards and Technology (formerly National Bureau of Standards), where he has been working since 1973. Albus was previously with NASA, where he designed electronic systems for over 15 spacecraft, conducted research in advanced computer technology, directed the NASA Artificial Intelligence Program, and studied the information-processing and computational capabilities of the human brain. He was elected Fellow of the Washington Academy of Science and awarded the Department of Commerce Silver Medal for his paper "A Theory of Cerebellar Function." In 1976, Albus' model of the cerebellum was recognized by *Industrial Research* magazine as one of the most significant inventions of the year. In addition to over 30 journal articles, Albus has written popular articles on robotics for *Scientific American*, *Encyclopedia Americana*, *Omni*, and *BYTE*. His book *People's Capitalism: The Economics of the Robot Revolution* (New World Books, 1976) outlines a series of dramatic proposals through which advances in robot technology could benefit humanity. His book *Brains, Behavior, and Robotics*

(BYTE books/McGraw-Hill, 1981) examines the state of the art of robotics and the difficulties of equipping a robot to do its designated tasks, and discusses potential applications ranging from industrial uses of robotics to space exploration, energy production, and home services.

EGON E. LOEBNER. Loebner is Counselor for Science and Technology at Hewlett-Packard Laboratories in Palo Alto. He joined the Hewlett-Packard Company in 1961 as manager of Optoelectronics of HP Associates after heading optoelectronics research for six years at RCA Laboratories and prior to that for three years at Sylvania Electric Products, Inc. From 1974 to 1976 he served as Chief of the Science Section at the American Embassy in Moscow. He received a PhD in Physics from SUNY at Buffalo in 1955. His thesis dealt with the electronic properties of carbons and graphite. In 1959 Loebner received an RCA Achievement Award for research on the chemistry of high band-gap semiconductors. He has authored over 40 patents and 50 papers in bionics, biophysics, chemistry, cognitive science, computational linguistics, electronics, human factors, information displays, material science, optics, physics, and telecommunications. Loebner has been elected Fellow of the IEEE in recognition of contributions to optoelectronic materials, devices, and networks, and for transdisciplinary research across life sciences, social sciences, and computer sciences. He is listed in *Who's Who in the West* and in *Who's Who in the World*. His main contributions are the establishment of optoelectronics as a technology, pioneering of LED displays, design of a functional model of the frog retina, and construction of a psychophysical model of discovery and invention.

ORIGINAL PAGE IS
OF POOR QUALITY

RIACS

Mail Stop 230-5
NASA Ames Research Center
Moffett Field, CA 94035
(415) 694-6363

The Research Institute for Advanced Computer Science
is an Institute of the
Universities Space Research Association
The American City Building
Suite 212
Columbia, MD 21044
(301) 730-2656