Hybrid Analog-Digital Architectures For Neuromorphic Systems

Rodney J. Douglas, Misha A. Mahowald, and Kevan A. C. Martin

Abstract— Signal restoration is necessary to perform computations of significant complexity. In digital computers each state variable is restored to a binary value, but this strategy is incompatible with analog computation. Nevertheless, cortical neurons, whose major mode of operation is analog, are able to perform prodigious feats of computation. Our research on visual cortex suggests that cortical neurons are able to compute reliably because they are organized into populations in which the signal at each neuron is restored to an appropriate analog value by a collective strategy. The strategy depends on feedback amplification that restores an input signal towards a stored analog memory. This principle is similar to recall by autoassociative neural networks. Networks of cortical amplifiers can solve simple visual processing tasks. They are well-suited to sensory processing because the same principle that restores their analog signals can also extract meaningful features from ambiguous sensory input. We describe a hybrid analog-digital CMOS architecture for constructing networks of cortical amplifiers. This neuromorphic architecture is a step towards exploring analog computers whose distributed signal restoration permits them to perform reliably sequential computations of great depth.

I. INTRODUCTION

Computation in physical systems is fundamentally limited by noise [1]. Digital computers attempt to overcome the problem of noise by fully restoring all their binary state variables at each step of a computation. The cost of this strategy is high, because the representational capacity of each state variable is reduced to a single bit. But full restoration at the hardware level is not the end of the problem. Noise enters again at the representational level. Even with

Authors are with the Medical Research Council Anatomical Neuropharmacology Unit, Mansfield Road, Oxford OX1 3TH, England

fully-restored signals, digital computers must use many bits of precision to represent real world data that are often only accurate to about one part in a hundred. They require high precision, not only to accommodate differences of scale, but because computational errors accumulate, or even multiply, in each step of the computation [2].

The brain adopts a different strategy for controlling noise than does the digital computer. Although neuronal action potentials could be thought of as fully restored digital signals, they are not used to represent binary numbers. Instead, the brain uses a hybrid analog-digital signal representation in which trains of pulses transmit analog information in the timing of the events, and these are translated back into analog signals in the dendrites of the neuron. How does the brain cascade large numbers of such imprecise analog processing steps without losing the signal in the noise? One possibility is that the brain is able to restore analog signals by using structured data representations stored in its connectivity. Such a mechanism is illustrated in Figure 1. We call this mechanism a *smart amplifier*, based on the observation that an amplifier can only amplify the signal without amplifying the noise if it already knows what the signal is [3].

II. CORTICAL AMPLIFIER

Our work on the anatomy and physiology of visual cortex has shown that positive feedback is a major feature of cortical processing [4, 5]. Even in the input layer of visual cortex, only 6% of the excitatory synapses are feedforward inputs derived from the principle sensory afferents, while 70% are derived from other intracortical neurons [5]. And we estimate that at least 10% of the excitatory synapses on the input neurons are di-synaptic reciprocal connections from cells of the same class. These feedback synapses alone are capable of driving the neuron to its maximum output.

Simplified neural networks with positive feedback Positive feedback have been studied in the context of content-addressable memory [6] and of computational mechanisms [7], but the behaviour of the more complex cortical networks is still poorly understood.
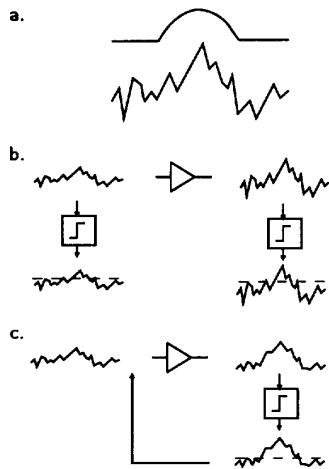
Figure 1: A general method for signal restoration in analog computers. a. An ideal analog signal profile is shown in the upper trace. This signal will be degraded by noise (lower trace) in the process of computation. b. Thresholding (left) is a common method for removing noise. The corrupted input signal from part (a) is passed through a threshold element. The lower trace shows the signal and a superimposed threshold level (dotted line). All the parts of the signal below threshold are multiplied by zero while all parts above are multiplied by one. The result resembles poorly the original signal. Linear amplification of the signal (right) improves the threshold's discrimination of the peak, but it does not improve the quantitative matching of the resulting signal to the original. c. A *smart* amplifier in a positive feedback loop is able to restore the analog shape of the signal because it has stored an analog memory. The parts of the signal that match the memory are preferentially amplified. The amplified signal is separated from the noise by a threshold element.

