

Seeing Chips: Analog VLSI Circuits for Computer Vision

Christof Koch

*Computation and Neural Systems Program,
Divisions of Biology and Engineering and Applied Science, 216-76,
California Institute of Technology, Pasadena, CA 91125, USA*

Vision is simple. We open our eyes and, instantly, the world surrounding us is perceived in all its splendor. Yet Artificial Intelligence has been trying with very limited success for over 20 years to endow machines with similar abilities. A large van, filled with computers and driving unguided at a mile per hour across gently sloping hills in Colorado and using a laser-range system to "see" is the most we have accomplished so far. On the other hand, computers can play a decent game of chess or prove simple mathematical theorems. It is ironic that we are unable to reproduce perceptual abilities which we share with most animals while some of the features distinguishing us from even our closest cousins, chimpanzees, can be carried out by machines. **Vision is difficult.**

1 Introduction ---

In the last ten years, significant progress has been made in understanding the first steps in visual processing. Thus, a large number of well-studied algorithms exist that locate edges, compute disparities along these edges or over areas, estimate motion fields and find discontinuities in depth, motion, color and texture (for an overview see Horn 1986 or the last reference at the end of this article). At least two major problems remain. One is the integration of information from different modalities. Fusion of information is expected to greatly increase the robustness and fault tolerance of current vision systems as it is most likely the key towards fully understanding vision in biological systems (Barrow and Tenenbaum 1981; Marr 1982; Poggio et al. 1988). The second, more immediate, problem is the fact that vision is very expensive in terms of computer cycles. Thus, one second's worth of black-and-white TV adds up to approximately 64 million bits which needs to be transmitted and processed further. And since early vision algorithms are usually formulated as relaxation algorithms which need to be executed many hundreds of times before convergence, even supercomputers take their time. For instance, the 65,536

processor Connection Machine at Thinking Machines Corporation (Hillis 1985), with a machine architecture optimal from the point of view of processing two-dimensional images, still requires several seconds *per image* to compute depth from two displaced images (Little 1987). Performance on microprocessor based workstations is hundreds of times slower.

Animals, of course, devote a large fraction of their nervous system to vision. Thus, about 270,000 out of 340,000 neurons in the house fly *Musca domestica* are considered to be "visual" neurons (Strausfeld 1975), while a third of the human cerebral cortex is given over to the computations underlying the perception of depth, color, motion, recognition, etc. One way for technology to bypass the computational bottleneck is to likewise construct special-purpose vision hardware. Today, commercial vendors offer powerful and programmable digital systems on the open market for a few thousand dollars. Why, however, execute vision algorithms on digital machines when the signals themselves are analog? Why not exploit the physics of circuits to build very compact, analog special-purpose vision systems? Such a *smart sensor paradigm*, in which as much as possible of the signal processing is incorporated into the sensor and its associated circuitry in order to reduce transmission bandwidth and subsequent stages of computation, is starting to emerge as a possible competitor to more general-purpose digital vision machines.

2 Analog circuits for vision: the early years _____

This idea was explicitly raised by Horn at MIT (1974), where he proposed the use of a 2-D hexagonal grid of resistances to find the inverse of the discrete approximation to the Laplacian. This is the crucial operation in an algorithm for determining the lightness of objects from their image. An attempt to build an analog network for vision was undertaken by Knight (1983) for the problem of convolving images with the Difference-of-two-Gaussians (DOG), a good approximation of the Laplacian-of-a-Gaussian filter of Marr and Hildreth (1980). The principal idea is to exploit the dynamic behavior of a resistor/capacitor transmission line, illustrated in figure 1. In the limit that the grid becomes infinitely fine, the behavior of the system is governed by the diffusion equation:

$$RC \frac{\partial V(x, t)}{\partial t} = \frac{\partial^2 V(x, t)}{\partial x^2} . \quad (2.1)$$

If the initial voltage distribution is $V(x, 0)$ and if the boundaries are infinitely far away, the solution voltage is given by the convolution of $V(x, 0)$ with a progressively broader Gaussian distribution (Knight 1983). Thus, a difference of two Gaussians can be computed by converting the incoming image into an initial voltage distribution, storing the resulting voltage distribution after a short time and subtracting it from the voltage distribution at a later time. A resistor/capacitor plane yields the same

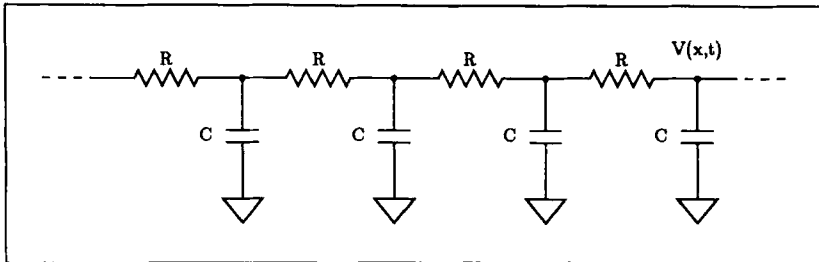


Figure 1: One-dimensional lumped-element resistor/capacitor transmission line. The incoming light intensity is converted into the initial voltage distribution $V(x, 0)$. The final voltage $V(x, t)$ along the line is given by the convolution of $V(x, 0)$ with a Gaussian of variance $\sigma^2 = 2t/RC$ and is read off after a certain time t related to the width of the Gaussian filter. From Koch (1989).

result in two dimensions. Practical difficulties prevented the successful implementation of this idea. A team from Rockwell Science Center (Mathur et al. 1988) is reevaluating this idea by using a continuous 2-D undoped polysilicon plane deposited on a thick oxidized silicon sheet (to implement the distributed capacitor). The result of the convolution is read out via a 64 by 64 array of vertical solder columns.

