THE FUTURE OF PARALLEL, ANALOG AND NEURAL COMPUTING ARCHITECTURES IN THE GRAPHIC ARTS

Richard A. Holub
EIKONIX - Advanced Electronic Systems
Electronic Pre-Press Systems, Eastman Kodak Company

Continuing automation requires that we confer some distinctively human capabilities on the machinery of production. Striking advances in microcircuitry, Neuroscience and parallel processing technology will make this increasingly possible in years to come. This paper will *not* attempt a comprehensive survey of the subject. Rather, it will begin by describing several applications simulating human intelligence which demand unconventional computing. The applications will serve to introduce some approaches to the study of new computing architectures; in reviewing them we will try to distinguish among the terms "parallel", "analog" and "neural". Computational properties of the neuronal synapse will be outlined. Then a simple, hybrid, electronic model of neuron-like computing element will be presented. Important aspects of its performance will be compared to data from biological neurons involved in visual information processing (single cells in cat visual cortex). Then the question of how large numbers of the computing elements might be organized will be revisited. The paper will conclude with a consideration of some of the hurdles which must be passed to bring the new technologies out of the laboratory and into cost-effective commercial applications.

As the industrial revolution concludes in bigger and better bombs, an intellectual revolution opens with bigger and better robots. The former revolution replaced muscles by engines and was limited by the law of conservation of energy, or of mass-energy. The new revolution threatens us, the thinkers, with technological unemployment, for it will replace brains with machines limited by the law that entropy never decreases.

- W. S. McCulloch

INTRODUCTION

I have chosen an opening quotation from one of the remarkable papers (McCulloch, 1948) of the Hixon Symposium, held in 1948, the year in which TAGA was organized. The author, with W. Pitts, was among the first to model nerve cells (neurons) as quasi-digital logical computing elements. That model is now considered naive because of what has been learned about how the brain functions. However, the recent explosion of interest in "neural networks" is due as much to advances in circuit miniaturization as to progress in Neuroscience. Yet a third factor must be considered, namely efforts in artifical intelligence laboratories to teach computers to perform tasks which humans take for granted, such as seeing.

The realization of just how difficult it is to build general, robust vision machines has filled a whole new group of scientists with awe at how remarkable the human brain is. The inability of a conventional, "von Neumann" machine to perform even "elementary" visual tasks in real time has lead many computer scientists to consider what is known of how the brain is put together and to fashion some aspects of new computer architecture after it.

von Neumann's name has become associated with a computing architecture in which one (or a small number of) processors execute our programs although it may have many slave processors performing low level functions. The logic of these processors is digital, or two-state (in contrast to analog circuits which may represent signals in a continuous way) and they process algorithms sequentially.

von Neumann (1948) also participated in the Hixon Symposium and realized that the digital method used in computing might be very unlike neural computation, which was characterized by ample redundacy (fault tolerance) and adaptiveness (self organizing properties). His contribution to the symposium was also remarkable because it outlined the proof of a theorem which stated that a processor of sufficient complexity could contain a program for building processors more sophisticated than itself. This explains the opening quotation of McCulloch.

Figure 1 is a photomicrograph of monkey retina prepared by the author almost 20 years ago. Key aspects of biological computation are suggested by it. It is not enough to point to large numbers of cells and argue that each might be a processor. Digital circuits may eventually have as many gates or logic circuits even though large subassemblies of the gates would not necessarily constitute a processor. What suggests that something different might be going on in retina is the proliferation of wires or connections among cells (see Figure 2 as well).

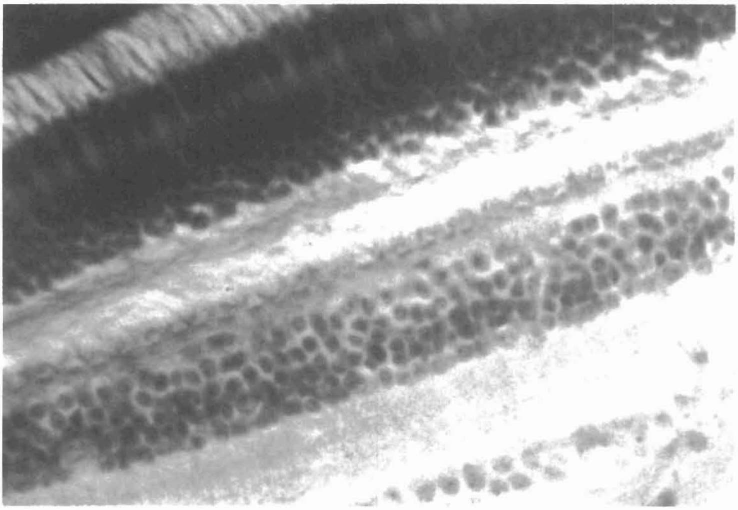

FIGURE 1: Cross-section through the retina of Rhesus macaque monkey at 400 magnifications. Upper-left: rod and cone photoreceptors; lower-right: cell bodies of ganglion cells (output of retina). In between are two layers of nuclei ("inner" and "outer") and of interconnecting processes (transmission wires) among cells. The latter are called inner and outer plexiform layers.

Carver Mead, a pioneer in Very Large Scale Integrated circuitry has described (1987) a design for a chip intended to simulate some retinal properties. He commented that only recently has the analog nature of the processing carried out in retina been recognized and come to serve as a model for electronic designs. In fact, it has been known for almost twenty years that processing in retina is entirely analog and that spatiotemporal convolutions are computed at individual synapses (junctions among cells), instantaneously, by virtue of the extensive connectivity. We will show evidence for comparable processing in visual cortex and present a model for how it is implemented.