Our approach to this problem is to explore the behavior of networks of non-linear neurons by making simple circuit approximations (Figure 2).

To understand cortical computations is to understand how reexcitation is controlled to produce selective neuronal responses while maintaining electrical stability in the cortex. Stabilitiy can be analyzed in terms of the abstraction of a cortical amplifier. A cortical amplifier is not a discrete, physical object; rather, it refers to a population of interconnected neurons that are activated together in a particular situation. A particularly simple cortical amplifier consists of a population of identical pyramidal neurons. These neurons receive the same input, they are connected to each other with the same synaptic strength, and they share a common inhibitory feedback signal. Since the outputs of all of the neurons are equal, the network can be reduced to the circuit shown in Figure 2b.

In this abstraction, the various active and passive conductances of the neuron, other than those
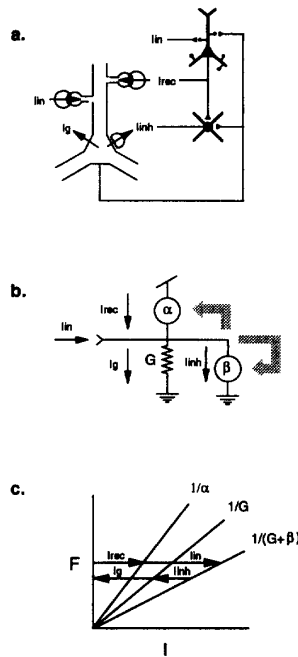

Figure 2: behaviour of the cortical amplifier composed of pyramidal neurons embedded in a network. a. The output of a cortical pyramidal cell drives other pyramidal and smooth (inhibitory) neurons (shown in black). b. The electrical equivalent circuit of the cortical neuron in its embedding circuit. The recurrent excitatory circuit creates a network conductance, α, while the recurrent inhibitory circuit creates another network conductance, β. The effective conductance of the neuron is thus $G_{eff} = G + \beta - \alpha$, and its current gain is $(G + \beta)/G_{eff}$ (see text). c. Current-discharge relations that explain the behaviour of the cortical amplifier. The $1/G$ curve expresses the current-sink, I_g, offered by the passive membrane conductance and the averaged spike conductances. The $1/(G + \beta)$ curve indicates the increased current sink, $I_g + I_{inh}$, caused by inhibition that is proportional to the output of the neurons. The $1/\alpha$ curve expresses the dependence of the feedback current measured in a particular neuron, I_{rec}, on the average output of neurons in the population. The steady-state discharge rate, $F(I_{rec} + I_{in})$, occurs were $I_{rec} + I_{in} + I_g + I_{inh} = 0$ is satisfied.

associated with synaptic inputs, are collected into a single input conductance G. The action potential discharge frequency, F, of a single neuron in the absence of feedback is an analog variable that is approximately linear in the input current $I_g = I_{in}$ [5]. So the rate output of the neuron can be equated with the voltage I_g/G. The output of each cell contributes some excitatory current to each of its fellows, and receives from all of them a total feed-

back component I_{rec}. All the pyramidal neurons are the same, so the feedback current to each neuron is proportional to its own output, $I_{rec} = \alpha F$, and so creates an effective conductance, α, which we call a *network conductance*. Since the excitatory feedback is positive, the excitatory network conductance is negative. By a similar argument, the feedback inhibitory connections contribute a positive network conductance, β.

Conservation of current dictates that in the steady state the sum of the excitatory currents must equal the sum of the inhibitory currents, as indicated in Figure 2c. The output of all the cells in feedback amplifier are proportional to the input current:

$$\frac{F}{I_{in}} = \frac{1}{G + \beta - \alpha}. \qquad (1)$$

The amplifier is stable in the sense that its response remains bounded without the restraint of saturation provided that $\alpha < (G+\beta)$. Equivalently, the gain of the feedback loop is less than one if perturbation of the output by an amount δF results in a greater increase in current sink than it does in current source, $\delta F(G + \beta) > \delta F\alpha$. Under these conditions, the amplifier will relax in the absence of input current.