A different approach — exploiting CCD technology — for convolving images was successfully tried by Sage at MIT's Lincoln Laboratory (Sage 1984), based on an earlier idea of Knight (1983). In this technology, incoming light intensity is converted into a variable amount of charge trapped in potential "wells" at each pixel. By using appropriate clocking signals, the original charge can be divided by two and shifted into adjacent wells. A second step further divides and shifts the charges and so on (Fig. 2). This causes the charge in each pixel to spread out in a diffusive manner described accurately by a binominal convolution. This represents, after a few iterations, a good approximation to a gaussian convolution. Sage extended this work to the 2-D domain (Sage and Lattes 1987) by first effecting the convolution in the x and then in the y direction. Their 288 by 384 pixel CCD imager convolves images at up to 60 times per second. Since CCD devices can be packed extremely densely — commercial million-pixel CCD image sensors are available — such convolvers promise to be remarkably fast and area-efficient.

3 Analog VLSI and Neural Systems

The current leader in the field of analog sensory devices that include significant signal processing is undoubtedly Carver Mead at Caltech

(Mead 1989). During the last years he has developed a set of subcircuit types and design practices for implementing a variety of vision circuits using subthreshold analog complementary Metal-Oxide-Semiconductor (CMOS) VLSI technology. His best-known design is the "Silicon retina" (Sivilotti et al. 1987; Mead and Mahowald 1988), a device which computes the spatial and temporal derivative of an image projected onto its phototransistor array. The version illustrated schematically in figure 3a has two major components. The photoreceptor consists of a phototransistor feeding current into a circuit element with an exponential current-voltage characteristic. The output voltage of the receptor \tilde{V} is logarithmic over four to five orders of magnitude of incoming light intensity, thus performing automatic gain control, analogous to the cone photoreceptors of the vertebrate retina. This voltage is then fed into a 48 by 48 element hexagonal resistive layer with uniform resistance values R . The photoreceptor is linked to the grid by a conductance of value G , implemented by a transconductance amplifier. An amplifier senses the voltage difference across this conductance and thereby generates an output at each pixel proportional to the difference between the receptor output and the network potential. Formally, if the voltage at pixel i, j is V_{ij} and the current being fed into the network at that location $I_{ij} = G(\tilde{V}_{ij} - V_{ij})$, the steady state is characterized by:

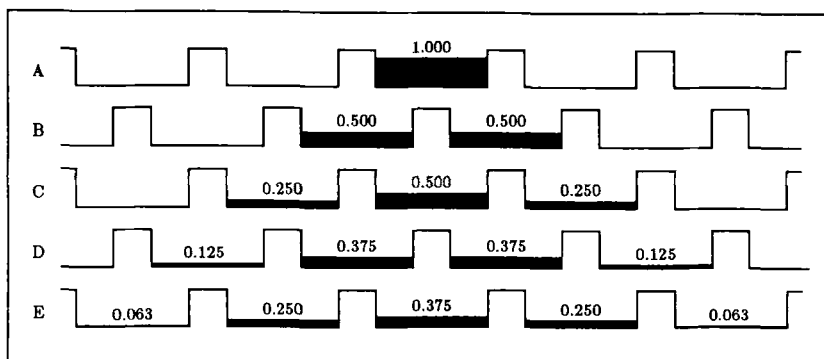


Figure 2: Schematic of a potential "well" CCD structure evolving over time. The initial charge across the 1-D array is proportional to the incoming light intensity. The charge packet shown in (A) is then shifted into the two adjacent wells by an appropriate clocking method. Since the total charge is conserved, the charge per well is halved (B). In subsequent cycles, (C, D and E) the charge is further divided and shifted, resulting in a binominal charge distribution. After several steps, this distribution is very similar to a Gaussian distribution. From Koch (1989).

$$6V_{ij} - V_{i-1j} - V_{i+1j+1} - V_{ij+1} - V_{i-1j} - V_{ij-1} - V_{i+1j-1} = RI_{ij}. \quad (3.1)$$

On inspection, this turns out to be one of the simplest possible discrete analogs of the Laplacian differential operator ∇^2 . In other words, given an infinitely fine grid and the voltage distribution $V(x, y)$, this circuit computes the current $I(x, y)$ via

$$\nabla^2 V = RG(\tilde{V} - V) = RI. \quad (3.2)$$

The current I at each grid point — proportional to $\tilde{V} - V$ and sensed by the amplifier — then corresponds to a spatially high-passed filtered version of the logarithmic compressed image intensity. Operations akin to temporal differentiation can be achieved by adding capacitive elements (Sivilotti et al. 1987). The required resistive elements of this circuit are designed by exploiting the current-voltage relationship (Fig. 3b) of a small transistor circuit, instead of using the resistance of a special metallic process. As long as the voltage across the device is within its linear range (a couple of 100 mV's), it behaves like a constant resistance whose value can be controlled over five orders of magnitude. The current saturates for larger voltage values, a nonlinearity with very desirable effects (see below). This, then, is the basic circuit element used for most vision chips coming out of Caltech.

