Committee 1
The Information Revolution,
Higher Education and Research

DRAFT--Oct. 15, 1997
For Conference Distribution Only

PROSPECTS FOR INTELLIGENT SYSTEMS

by

Edward J. Finn
Professor of Physics
Georgetown University
Washington, D.C. USA

The Twenty-first International Conference on the Unity of the Sciences
Washington, D.C. November 24-30, 1997

Prospects For Intelligent Systems

Edward J. Finn
Professor of Physics
Georgetown University
Washington, DC 20057-0995

What Constitutes an Intelligent System?

If the purpose of this paper is to discuss prospects for intelligent systems as related to information technology and research, it would be best to have an operational definition for "intelligence" or an "intelligent system". I find this a rather slippery task because, to paraphrase one of our country's Supreme Court Justices: I can't <u>define</u> intelligence, but I know it when I see it. Nevertheless, let me try, first using standard sources. The Random House Dictionary defines intelligence as "capacity for reasoning, understanding, and for similar forms of mental activity; aptitude for grasping truths, facts, meanings, etc." plus a lot of others. The dictionary further suggests that we look under "mind" for synonyms for intelligence. The Columbia Encyclopedia has a full column on the topic describing it as a psychological attribute " ...variously described as the general ability of the organism acting as a whole to utilize understanding gained in past experience in dealing with a similar or new situation, to adjust or adapt quickly and readily to the environment, to learn without difficulty, or to form new behavior patterns to meet a new situation by the modification or readjustment of those already acquired." My Encyclopedia of Science simply sends me off to read about Animal Intelligence, dismissing any other possibility. What both you and I want from this discussion, rather, is associated with what is often described as Artificial Intelligence; that is, a device capable of performing some (preferably, all) of the feats attributable to a living being. Notice that I have *not* predicated the intelligence desired as that of a human mind/brain. It would be a marvelous feat when we produce a machine that "thinks like a human". I want to consider the possibility of going *beyond* that point!

Let's start with a little historical perspective. Alan Turing[1] was responsible for what we all know as the Turing Test or the Turing Machine, whereby it may prove impossible to recognize the **behavior** of a machine/robot/computer as different from that of an "intelligence", however that term is defined. More than one person has found such a test inadequate in identifying intelligence. One of the earliest examples is that where Kurt Gödel[2] in his famous theorem showed that for every consistent system, there exist truths that are unprovable within that system. For those of us unfamiliar with this theorem, which is mathematical in nature, let me say that the key word in its statement above is *consistent*. In Gödel's use, a system is consistent if any given statement, *S*, and a statement that is the negation of *S* are both not theorems of the system. According to Gödel then, we can never get to all the truth by simply following a set of rules; the real world simply refuses to be constrained by deductive reasoning and that is the best any rule-based system can be expected to follow. As food for thought, however, consider that Gödel's theorem implies machines cannot be intelligent by placing limitations on *consistent* machines. What if the machines were *in*consistent in the same way that humans are often inconsistent?[3]

The reason why rule-based Artificial Intelligence, AI, though adequate for certain (mindless [?!]) tasks, constantly runs into failure has to do with an inability to *adapt* - or to intuit. As Stuart and Herbert Dreyfus[4] described the stages of learning in becoming proficient to drive a car, a person first follows rules that are slowly and sometimes painfully acquired up to the point where a driver performs almost automatically. We, intelligent people, can drive too close to the car in front of us, "sense" a car beginning to pass on the right and all the other remarkable things we do when behind the wheel and properly attentive. The "automatic" cars we saw recently on TV traveling like a group of ducklings one behind the other, required more than a few control devices to keep them in line - including buried magnets. The display may have looked impressive, but a careful inspection would have uncovered numerous causes for concern.

So, it's the ability to adapt that really sets the goal for an "intelligent system". Chief Administrator Dan Goldin of NASA understands this. He convened a seminar in 1996 because he envisions sending an unmanned craft to the planet of another star. If the "intelligent system" sent is not adaptable and runs into a situation totally unexpected - for instance, life forms based on silicon rather than carbon - we are looking at a different atmosphere, different spectral regions for information, lots of different problems. Being light-years away, the "intelligent system" may have to reconstruct itself (if it can) even before it descends to the surface. It certainly will not have the choice of calling home for instructions. Let me suggest we are at the beginnings of building "intelligent systems" that will have this capacity. Their generic name is Neural Networks.

