Photoneural systems: an introduction

R. Victor Jones

Recent developments in technology make feasible the implementation of hybrid optoelectronic processors designed to mimic directly neurological processes of cognition. This paper discusses the practicability and potential utility of a particular class of photoneural device and architecture which is based on neural models of sensory perception. Perception is mediated by a vastly complex hierarchy of processes, but each step in the hierarchy seems to involve relatively simple antagonistic responses which remain topographic coherent when mapped to various subsystems of the brain. The topographic invariance of cognitive data transmission is suggestive of optical imaging operations and is the essential rationale for the design strategy discussed here.

I. Introduction

Any effort to develop artificial intelligence is in some sense an attempt to imitate nature. The approach introduced here is a more specifically imitative effort to draw on the lessons of evolutionary development in the design and implementation of intelligent processors. We characterize this approach as *photoneural* since it builds on the premise that optic or optoelectronic operations can be configured to mimic elemental mechanisms of the central nervous system. Photoneural processors are envisaged as realizations of cognitive functions within which optical propagation provides adaptable neural-like interconnections to link a manifold of synaptic-like processors which are implemented as autonomous elements of monolithically integrated optoelectronic arrays.[1]

At the moment, photoneural design is but a speculative concept. In this paper, we undertake to establish a case for the development of photoneural devices and systems. There are two aspects of the case. First, we argue that recent advances in optoelectronic technology make practicable photoneural device structures eminently feasible. Second, we argue that photoneural processors could be useful as front-end processors to reduce the inherent computational burden in machine perception and that the study of photoneural systems could have an impact on cognitive research.

Section II sets a conceptual framework for photoneural design. While a superficial view of a vast body of research in psychophysics and neurobiology, it does introduce several key metaphors or constructs widely used in the neurosciences. The basic photoneural device concepts are developed in Sec. III with an emphasis on heterojunction realization of the optoelectronic equivalent of nerve cells. Section IV is a brief overview of how basic photoneural devices might be integrated into a photoneural system to accomplish perceptual tasks. The spirit of this paper is didactic and stresses simple ideas, since the photoneural concept is an attempt to integrate relatively well-developed precepts of disparate fields of study, namely, computer science, microelectronics, neurophysiology, psychophysics, optical communication, and optical information processing. A complete citation of all germane work is impractical, and thus only the most relevant or seminal references are included.

II. Perspective from the Neurosciences

At a biological level, much is known about the vastly intricate cellular mechanisms and patterns of interconnection in the central nervous system (CNS).[2,3] The profound scientific challenge is to understand how these constituent processes globally aggregate to effectuate mentation (perception, learning, memory, consciousness, self-awareness). Over the last three decades, ingenious experiments in the neurosciences have established an increasingly rich conceptual framework for the description and analysis of neural operations. Important elements in this framework are detailed biological models of neuronal signal propagation and interneuronal synaptic interactions, refined cytoarchitectures of significant subsystems of the CNS, resolved patterns of cellular interconnections between and within neural subsystems, psychophysi-

---

The author is with Harvard University, Division of Applied Sciences, Cambridge, Massachusetts 02138.

Received 23 July 1986.
0003-6935/87/101948-11$02.00/0.

cal and neurophysiological models of stimulus/response patterns in sensory perception, and a growing appreciation of both phylogenetic and ontogenetic factors in the development of neural structures and interconnections. With an audacity born of partial knowledge, we venture to reduce this great and subtle body of knowledge to a few metaphorical principles of neural organization. Furthermore, we assert that these architecture metaphors can serve as a valid, indeed compelling, basis for a cognitive processor design strategy.

The emerging picture of the neural organization of the CNS is not readily characterized in the argot of the computer architect. Information streams converge and diverge in exquisitely complex patterns so that the designations, serial, parallel, and distributed, all seem to apply to some limited extent but are not particularly helpful as design metaphors. We suggest here that the designation permeative architecture better describes neural organization, wherein the execution of several distinct processes simultaneously permeates a conjoint set of processing elements.

Visual and somatic sensory perceptions are the most comprehensively understood cognitive functions. In this brief commentary, we highlight insights drawn from the study of the early phases of vertebrate vision.[4,5] In particular, consideration of the cytoarchitecture of the vertebrate retina and of retinal signal transformations reveals the operation of cellular mechanisms and organizations which may be inferred to operate in other less accessible structures of the brain. As schematized in Fig. 1 (a), the initial light intensity distribution incident on the eye is transformed at the receptors (R) into a spatially encoded neural signal.[6–8] The encoding of the visual stimulus is transformed as the information passes through a hierarchical sequence of cellular layers from receptors to the axons of the ganglion cells (G), which aggregate to form the optic nerve that carries the retinal output to the brain. Although the code transformation at each intercellular synapse is simple, mechanisms of *lateral inhibition* among neighboring cells lead to output responses which accentuate contrast. An overwhelming body of evidence in both psychophysics and neurophysiology supports the premise that visual perception builds on fundamental neural processes which are responsive to stimulus change and relatively insensitive to stimulus level.[4,5] The first perceptual steps, in particular, are described as responses to spatial antagonism. At the lowest level, lateral inhibition is the cellular locus of the perceptually described *center/surround* antagonism in which a given ganglion cell is sensitive to the contrast in excitation between a particular central field and its enveloping surround region. At higher levels, lateral inhibition leads to discriminatory responses of increasing spatial and motional complexity.

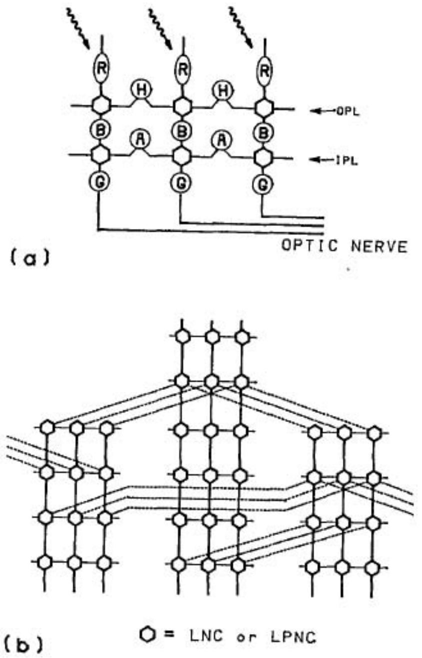

Fig. 1. Neural architecture. Schematic representations of interconnection patterns linking local neural circuits (LNC) or local photoneural circuits (LPNC): (a) cytoarchitecture of retinal subsystem; (b) three subsystems of a large sensory system.