Although the feedback loop gain is less than one, the closed loop current gain of the amplifier can be much greater than one. The current gain is the ratio of the total excitatory current to the input current:

$$\frac{I_{in} + I_{rec}}{I_{in}} = \frac{G + \beta}{G + \beta - \alpha}. \qquad (2)$$

As α approaches $G+\beta$, the current giving rise to the output is due mainly to recurrent excitation rather than the input current. If a noise component is added to the inputs of each of the identical neurons in the amplifier, the responses of the neurons are nevertheless nearly equal, since the input current is dominated by the recurrent connectivity pattern, which averages over the noise in the input.

A. Noise Suppression in the Computation of Orientation Selectivity

Computation can be performed by a network of cortical amplifiers in which each neuron participates in the amplification of a whole range of input patterns. We have simulated such a network computing just orientation selectivity, one of the most studied problems in primary visual cortex. The problem is to explain how the cortical neurons acquire an exquisite sensitivity to the orientation of the visual stimuli, when the primary inputs to visual cortex show only a coarse orientation bias at best. The model is illustrated schematically in Figure 3. Geniculocortical connections have imprecise mapping onto
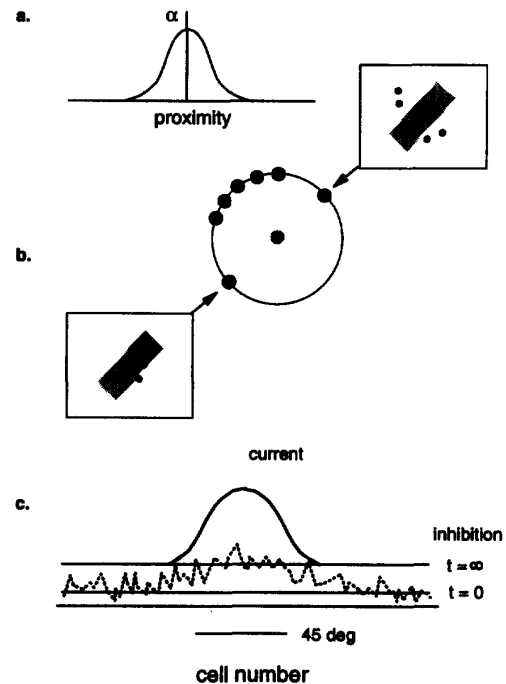

Figure 3: Orientation tuning in the presence of noisy signal and connections. b Forty pyramidal cells (large black filled circles, examples shown) each receive input from a group of LGN afferents (small dark filled circles, examples shown), whose receptive fields are organized along an axis in the visual space (rectangles). A pyramidal neuron is optimally stimulated by a light bar (shaded rectangle) that excites the receptive fields of, and has the same orientation as, the group of afferents that supply it (bottom left). If the orientation of the bar is orthogonal to that of the afferents (top right), then the pyramidal cell receives sub-optimal stimulation and so respond less strongly. Pyramidal neurons that encode similar orientations are reciprocally connected with weights, α_{ij} that are a gaussian function of the similarity of their preferred response, as indicated by ring connected neurons and connection function on the ring (part a.). c. Geniculate current input to pyramidal populations (dashed line) is noisy because of imprecise connections and intrinsic signal noise. The initial level of inhibitory current ($t = 0$) cannot suppress all of the noisy inputs. Nevertheless, most of the neurons with receptive fields oriented close to the stimulus receive more excitation than inhibition, and the amount of recurrent excitation increases there. The increased activity of preferred neurons increases the average inhibition, and suppresses more of the outliers. Elimination of the outliers improves the correlations amongst the survivors, and enhances their response. After convergence ($t = \infty$), inhibition suppresses all of the cells whose receptive fields are oriented more than 45/2 degrees from the stimulus (solid, guassian-shaped line).

cortical cells [8]. The statistics of this mapping are approximated by assigning geniculo-cortical connection values from a binormal spatial distribution whose principle axis is rotated progressively through 180 degrees, across the 40 cortical cells. Thus each cell responds to a slightly different principle axis. The cortical cells form reciprocal positive connections with each other. These α_{ij} are a Gaussian function of the angle between the preferred orientations of the coupled cells, i and j. In addition, all forty pyramidal neurons stimulate a single inhibitory interneuron, which inhibits them all equally.