However, the thesis of this paper is that hybrid circuitry most properly simulates neural information processing, which, everywhere outside the retina, has both analog and digital components. It is only recently that hybrid silicon microcircuitry has become available, making truly neural simulations feasible.

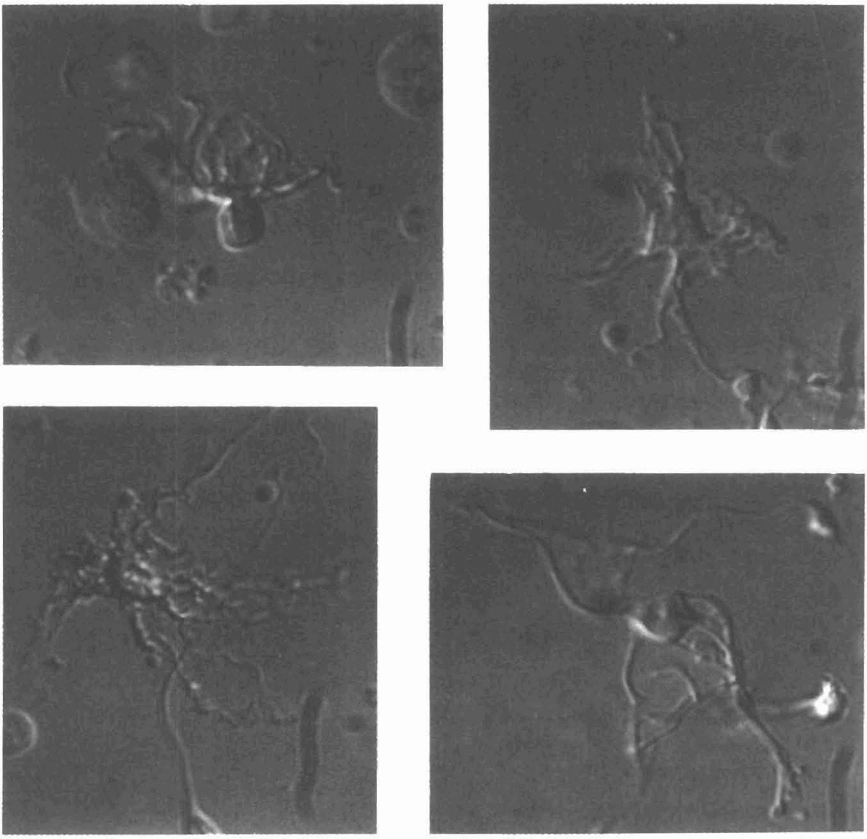

FIGURE 2: Shows a.) an isolated cell whose processes reach up into outer plexiform layer (closer to receptors); b.) an isolated cell, two of whose processes reach down into inner plexiform layer (closer to ganglion cells); c and d.) two isolated ganglion cells. All cells were photographed at about 600 magnifications using phase contrast microscopy. The cells were alive and dissociated from a turtle retina in tissue culture medium.

In summary, a retina contains millions of analog, parallel processors. Elsewhere in the brain the units are more like hybrid processors. Before considering circuit models of a synapse, we will cite several visual (broadly defined) tasks of industrial relevance which are beyond the abilities of contemporary, von Neumann super-minicomputers. These applications are machine simulations of binocular fusion (stereo vision) and color constancy and the transformation from color 3-space to 4-colorant specifications. These are instances of problems in which n-1 dimensional data must be combined with physical constraints to arrive at n dimensional solutions.

Stereo Vision

From two-dimensional images on our retinas, we construct three-dimensional perceptions of the world. One strategy for doing so exploits the fact that features of external objects at different distances from us than the fixation point are imaged at different distances on our retinas from the image of the fixated point. This "disparity" in the two retinal images can be used to compute depth relationships. I will briefly describe a machine vision solution (see Marr, 1982) to this problem and its implementation on a new computer developed at MIT and at Thinking Machines Corporation (Hillis, 1985).

Computing depth relationships is relatively simple once points in the two retinal images which correspond to the same point in the world are identified. The latter involves computing image features, such as the sign at a given pixel after suitable spatial filtering of the image. Next, image feature maps for the two cameras (eyes) are translated back and forth across one another in order to see which alignments produce matches between like features in the two images - possible matches are tallied. One-dimensional, horizontal translations are assumed adequate because the eyes scan horizontally.

Lastly, false matches are eliminated and correct ones selected by enforcing physical constraints and consistency checks, i.e., a match is enhanced if it is consistent with a match between neighboring features. A match is rejected if it is inconsistent with knowledge or assumptions about the relationships among object surface points in the real world. Such an algorithm results in disparity maps which can be used to interpret the depth relationships of objects in a scene and requires 10-15 minutes of computation for a single image on a VAX. Drumheller and Poggio (1986) report run times of the order of 1.5 seconds on a Connection Machine^{TM}. This computer can be viewed as a two-dimensional array, or mesh, of one-bit processors, each having some number of one-bit "words" of memory. The column vector of memory cells associated with each processor constitute a third, spatial dimension. For the parallel stereo implementation, a 256X256 array of processors was used, each with 4096 bits of memory (see Figure 3).

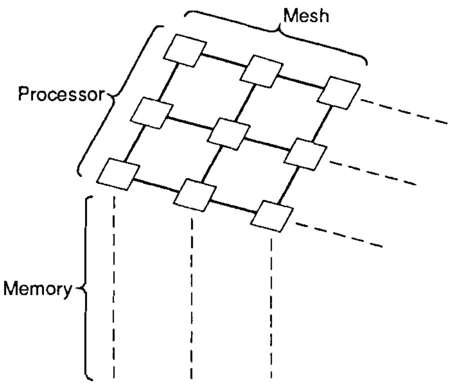

FIGURE 3: Schematic of topology of processors and memory in the Connection Machine™, manufactured by Thinking Machines Corp.