A Neural Networks Primer

Neural networks are models for computation that take their inspiration from the way the brain is supposed to be constructed. They often try to solve the problems that the brain seems to try to solve. **Biological neural networks** in mammals are built from neurons, nerve cells, which are themselves remarkably complex biological units (Figure 1). Huge numbers of neurons,

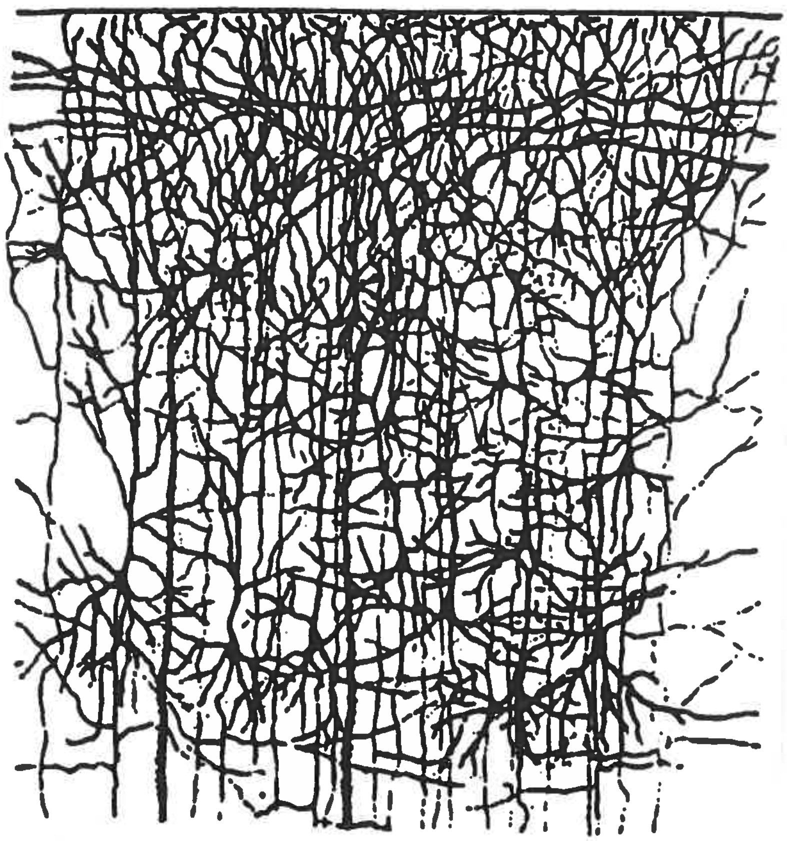

Figure 1. Reproduction of Ramón y Cajal's drawing of neuronal structure.

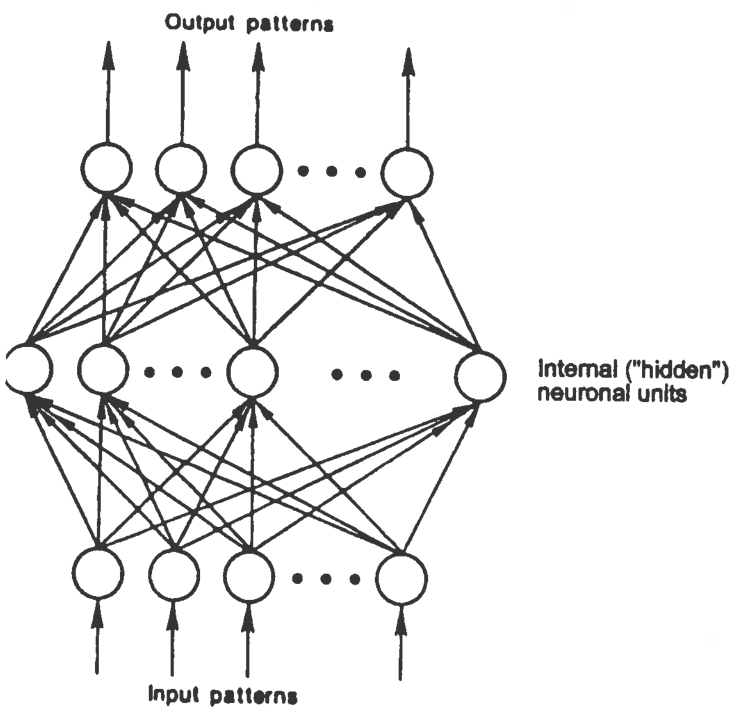

Figure 2. Generic structure of a prototypical artificial neural network.

connected together and cooperating in poorly understood ways, give rise to the complex behavior of organisms. **Artificial neural networks** are smaller, simpler, and more understandable than the biological ones, but still able to do some remarkably interesting things (Figure 2). Some of the operations that artificial networks are good at -- pattern recognition, concept formation, association, generalization, some kinds of inference -- seem to be similar to things that brains do well. It is fair to say that artificial neural networks (ANNs) behave a lot more like humans than digital computers do, even to the fact that both require time to learn and neither are very good at balancing a checkbook without help.

There are two ways of "doing" neural networks; one is by building electronic circuits and the other is to *simulate* the network. In the first case silicon chips are used that have been organized so that they may recognize images or sounds (such as Carver Mead's retina and cochlea)[5]. There have been many simulations used for basic research as well as programs, called neural networks, actually used to perform such high level tasks as decision-making on loan applications and translations of written material to audible text. The Appendix will delineate examples of these two techniques.