Ultimately, the basic datum of neural information is the value of the electrical potential across a cell membrane at a particular location in the neural network. In the retina, as in all parts of the CNS, there are two distinct formats for representing and communicating information, namely, *slow* and *action* potentials.[6–8] The receptors are, of course, the basic photoelectric transducers which transform the light stimulus into neural excitation. An increase in incident light intensity within the receptor field invokes an analog or graded hyperpolarization (decrease) in the level of the receptor membrane potential. Any changes in receptor excitation spread conductively along the membrane to all parts of the cell, but the level of excitation decays with the distance from the point of excitation. By convention, this diffusive analog mode of data transmission is characterized as the electrotonic spread of slow potentials. At the other end of the retina chain, as is more common in the nervous system, information in the ganglion cells is represented and transmitted as nerve spikes or short pulses of fixed magnitude. Although the pulses occur in bursts, the spiking rate is a direct measure of the excitation level. As the spikes or action potentials propagate along the nerve, mechanisms of electrochemical regeneration act to maintain constant pulse shape and amplitude. The action potential mode of data transmission is a kind of pulse interval coding and has obvious advantages for robust communication over long distances. Ironically, information is transferred at a very much faster rate by spreading sustained potentials in the slow potential mode, since electrochemical regeneration significantly slows the speed of neural propagation.

At the simplest perceptual level, the receptors show a maximum response for spot illumination centered on the cell and relative little response for an annular illumination surrounding the cell; i.e., the incident light distribution is spatially encoded or sampled. In contrast, the interconnecting horizontal cells (H) have very broad receptive fields with spot and annulus producing equivalent hyperpolarizing responses. The responses of two classes of bipolar cells (B) exhibit the first hints of significant stimulus differentiation. In one class, the cell is hyperpolarized by the spot stimulus to signal an *on* response, and in the other the cell is depolarized to signal an *off* response. The crucial point is that in both cases a simultaneous annular illumination antagonizes the potential change and reduces the effect of the center illumination. In hierarchical terms, the outer plexiform layer (OPL) seems to mark the first significant step in the process organization. Receptor, bipolar, and horizontal cells are connected by a complex network of presynaptic and postsynaptic processes within a relative thin physical layer, and this synapse generates a clearly identifiable center/surround response in the bipolar cells. At this first layer of the visual hierarchy, the temporal character of the stimulus does not appear to be significant. In striking contrast, the amacrine cells (A) respond only to temporal changes in illumination and are the first visual processes to exhibit temporal antagonism and adaptivity. The inner plexiform layer (IPL) marks the second significant step in the visual hierarchy. Again, the bipolar, amacrine, and ganglion cells are connected by presynaptic and postsynaptic processes within a relatively thin physical layer. In most vertebrates, there are two distinct types of ganglionic response. The cells of one population (X-cell) synapse mainly with bipolar cells and mirror the spatial and temporal response of those cells. The synaptic process is essentially an analog-to-digital converter that encodes the slow potential presynaptic excitation as an action potential postsynaptic excitation for transmission to the brain. The ganglion cells of other populations (Y-cell) synapse to some degree with amacrine cells and hence mirror the transient temporal behavior of those cells.

Many of the axons of retinal ganglion cells project to the various layers of the thalmic lateral geniculate nuclei (LGN). Cellular recordings of potentials in the LGN show responses akin to the ganglion responses. In particular, the concentric center/surround antagonism is the dominant response but with a heightened contrast. The LGN functions essentially as a neural relay center, and the principal cells of the LGN radiate to various layers of the visual cortex. As was first shown in the famous work of Hubel and Wiesel, cortical cells continue to exhibit responses of spatial antagonism.[5,9] However, the patterns of sensitivity are quite diverse and not unambiguously resolved. The least ambiguous responses are found in the so-called simple cells which are most sensitive to bar or line patterns of specific angular orientations and at particular retinal locations. The complex cells of the visual cortex are also sensitive to lineal properties, but the position of stimulus in the retinal field is less critical.

This sort of detailed neurobiological analysis of perceptual processes provides the context for framing the metaphorical principles of neural architecture. An important, elegant, and yet elusive metaphor is the concept of local neural circuits or LNCs.[10] The term is defined and used so broadly it appears at the outset to convey little of value. The physical boundaries of a LNC are functionally defined and might encompass a portion of a single nerve cell, an aggregate of many cells, or a module of many identical local circuits. The OPL and IPL synapses are prototypical of the simplest type of LNC. However, the scale of LNC aggregation ranges from such simple intercellular synaptic junctions to columnar organized components of large entities of the brain (e.g., neocortex, dorsal thalamus, limbic lobe). A LNC functions as an independent integrative unit which processes information available at a specific set of convergent afferent (inputs) terminals, but the character of the response available at a set of divergent efferent (output) terminals depends on the state or condition of the component elements. The internal wiring of a LNC reflects the influences of both genetic and epigenetic cell development.[3] When a particular neural element (e.g., synapse, neuron, or local circuit) is incorporated in more than one LNC, activity in one circuit *conditions* the response of others, and it thereby provides a mechanism for cross-linking signals in independent neural pathways. This intricate cross-linkage among distinct processes is what suggests a permeative model of process organization, wherein distinct and numerous processing sequences share and interact within a conjoint set of cellular elements (i.e., the processes permeate the cells).

The metaphor of permeative neural organization is an alternate way of expressing the prevalent view that cognitive processes are mediated by multiple and dynamic interactions of neural subsystems distributed over many parts of the brain.[2,3] Figure 1(b) schematically depicts a portion of such a multiply interconnected network, wherein LNCs are symbolized as hexagonal cells. Although intended as a very general representation of neural architecture, the center subsystem is drawn to suggest a retina interconnected to higher-order brain centers (e.g., to the lateral geniculate nucleus and on to the visual cortex). The figure emphasizes the kind of organizational dichotomy between centrally (vertical) and laterally (horizontal and diagonal) directed neural pathways found in retinal organization. It is a striking and powerful observation that this dichotomy seems to be valid and useful in the study of cellular organization in other subsystems and for analysis of processes at different levels of abstraction. The interpretation at a high level of abstraction takes the local circuits as large entities of the CNS, the centrally directed pathways as the flow of information in the various sensory modalities through a hierarchy of perceptual stages of ever increasing complexity, and the laterally directed pathways as intermodality or interentity interactions. An interpretation at a lower level takes, for example, the local circuits as neural synapses,[8] the centrally directed chains as signal pathways of the receptive field of a particular sensory receptor, and the laterally directed chains are cross-comparative operations.

One aspect of the dichotomy is that centrally directed processes maintain some kind of topological invariance. The first steps in visual perception (i.e., responses of ganglion cells) preserve strict topographic relationships. A stimulus attribute (e.g., center/surround-on) activates a ganglion cell in close proximity to the initial stimulus. Given the structure of the eye, this organization does not seem surprising. It is surprising that organization continues in the lateral geniculate nucleus and on into areas of the visual cortex.[10] The attributes of the stimulus are refined, but they continue to be organized topographically.