The response of the silicon retina to a 1-D edge projected onto the phototransistors is shown in figure 3c. The voltage trajectory can be well approximated by the second spatial derivative of the smoothed brightness intensity. In 2-D the response is similar to that obtained by convolving the image with the DOG edge detection operator (Marr and Hildreth 1980). A different circuit (Tanner 1986) computes the optical flow field induced by a spatially homogeneous motion, such as moving a pointing device over a fixed surface (for example, an optical mouse).

A serious practical problem in designing the type of networks discussed here is that unwanted oscillations can spontaneously arise when large populations of active elements are interconnected through a resistive grid. These oscillations can occur even when the individual elements are quite stable. Using methods from nonlinear circuit theory, Wyatt and Standley (1989) at MIT have shown how this flaw can be circumvented. They have proven that if each linear active element in isolation is designed to satisfy the experimentally testable Popov criterion from control theory (which guarantees that a related operator is positive real), then stability of the overall interconnected nonlinear system is guaranteed.

Mead's principal motivation for this work comes from his desire to understand and emulate neurobiological circuits (as expressed in his new textbook, Mead 1989). He argues that the physical restrictions on the density of wires, the low power consumption of the CMOS process in the subthreshold domain, the limited precision and the cost of communication imposed by the spatial layout of the electronic circuits are similar to the constraints imposed on biological circuits. Furthermore, the

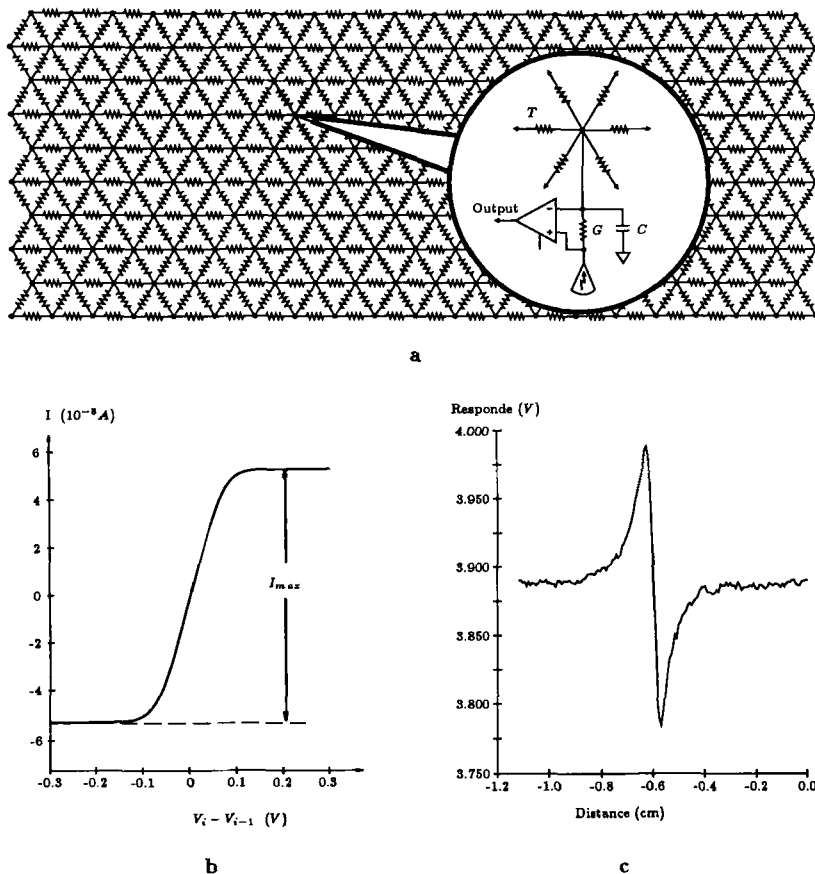


Figure 3: The "Silicon retina." (a) Diagram of the hexagonal resistive network with an enlarged single element. A photoreceptor, whose output voltage is proportional to the logarithm of the image intensity, is coupled — via the conductance G — to the resistive grid. The output of the chip is proportional to the current across the conductance G , or in other words, to the voltage difference between the photoreceptor and the grid. (b) The current-voltage relationship for Mead's resistive element. As long as the voltage gradient is less than ≈ 100 mV, the circuit acts like a linear resistive element. The output current saturates for larger gradients. (c) The experimentally measured voltage response of a 48 by 48 pixel version of the retina when a step intensity edge is moved past one pixel. This response is similar to the one expected by taking the second spatial derivative of the smoothed incoming light intensity. Adapted from Mead and Mahowald (1988). From Koch (1989).

silicon medium provides both the computational neuroscience and the engineering communities with tools to test theories under realistic, real-time conditions. To further the spread of this technology into the general academic community, all circuits are fabricated via the silicon foundry MOSIS.