The modern history of artificial neural networks is usually considered to have begun with an often reprinted 1943 paper by Warren McCulloch and Walter Pitts[6]. They were making models for brain function; that is, what does the brain compute and how does it do it? However, only two years after the publication of their paper, in 1945, John Von Neumann[7] used their model for neuron behavior and neural computation in an influential discussion of the proper design to be used for future generations of digital computers.

There are two related but distinct goals that have driven neural network research since its beginnings. The idea is to construct and analyze artificial neural networks because that may allow us to: **First** begin to understand how the biological neural networks in our brains work (this is the domain of neuroscience, cognitive science, psychology, and perhaps philosophy) and **Second** build more intelligent machines (this is the domain of engineering and computer science). These two goals -- understanding the brain and making smart devices -- are mixed together in varying proportions, though the bias is toward the careful analysis and application of artificial networks. Although there is a degree of creative tension between these two goals, there is also synergy.

The creative tension arises from the following observation. Consider an engineer who wants to use biology as inspiration for an *intelligent adaptive device*. Why should the engineer be bound by biological solutions? If the biological hardware is slow and unreliable, perhaps only intrinsically undesirable algorithms are possible. Ample evidence suggests that our lately evolved species-specific behaviors, like language, are simply not very well constructed. After only a few tens of thousands of generations of talking ancestors, human language is still no more than an indispensable kludge, grounded in and limited by the circuitry that nature had to work with in the primate brain. Perhaps after several million more years of evolution our descendants will finally get it right. Perhaps there are better ways to perform the operations of intelligence and so why stick with the second rate?

The synergy between biological neural networks and artificial neural networks arises in several ways: **First**, precise analysis of simple, general neural networks is intrinsically interesting and can have unexpected benefits. The McCulloch-Pitts paper developed a primitive model of the brain, but a very good model for many kinds of computation. One of its side effects was to originate the field of finite state automata[8]. **Second**, to make intelligent systems usable by humans perhaps we must make artificial systems that are conceptually, though not physically, designed like we are. We would have difficulty communicating with a truly different kind of intelligence. The current emphasis on user friendly computer interfaces is an example. Large amounts of computer power are spent to provide a translator between a real logic processor and our far less logical selves. For us to acknowledge a system as intelligent perhaps it has to be just like us, perhaps even with arms, legs and a warm smile! As Xenophanes commented 2,500 years ago, "... horses would draw the forms of gods like horses, and cattle like cattle, and they would make the gods' bodies the same shape as their own." **Third**, neural networks provide a valuable set of examples of ways that a massively parallel computer could (should?) be organized. Current digital computers will soon run up against limitations imposed by the physics of electronic circuitry and the speed of light. One way to keep increasing computer speed is to use multiple central processing units (CPUs); if one computer computes fast, then two computers should compute twice as fast. Unfortunately, coordinating many CPUs to work fast and effectively on a single problem has proven to be extremely difficult. Neurons have time constants in the millisecond range (10^{-3} s); present-day silicon devices have time constants in the nanosecond range (10^{-9} s). Yet somehow the brain has been able to build exceedingly powerful computing systems by summing the abilities of huge numbers of biological neurons, even though each neuron is computing several orders of magnitude more slowly than an electronic device constructed from silicon. The best known example of this design is the mammalian cerebral cortex, where neurons are arranged in parallel arrays in a highly modular structure (recall Figure 1). Most neural networks are abstractions of the architecture of mammalian cerebral cortex.

Knowing, in detail, how this parallel architecture works would be of considerable practical value. However, the study of human cognitive abilities suggests a price may be paid for using it. The resulting systems, both biological and artificial, may be forced to become very special purpose and almost surely will lack the universality and flexibility that we are accustomed to in digital computers.

The things that make neural networks so interesting as models for human behavior [for example, (a) good generalization, (b) easy formation of associations, and (c) the ability to work with inadequate or degraded data], may appear in less benign form in artificial neural networks as: (a) loss of detail and precision, (b) inexplicable prejudice, and (c) erroneous and unmotivated conclusions. Making effective use of artificial neural networks requires a different kind of computing than we are used to, one that solves different problems in different ways but one with great power in its own domain (See the Appendix for one of the methods followed by ANNs to "compute").

Where are Neural Networks Leading Us?