Neurophysiologists over the last decade have been able to identify many different examples of columnar structures of cortical cells which function as modular elemental sensory processors.[3] Of particular significance are the much studied columnar structures in area 17 of the visual cortex. These structures are the examples par excellence of the permeative model neural of organization. In brief, a column is a cortical representation of a particular stimulus attribute (e.g., ocular dominance, line orientation, or color) which preserves topologic or retinotopic relationships. These columns are LNCs and as such function as input–output processors. However, columns representing different attributes infuse or permeate the same cortical space, and hence local lateral processes lead to higher level responses which preserve topological relationships. The processing of a particular attribute may also occur in parallel at other cortical locations or in other entities (e.g., lateral geniculate nucleus). Again at each location it appears that the processes are organized into patterns which reflect the topographical relationships within the initial sensory stimulus. Since the parallel process seems to be dynamic and highly interactive, it may be inferred that the distribution network plays a key role in maintaining this topographic order throughout the system.

In the sensory realm at least, mapping or topographical imaging is an important characteristic of neural communication.[11] It is this observation which leads us to believe that optics offers some unique opportunities in dealing with tasks in artificial sensory perception.

III. Photoneural Device Structures

The photoneural strategy developed here is based on the proposition that the permeative metaphor of neural process organization can be effectively interpreted in an adaptive optical system.[1] Accordingly, Fig. 1(b) is reinterpreted as a schematic representation of a portion of a multiply interconnected distributed photoneural system of local information processors interacting via lightwaves propagating through a progression of encoded optical channels. The processors are hybrid optoelectronic interpretations of the previously discussed local neural circuits and are henceforth referred to as local photoneural circuits or LPNCs.

In the most direct interpretation, the channels are spatially encoded operations by which the output of one LPNC is imaged at or broadcast to the input of another. The basic data type in a photoneural system is a spatial-encoded light intensity distribution, which is taken to represent a sensory or cognitive response at a given hierarchical level within a specific processing module. A basic data transfer involves the mapping of a complete representation of the response at one level to other levels of the same module or to other processing modules by means of optical imaging techniques. The word mapping is used to emphasize the notion that imaging operations maintain topological relationships within the data type. It is essential to note that a single mapping operation may be multiplexed to carry many topological representations in parallel, since the light intensity distribution may be independently wavelength (color), polarization, and/or time encoded. Thus light-mapped cognitive data structures may carry several stimulus attributes in parallel. This aspect of photoneural organization is of critical importance, since, as noted in Sec. II, such multidimensional mappings seem to be precisely the kind of permeative process organization found in the columnar structures of the cerebral cortex.

A. LPNC Design Considerations

Clearly, the local processing unit—the LPNC—is the crucial element in the photoneural concept. It is the optoelectronic analog of a single neural synapse. Specifically, an individual LPNC unit performs three functionally distinct operations. The initial or input operation includes both the decoding and photoelectric transduction of the incident optical signals. The intermediate or processing operation is a simple electronic circuit which mimics the integrative transformation of an analogous local neural circuit. Following the synaptic analogy, the output of the electronic circuit in most cases is a discriminatory or antagonistic response but is adaptable or programmable to allow for epigenetic development of the overall system response.[3] The final or output operation involves the reencoding of the electronic response as lightwaves for transmission to other LPNCs.

Given this somewhat abstract general specification of LPNC function, one might imagine at outset many equally attractive realizations. However, the large-scale and complex connectivity of any potentially interesting photoneural system imposes critical constraints on the LPNC design. Above all, LPNCs must be simple to produce in very large numbers and accessible by standard optical imaging techniques. Both requirements suggest realizations based on planar integrated organization and microfabrication methodology. Indeed the photoneural concept has been inspired and is made practicable by the recent advances in heterojunction technology. A heterojunction is an abrupt epitaxial interface between two different semiconductors or semiconductor alloys, where, in most cases of importance, the band gap energies of the two semiconductors are substantially different.[12] Of primary interest here are heterojunctions formed by quaternary alloys of group III–V compound semiconductors with matched lattice spacings. Wide interest in the physics and device potentialities of heterojunctions materials has spawned and continues to motivate major efforts to produce layered heterostructures. Of particular importance has been the development of sophisticated thin-film epitaxial crystal growth techniques.[13,14] Given the technology base, the options in integrated optoelectronic device design have become exceedingly flexible. Following Kroemer,[15,16] we may take as a technological premise the availability of almost any kind of multilayered structure of multicomponent III–V compounds of high crystal perfection with precise control over composition, composition profile, doping levels, and layer thickness if there is sufficient reason to build it.

There is a further general comment on LPNC design. In various potentially feasible LPNC realizations, the particulars of input and processing operations may differ in significant ways, but it is the output operation which poses the most difficult technical problems and which has the greatest impact on overall system design. In one class of device, light generating local photoneural circuits (LG-LPNCs), the output electrical response controls the drive current of a light emitting diode (LED) to generate a light response directly. In a second class, light switching local photoneural circuits (LS-LPNCs), the output electrical response controls the voltage across an electrooptic switch element, and hence the response is read by an externally generated light probe. The advantages of LG-LPNC realization are twofold: design and fabrication require a minor extension of current integrated circuit capabilities, and the individual units are highly autonomous in the sense of requiring only power supply connections. LG-LPNC structures are technically elegant, since all three operations are implemented within a single microcircuit of a simple heterostructure configuration. Of course, the disadvantages of LG-LPNC realization arise from the inherent inefficiency of injection electroluminescence.[17] The heat generated by an individual unit is proportional to but orders of magnitude greater than signal levels. Thus systems incorporating LG-LPNCs must be designed to run at low average light levels, to minimize thermal degradation, and to maximize electroluminescent efficiency.

Since electrooptic effects are voltage dependent, light switching requires minimal control power and hence obviates most of the thermal dissipation problems associated with light generation. The major problem, of course, is that the technology for an effective fast integrated semiconductor electrooptic switch is not presently available. In the long term, the development of quantum-well or excitonic switches may make a truly integrated LS-LPNC feasible.[18] In the near term, however, an eclectic mix of semiconductor and liquid crystal technologies affords a reasonable compromise. Discussion of LS-LPNC realization is complicated by the fact that the overall system design must accommodate light beams for reading the switch outputs. However, there are several convenient ways to integrate semiconductor laser read beams into LS-LPNC arrays which would greatly facilitate spectral multiplexing of photoneural operations. In light of the rapid developments in LCD technology, the electrooptic option appears promising and is discussed in detail elsewhere. However, the discussion here stresses systems concepts, and thus the emphasis is on electroluminescent devices.