4 Regularization theory and analog networks

Problems in vision are usually inverse problems; the two dimensional intensity distribution on retina or camera must be inverted to recover physical properties of the visible three dimensional surfaces surrounding the viewer. More precisely, these problems are ill-posed in that they either admit to no solution, to infinitely many solutions or to a solution that does not depend continuously on the data. In general, additional constraints must be applied to arrive at a stable and unique solution. One common technique to achieve this, termed "standard regularization" (Poggio et al. 1985), is via minimization of a given "cost" functional (for earlier examples of this see Grimson 1981; Horn and Schunck 1981; Ikeuchi and Horn 1981; Terzopoulos 1983; Hildreth 1984). The first term in these functionals assesses by how much the solution diverges from the measured data. The second term measures how closely the solution conforms to certain *a priori* expectations, for instance that the final surface should be as smooth as possible. Let us briefly consider the problem of fitting a 2-D surface through a set of noisy and sparse depth measurements, a well-explored problem in computer vision (Grimson 1981). Specifically, a set of sparse depth measurements is given on a 2-D lattice, d_{ij} , which are corrupted by some noise process. It is obvious that infinitely many surfaces, f_{ij} , can be fitted through the sparse data set. One way to regularize this problem is to find the surface f that minimizes

$$\sum_{i,j} ((f_{i+1j} - f_{ij})^2 + (f_{ij+1} - f_{ij})^2) + \alpha \sum_{i,j} (f_{ij} - d_{ij})^2, \quad (4.1)$$

in which α depends on the signal-to-noise ratio and the second sum only contains contributions from those locations i where data exists. Equation (4.1) represents the simplest possible functional, even though many alternatives exist (Grimson 1981; Terzopoulos 1983; Harris 1987). This and all other quadratic regularized variational functionals of early vision can be solved with simple linear resistive networks by virtue of the fact that the electrical power dissipated in linear networks is quadratic in the current or voltage (Poggio et al. 1985; Poggio and Koch 1985). The resistive network will then converge to its unique equilibrium state in which the dissipated power is at a minimum (subject to the source constraints). The static version of this statement is known as *Maxwell's Minimum Heat Theorem*. The steady-state of the resistive network in figure 4a minimizes

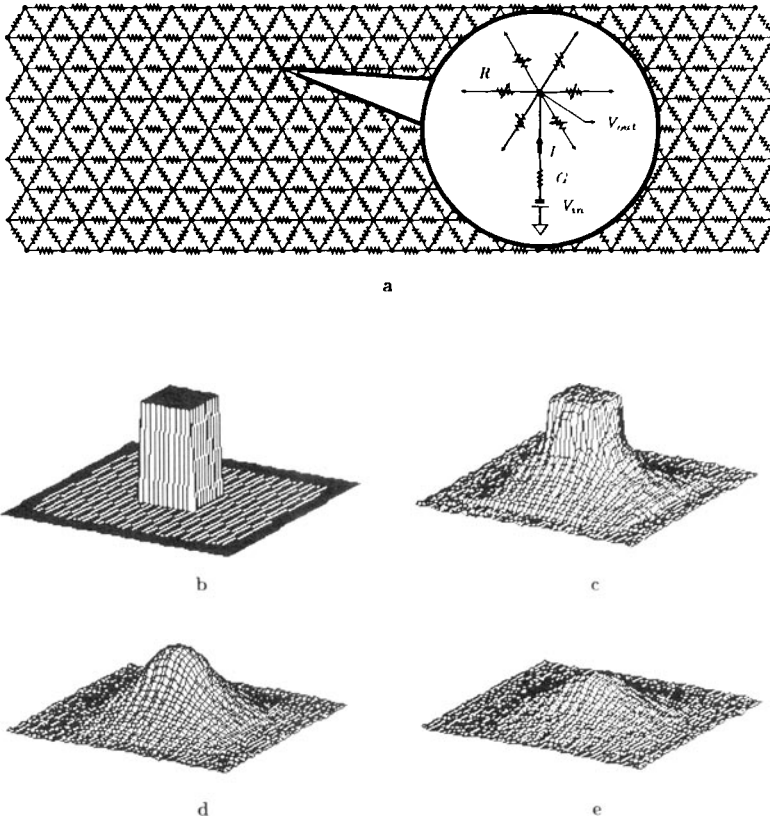


Figure 4: Surface interpolating network. (a) At those locations where depth data are available, the values of the battery V_{in} and of the conductance G are set to their appropriate values, via additional sample-and-hold circuitry. The output is the voltage V_{out} at each location. This circuit solves, for small enough voltage gradients, a modified form of Poisson's equation, via minimization of equation (4.2). Experimental results from a 48 by 48 subthreshold, analog CMOS VLSI circuit are shown next (Luo et al. 1988). (b) The input voltage V_{in} , corresponding to a flat, 2 pixel wide, strip around the periphery and a central 4 pixel wide tower (solid coloring). At these locations, the conductance G is set to a constant, fixed value, while G is zero everywhere else. Thus, no data are present in the area between the bottom of the tower and the outside strip. (c), (d), and (e) show the output voltage for a high, medium, and low value of the transversal resistance R . If R is small enough, the resulting smoothing will flatten out the central tower.

expression (4.1) if the voltage V_{ij} is identified with the discretized solution surface f_{ij} , the battery E_{ij} with the data d_{ij} and the product of the variable conductance G_{ij} , connecting the node to the battery and the constant horizontal resistance R with α . The power minimized by this circuit is then formally equivalent to the functional of equation (4.1). The performance of an experimental 48 by 48 subthreshold, analog CMOS VLSI circuit is illustrated in figure 4 (Luo et al. 1988). For an infinitely fine grid and a voltage source $E(x, y)$ the surface interpolation chip computes the voltage distribution $V(x, y)$ according to the modified Poisson equation

$$\nabla^2 V + RGV = RGE, \quad (4.2)$$

with either an arbitrary Dirichlet boundary condition (such as zero voltage along the boundary) or a zero voltage slope (that is, no current across the boundary) Neumann boundary condition. If RG is a constant across the grid, this equation is sometimes known as the Helmholtz equation. Note that the difference to equation (3.2) lies in the choice of observable, current $I(x, y)$ versus voltage $V(x, y)$. A large number of problems in early vision, such as detecting edges, computing motion or estimating disparity from two images have a similar architecture, with resistive connections among neighboring nodes implementing the constraint that objects in the real world tend to be smooth and continuous.