When neural networks regained popularity, in the mid 1980's, a term that was sometimes used to describe systems containing them was "neuromorphic." "Brain-like computing" was another way of saying about the same thing. When one of these terms was used in engineering the implication was that the artificial devices being built were following at least some of the design principles of the mammalian brain (recall Figure 1, again). To those professionally concerned with behavior, a parallel set of names might be proposed: "Psychomorphic" systems and "mind-like computing." Artificial intelligence (AI), as classically defined, is describable by these names, though when AI first developed in the 1950's and 60's it deliberately paid little attention to the substantial amount known about the facts of human behavior, believing that sheer cleverness was capable of overcoming ignorance. Day by day we are learning more how the brain functions, how and where it stores information and numerous details with respect to the purpose of the various types of neurons. At least some of this knowledge will be useful in building neural networks in the future. However, a major conceptual problem of neural networks is that, even if they are in some vague architectural sense neuromorphic, they are rarely psychomorphic. Even though there is a large body of lawful, regular, and reproducible experimental results in the behavioral sciences, these ideas have rarely had much influence in the neural network community, outside of a small number of researchers who specifically try to model human cognition.

Let me state several reasons for this neglect: **First**, there are missing levels of organization on the neuroscientific level. Neural network models are built from elementary computing units. The largest neural network simulations used in practice contain perhaps a few thousand units. The human brain contains billions of neurons, 10^{12} or so, with as many as 10^{15} connections (or *synapses*). Current neural network models have a severe problem using, or even acknowledging, the intermediate levels of organization that must exist in this numerical gap in scale between the properties of single units and coordinated activity of the whole brain. As an example of the problem, consider a large business organization, say IBM. We can follow an individual employee during the course of a day. Or we can follow the health of the company as a whole by looking at the annual report. It would be difficult to infer from either of these sources of information the presence of workgroups, departments, and divisions; that is, groups of employees and groups of groups of employees, where in fact most of the work of the company is organized and performed. Similarly, government has complex and essential intermediate level structures, for example, in rough order of size, neighborhood, city, county, state, and federal. Neuroscience currently allows us to look at single unit recordings for the behavior of single neurons and gross electrical activity (EEG, evoked responses, imaging) for overall activation levels, roughly the lowest and highest levels of neural organization. It is clear that there are several orders of magnitude of grouping that must exist, have been conjectured to exist, are felt to be important, but about which almost nothing is known as yet. We wait patiently and with great expectation for experimental evidence to fill in the gaps. **Second**, there are missing levels of organization at the Cognitive Science level. The most commonly used formulations of network learning are limited and often misleading from the point of view of a psychomorphic system. Neural network theory has been strongly influenced, for better and worse, by the mathematics of classic *pattern recognition*[9]. Typically, pattern recognition assumes that sensors have provided a set of input data connected to a classification, say a set of pixels corresponding to the written letter "A." A network is presented with a number of examples of the classification in a training set and the weights in the network are adjusted by various learning algorithms so as to make it classify more accurately in the future. It can be shown that properly designed neural networks can do this operation effectively enough for many useful applications. However, a psychomorphic engineer might ask: is this all that we want to do? This structure, with an input pattern transformed in the network to an output pattern, reproduces in form classic Stimulus-Response (S-R) learning from psychology. S-R learning was proposed by the behaviorists in the 1920's and 30's as the only true basis of a scientific psychology[10]. Essentially, we can solve the problem of animal behavior when we make lists of externally observable stimuli, the associated observed responses, and assume the brain is there to make links between them. No hidden mental processes need be invoked. Clearly there is *some* truth behind this analysis. *Association* has been known to be a primary mechanism of learning since Aristotle. Even Aristotle, however, was quite aware, and it has been amply confirmed by work in psychology and cognitive science over the past decades, that such a limited definition of association cannot explain many aspects of behavior. It is therefore distressing to see neural network theorists deliberately, or even worse, unconsciously, reproduce a severely limited and inadequate view of mental operation. **Third**, focus on the *formation* of accurate associations has distracted attention from a number of other important requirements for a psychomorphic system. *Controllability*, *flexibility*, and *teachability* are at least as important in human cognition as accuracy in retrieval, probably more so. For example, consider (Figure 3) the pixel pattern that a letter recognizer classifies as a letter "A." Depending on the context, this pattern can be labeled

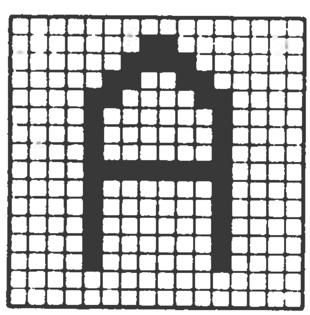

Figure 3. A set of pixels displaying a pattern giving an "A".