B. LG-LPNC Realization

Examination of a specific LPNC realization best illustrates key photoneural design issues. One promising possibility is a device structure inspired by reports of a new and innovative class of image processing devices.[19–24] These devices are monolithically integrated planar arrays of optoelectronic image elements within which incident light signals are amplified and, if need be, shifted in wavelength. Each image element of the array is an electrically isolated multilayered semiconductor heterostructure. The top and bottom layers of the heterostructure are optically transparent electrical conductors connected to effectuate a voltage bias across the element. A sequence of three layers within the structure form in turn the emitter, base, and collector of a bipolar heterojunction phototransistor (HPT).[25] Light incident on the top surface of the element passes through the emitter into the base region of the HPT where it is absorbed and generates excess electron–hole pairs. The excess photogenerated minority carriers are drawn off by a reverse bias across the base–collector junction, and the resultant charge separation induces a small photovoltaic shift in the base potential. Since the base is floating, this shift effects an exponential increase in the forward current across the base-emitter. If emitter injection efficiency is high and the base region thin, most of this light-induced forward current is collected at the base–collector junction to yield a high-net photoelectric gain. The photocurrent generated and amplified within the HPT sequence flows directly into an adjacent second sequence of layers which form a double-heterojunction light emitting structure designed to maximize the overall injection electroluminescent efficiency and to achieve specific output spectral characteristics. Finally, the amplified and spectrally shifted optical signal is reradiated out through the bottom surface of the element.

Successful implementation of these devices demonstrates at least two significant aspects of heterostructure physics. A fundamentally important transport property is the marked asymmetry in majority carrier injection across anisotypical heterojunctions (i.e., *pn* heterojunctions). The asymmetry varies exponentially with the difference in band gap energies of two semiconductors in the sense that the forward current is dominated by the component injected from the wide-gap region into the narrow-gap region. In the heterojunction version of a phototransistor with a narrow-gap base region the asymmetry leads to a high-emitter-injection efficiency which is relatively insensitive to other design parameters. Taken together with the dimensional and compositional control afforded by epitaxial methods, this enhanced emitter efficiency has generated a resurgent interest in bipolar technology in general and phototransistors in particular.[15,25]

As these devices demonstrate, the shift in intrinsic optical absorption edge at a heterojunction provides the means for spectral control of optical penetration and, thus in principle, the means for injecting signal information at any depth within a heterostructure. In this sense, HPTs afford a unique opportunity for 3-D process integration. Specifically, the spectral response or sensitivity of a HPT sequence is relatively constant over an activation band bounded by two band gap energies. Incident light with photon energy greater than the band gap of the emitter layer is absorbed within an absorption length from the surface of the device, and thus a device with a moderately thick emitter region has a minimal photo response. Light with photo energy less than the emitter band gap, but greater than the base band gap, has maximal response, since the light penetrates through the emitter region, and photogeneration occurs close to the base–collector junction in the thin base region. Light with photon energy less than the base band gap passes through the complete sequence without appreciable interaction.

These high-optical-gain amplifiers demonstrate the feasibility of the input/output operations of a LG-LPNC. In addition, they exhibit a mode of vertical organization akin to that found in the retina and in cortical tissue. The major design issue is to find a means to simply effectuate the processing operation within such vertically organized structures. As previously noted, the essential processing operation is an integrative response activated by antagonistic or contrasting stimuli, e.g., the center/surround response. In particular, we need an optoelectronic means for realizing lateral inhibition. Many schemes for doing logic have been discussed. For example, basic photodetector–photoemitter combinations have been used to realize logical AND and OR operations.[26–28] However, inhibition requires in effect the EXOR or NOT-AND operation. Said another way, inhibition requires effectively the INVERSE operation to produce a subtraction. In the realm of incoherent optics, the only variable is the unipolar distribution of the light intensity, and thus real-time subtraction of images is a difficult problem requiring some mechanism of contrast reversal.

We have developed,[1] however, several simple optoelectronic modes for realizing lateral inhibition based on the paradigm of inversion by light activated current stealing. Figure 2(a) illustrates the operation of the most promising version of the paradigm in which input photodetectors and an output photoemitter are arrayed in a single-node star configuration. One limb of the star is the conduction path from node to ground through a LED. The other limbs are paths through a set of HPTs to various supply voltages. Some of the HPTs are biased to provide, when light activated, a source of drive current for the LED, while others are biased to sink or steal the excitation current when light activated. The light incident on the first group may be characterized as an excitatory stimulus and that on the second as an inhibitory stimulus. When the net stimulus invokes equal currents in the two HPTs, the voltage at the central node is close to ground (resting potential), and no current flows in the LED. As the excitatory light increases above this quiescent level, the voltage at the node rises to the threshold cut-in voltage of the diode. Above threshold, the light output of the LED is a direct measure of the net excitatory stimulus. The biasing voltages may be separately controlled to compensate for asymmetries, to modify thresholds, or to change overall functional characteristics.

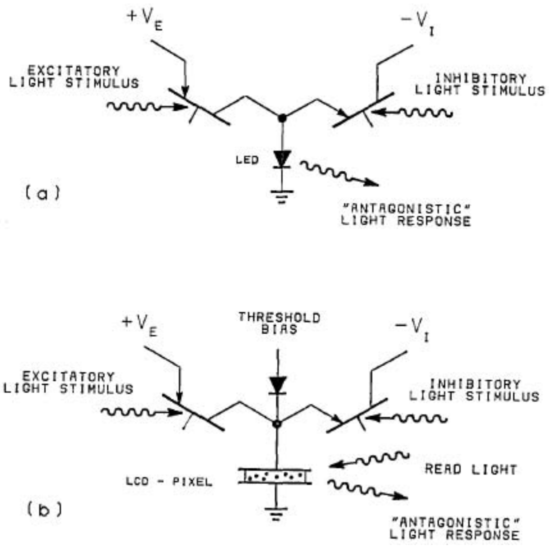

Fig. 2. Discrete circuit representation of star configured local photoneural circuit or photoneural synapse: (a) light generating version (LG-LPNC); (b) light switching version (LS-LPNC).

Figure 2(b) depicts an obvious light switching variant of the basic star configuration wherein the LED is replaced by a single element or pixel of a liquid crystal display. For net excitatory activation, current flows to the electrode of the pixel, and the stored charge determines the state of the optical switch. A net inhibitory excitation pulls the nodal voltage below a preset threshold level, and the inhibitory current leaks off through the diode leaving the pixel uncharged. Since most modes of liquid crystal switching are functions of the magnitude of field across the electrode, the diode provides an essential nonlinearity in the LS-LPNC response. The circuit has a short-term memory in the sense that the charge on the electrode is the integral of the net excitatory current over a time period determined by the leakage RC constant of the switching element.