Neighboring processors communicate with one another by direct links, although a router is also available for communications between processors at a greater distance in the mesh. A host computer broadcasts instructions which local one-bit processors may execute or ignore, depending on what's in local memory. It is apparent that left and right images, in the stereo example, can be stacked in memory, with a single processor devoted to each image pixel. It is easy to see how filtering (i.e., convolution with a spatial weighting function), which is a local operation replicated over the entire image, can be implemented in parallel on such an architecture. The value of a pixel simply becomes a weighted sum of neighboring pixels once or twice removed in each direction. However, unlike a synapse, where weighted inputs are combined simultaneously, neighboring inputs are cumulated sequentially in this architecture.

Horizontal shifts of images are readily accomplished using interprocessor links and potential matches can be tallied in memory planes below the images as they are found. In other words, a set of bit planes in memory can be treated as a disparity axis projecting through the image coordinate plane. Functions embodying the physical constraints and consistency criteria can also be mapped onto other available bit planes of memory and used by the processors to identify locally good matches and disparity values. The gain in speed is tremendous because each small region of the image is treated in parallel. The Connection Machine is more closely related to a network of McCulloch-Pitts neurons than to a contemporary neural network because processing is entirely digital.

Color Constancy (Chromatic Adaptation)

It is a matter of common experience that we use the colors of objects as an aid in identifying them. Not so obvious, perhaps, is the fact that a building we call red at noon presents us with very different, spectrally, stimulation near sunset when we also call it red. This is not a matter of semantics; an observer who has never seen the building before at either time of day will use appropriate color names at each time and will probably recognize the structure in one viewing condition, having seen it in another.

Teaching a machine to exploit color in recognizing objects under varied illuminations could benefit from an understanding of how humans do it. In the Graphic

Arts, we would like to have machines which can not only "factor out" the illumination of the input, but factor back in the viewing conditions of the final copy on output. There are several schools of thought regarding chromatic adaptation, those of Hunt and Nayatani and their respective co-workers and Land and his co-workers. Land's Retinex model (1983 version) will be reviewed here because it has been mapped onto a parallel processor (Hurlbert and Poggio, 1987) and comparative execution times are available for both parallel and traditional processor implementations.

Land's notion is that approximate color constancy can be achieved by a mechanism having three channels, each of which compute the average relative reflectance of each point, i, in the image. The relative reflectance, r(i,j) at i is defined in terms of reflectances relative to neighboring points, j, in the same chromatic channel:

$$r(i,j) = \sum_k f(I_k)$$

where

$$f(I_k) = \log \frac{I_{k+1}}{I_k} \qquad \text{if f > a threshold}$$

and

$$f(I_k) = 0 \qquad \text{otherwise}$$

The k's represent image points sampled along the path between points i and j on the image plane. The use of the threshold is intended to filter out the effects of slow gradients in the illumination and restrict consideration to larger regional changes which are presumed to signal changes in the reflectance of the objects viewed.

The color perceived by the mechanism, then, is given by three values of the average relative reflectance, one for each chromatic channel. Each value is merely the average of relative reflectance values of i with respect to a suitably large (hundreds) number of j's taken over a wide extent of the image. Hurlbert and Poggio (1987) reformulated this problem in more general terms. By their view, what the sensor sees is

$$s(x,y) = r(x,y) + e(x,y)$$

where s is the log of image irradiance, r, the log of the surface spectral reflectance and e the log of the illumination intensity and x and y are the coordinates of the point in the image.

On the assumption that r varies much more rapidly than e (has significantly higher spatial frequency content), the machine, or organism, can use this relationship to recover the "constant colors" of objects, given by r. But they went further than this; they sought to teach a Connection Machine to extract r, given s. They presented the machine with a large number of corresponding r and s values for each point in the image plane for each of three channels. All the pairs of values can be viewed as a pair of "training vectors" **S** and **R** whose relationship is given by the linear matrix **L**:

$$\mathbf{LS} = \mathbf{R}$$

L is what the machine learns by "experience" with the training vectors.

The elements of **L** have the following interpretation: They are the weights to be applied to the activity of some jth pixel in computing its influence on the pixel, i, whose color percept is being computed. In other words, the matrix, **L**, is a large (in spatial extent and numbers of elements) convolution mask which gets moved over the image. The learning of **L** requires 20 hours for a Symbolics 3640 computer (a conventional, serial computer optimized for the LISP language), but less than 10 minutes on a CM-1 Connection Machine.

The form of the learned filter, **L**, is suggested by Figure 4, adapted from Hulbert and Poggio. It can be modelled as the difference of excitatory and inhibitory Gaussian weighting functions, in which the sigma of the inhibition is much greater. This functional form is very typical of biological receptive fields, except for the unusually large extent of the inhibition. A "receptive field" maps a nerve cell's relative sensitivity to points in visual space. In the frequency domain, the function depicted in Figure 4 represents a bandpass filter which rejects very slow changes (such as those due to gradients of illumination, e, in equation 4) and very rapid changes due to noise. The purpose of Land's

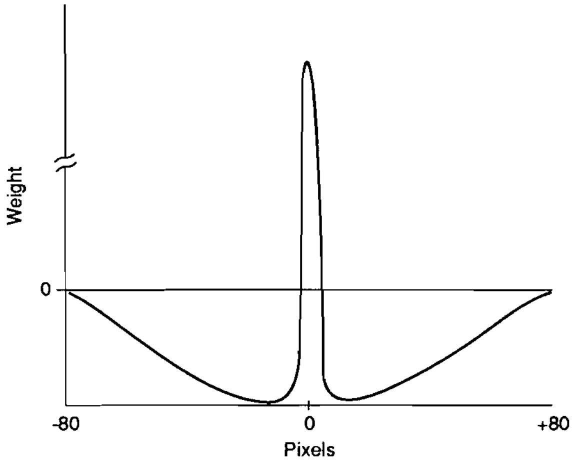

FIGURE 4: A schematic depiction of the color constancy filter learned by a computer in Hurlbert and Poggio's work. The pixel at relative position 0 is the one whose "color percept" is being computed. Plotted are the weights applied to the activity of neighboring pixels, as a function of distance from 0, in performing the computation. The central excitatory region of the plot is even narrower than depicted.

threshold mechanism is served by the bandpass characteristic of this model.