5 Discontinuities

However, the most interesting locations in any scene are arguably those locations at which some feature changes abruptly, for instance the 2-D optical flow at the boundary between a moving figure and the stationary background or the color across the sharp boundaries in a painting. Geman and Geman (1984) (see also Blake and Zisserman 1987) introduced the powerful concept of a binary line process l_{ij} which explicitly codes for the absence ($l_{ij} = 0$) or presence ($l_{ij} = 1$) of a discontinuity at location i, j in the 2-D image. Further constraints, such that discontinuities should occur along continuous contours (as they do, in general, in the real world) or that they rarely intersect, can be incorporated into their theory, which is based on a statistical estimation technique (see also Marroquin et al. 1987). In the case of surface interpolation and smoothing, maximizing the *a posteriori* estimate of the solution can be shown to be equivalent to minimizing:

$$\begin{aligned} \sum_{i,j} (1 - l_{ij}^h)(f_{i+1j} - f_{ij})^2 &+ \sum_{i,j} (1 - l_{ij}^v)(f_{ij+1} - f_{ij})^2 \\ &+ \alpha \sum_{i,j} (f_{ij} - d_{ij})^2 \end{aligned} \quad (5.1)$$

$$+ \beta \sum_{i,j} V(l_{ij}),$$

where l_{ij}^h and l_{ij}^v are the horizontal and vertical depth discontinuities, β a fixed parameter and V a potential function containing a number of terms penalizing or encouraging specific configurations of line processes. In the case of surface interpolation, a simple example is $V(l_{ij}^v) = l_{ij}^v$; that is, the line process l_{ij}^v between i, j and $i + 1, j$ will be set to 1 if the "cost" for smoothing, that is $(f_{i+1j} - f_{ij})^2$, is larger than the parameter β . Otherwise, $l_{ij}^v = 0$. Discontinuities greatly improve the performance of early vision processes, since they allow algorithms to smooth over unreliable or sparse data as well as account for boundaries between figures and ground. In fact, it can be argued that the introduction of discontinuities represents the single biggest advance in machine vision in the last 5 years. They have been used to demarcate boundaries in the intensity, color, depth, motion and texture domains (Geman and Geman 1984; Terzopoulos 1986; Blake and Zisserman 1987; Marroquin et al. 1987; Gamble and Poggio 1987; Hutchinson et al. 1988; Poggio et al. 1988; Chhabra and Grogan 1988).

Line discontinuities can be implemented in various ways. In a *hybrid* implementation, each line process is represented by a simple binary switch. When the switch is open, no current flows across the connection between the two adjacent nodes i, j and $i + 1, j$. The network operates by switching between distinct modes. In the analog cycle the network settles into the state of least power dissipation, *given a fixed distribution of switches*. In the digital phase, the line processes are evaluated using expression (5.1); that is, the switches are set to the state minimizing this expression. Such a hybrid implementation is illustrated in figure 5 for the case of computing the optical flow in the presence of motion discontinuities. The flow field — induced by the time-varying image intensity $I(x, y, t)$ — is regularized using a smoothness constraint (Horn and Schunck 1981). The amount of smoothing is governed by the constant resistance value R of the upper and lower horizontal grids. In a complete *analog* implementation, each line process is represented by a "neuron" whose output varies continuously between 0 and 1 (Koch et al. 1986), similarly to Hopfield and Tank's (1985) use of such continuous variables to solve the "traveling salesman problem." Another possibility exploits the saturation inherent in Mead's design for resistances (Mead 1989). As illustrated in figure 3b, the current-voltage relation of the resistive element is linear for voltage differences on the order of 100 mV, while the element saturates for larger voltage differences. In other words, the peak current is independent of the size of the voltage gradient (as long as it is larger than some threshold), implementing a first approximation of a binary line process. This occurs in figure 4c, where segmentation starts to occur due to the large voltage gradient between the base and the top of the "tower."

A very promising implementation is via “resistive fuses” (Harris et al. 1989): in such a two-terminal nonlinear resistor, the current flowing through is proportional to the voltage difference across as long as that difference is less than a threshold. If the voltage gradient exceeds the threshold, the current decreases and, for large enough voltage gradients, is set to zero. The experimentally determined voltage-current relationship of this device (Harris et al. 1989) is closely related to the cost function used with the “analog” line discontinuities (Koch et al. 1986). It can also be derived from the cost function used in the “graduated non-convexity” method of Blake and Zisserman (1987). The notion of minimizing power in linear networks implementing quadratic regularization algorithms must be replaced by the more general notion of minimizing the total co-content J for linear networks with “resistive fuses” (where $J = \int_0^V f(V')dV'$ for a resistor defined by $I = f(V)$).

Although the method proposed by Geman and Geman (1984) requires stochastic optimization techniques and complicated potential functions (V in expression (5.1)) to implement the various constraints under which line discontinuities operate, computer simulations have shown that various deterministic approximations as well as much simplified potential functions can be used (see also Blake 1989).