The antagonistic response of the basic star configuration closely mimics the operation of a neural synapse.[2,3] In fact, the physical mechanisms active in this device are quite analogous to mechanisms operating in neuronal membranes. The transistors as back-to-back diodes electrically isolate the central node in much the same way that the mechanism of selective ionic permeability in the cellular membrane electrically isolates the interior of a neuron. The stimulus-dependent conductive response of the phototransis-

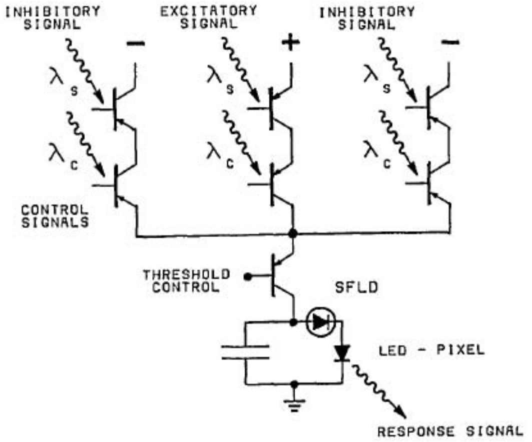

Fig. 3. Discrete circuit representation of an enhanced version of the basic, LG-LPNC star configuration. This configuration has a pulse-rate encoded response which may be progammed dynamically by means of external electrical and/or optical signals.

tors operates akin to the chemical gating of ion channels by neural transmitters. The nonlinear response of the LED (or the bias diode in the light switching variant) to changes in nodal potential is similar to the characteristic excitable response of the voltage-gated ionic channels to changes in the membrane potential.

Building on this rather direct neural analog, Fig. 3 shows a discrete circuit representation of a modification to the basic configuration that incorporates several important functional enhancements. One enhancement makes possible multispectral control of the LPNC response. As noted above, the response of a given HPT sequence is spectrally bounded by the band gaps of the emitter and base regions. However, by judicious selection of semiconductors two or more separately addressable HPT sequences may be stacked one above the other. A light signal penetrates a HPT sequence if its photon energy is less than any band gap in the sequence. In Fig. 3 two spectrally distinct HPT sequences are schematized as series elements, and the response of the particular star limb is essentially the product or correlation of the excitations at two distinct wavelengths. It is assumed that the signals labeled excitatory/inhibitory are at wavelengths within the activation band of the upper HPT and are completely absorbed there. The signal labeled control is assumed to be at a wavelength within the activation band of the lower HPT but long enough to allow for penetration of the upper layers. Thus the longer wavelength control signal acts as a switch or gate to activate the circuit's response at shorter wavelengths. This easily achieved multispectral response characteristic adds greatly to the potential utility of heterojunction LPNC realizations. The control signal may be used to activate a single device, to program dynamically the operation of a whole array, to carry in parallel distinct stimulus attributes, to produce contrast reversal, or to effectuate adaptive feedback responses.

In the simple star configuration the LPNC threshold is fixed at the cut-in potential of the LED. A second enhancement in Fig. 3 is a transistor switch in series with the LED which provides the option of independent control of the threshold level determined by the voltage on the externally available base lead. This control may be used in a variety of ways to facilitate adaptive operations. For example, if the base voltage is derived from a RC voltage divider connected between the central node and ground, the switch will be cut off for steady-state excitations, and the circuit will only respond to transient excitations.

The operation of the simple star configuration much resembles the slow potential operation of the retinal receptor, horizontal, and bipolar cells as discussed in Sec. II. However, a third enhancement, in the form of a capacitor shunted by a Shockley four-layer diode (SFLD)[29] in series with the LED, transforms the operation to an astable mode which resembles the more common neural spiked potential operation of the amacrine and ganglion cells. Initially a net excitation results in a charging current to the capacitor, and the LED is isolated by the SFLD in its high impedance state. When the voltage across the capacitor exceeds the SFLD threshold, the SFLD switches to its low impedance state, and the capacitor discharges rapidly through the LED. After discharge the SFLD returns to the high impedance state, and the capacitor is recharged. Thus the light pulse rate is a direct measure of LPNC excitation.

Figure 4 suggests an integrated realization of the LG-LPNC represented as a discrete circuit in Fig. 3. The fabrication of the array might start with a substrate composed of a matrix of highly conducting islands (p^+) formed by conventional diffusion of acceptor impurities through a semi-insulating (SI) wafer of, say, intrinsic GaAs. All other elements are deposited in a single epitaxial growth run. (Unfortunately, a two-sided deposition is required.) The inset indicates the essential layer sequences where the notation WG (wide gap) designates a III–V alloy with a larger band gap energy, and NG (narrow gap) designates a smaller band gap material. The upper five layers form two HPT sequences with the band gap of NG'' less that of NG'. The incident light excitation passes through the transparent contact pads and is absorbed in the appropriate layer. Each LED pixel is electrically isolated from the rest of the array by the ion beam damaged regions DR_2 and DR_3. The HPT sequences are electrically isolated by the damaged regions DR_1 into, for example, nine independent elements. As indicated, electric supply buses are superimposed on the transparent conducting layer, and contact patterns are etched in that layer. In this case, the pattern is a center/surround-on (or -off) configuration with the center HPT biased positive (or negative) and the surrounding eight HPTs biased negative (or positive). The lower surface is etched so that the layer sequence associated with the SFLD and LED are on a small mesa. All the lower surface, except for the top of the mesa, is covered with an insulating layer, and a capacitor is formed between the p^+ layer and the grounded

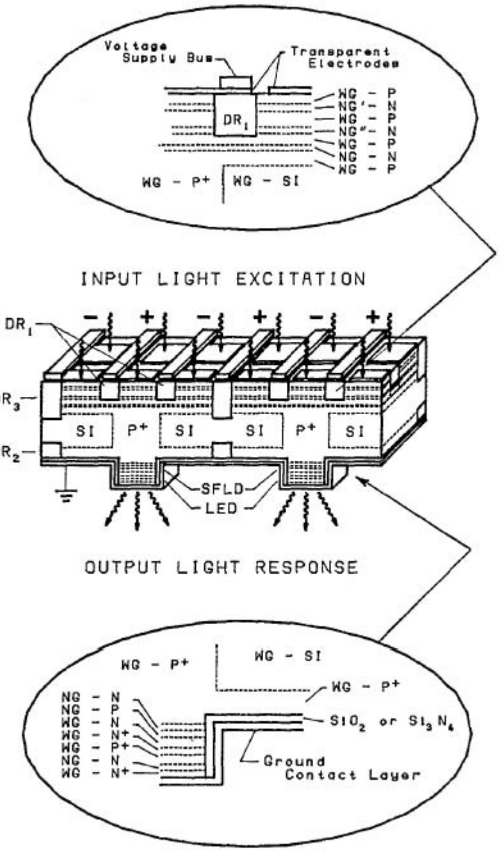

Fig. 4. Two pixels of an integrated heterojunction version of a circuit in Fig. 3: *NG*, narrow band gap material; *WG*, wide-band gap material; *SI*, semi-insulating material; *DR*, ion beam damaged region of high-resistance material; SFLD, Shockley four-layer diode.

contact layer. Note that several pixels may be aggregated into a larger unit by merely reducing the depth of the ion damage in the appropriate DR_3 regions.