In this case, I_{rec} is not simply αF, but the sum over the $\alpha_{ij} F_j$. Furthermore, some of the neurons in this network will be below the threshold nonlinearity. However, the thinking in the previous, simple case provides intuition for the operation of the orientation circuit. The equation that the ith neuron in the network obeys is:

$$F_i = max(0,$$

$$\frac{1}{G}(I_{in_i} + \Sigma_j \alpha_{ij} F_j - \frac{\beta}{N} \Sigma_{j=1}^{N} F_j)), \quad (3)$$

where N is the total number of neurons in the network.

The effective recurrent current is maximal when all of the cells that are connected to the ith neuron are above threshold. If $\Sigma_j \alpha_{ij}$ meets the condition for stability describe in the previous section, then the response of the neuron cannot grow without bound. The gain of the recurrent amplifier of a given neuron changes as neurons to which it is connected pass through the threshold point. The amplifier's gain is highest when all its neurons are above threshold, and its gain is zero when all of the neurons are below threshold. Similarly, the degree to which the inhibition tracks the output of the winning amplifier increases as convergence proceeds, because the outputs of the neurons that are loosing the competition stop contributing to inhibition when they fall below threshold. In this way inhibition changes its role from thresholding to proportional, as convergence proceeds. Thresholding is required initially to extract the best estimate of the signal, and proportional inhibition is required later to stabilize the cooperative excitation of the winning neurons. The maximal value of proportional (stabilizing) inhibition that the amplifier can experience is $\beta(n/N)$, where n is the number of neurons above threshold, participating in the active amplifier.

The model network amplifies the correlated signal in a pattern that was stored in the connectivity of the network, without amplifying random noise (Figure 3c). It does this by expressing the tuning curve inherent in the cortical connectivity. It is likely that the same architectures and adaptive processes used by neural systems to generate coherent action in the presence of imperfect components also enable them to extract precise information from a noisy and ambiguous environment.

B. Control of Amplifier Activation

Previous models of the selective responses of neocortical neurons depended on specific spatial patterns of strong inhibition [9, 10]. By contrast, cortical amplifier networks proposed here are able to achieve selective responses that are shaped by patterns of excitatory connection. This excitation can and be controlled by small levels of inhibition, provided that the inhibition is supplied to all of the neurons participating in the amplifier.

In a feedforward network, complete inhibition requires that an inhibitory current at least as large as I_{in} to prevent a neuron from firing. When the neurons are embedded in a recurrent circuit, the same value of F is achieved with a much smaller I_{in}, because a substantial fraction of the total current is due to a recurrent current, I_{rec}. Under these conditions complete inhibition of discharge also requires that the inhibitory current be at least as large as I_{in}, but because I_{in} is smaller than the feedforward case, so is the required inhibitory current. The inhibition, $I_{inh_{amp}}$, required to suppress neurons participating in a cortical amplifier, compared to the inhibition, $I_{inh_{iso}}$, required to suppress an isolated neuron is

$$\frac{I_{inh_{amp}}}{I_{inh_{iso}}} = \frac{G - \alpha}{G}. \quad (4)$$

The higher the gain of the amplifier, the less input current is needed to achieve a given level of activation, and the less inhibition is needed to suppress it.

C. Latching an Analog Memory

The cortical amplifier can be hysteretic in its behavior. This property allows the amplified state to be latched as a stable analog pattern by a simple control signal until it is ready for the next step in the computation. Persistence of patterns of cortical activity has been observed in monkeys waiting for a cue to make a requested eye movement [11]. The spatial location of the requested eye movement is indicated by a visual target. At the onset of the target some neurons in cortex are activated, and they remain so even after the position indicator is switched off. They return to rest only after the animal is cued to make the eye movement.