6 Analog chips versus digital computers

As we have seen, all of the above circuits exploit the physics of the system to perform operations useful from a computational point of view. Thus, the transient voltage or charge distribution at some time in the networks of figures 1 and 2 corresponds to the solution, in this case convolution of the image intensity with a Gaussian. In the networks derived from the appropriate variational functionals, the stationary voltage distribution corresponds to the interpolated surface (Fig. 4) or to the optical flow (Fig. 5). These quantities are governed by Kirchhoff's and Ohm's laws, instead of being symbolically computed via execution of software in a digital computer. Furthermore, the architecture of the analog resistive circuits reflects the nature of the underlying computational task, for instance, smoothing, while digital computers — being Turing universal — do not. One of the advantages of these non-clocked analog circuits is that their operating mode is optimally suited to analog sensory data since they avoid temporal aliasing problems caused by discrete temporal sampling. Furthermore, their robustness to imprecisions or errors in the hardware, their processing speed and low power consumption (Mead's retina requires less than a mW , most of which is used in the photo-conversion stage) and their small size make analog smart sensors very attractive for tele-robotic applications, remote exploration of planetary surfaces and a host of industrial applications where their power hungry, heat producing, bulky and slow digital cousins are unable to compete.

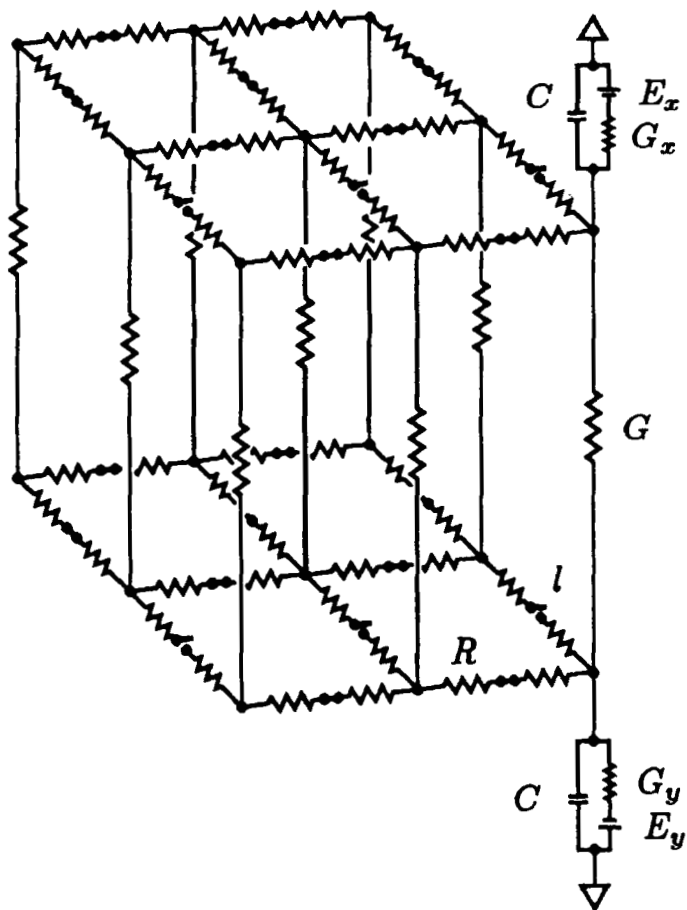


Figure 5: Hybrid resistive network to compute the optical flow in the presence of discontinuities. The algorithm computes the smoothest flow field compatible with the measured motion data (Horn and Schunck 1981; Hutchinson et al. 1988). The steady-state voltage distribution in the upper grid is equivalent to the x component and the stationary voltage distribution in the lower grid is equivalent to the y component of the optical flow. A high voltage at location i, j will spread to its four neighboring nodes. The degree to which voltage spreads, and thus the degree of smoothness, is governed by the value of the constant horizontal resistance R . The value of the batteries E_x and E_y and the conductances G_x , G_y (for clarity, only two such elements are drawn) and G depend on the measured spatial and temporal intensity gradients ∇I and I_t and will be set by on-chip photoreceptors. Binary switches l implement motion discontinuities, since an arbitrary high voltage, that is, velocity, will not affect the neighboring site across the discontinuity. Adapted from Hutchinson et al. (1988).

The two principal drawbacks of analog VLSI circuits are their lack of flexibility and their imprecision. The above circuits are all hard-wired to perform very specific tasks, unlike digital computers which can be programmed to approximate any logical or numerical operation. Only certain parameters associated with this algorithm, for instance the smoothness in the case of figures 4 and 5, can be varied. Thus, digital computers appear vastly preferable for developing and evaluating new algorithms; analog implementations should only be attempted after such initial exploration of algorithms. Furthermore, although 12 and even 16 bit analog-to-digital converters are commercially available, it seems unlikely that the precision of analog vision circuits will exceed 7 to 8 bit of resolution in the next few years. However, for a number of important tasks, such as navigation or tracking, the incoming intensity data are rarely more accurate than 1% in any case.

7 The Future

Within the last year, a number of potentially very exciting developments have occurred which bode well for the future of analog vision circuits. Mahowald and Delbrück (1989) from Mead's laboratory have built and tested an analog CMOS VLSI circuit implementing a version of Marr and Poggio's (1976) cooperative stereo algorithm. Two 1-D phototransistor arrays, with 40 elements each, located next to each other on the chip provide the input to the circuit. A winner-take-all circuit selects the most active node among the seven possible disparity values at each pixel, replacing the inhibitory interaction in the original algorithm.