Color to Colorant Transformations

We have dwelt on this problem elsewhere (Holub, Pearson and Kearsley, 1988) as one involving the inversion of equations giving color as a function of colorant. In 4-color printing, the inversion must be constrained by means of UCR, GCR, and heuristic considerations. We merely mention, here, that this problem is conceptually similar to the extraction of 3-D depth information from 2-D retinal image information with the help of constraints. The problem is computationally very intensive on conventional, serial processors, but might be handled expeditiously by networked processors of the sort applied to the stereo and color constancy problems.

ELECTRONIC AND NEURAL CIRCUIT ANALOGIES

In this section I will discuss some electronic circuit models for small nerve cell assemblies. The circuits are models at the level of functional and organizational properties, not necessarily at the level of physiological mechanisms. Any adequate model is *hybrid*, incorporating digital and analog attributes. As we noted above, with Figures 1 and 2, neurobiological systems have numerous connections among cells. There are estimates (Resnikoff, 1986) of between 10 and 16 inputs to the average cell, probably almost an order of magnitude greater than in conventional electronic devices. This is feasible, in part, because there is only one wire per processor. Except within retina, the wires in the nervous system transfer information using a serial pulse code; this is the quasi-digital aspect of biological computing which has been known for over 50 years.

There are two reasons why transmission by a serial pulse code is inevitable in biological systems: 1) Biological wires are filled with salt solution and travel through salty body fluids. Therefore, the capacitance along the wire's insulation is considerable, making the wire act like a filter which only transmits high frequency information. 2) If this weren't enough, two target cells contacted by branches of the wire of different length would see different potentials if there were not a uniform pulse code for the potential level of the transmitting nerve cell. Accordingly, the transmitting neuron functions as a voltage to frequency converter and the wire has mechanisms for sustaining regenerative potentials known as "spikes" or action potentials.

In silicon devices, "poly" (polysilicon), diffusion and metal wires, the building blocks of diodes and transistors, travel in a non-conducting substrate. Wire capacitance is not so significant a problem as it is biologically. Nevertheless, for moderate to long runs it is not negligible and these circuits might be less prone to cross talk and might have better bandwidth if a pulse code were used for interdevice communication.

Coon and Perera (1987) have discussed very low temperature experiments with "injection mode" diodes which show linear pulse coding of applied voltage very similar to a nerve cell's. In effect, they have found very compact voltage-to-frequency conversion devices. They argue that such circuits may provide more nerve-like implementations of "neural" computers than analog networks of operational amplifiers used by other modellers. I agree that this is true of the transmission aspect of nervous system function, i.e., the aspect in which one processor relays the results of its computations to another. However, the computations themselves are analog, as described in what follows.

The wire from one biological processor contacts another processor at a synapse. Biological synapses can perform addition, subtraction and multiplication. A functionally equivalent circuit to the synapse is a D/A converter. Figure 5 is a schematic of an operational amplifier wired to add or subtract digital code words. The op amp sums, more or less linearly, the weighted (by the resistor values) contributions from each bit position of the code word. The switches indicate that a bit can be on or off (multiplying the resistive weight by 0 or 1). Commercially, the weights increase by factors of two from least significant to most significant bit.

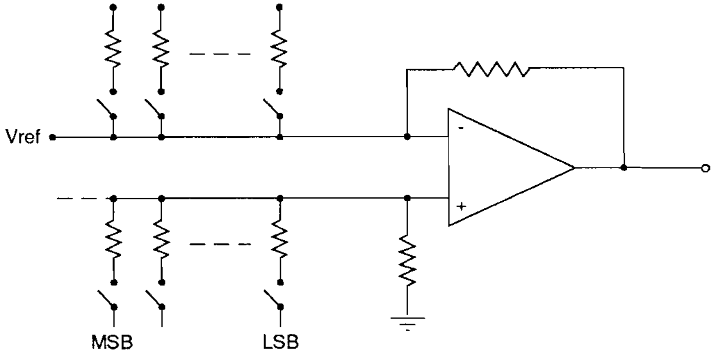

FIGURE 5: Schematic of a D/A-like circuit which sums inputs like a synapse. "Inhibitory" inputs may be applied. The schematic is conceptual; a D/A wouldn't be implemented this way.

However, binary weightings are not necessary. One may easily imagine individual members of an array of photodiodes connected with arbitrary positive (or negative) weights to inverting (or non-inverting) inputs of the D/A so as perform spatial convolutions in the image plane in a completely parallel way. In contrast to the Connection Machine, the only digital aspect of this arrangement would be the quantization of the image plane. We will see that this is, functionally, a very good model of the formation of visual "receptive fields", functions defining the relative sensitivity of visual cells to regions of external space.