IV. Photoneural Systems

There is little doubt that LPNC arrays could be realized with current technology. It is equally evident that such arrays could find direct application as simple front-end processors in machine vision systems. In particular, the device structure discussed above could be configured statically or programmed dynamically to accomplish spatial filtering operations or to function directly as a 2-D motional, depth, or feature (e.g., edge, line, line termination, blob) detector. In this relatively modest context, one particularly attractive possibility is the construction of fast optoelectronic focal plane Gabor filters for optical data compression. Neurophysiological measurements in the cat striate cortex and psychophysical tests with human subjects have established the importance of representations of the visual field which involve localization in both space and spatial frequency, i.e., Gabor representations.[30,31] In the cat, there is a class of cortical cell which individually responds only to light patterns localized in a small spatial field and is characterized by a narrow band of spatial frequencies within a narrow orientational domain. A Gabor LPNC could be realized simply as a set of alternating strips of excitatory and inhibitory biased phototransistors driving a single output device. The extent of the set determines the spatial localization, the strip orientation affords orientational sensitivity, and the strip spacing sets the spatial frequency response. The Gabor filter could be permanently configured by burning-in appropriate electrical contacts or dynamically configured with control light patterns, as discussed in Sec. III. Thus photoneural systems involving a few layers of LPNC processing are a feasible and potentially useful means for accomplishing a variety of preprocessing image transformations in computer vision and to simplify down-stream computational tasks. In a sense, such a system functions as a fast, sensitive, simple, and compact artificial retina.

Assessment of the potential utility of larger-scale photoneural systems is at the moment a matter of pure speculation. As suggested, photoneural interpretations of known cognitive processes appear to be feasible. It is, however, difficult to gauge the potential promise of the photoneural approach *vis-a-vis* more conventional all-electronic computational methods. Computer implemented cognitive research, e.g., machine vision, speech recognition, associative memory, is currently an exceedingly active endeavor. At a theoretical level, this research has stimulated wide interest in learning algorithms, cognitive data representations, and cooperative phenomena in neural networks. At a hardware level, the computational burden inherent in these problem areas has been one of the driving forces in the development of high-speed parallel computational systems. Of great importance and promise is the development of special purpose VLSI structures to support new parallel processing algorithms.

While photoneural realization of certain cognitive functions has, in principle, distinct advantages, the issue is one of technological catch-up. The maturing of photoneural ideas requires the maturing of a new technology and, perhaps more important, the development of a new mind set. The practical incentive for photoneural development lies in the direct application of simple systems in low-level visual perception. But given the availability of LPNC arrays, more complex processes would be uncomplicated to model and fascinating to study in an exploratory mode. To make photoneural system concepts more concrete, we conclude with a brief outline of some optical interconnection strategies and comments on possible applications in vision systems.

Figure 5 schematically suggests a possible approach to photoneural implementation. For simplicity, the organization is illustrated with planar arrays of autonomous LG-LPNC elements (labeled *AR*-1 etc. in the figure). As previously noted, the necessity of incorporating a read light in the LS-LPNC-based realizations slightly complicates the system description, if not the implementation, and is not discussed further here. The optics of the implementation is based on, what is now, relatively standard practice.[32] To keep the overall system compact, mechanically stable, and yet flexi-

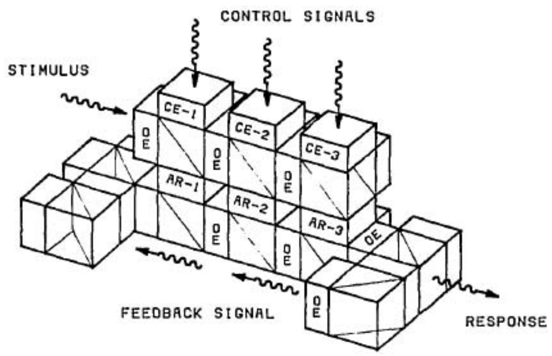

Fig. 5. Schematic view of optical flow in part of a photoneural system: *AR*, a LPNC processing array; *CE*, a LPNC or LCD control element; *OE*, an imaging optical element.

ble, the paths of possible optical data flow are determined by a configuration of beam-splitting cubes and confined within this configuration by appropriate optical imaging elements (labeled *OE* in the figure). All necessary processing, imaging, and control elements, such as LPNC arrays, holographic optical elements, graded-index optics relay lenses, and conventional LCDs may be mounted conveniently at the interfaces between the cubes. In some parts of the system LPNC arrays might be configured to accomplish fundamental data transformations. At other locations LPNC arrays as well as conventional LCDs function as control elements (labeled *CE* in the figure), relay cells, or switching matrices to control signal flow dynamically.

To illustrate some possible processing sequences, suppose a 2-D sensory data pattern enters the system as a signal labeled as stimulus in the upper left corner of the figure. The stimulus is divided by the upper sequence of beam splitting cubes into three copies, which in turn are processed in parallel by the LPNC arrays labeled *AR*-1,2,3. The individually processed copies pass out of this part of the system through the bottom faces of the lower sequence of cubes, while a composite of the three processed signals, labeled response in the figure, flows to other parts of the system or is routed back as a possible feedback signal. If the input is a visual stimulus, the three arrays, for example, might be configured to process simultaneously center/surround responses at three different spatial resolutions or feature responses for three distinct features. The control elements labeled *CE*-1,2,3 control the biasing light distributions incident on the individual processing arrays labeled *AR*-1,2,3. As discussed in Sec. III, the output response of the array is essentially the product of the incident stimulus and control distributions. The composite response and associated feedback signal is thus a complex convolution of the product responses of three arrays and the image transformations associated with the intervening optical devices. The control light distributions are a means to program dynamically the spatial response of each LPNC array. There are many possible ways to use this mode of real-time control of array response, but the most interesting possibility is the use of control signals derived from internal feedback to effectuate epigenetic development of the complete system response. As research in neural networks elegantly demonstrates,[33,34] control by suitably convolved feedback signals affords a mechanism for learning and adaptive behavior.