A problem plaguing analog subthreshold circuits are random offsets which vary from location to location and are caused by fluctuations in the process accuracy as well as dark currents. Such offsets, while usually not problematic for digital circuits, can be very disruptive when operating in the analog domain, in particular when spatial or temporal derivatives are required. Mead (1988) has recently developed a variant of the "floating gate technology" used for a long time for resetting programmable read-only-memory cells (EPROM) by means of ultra-violet light. While previously the chips were bombarded with UV radiation to erase memory, Glasser (1985) of MIT demonstrated how this technology could be used to selectively write a "0" or a "1" into the cell. Mead is the first to have applied this technique to the analog domain, by building a local feedback circuit at every node of the retina (Fig. 3) which senses the local current and attempts to keep it at or near zero by charging up a capacitor located between two layers of poly positioned above each node. Exposure to UV light excites electrons sufficiently to enable them to surmount the potential barrier at the silicon/silicon dioxide interface. In order to adapt the retina, a blank, homogeneous image is projected for a fraction of a minute onto the chip — in the presence of the UV light. This ef-

fectively creates a “floating” battery at each location, which induces a current exactly counteracting the effect of the offset current at that pixel. Mead (1988) has even been able to show after-image-like phenomena.

Another problem with most resistive networks for early vision problems is that the values of the individual circuit elements, such as conductances or voltage sources, depend on the measured data or can even be negative in value (the associated operator is not, in other words, of the convolution type) raising problems with network stability. Harris at Caltech has shown how this problem can be circumvented via the use of so-called “constraint boxes”, which impose a generalized constraint equation (Harris 1987; 1989). For the case of reconstructing surfaces using a smoother functional than the one of equation (4.1) (so-called cubic spline or thin plate interpolation), his circuit implements an equation of the form $V_a - V_b - V_c = 0$. The VLSI circuit has been tested successfully (Harris 1989) and is unusual in that all of its terminals can act as input or output nodes. Thus, if nodes a and b are held constant, then the c node is fixed to $V_a - V_b$. Using these constraint boxes in the case of computing smooth optical flow (Horn and Schunck 1981; Hutchinson et al. 1988), all resistance values are positive and data independent, a considerable advantage when building these circuits.

A team at MIT headed by J. Wyatt, and including B. Horn, H.-S. Lee, T. Poggio, and C. Sodini, is initiating an ambitious effort to fabricate analog, early vision chips exploiting different circuit technologies, such as CCD or mixed bipolar and CMOS devices. They plan to build various 2-D spatial correlator and convolver circuits, analog image memories and single-chip moment calculators and motion sensors. As part of this effort, new methods for estimating first and second image moments or computing optical flow under various constraints (for example, rigid environment) are being developed purely for such analog implementations (Horn 1989). Fusion of information on-chip is being attempted in Koch’s laboratory at Caltech, by integrating a set of simple resistive networks computing depth and depth discontinuities, as well as edges and optical flow onto a small, autonomous moving vehicle. A number of other laboratories are also engaged in efforts to build vision sensors. In particular, a group at UCLA and Rockwell International (White et al. 1988) is designing a 2-D network for edge detection on the basis of Poggio, Voorhees and Yuille’s (1985) proposal via a set of four 1-D resistive lines. Thus, it appears that the analog computers of the 1940s and 1950s (Karplus 1958), until recently considered extinct, are making a sort of comeback in the form of highly dedicated smart vision chips.

Acknowledgments

The author thanks John Harris, Berthold Horn, Andy Lumsdaine, and in particular John Wyatt for a careful reading of the manuscript. Research

on analog circuits for vision is supported by a Young Investigator Award and grant IST-8700064 from the Office of Naval Research, a Presidential Young Investigator Award from the National Science Foundation, as well as by DDF-II funds from the Jet Propulsion Laboratory and by Rockwell International.

References

- Barrow, H.G. and J.M. Tenenbaum. 1981. Computational vision. *Proceedings of the IEEE*, **69**, 572–595.
- Blake, A. 1989. Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 2–12.
- Blake, A. and A. Zisserman. 1987. *Visual Reconstruction*. Cambridge: MIT Press.
- Chhabra, A.K. and T.A. Grogan. 1988. Estimating depth from stereo: Variational methods and network implementation. *IEEE International Conference on Neural Networks*, San Diego.
- Gamble, E. and T. Poggio. 1987. Integration of intensity edges with stereo and motion. *Artificial Intelligence Laboratory Memo*, **970**. Cambridge: MIT Press.
- Geman, S. and D. Geman. 1984. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.
- Glasser, L.A. 1985. A UV write-enabled PROM. In: 1985 Chapel Hill Conference on VLSI, ed. W. Fuchs, 61–65. Rockville, MD: Computer Science Press.
- Grimson, W.E.L. 1981. *From Images to Surfaces*. Cambridge: MIT Press.
- Harris, J.G. 1987. A new approach to surface reconstruction: The coupled depth/slope model. *Proceedings of the IEEE First International Conference on Computer Vision*, 277–283, London.
- . 1989. An analog VLSI chip for thin plate surface interpolation. In: *Neural Information Processing Systems*, ed. D. Touretsky. Morgan-Kaufmann Publishers, 687–694.
- Harris, J., C. Koch, J. Luo, and J. Wyatt. 1989. Resistive fuses: Analog hardware for detecting discontinuities in early vision. In: *Analog VLSI Implementations of Neural Systems*, eds. C. Mead and M. Ismail. Norwell, MA: Kluwer.
- Hildreth, E.C. 1984. *The Measurement of Visual Motion*. Cambridge: MIT Press.
- Hillis, W.D. 1985. *The Connection Machine*. Cambridge: MIT Press.
- Hopfield, J.J. and D.W. Tank. 1985. Neural computation in optimization problems. *Biological Cybernetics*, **52**, 141–152.
- Horn, B.K.P. 1974. Determining lightness from an image. *Computational Graphics Image Processing*, **3**, 277–299.
- . 1986. *Robot Vision*. Cambridge: MIT Press.
- . 1989. Parallel networks for machine vision. *Artificial Intelligence Laboratory Memo*, **1071**. Cambridge: MIT Press.
- Horn, B.K.P. and B.G. Schunck. 1981. Determining optical flow. *Artificial Intelligence*, **17**, 185–203.