Instantaneous multiplication of two signals may be implemented in commercial D/As by allowing the reference voltage shown in Figure 5 to vary. The output is the product of V_{ref} and the digital code word, which is some fraction of 1. A biological analogy for this is the well-documented phenomenon called presynaptic inhibition. In it, one cell modulates the drive which a second cell applies to a third, multiplying a single input to a nerve cell by a variable fraction. The geometry is suggested by Figure 6.

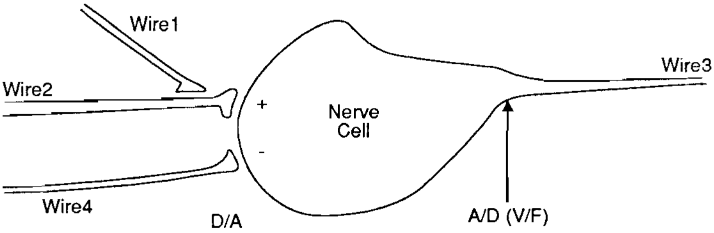

FIGURE 6: Drawing shows synaptic contacts, excitatory (+) and inhibitory (-) made by "wires" from cells two and four onto nerve cell three. Cell One exercises presynaptic inhibition over the input from two to three and is capable of silencing it altogether. D/A conversion occurs at the synapse while the output of the summing element (nerve cell) is subjected to a form of A/D conversion known as Voltage- Frequency Conversion.

The single input should not be construed as a single bit although modellers of two-state neurons think of it in this way and this is the situation in Connection Machine. Rather, the digital word that is transmitted down a wire arrives at the synapse for D/A conversion as a serial bit stream. One can give the word great precision in terms of number of bits, but only at the price of lowering the number of words per unit time. In any case, presynaptic inhibition variably scales this word. Figures 5 and 6 show only one excitatory and one inhibitory input to the summing element; obviously, this simplification is not necessary.

One last point should be made regarding bandwidth/resolution trade-off. Biological wires can support at most 1000 impulses/sec. A more reasonable figure for most cells in mammalian central nervous system may be 500/s. If it were desired, with consideration to Nyquist sampling, to transmit a signal band-limited to 10 Hz or less, we would get about 25 levels per sample, or 4-5 bit resolution. These numbers give an approximate sense of capability of one biological processor.

Voltage-to-Frequency conversion is a very slow method of performing A/D, but it has the advantages of noise immunity and of being adaptive with respect to the resolution/bandwidth product. It also requires only one transmission wire, making redundancy more feasible. Silicon implementations could support a much higher resolution/bandwidth product, possibly compensating for lower densities of devices per cubic mm.

The thoughts outlined in this section were conceived several years ago in connection with the interpretation of data to be summarized in the next section. Circuit models of the sort considered will be shown to simulate some physiological properties. It is of interest to note that a group at MIT's Lincoln Laboratories (Mann, Lippmann, Berger and Raffel, 1988) have made a self-organizing "neural chip" based on circuit constructs not unlike some of those outlined above. The device is to be used in work on machine speech recognition and acquisition.

SOME VISUAL DATA

In the introductory section on stereo vision, it was noted that the first step in fusing images from the two eyes was the computation of features. This involves filtering the images of noise and producing "markers" of significant events in the image such as a change in the sign of a convolution result. Some years ago, the author (1977) conducted a study of individual nerve cells in the visual area of the cat brain to see if they acted as linear, spatio-temporal filters. In other words, were the cells bandpass filters in the frequency domain and were they sufficiently linear to support predictions of their responses to various visual patterns? And, what was the nature of their nonlinearities?

We began by verifying that, for at least a subset of cells, the responses preserved the waveform of a sinusoidal input and increased in direct proportion to the amplitude of the driving sine wave. Very few cells exactly preserved the input; rather, most clipped or half-wave-rectified the input due to the absence of a maintained activity level in the absence of stimulation. In spite of a rectifying nonlinearity, we examined adherence to the superposition principle with the following experiment. We stimulated the cell with a moving spatial step (boundary between regions of homogeneous light and dark) or impulse (narrow slit or bar) functions. Elongated stimuli are appropriate for cortical cells because their sensitivity profiles have a long axis of bilateral symmetry.

We Fourier transformed the derivative of the response to the step or the response to the impulse itself (since these should be the same in a linear system) for comparison to the responses to an equal energy spectrum of sinusoidal spatial frequencies presented one at a time, moving at the same velocity as the composite edge or slit stimuli. In other words, we measured something like a spatio-temporal Modulation Transfer Function for the cells directly. This was compared to an estimate derived by Fourier transformation of the response to a stimulus having an equal energy spectrum, namely the impulse. Figure 7 shows the results for one cell, where data points represent measurements with individual spatial frequencies and the solid function the transform of the derivative of the step response.

We rescaled the data points (in general, we did not need to resort to arbitrary scalings exceeding a factor of two) to have the same peak amplitude as the transform and restricted our attention to the amplitude component of the complex transform. The two

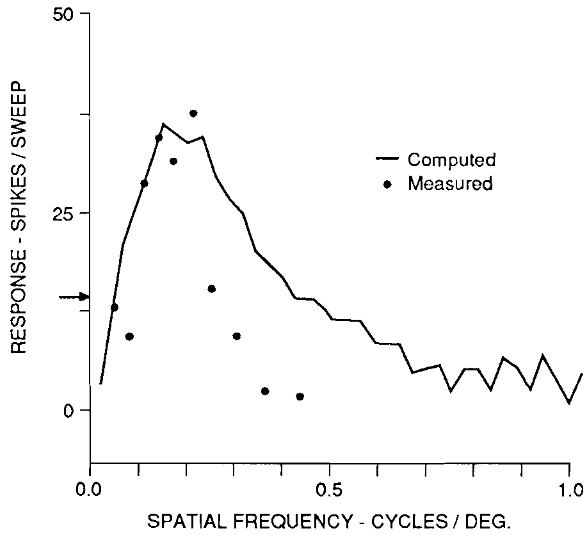

FIGURE 7: Plot of "Spikes" (pulses in the uniform pulse code transmitted by nerve cells) per stimulus sweep vs. spatial frequency. Compares measured and estimated (by Fourier transformation of differentiated step response of cell) Modulation Transfer Function of cell.

functions differ most markedly at high spatial frequencies, suggesting a nonlinearity which introduces high spatial frequencies into responses, as would rectification. Nevertheless, the most effective spatial frequency and the amplitude of response at that frequency were reasonably well predicted and the cell can be considered an approximately linear filter. It is also clear that the cell acts like a bandpass filter, rejecting very low and very high frequencies.