as a capital "A", a grade in a college class, an indefinite article in English, and so on, while in French, Spanish or Italian it is a preposition. Further, the switch between one possible association of the pattern and another is extremely rapid. For example, in a psychological experiment, an "A" can be first associated with, say, pressing a button on the left. Time to respond to the presentation of the "A" will become faster with repeated presentations, even though responses have been error free since the beginning of the experiment. Suppose a verbal instruction now tells the subject to respond to an "A" by pressing the button on the right. Suddenly the subject is making a different response. The responses may be a little slower at first, but performance is still error free. This flexibility is common and so trivial that we hardly even think about how difficult it will be to get an artificial neural network to completely and correctly shift its input-output relationships in a matter of milliseconds. My guess is that need for this flexibility places much more stringent constraints on possible neuromorphic architectures than accurate learned association. The psychomorphic system can constantly and quickly reprogram itself. The previous example also suggests the importance of "teachability" for network operation. Somehow presentation of properly structured inputs can speed up learning by orders of magnitude. In the example above, the inputs causing a change in association were not even examples of the association but verbal instructions recombining past learning. Learning in school would be a painful and slow process if it were purely associative. Learning does not proceed by a random pairwise accretion of facts in knowledge space. Something much more complex is occurring, involving the formation of mental structures, use of interlocked concepts and detailed mental models and the presentation of specific factual examples that are explained by a teacher. The time course of real learning is often strikingly unlike the time course of simple neural network learning. Neural network learning typically starts with a *tabula rasa*, learns the first associations quickly and accurately, and then gets slower and less accurate as it learns more and more. Real learning often starts slowly -- for example, learning the multiplication tables in grade school -- and then accelerates, so college mathematics courses provide an immense amount of information very rapidly, once the foundations are built. As William James[11] commented:

> ... the more other facts a fact is associated with in the mind, the better possession of it our memory retains ... Let a man early in life set himself the task of verifying such a theory as that of evolution and facts will soon cluster and cling to him like grapes to their stems. Their relations to the theory will hold them fast ...

The point here is that real memory has strong high level structure that uses simple association as an elementary mechanism. Past information can aid in the learning and retrieval of later information. One of the best critiques of simple neural networks is in the paper by Jerry Fodor and Zenon Pylyshyn[12], who, among other points, observed that simple association is such an inefficient way to build an information processing and retrieval system that an engineer would be strongly advised to use something else if the system was to be in any way useful. An obvious and practical task for future research is to take today's relatively well understood simple neural network systems and try to combine them in such a way as to reproduce at least a little of the flexibility and controllability observed in human memory. **Fourth**, because the history of neural networks is tied to pattern recognition and computer science, there is a tendency to believe that neural networks form general computing systems in the sense that Turing machines form universal computers. There is absolutely no reason to believe that this is true. The biological nervous system is concerned with *specificity* and not *generality*: specific sensory systems, specialized structures, specific kinds of computations. Although we like to think the human brain is very general, when mental operations are looked at in detail striking limitations appear. For example, the simple logic operation known as the Exclusive OR, the *bête noire* of neural nets, can be incorporated into a puzzle. This puzzle can be solved by humans, though often with some difficulty. The same logical structure when it appears in what seems to be a different problem often does not generalize. There is a substantial body of research on this observation in cognitive science.

Successful computation in neural networks is dependent on details of the data representation, that is, on how the pattern of input and output unit activation relates to the world. Neural networks are extremely sensitive to representations. In a real sense, the data representation is the mechanism by which networks are programmed. The choice of a good data representation is of far more value toward the solution of a problem than is the choice of the learning rule or network. For various reasons, including the fact that neural structures tend to be noisy (just as biological neurons are), and that small errors can propagate and amplify, it is not possible to have psychomorphic computers perform in sequence the very large number of accurate elementary computational steps that characterize the operation of digital computers. A small sequence of computational operations combined with an effective input and output neuromorphic data representation comprises the entire psychomorphic computation. John von Neumann pointed out this essential characteristic of neural computation in 1958[13].

The biological brain contains true marvels of data representation, using details of neuroanatomy and neurophysiology to respond to useful properties of the world. However, data representations tend to be very problem specific. The more that is known about a given problem, the less general adaptabilty is needed. Learning requires ignorance; if everything is known, nothing need be learned. Learning and adaptation are dangerous for an animal because they involve rewiring the nervous system and should be used only when necessary. It has been suggested that normal learning is one end of a continuum, with pathology lying at the other. Here, perhaps more than in many fields, God is in the details.

I suppose the point of this discussion is that we know only a little about the earliest stages of intelligent system design. The outlines of intermediate level network organization and the rules, if there are any, for designing data representations for specific problems remain to be discovered. It is not even clear what is the best way to analyze complex intelligent systems; proper analysis may start with traditional statistics and its extensions to pattern recognition but is unlikely to end that way. The most important future developments for both intelligent machines and for the understanding of our own mental processes may arise when the constraints and the abilities seen at the highest levels of cognitive function can be connected with low and intermediate level neural network architectures.