- Hutchinson, J., C. Koch, J. Luo, and C. Mead. 1988. Computing motion using analog and binary resistive networks. *IEEE Computers*, **21**, 52–63.
- Ikeuchi, K. and B.K.P. Horn. 1981. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, **17**, 141–184.
- Karplus, W.J. 1958. *Analog Simulation: Solution of Field Problems*. New York: McGraw-Hill.
- Knight, T. 1983. *Design of an integrated optical sensor with on-chip preprocessing*. Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.
- Koch, C. 1989. Resistive networks for computer vision. A Tutorial. In: *An Introduction to Neural and Electronic Networks*, eds. S.F. Zornetzer, J.C. Davis, and C. Lau. Academic Press, in press.
- Koch, C., J. Marroquin, and A. Yuille. 1986. Analog 'neuronal' networks in early vision. *Proceedings of the National Academy of Science USA*, **83**, 4263–4267.
- Little, J. 1987. Parallel algorithms for computer vision on the connection machine. *Artificial Intelligence Laboratory Memo*, **928**. Cambridge: MIT Press.
- Luo, J., C. Koch, and C. Mead. 1988. An experimental subthreshold, analog CMOS two-dimensional surface interpolation circuit. Oral presentation at the Neural Information Processing Systems Conference, Denver.
- Mahowald, M. and T. Delbrück. 1989. Cooperative stereo matching using static and dynamic features. In: *Analog VLSI Implementations of Neural Systems*, eds. C. Mead and M. Ismail. Norwell, MA: Kluwer.
- Marr, D. 1982. *Vision*. New York: Freeman.
- Marr, D. and E.C. Hildreth. 1980. Theory of edge detection. *Proceedings of the Royal Society of London B*, **207**, 187–217.
- Marr, D. and T. Poggio. 1976. Cooperative computation of stereo disparity. *Science*, **194**, 283–287.
- Marroquin, J., S. Mitter, and T. Poggio. 1987. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, **82**, 76–89.
- Mathur, B.P., H.T. Wang, C.T. Tsen, and E. Walton. 1988. Variable scale edge detection of images. *IEEE Conference on Circuit Design*, Rochester.
- Mead, C.A. 1988. Plenary talk. *Proceedings of the International Neural Network Society*, Boston.
- . 1989. *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley.
- Mead, C.A. and M.A. Mahowald. 1988. A silicon model of early visual processing. *Neural Networks*, **1**, 91–97.
- Poggio, T., E.B. Gamble, and J.J. Little. 1988. Parallel integration of vision modules. *Science*, **242**, 436–440.
- Poggio, T. and C. Koch. 1985. Ill-posed problems in early vision: From computational theory to analogue networks. *Proceedings of the Royal Society of London B*, **226**, 303–323.
- Poggio, T., V. Torre, and C. Koch. 1985. Computational vision and regularization theory. *Nature*, **317**, 314–319.
- Poggio, T., H. Voorhees, and A. Yuille. 1986. A regularized solution to edge detection. *Artificial Intelligence Laboratory Memo*, **833**. Cambridge: MIT Press.

- Proceedings of the First International Conference on Computer Vision*. 1987, Washington, D.C. London: IEEE Computer Society Press.
- Sage, J.P. 1984. Gaussian convolution of images stored in a charge-coupled device. *Quarterly Technical Report*, August 1–October 31, 1983, 53–59. Lexington: MIT Lincoln Laboratory.
- Sage, J.P. and A.L. Lattes. 1987. A high-speed two-dimensional CCD gaussian image convolver. *Quarterly Technical Report*, August 1–October 31, 1986, 49–52. Lexington: MIT Lincoln Laboratory.
- Sivilotti, M.A., M.A. Mahowald, and C.A. Mead. 1987. Real-time visual computation using analog CMOS processing arrays. In: 1987 Stanford Conference on VLSI, 295–312. Cambridge: MIT Press.
- Strausfeld, N. 1975. *Atlas of an Insect Brain*. Heidelberg: Springer.
- Tanner, J.E. 1986. *Integrated optical motion detection*. Ph.D. thesis, Department of Computer Science, 5223:TR:86, Caltech.
- Terzopoulos, D. 1983. Multilevel computational processes for visual surface reconstruction. *Computer Vision Graphics Image Proceeding*, **24**, 52–96.
- . 1986. Regularization of inverse problems involving discontinuities. *IEEE Transactions Pattern Analysis Machine Intelligence*, **8**, 413–424.
- White, J., B. Furman, A.A. Abidi, R.L. Baker, B. Mathur, and H.T. Wang. 1988. Parallel analog architecture for 2D gaussian convolution of images. *Proceedings of the International Neural Network Society*, Boston.
- Wyatt, J.L. and D.L. Standley. 1989. Criteria for robust stability in a class of lateral inhibition networks coupled through resistive grids. *Neural Computation*, **1**, 58–67.