Figure 6 suggests in a highly idealized way some aspects of visual perception which might be well addressed in such a photoneural system. The initial visual scene in Fig. 6(a) is drawn to emphasize some lineal properties, namely, lineal spatial frequency, line orientation, edges, and terminations. Figure 6(b) illustrates the lowest-level processing of the original scene by an array of LPNCs configured to sense the concentric center/surround-on antagonism. The size of the resolution cell is indicated in the lower left-hand corner of the figure. Consideration of Fig. 6(b) highlights several important issues. Although the external edges of the octagon are resolved fairly well, the accident of fixation is critical. A half-pixel displacement loses these edges but resolves more clearly the internal ones. Again, while some of the vertical and horizontal lines are resolved, the global regularity of the pattern is apparent only for the oblique set which has orthogonal spatial frequencies accidentally matching the cell fundamental. As Fig. 6(c) demonstrates, a map of both center/surround-on and -off responses helps to resolve some of the ambiguities.

Figure 6(c) is in essence one of the zero-crossing maps which figure as fundamental data structures in the famous computational theory of vision formulated by Marr and Poggio.[35] In the Marr algorithm zero-crossing maps are computed by application of spatial antagonistic-like nonlinear filters to the input light intensity distribution. These primal sketches are in turn computationally cross-correlated to obtain new representations of the visual scene which express particular stimulus attributes, e.g., boundaries, lines, and discontinuities. Again these representations are cross-correlated to produce representations of new attributes, e.g., surface orientation, texture, and depth. The process continues until some given level of perception is obtained. As demonstrated, the LPNC arrays afford a rapid means for directly obtaining the primal sketches.

There is abundant evidence that multiscale primal sketches or representations are essential in perception.[4] The pixel size in Fig 6(c) has been selected to register most of the small details in the original scene. Figure 6(d) shows a second representation of the scene taken with a fourfold increase in the area of the response field. The scaling of the visual field could be accomplished in a photoneural system in several different convenient ways. If all arrays in the system are identically configured, optical magnification may be used to adjust the scale of the scene image. However, the LPNC realization discussed in Sec. III is easily scaled by modifying the depth of the isolating damage regions. The important notion is that scaled primal sketches are available in parallel for cross-comparison. For example, the comparison of Figs. 6(c) and (b) suggests that the various line elements are indeed clus-

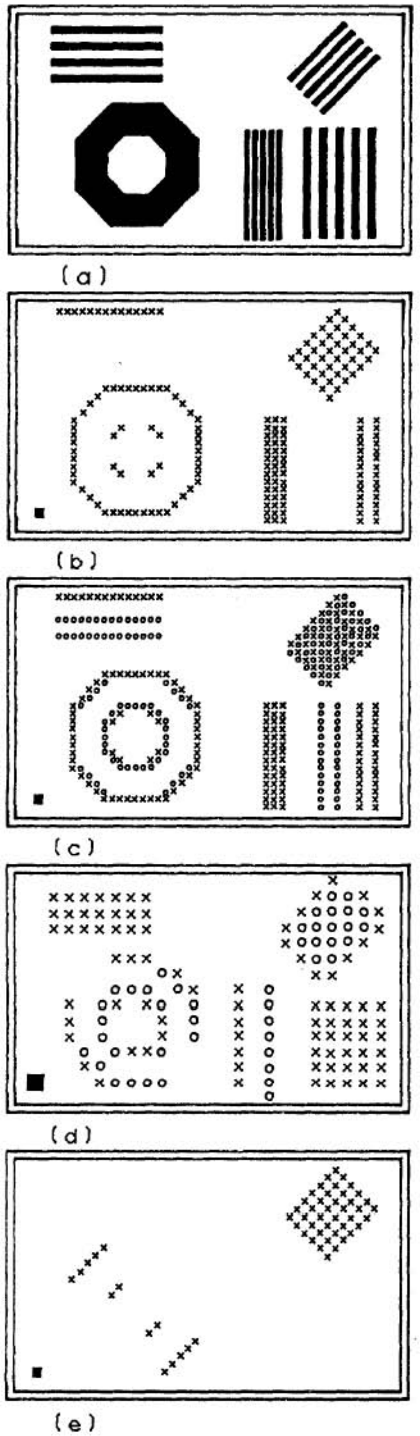

Fig. 6. Some early steps in a visual perception sequence: (a) original visual scene; (b) high-resolution center/surround-on × reduction of scene; (c) high-resolution primal sketch, i.e., center/surround-on × and -off o; (d) low-resolution primal sketch, i.e., center/surround-on × and -off o; (e) analog of cortical simple cell response.

tered into coherent entities. In general, the low-resolution primal sketches help to sort out the regions of the visual field which require special attention.

The orientational properties of the simple and complex cortical cells could be accomplished directly with anamorphic optics. For example, the light pattern illustrated in Fig. 6(e) is the analog of simple cell response to the initial scene. It is obtained by first anamorphically transforming the field in Fig. 6(b) so that a point source is imaged as an oblique line and then detected with a concentric center/surround LPNC. The analog of complex cell behavior is obtained in turn by optically integrating the light distribution in Fig. 6(e).

These few examples demonstrate that many of the constructs of visual perception could be mapped into the photoneural context. At the moment further elaboration would be idle. The essential point is that photoneural arrays if developed would be interesting devices for exploring cognitive processes.

Early work on the photoneural concept was supported by the U.S. Army Research Office as a short-term exploratory study (contract DAAG-84-K-0119, USARO).