To conclude, consider the similarity in the following two concepts. One or two gas molecules in a box move around on carefully identifiable paths. Two hundred or so also can be thought to follow already defined paths, even though they may collide infrequently and keeping track of all their path would be difficult but not impossible - just get a bigger and faster computer and you can simulate such a process "almost in real time". When millions of gas molecules are in the box, entirely different *macro*scopic phenomena are apparent; we use such concepts as pressure, temperature and even entropy to describe them. A gas even supports wave phenomena, where large groups of molecules move in unison and have different macroscopic values at different times. The same may be said when we consider first a few neurons where we can follow the creation of an action potential that travels down an axon, which results in the release of a cloud of neurotransmitters at one (or a number) of its synapses and, after passing across the cleft come to receptors on the dendrite of a different neuron. All that is "easy" to follow and predict. Get many, many neurons to interact and macroscopic phenomena must appear that are vastly different from the microscopic working of individual neurons.

The number of synaptic connections in the human brain is *MILLIONS* of times greater than the number of human genes. In *aplysia* (a common sea slug often used for neuroscientific studies), there are only 20,000 neurons and each is unique, but identical in every *aplysia*. In cloning, the clones are identical in that they have identical genes; but they *do not* have the same synaptic connections.

Bibliography

J. Casti (1994) *Complexification: Explaining a Paradoxical World Through the Science of Surprise* New York: HarperCollins Publishers.

R. Penrose (1990) *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics* Oxford: Oxford University Press.

F. Crick (1995) *The Astonishing Hypothesis: The Scientific Search for the Soul* New York: Simon and Schuster.

I. Aleksander and P. Burnett (1987) *Thinking Machines: The Search for Artificial Intelligence.* New York: Alfred A. Knopf.

S. Haykin (1994) *Neural Networks: A Comprehensive Foundation.* New York: Macmillan College Publishing Company.

J. A. Anderson (1995) *An Introduction to Neural Networks.* Cambridge, MA: MIT Press.

E. L. Schwartz, ed. (1990) *Computational Neuroscience.* Cambridge, MA: MIT Press.

APPENDIX

The information-processing unit fundamental to the operation of an artificial neural network (ANN) is often called a *neuron* due to the influence on its structure by the biological neuron that appears in animals. Figure A-1 depicts a *model* for the artificial neuron, which contains three basic elements:

1. A set of *synapases* or *connections*, each of which has a *weight* or *strength*.
2. A *summing element* that adds each input signal after it has been "weighted".
3. An *activation function* that limits the output of the neuron.

In particular, a *signal*, x_i, is input at synapse *I* of neuron *k* and is modified by the weight w_{ki}. The weight may be *excitatory* (when its value is positive) or *inhibitory* (when it is negative). The modified signals are added by the summing element, resulting in a signal u_k, which is modified by a *threshold* term and then passes into the activation function. The threshold term is applied to mimic the manner whereby a biological neuron "fires" only when its electric potential is a certain "threshold" value above its quiescent state. The activation function used depends on the modeler, but typically is some sort of a sigmoid function (Figure A-2). Mathematically the process is described by:

$$u_k = \sum_{i=1}^{n} w_{ki} x_i$$

and

$$y_k = \phi_k(u_k - \theta_k)$$

where $x_1, x_2, x_3, \ldots x_n$ are the input signals, $w_{k1}, w_{k2}, w_{k3}, \ldots w_{kn}$ are the synaptic weights of the neuron k, u_k is the output of the linear summing device, θ_k is a threshold quantity and ϕ_k is the function that transforms the summer output into the neuron's output, y_k. A neural network, then,

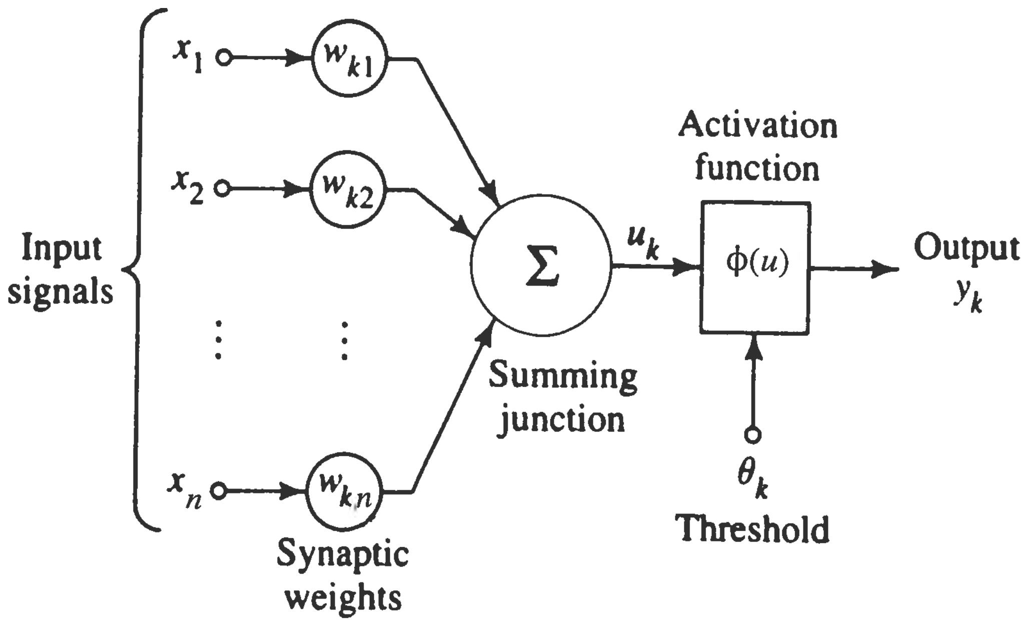

Figure A-1. A nonlinear model for an artificial neuron.