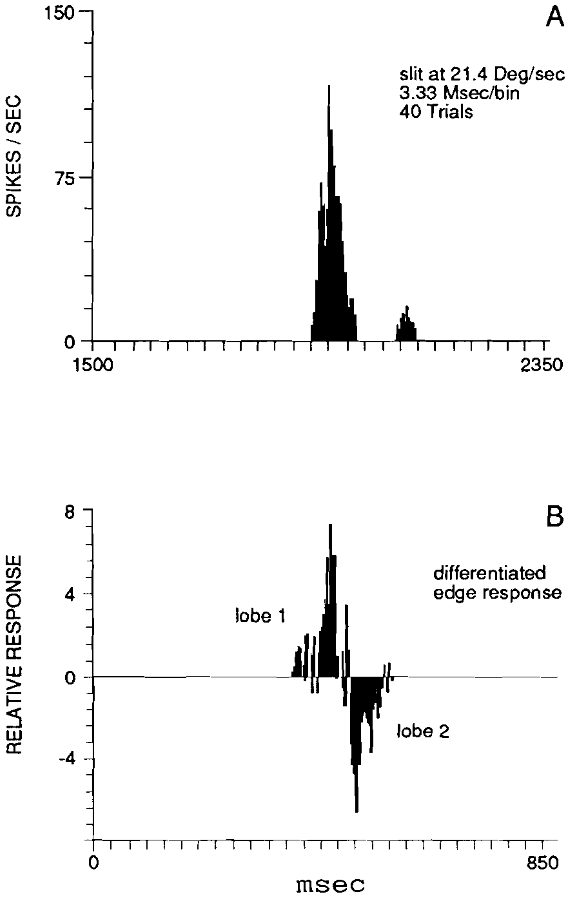

FIGURE 8 A: Pulses per second vs. time. Response of nerve cell in visual cortex to slit (elongated spatial impulse) stimulus moving across its receptive field at velocity of 21.4 degrees visual angle per sec. B: Time derivative of response of same cell to spatial step input, moved at a different velocity.
Although the plots are functions of time, they can be viewed as a map of the relative sensitivity of the cell to a slit stimulus presented at different positions in visual space. In B, note the "zero-crossing" between lobes 1 and 2, where the sign of convolution result changes; this is the kind of feature exploited by Drumheller and Poggio for stereo vision by machine.

We examined these notions in more detail. Figure 8a shows the response of another cell to a moving slit (impulse) and 8b shows the derivative of the response to the step moved at a different velocity. a and b are similar except that the negative lobe seen in the derivative is missing in the impulse response due to rectification. Figure 9 shows four MTFs; data points denote measurements for individual sinusoidal inputs, the solid curve the transform of the derivative of the step response, the coarse, dashed curve the transform of the impulse response and the fine, dashed curve the transform of the impulse response "corrected" for rectification. Figure 9 shows disagreement between measurement and both step and impulse transforms at low and at high frequencies. However, fudging in a negative lobe to correct for rectification produces a much closer correspondence between measurement and impulse transform. The time axis in plots of step or impulse responses can be related to the spatial abscissa in frequency domain plots by the velocity of movement of the stimulus patterns.

The approach employed for estimating the unrectified impulse response (addition of a heuristic, negative lobe or inhibited sideband to the receptive field) cannot be used to decide if the discrepancies observed with edge data are likely to result from rectification. To examine this possibility, we assembled an active, second-order, bandpass filter whose temporal characteristics roughly approximated those of a cortical neuron except that its center frequency (530 Hz) was about two decades higher. The circuit has a gain of 2.0, Q of 0.82 and bandwidth of 1.75 octaves.

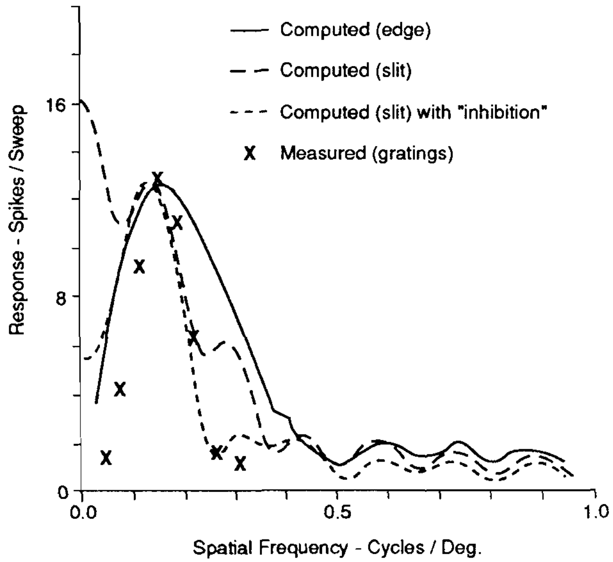

FIGURE 9: Data points represent the spatio-temporal "MTF" measured with stimuli which were sinusoidal in one spatial dimension and uniform in the other. Each individual spatial frequency was moved over the neuron's receptive field at a velocity common to all; it is the velocity of drift which relates abscissae of time domain plots to those of spatial frequency domain plots. The solid curve denotes the transform of the derivative of the response to a moving step (light-dark edge), the dashed curve, the transform of an impulse (slit) response and the finely dotted curve, the transform of the impulse response "corrected" for rectification. It is legitimate to plot impulse and step transforms in the same plot even though the stimuli were moved at different velocities (cf, author's dissertation).