The principles underlying latching of output can be seen in a system composed of two cortical amplifiers, A and B, each similar to the population

1851

of neurons illustrated in Figure 2b. The two amplifiers compete via common inhibition, in which the inhibitory current for each is given by $I_{inh} = \frac{\beta}{2}(F_A + F_B)$. Assume that $I_{in_A} > I_{in_B}$ has led to the state in amplifier A is the winner with F_A above zero, and F_B zero. Under these conditions both amplifiers are inhibited by an amount,

$$I_{inh} = \frac{\beta}{2}F_A = \frac{I_{in_A}}{1 + \frac{G-\alpha}{\frac{\beta}{2}}}. \tag{5}$$

I_{in_B} does not contribute to this inhibition, because F_B is zero. To raise F_B above zero to contend for a new winner, I_{in_B} must be greater than the inhibition. If the denominator of the above equation is less than one ($\alpha > G$), then I_{in_B} will have to be larger than I_{in_A} before inhibition is exceeded and a new competition can begin. At any I_{in_B} value beneath this I_{inh}, the state of the system will remain unchanged. Therefore, provided the gain of the populations is sufficiently large, the winning state of the system can be latched by supplying the same input currents to all of the populations equally. Although the system is hysteretic, the activity of the winning population will still decay when the input currents are removed.

III. Hybrid Analog-Digital CMOS Implementation of Cortical Amplification

A multi-chip hybrid analog-digital system that uses the principle of smart amplification for computing the stereocorrespondence of simple one-dimensional images in real time has been reported previously[12]. This system comprised two silicon retinae, each containing 4096 pixels, and a stereocorrespondence chip, with 969 processing nodes. The system used the address-event communication method described in the next section. We are generalizing this architecture to make an emulator for networks of biologically-realistic neurons.

A. Communication

The action potentials of the brain, necessitated by the the poor conduction of neurons, seem to be a specialization wholly inappropriate for the modern VLSI medium. However, robust communication of analog information between chips is difficult even though aluminum is a good conductor. Part of the difficulty rests in the inherent systematic differences between subthreshold transistor[13] properties on different chips. This variability implies that the effect of a voltage on the transistors of one chip is different than the effect on similar transistors of another. A more profound impediment to building

neuromorphic systems is the small number of interface ports to a chip. Chips with many neurons must time multiplex their outputs. Accurately multiplexing analog signals is slow due to the settling times of the analog amplifiers. Analog VLSI neural network chips have overcome these difficulties by using digital pulse-based communication [14, 15].

The address-event representation (AER) is an asynchronous digital pulse-based encoding method that captures the temporal structure of action potential discharge. The silicon neurons generate trains of action-potential-like events. When a particular neuron generates an action potential, this event is detected by an arbiter that immediately places the address of the active neuron on the bus. The address event is passed on to all the synapses in the network. Those synapses that are 'connected' to the source neuron detect that their source neuron has generated an action potential, and they initiate a synaptic input to the post-synaptic neuron to which they are attached. The mappings between source neurons and recipient synapses can be established by digital circuitry on the AE bus that map the source-encoded output addresses to different destination-encoded synaptic positions.

The bandwidth of communication does limit the number of neurons that can share the same bus, but a number of factors act to conserve bandwidth. Like their biological counterparts, only a small fraction of the silicon neurons embedded in a network are active at any time. Even the active ones generate action potentials at less than about 100 Hz. However, the AE bus is able to transmit events at approximately 1 MHz. This means that the same communication bus can be shared by thousands of neurons. Unlike sequential scanning techniques for multiplexing, which poll neurons on every frame even if the neurons have no new data, the (AER) devotes all its bandwidth to active neurons and so reduces the temporal aliasing problem. The bandwidth limitation of a global bus will foster the development of hierarchical traffic routing schemes for constructing networks with mainly local connectivity.

B. SCX1

The SCX1 project currently underway is a general-purpose reconfigurable neural emulator that will help us to investigate the computational power of a network of biologically realistic analog VLSI neurons [16]. The SCX1 board uses AER communication. It is designed to be integrated in a multi-board system that also uses AER. The SCX1 provides a reconfigurable network of neurons that can function in real time, interacting with the real world via silicon sensors, and actuators. These networks will be

used to explore the performance of cortical amplifier networks on sensorimotor tasks.

Each SCX1 board is composed of combination of analog and digital chips. The neurons reside on a number of analog multi-neuron chips. The prototype chip analog chip is a 6.8×4.6 mm^2 and has 36 neurons. Each neuron has 6 compartments with active and passive channel emulator circuits and a total of 20 synapses. A single synapse may be shared by multiple presynaptic inputs. Both the biophysical properties of the neurons, and the synaptic weights are stored on volatile capacitors which are refreshed by a DSP chip. In addition to this maintenance function, the DSP participates in the network computation by routing presynaptic address events onto the synapses of neurons.