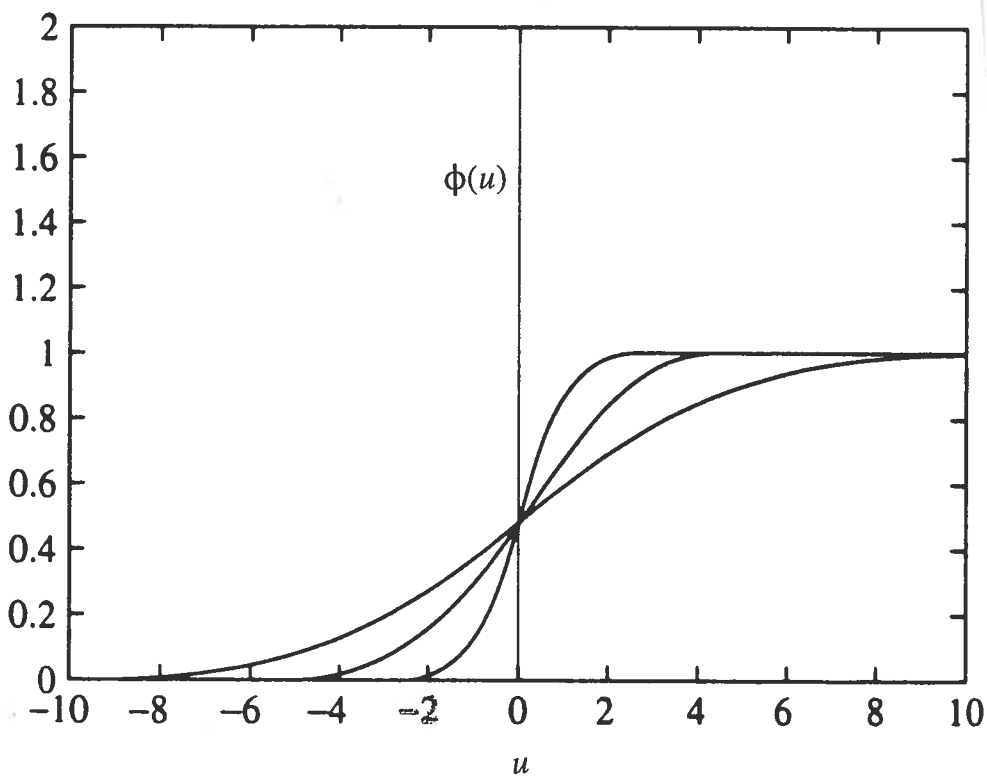

Figure A-2. Three examples of the activation function $\phi(u)$.

is created by connecting a number of these neurons together, where the output of one neuron may be transmitted to one, two or two thousand other neurons (including itself for the purpose of self-inhibition). In computer simulation a mathematical model of Figure A-1 is written and "layers" of each neuron are set up in a format similar to that as given in Figure 2 for a representative neural network consisting of three "layers" of neurons, typically called (I) input layer; (ii) hidden layer; (iii) output layer.

The digital approach would typically use a reduced instruction set computer (RISC) processor, which is designed to execute a small number of simple instructions. Such RISC systems are often fast enough to be satisfactory for such applications as a decision-maker for loan applications or a Pap smear analysis. However, for complex applications such as speech recognition or optical character recognition, process control or adaptive noise cancellation, a system that can not only learn rapidly but must also adapt rapidly is needed. To meet these demands very large scale integrated (VLSI) circuits are beginning to provide an ideal medium for hardware implementation of neural networks. In VLSI technology, tens of millions of transistors can be fabricated into integrated circuits on a single silicon chip.

The chips available today are of three varieties:

(a) *digital*, where all processes are done in a manner quite similar to the methods used in present-day digital computers and the output is a digital signal. This technology allows for ease of design and manufacture with high precision as well as flexibility in the use of complex algorithms, despite the fact that digital implementation of multiplication requires relatively large quantities of both area and power.

(b) *analog*, where the information-bearing signals have a continuous amplitude throughout the "computation". Although analog circuits suffer from lack of precision, this shortcoming is compensated by the efficiency of the computations based on the principles of classical circuit theory (and the fundamental laws of physics). Such circuits, based on metal oxide silicon (MOS) transistors, can do computations that are either difficult or time-consuming (or both) in the digital mode with ease and perform them with much less power required to operate the chip.