We compared MTFs of the responses of the filter to a step both at the input and the output of a diode (half-wave rectifier).

As a check on our procedures, we measured the responses of the filter to sinusoidal inputs. The MTF measured in this way and the transform of the

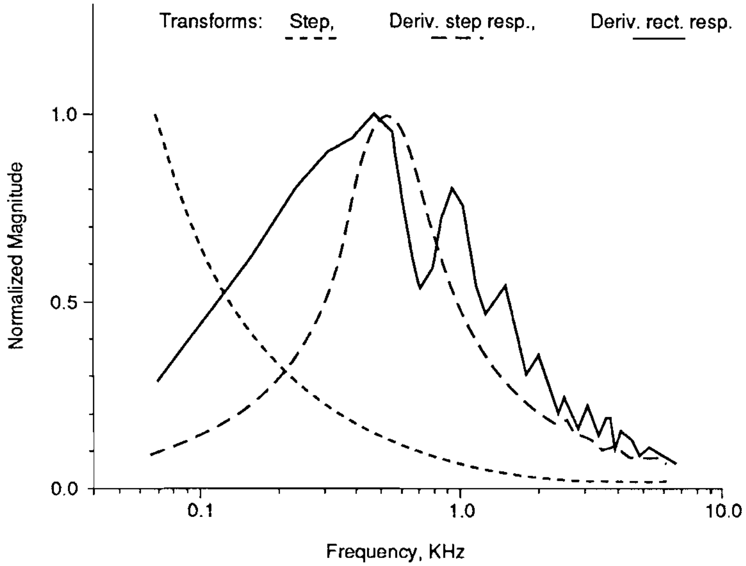

FIGURE 10: All functions are normalized to 1. The finely dashed curve represents the magnitude component of the Fast Fourier Transform of the the temporal step response of the electronic filter described in the text. The coarse, dashed curve is the FFT of the derivative of the step response just described; this curve does not vary with the frequency of sampling (for A/D conversions) if the rate exceeds 12.8 KHz. The solid curve is the FFT of the derivative of the rectified response to a step digitized at a rate of 20 KHz; the transform depends on the sampling frequency.

differentiated, unrectified, edge response (shown by the dashed curve in Figure 10) agreed perfectly. The solid curve represents the transform derived from the rectified response to the edge sampled at 20 KHz. The high frequency limb bears considerable resemblance to Figure 7 supporting the notion that high frequency contamination of step transforms was also due to rectification. The exaggeration of low frequencies (see Figure 9) is suggestive of aliasing, an idea we tested by varying the frequency at which we sampled the output of our test circuit.

Indeed, transforms computed from rectified responses sampled at 12.5, 20.0, 25.6 and 50.0 KHz changed in shape as the sampling frequency increased (becoming more like those of unrectified responses), while those of unrectified rfesponses did not. By analogy to the electronic model, we posit that rectification which occurs in transmitting the pulse train engenders high frequencies in the transmitted signal which exceed the resolution/bandwidth product of the nerve cells. As a result, the signals become aliased and the low frequency portions of MTFs depart from what would be expected for truly linear filters.

In summary, the visual cells considered here are approximately linear. Their receptive fields to moving step or impulse stimuli represent a convolution of the image with a spatial weighting function which may be approximated by a difference of Gaussians. Accordingly, in the frequency domain, the cells are bandpass filters whose characteristics can also be modelled as a difference of Gaussians. Thus, the receptive fields have "antagonistic" excitatory and inhibitory zones and are characterized by zero-crossings, or changes in the sign of convolution (Figure 7b). These events are precisely the features or markers alluded to in the introductory section on stereo vision. Likewise, an inhibitory region of large extent surrounding a narrow excitatory region was the model used in retinex calculations by Hurlbert and Poggio (1987).

SUMMARY AND CONCLUSIONS

We distinguished among analog, parallel and neural processors. The Connection Machine^{TM} (of Thinking Machines Corporation) was cited as an example of a parallel, but entirely digital processor. Since all processors work on a given, broadcast instruction in tandem, the Connection Machine is classified as a SIMD architecture, or Single Instruction Multiple Data machine. I am aware of only one commercial installation of Connection Machine (Computer Pictures, May 1987, p. 10). General acceptance of this kind of architecture is probably ten or more years off, since expensive host computing environments and specially trained programmers are necessary to realize its potential.

Other entirely digital parallel processors include the Butterfly^{TM} (of BBN Advanced Computers, a subsidiary of Bolt, Beranek and Newman, Inc.) and a number of Inmos, Inc. Transputer-based processors. R. Cok (1987) of Kodak Research Labs has discussed one such implementation. The latter two machines are examples of MIMD (Multiple Instruction Multiple Data) architectures, in which there are far fewer processors than in the Connection Machine, but in which each processor is relatively powerful and capable of executing its own program independently of the others though they may all be working on a common problem. We can anticipate significant commercialization of this technology within the five-year frame. The most significant barrier to its use is the training of people in programming parallel computers.