The number of neurons on our neuron chip is small. However, simulation of this biophysical complexity on a SPARC2 workstation would take about 10,000 times real time. The number of neurons in this application is not limited by the architectural framework but by the space on the chip required to emulate full neuronal biophysics. We estimate that the SCX1 architecture could support a population of 1,000 neurons with an average firing rate of 50 Hz, and 100 inputs per neuron on a single board. As demonstrated by the stereocorrespondence system, a neuron pool of this size can already perform a useful computation.

IV. HYBRID ANALOG-DIGITAL COMPUTATION

The cortical architecture described in this paper suggests the possibility of a new kind of analog-digital hybrid computer. The physical interface between the digital world and the analog world is the action potential of the neurons. The representational interface between the abstract token of the digital world and the analog symbol [1] is the cortical amplifier. The cortical amplifier aggregates fragmentary information into a cohesive pattern. The cortical amplifier provides a framework of a creative memory device. Rather than the timeless and desiccated text of a digital memory, the hybrid analog-digital memory is alive and changing with the context of recollection. The cortical amplifier implements content addressable memory (CAM) but with an analog encoded output. The analog values indicate the confidence with which the symbol is recalled, making implementation of fuzzy logic rules natural.

REFERENCES

[1] C. A. Mead and L. A. Conway, *Introduction to VLSI Systems.* Reading, MA: Addison-Wesley, 1980.

[2] J. von Neumann, *The Computer and the Brain.* New Haven, CT: Yale University Press, 1958.

[3] J. V. Candy, *Signal Processing. The model-based approach.* New York: McGraw-Hill, 1986.

[4] R. Douglas and K. Martin, "A functional microcircuit for cat visual cortex," *J. Physiology,* vol. 440, pp. 735–769, 1991.

[5] B. Ahmed, J. C. Anderson, R. J. Douglas, K. A. Martin, and C. Nelson, "Polyneuronal innervation of spiny stellate neurons in cat visual cortex," *J. Comp. Neurol.,* vol. 339, p. In Press, 1994.

[6] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci. USA,* vol. 79, pp. 2554–2558, 1982.

[7] S. Grossberg, "Nonlinear neural networks: principles, mechanisms, and architectures," *Neural Networks,* vol. 1, pp. 17–61, 1988.

[8] B. Chapman, K. Zans, and M. Stryker, "Relation of cortical cell orientation selectivity to alignment of receptive fields of geniculocortical afferents that arborize within a single orientation column in ferret visual cortex," *J. Neurosci.,* vol. 11, pp. 1347–1358, 1991.

[9] P. Heggelund, "Receptive field organization of simple cells in cat striate cortex," *Exp. Brain Res.,* vol. 42, pp. 89–98, 1981.

[10] F. Wörgötter and C. Koch, "A detailed model of the primary visual pathway in the cat: Comparison of afferent excitatory and intracortical inhibitory connection schemes for orientation selectivity," *J. Neuroscience,* vol. 11, no. 7, pp. 1959–1979, 1991.

[11] R. Andersen, "Visual and eye movement functions of the posterior parietal cortex," *Ann. Rev. Neurosci.,* vol. 12, pp. 377–403, 1989.

[12] M. Mahowald, *An Analog VLSI System for Stereoscopic Vision.* Boston: Kluwer, 1994.

[13] C. Mead, *Analog VLSI and Neural Systems.* Reading, MA: Addison-Wesley, 1989.

[14] A. Murray, D. Del Corso, and L. Tarassenko, "Pulse-stream VLSI neural networks mixing analog and digital techniques," *IEEE Trans. Neural Networks,* vol. 2, pp. 193–204, 1991.

[15] J. Lazzaro, J. Wawrzynek, M. Mahowald, M. Sivilotti, and D. Gillespie, "Silicon auditory processors as computer peripherals," *IEEE Trans. Neural Networks,* vol. 4, pp. 523–528, 1993.

[16] M. Mahowald and R. Douglas, "A silicon neuron," *Nature,* vol. 354, pp. 515–518, 1991.

---

[1] Tokens are arbitrary, whereas symbols carry with them some aspect of their meaning.