(c) *hybrid* VLSI networks using pulse-frequency modulation have been developed because they have been found to be both compact, relying on analog computations, and reliable by using digital signals, which are known to be robust, easily transmitted and regenerated. The technology has been well-developed in the communication field. It should be pointed out that biological neurons signal one another using pulse-frequency modulation.

A human begins to put together relationships between the outside world represented by a visual sense through a physical image projected onto the rear of one's *retina*, which is an array of photosensitive receptors, connected to a neural network in the other layers. These neurons convert an optical image into a neural image for transmission along the optic nerve for final processing into a cognitive "signal" we call sight. The transformation involves three steps:

(a) Phototransduction by a layer of receptor neurons (the rods and cones);

(b) Transmission of those signals to a layer of *bipolar* neurons through synaptic processes;

(c) Passage of these signals for further processing to *ganglion* neurons and thence to the optic nerve.

Figure A-3 is a simplified circuit diagram of the *silicon retina* built by Mead; it is modeled on a portion of the vertebrate retina. The primary signal that falls on this retina proceeds through each photoreceptor and the circuitry representing a bipolar cell (as shown in the inset). The entire image signal is processed, in parallel, at each node of the network. The net result is that the silicon retina generates, in real time, outputs that correspond directly to signals observed in the corresponding layers of biological retinas. These chips demonstrate a tolerance for device imperfections that is characteristic of any collective analog system. They have already been used commercially to read the numerical code printed along the bottom of checks.

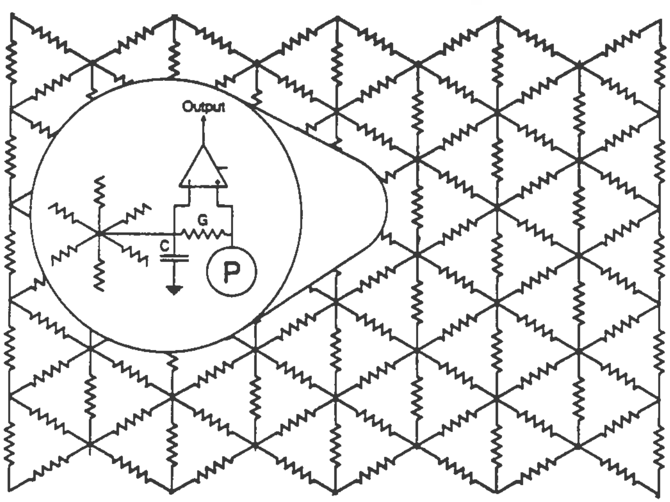

Figure A-3. Simplified circuit drawing of C. Mead's silicon retina. A single pixel element is shown in the circular window. In the diagram *G* is the conductance by which the photoreceptor *P* drives the resistive network; the output results from an amplification of the difference between the photoreceptor's output and the electric potential of the resistive network at the point where it is connected.

Footnotes

1. A. Turing (1936) "On Computable Numbers With an Application to the Entscheidungs-problem." Proc. Of the London Math. Soc., **42**, 230-265.

2. K. Gödel: Collected Works (1986), S. Feferman et al, eds. New York: Oxford Univ. Press.

3. See M. A. Arbib (1987) *Brains, Machines, and Mathematics*. (2nd Ed.) New York: Springer-Verlag for an interesting discussion in Chapter 8.

4. H. Dreyfus and S. Dreyfus (1986), *Mind Over Matter*. New York: The Free Press.

5. C. Mead (1987) *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley.

6. W. S. McCulloch and W. H. Pitts (1943), "A logical calculus of the ideas immanent in nervous activity." Bulletin of Mathematical Biophysics, **5**, 115-133.

7. W. Aspray and A. Burks, eds. (1986), *Papers of John von Neumann on Computing and Computer Theory*. Cambridge, MA: MIT Press.

8. An *automaton* is any machine or system that has a set of internal states, which may be concealed, whose output and/or next state is determined not only by the next input but also by its current state.

9. See, for example, Y. -H. Pao (1989), *Adaptive Pattern Recognition and Neural Networks*. Reading, MA: Addison-Wesley.

10. B. F. Skinner (1957), *Verbal Behavior*. New York: Apple-Century-Crofts. For a review, and rebuttal of Skinner's verbal behavior, see N. Chomsky (1959), Language, **35**, 26-58.

11. W. James (1892/1984), *Psychology: Briefer Course*. Cambridge, MA: Harvard Univ. Press.

12. J. A. Fodor and Z. W. Pylyshyn (1988), "Connectionism and cognitive architecture: A critical analysis." Cognition, **28**, 2-72.

13. J. von Neumann (1958), *The Computer and the Brain.* New Haven, CT: Yale University Press.