As with digital technology, analog computers can be based on single or multiple processors. Examples of single-processor analog computers which are well known within the printing industry are the colorimetric separation machine built by Hardy and Wurzburg (1948), also in the year of TAGA's convening, and the older rotary drum scanners. It might be argued that these machines were really multiprocessors, but I would prefer to think of them as such only to the extent that they processed several channels simultaneously. The computations these processors were specialized to carry out were complex, compared to those of a synapse, but they are correspondingly less general.

I am unaware of any parallel, entirely analog processors which are beyond the prototype stage. There is a very active literature on the modelling of analog networks which we have not reviewed for reasons of time and space. A significant contributor to this literature is S. Grossberg (1982, Cohen and Grossberg, 1983) of Boston University. Many of the theorems he has proven about the connectivity and functional stability of large assemblies of networked processors apply to hybrid or neural computing devices as well as purely analog ones. Network algorithms generally involve parallel implementations of "steepest descent" search algorithms which follow the gradient of an error metric, or cost function, of several variables to a minimum value. It is hoped that this minimum is a solution to the problem in some global sense.

The same algorithms have long been in use on serial processors where they execute very slowly. Hopfield (1985) published an interesting simulation of an analog network solving a classic minimization problem, that of the Travelling Salesman. Given a map, the problem is to find the shortest circuit in which a salesman visits each of a number of cities and ends up at the starting point. Endowed with the correct connectivity matrix (analogous to the **L** matrix used in the color constancy example), a good starting point and appropriate "cost" or error weighting functions, the simulated network was able to "look at" a map and quickly chart an efficient touring of a set of cities. I cite this as another problem analogous in various ways to finding a best solution to the realization of a target color with a set of colorants.

Properly neural processors (e.g., Mann, et al., 1988) are just emerging from the laboratory. For applications of sufficiently limited scope, these may be in fairly general use in the five- to ten-year time frame. Since these devices may be largely self-organizing (capable of learning) they may not require extensive programming. Nevertheless, interfacing these devices to host processors may be very challenging.

ACKNOWLEDGMENTS

The author thanks M. Liberacki, P. Norton, K. Penders and M. Taylor for skilled assistance in preparing the manuscript. He also is grateful to R. M. Jones and M. Morton-Gibson for valuable discussions and help with some of the experiments. Parts of this work constituted part of a dissertation submitted by the author to the University of Wisconsin in partial fulfillment of the requirements for the Ph.D. degree. Some phases of this work were supported by the following grants from the National Institutes of Health: 4FO1MH48872 (to the author), NS06225 and EY00308 to the Depts. of Neurophysiology and Ophthalmology (U.Tulunay-Keesey) at the University of Wisconsin and EY01812 to P. O'Bryan at Boston University.

LITERATURE CITED

Cok, R. S.
1987 "Case study: multiprocessing with transputers", *Electronic Engineering Times,* March 17, 1986, iss. no. 375, pp. T15-T16.

Cohen, M. A. and Grossberg, S.
1983 "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks", *I.E.E.E. Transactions on Systems, Man and Cybernetics,* vol. 13, pp. 815-826.

Coon, D. D. and Perera, A. G U.
1987 "Semiconductor electronic concepts for neural network emulation", *Int. J. Electronics*, vol. 57, pp. 61-69.

Drumheller, M. and Posggio, T.
1986 "On parrallel stereo", *IEEE Conf. Robotics, and Automation,* San Francisco, CA; April 1986, pp. 1439-1448.

Grossberg, S.
1983 *Studies of Mind and Brain,* (D. Reidel, Dordrecht).

Hillis, W. D.
1985 *The Connection Machine,* (MIT Press, Cambridge, MA).

Holub, R. A., Jones, R. M. and Gibson, M. M.
1977 "Fourier analysis applied to spatial properties of neurons in the cat's visual cortex", *Neurosci. Abs.,* vol. 3, p. 563.

Holub, R. A., Pearson, C. H. and Kearsley, W.
1988 "Color systems calibration for Graphic Arts: II. Output devices", *J. Imag. Technol.,* vol. 14, to appear in April issue.

Hopfield, J. J. and Tank, D. W.
1985 ""Neural" computation of decisions in optimization problems", *Biol. Cybern.*, vol. 52, pp. 141-152.

Hurlbert, A. and Poggio, T.
1987 "Learning a color algorithm from examples", MIT Artifical Intelligence Laboratory Memo 909, pp. 29.

Land, E. H.
1983 "Recent advances in retinex theory and some implications for cortical computations: Color vision and the natural image", *Proc. Natl. Acad. Sci.*, vol.. 80, pp. 5163-5169.

Mann, J., Lippmann, R., Berger, B. and Raffel, J.
1988 "A self-organizing neural net chip", *I.E.E.E. Custom Integrated Circuits Conference,* to be held in May in Rochester, NY

Marr, D.
1982 *Vision,* (W. H. Freeman, NY).

McCulloch, W. S.
1948 "Why the mind is in the head", in: L. A. Jeffress, Ed., *Cerebral Mechanisms in Behavior: The Hixon Symposium,* (orig. Wiley, NY, 1951, republished Hafner, NY, 1967), pp. 42-74.

Mead, C. A.
1987 "Neural hardware for vision", *Engineering &*

Science, (Cal. Tech., Pasadena, CA), June, pp. 2-7.

Resnikoff, H. L.
1985 "Concurrent computation and models of biological information processing", *Ann. Mtg. Amer. Assn. Advancement Sci.,* May 1985, Los Angeles, CA.

von Neumann, J.
1948 "The general and logical theory of automata", in: L. A. Jeffress, Ed., *Cerebral Mechanisms in Behavior: The Hixon Symposium,* (Orig. Wiley, NY, 1951, republished Hafner, NY, 1967), pp. 1-41.