References

1. R. V. Jones, *A New Architecture for Optoelectronic Information Processing*, Final Report contract DAAG-84-K-0119, USARO (1984).
2. G. M. Shephard, *Neurobiology* (Oxford U. P., New York, 1983).
3. E. R. Kandel and J. H. Schwartz, *Principles of Neural Science* (Elsevier, New York, 1985).
4. M. W. Levine and J. M. Shefner, *Fundamentals of Sensation and Perception* (Addison-Wesley, Reading, MA, 1981).
5. S. W. Kuffler and J. G. Nicholls, *From Neuron to Brain* (Sinauer Associates, Sunderland, MA, 1976).
6. J. E. Dowling, "How the Retina 'Sees'," Invest. Ophthalmol. **17,** 832 (1978).
7. J. E. Dowling, "Information Processing by Local Circuits: the Vertebrate Retina," in *The Neurosciences: Fourth Study Program* (TN:FSP), F. O. Schmitt and F. G. Worden, Eds. (MIT Press, Cambridge, 1979).
8. F. S. Werblin, "Integrative Pathways in Local Circuits Between Slow-Potential Cells in the Retina," in TN:FSP.[7]
9. D. H. Hubel and T. N. Wiesel, "Functional Architecture of Macaque Monkey Visual Cortex," Proc. R. Soc. London Ser. **198,** (1977).
10. V. B. Mountcastle, "An Organizing Principle for Cerebral Function: The Unit Module and the Distributed System," in TN:FSP.[7]
11. A. Cowey, "Why are There so Many Visual Areas?," in *The Organization of the Cerebral Cortex*, F. O. Schmitt, F. G. Worden, G. Adelman and S. G. Dennis, Eds. (MIT Press, Cambridge, 1979).
12. M. P. Shaw, "Properties of Junctions and Barriers," in *Handbook on Semiconductors, Vol. 4*, T. S. Moss and C. Hilsum, Eds. (North-Holland, Amsterdam, 1981).
13. K. Ploog, "Molecular Beam Epitaxy of III–V Compounds," in *Crystal: Growth, Properties and Application, Vol. 3*, H. C. Freyhardt, Ed. (Springer-Verlag, New York, 1980).
14. R. D. Dupuis, L. A. Moudy, and P. D. Dapkus, "Preparation and Properties of GaAlAs/GaAs Heterojunctions Grown by Metalorganic Chemical Vapor Deposition," Inst. Phys. Conf. **45,** (1979).
15. H. Kroemer, "Heterostructure Bipolar Transistors and Integrated Circuits," Proc. IEEE **70,** 13 (1982).

16. H. Kroemer, "Heterostructure Bipolar Transistors: What Should We Build?" J. Vac. Sci. Technol. B **1**, 126 (1983).
17. M.H. Pilkuhn, "Light Emitting Diodes," in *Handbook on Semiconductors, Vol. 4*, T. S. Moss and C. Hilsum, Eds. (North-Holland, Amsterdam, 1981).
18. T. H. Wood, C. A. Burrus, D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, and W. Wiegmann, "High-Speed Optical Modulation with GaAs/GaAlAs Quantum Wells in p-i-n Diode Structure," Appl. Phys. Lett. **44**, 16 (1984).
19. H. Beneking, N. Gote, W. Roth, and M. N. Svilans, "GaAs-GaAlAs Phototransistor/Laser Light Amplifier," Electron. Lett. **16**, 602 (1980).
20. A. Sasaki and M. Kuzuuhara, "InGaAsP-InP Heterojunction Phototransistors and Light Amplifiers," Jpn. J. Appl. Phys. **20**, L283 (1981).
21. H. Beneking, "Full Solid State Image Converter Based on Integration of Phototransistors and LEDs," IEEE Trans. Electron Device Lett. **EDL-2**, 99 (1981).
22. H. Beneking, N. Gote, and M. N. Svilans, "Monolithic GaAlAs/GaAs Infrared-to-Visible Wavelength Converter with Optical Power Amplification," IEEE Trans. Electron Devices **ED-28**, 404 (1981).
23. A. Sasaki, K. Matuda, Y. Kimura, and S. Fujita, "High-Current InGaAsP-InP Phototransistors and Some Monolithic Optical Devices," IEEE Trans. Electron Devices **ED-29**, 1382 (1982).
24. A. Sasaki, M. Taneya, H. Yano, and S. Fuijita, "Optoelectronic Integrated Device with Light Amplification and Optical Bistability," IEEE Trans. Electron Devices **ED-31**, 805 (1984).
25. J. C. Campbell, "Phototransistors for Lightwave Communications," Semicond. Semimetals **22D**, 389 (1986).
26. J. A. Copeland, "Triggerable Semiconductor Lasers and Light-Coupled Logic," J. Appl. Phys. **51**, 1919 (1980).
27. J. A. Copeland, J. C. Campbell, A. G. Dentai, and S. E. Miller, "Wavelength-Multiplexed AND Gate: A Building Block for Monolithic Optically Coupled Circuits," Appl. Phys. Lett. **39**, 197 (1981).
28. J. C. Campbell, A. G. Dentai, J. A. Copeland, and W. S. Holden, "Optical AND Gate," IEEE J. Quantum Electron. **QE-18**, 992 (1982).
29. S. M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981).
30. J. G. Daugman, "Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters," J. Opt. Soc. Am. **A2**, 1160 (1985).
31. J. G. Daugman, private communication.
32. J. W. Goodman, F. I. Leonberger, S. Y. Kung, and R. A. Athale, "Optical Interconnections for VLSI Systems," Proc. IEEE **72**, 850 (1984).
33. D. Psaltis and N. Farhat, "Optical Information Processing Based on Associative-Memory Model of Neural Nets with Thresholding and Feedback," Opt. Lett. **10**, 98 (1985).
34. N. H. Farhat, D. Psaltis, A. Prata, and E. Paek, "Optical Implementation of the Hopfield Model," Appl. Opt. **24**, 1469 (1985).
35. D. Marr, *Vision* (Freeman, San Francisco, 1982).

PATENTS PATTER

Franklin S. Harris, Jr.

In *NASA Tech Briefs* **10**, No. 6 (November–December 1986), some interesting developments are reported in optics-related fields, and a selection of these is given below. Further information can be obtained, where available, by writing to the Manager, Technology Transfer Division, P.O. Box 8757, BWI Airport, MD 21240, and giving the identifying number.

Circuit for lifetime and surface-recombination measurements

A new circuit increases the accuracy of measurements of the recombination lifetime and the effective surface recombination velocity in a silicon solar cell. Essentially a fast electronic switch, the circuit grounds a forward-biased cell so rapidly that the transient voltage to be measured is not affected significantly. Previously, the method of open-circuit voltage decay or of junction-current recovery was used. The solar cell was maintained initially under forward voltage and current. Then a reverse current was applied suddenly, and the voltage across the cell was measured as a function of time. For the open-circuit voltage-decay method, the applied reverse voltage was zero; for the junction-current recovery method, the reverse voltage was finite.

In either method, the transient voltage across the solar cell was interpreted by an idealized theory that neglected the effects of mobile holes and electrons in the space-charge region of the cell. This theory is acceptable for germanium devices, which have a low energy gap and a low proportion of mobile holes and electrons. In silicon devices, however, mobile holes and electrons play a prominent role in establishing the voltage transients, and their presence creates errors in the interpretation of the measurements. The new circuit avoids the problem by placing a short circuit across the solar-cell terminals. The short circuit removes mobile holes and electrons from the space-charge region, within ~1 ns.

As before, the solar cell is first subjected to forward current. At the designated switching time, a metal-oxide/semiconductor transistor is switched on to create the short circuit (see Fig. 1). The

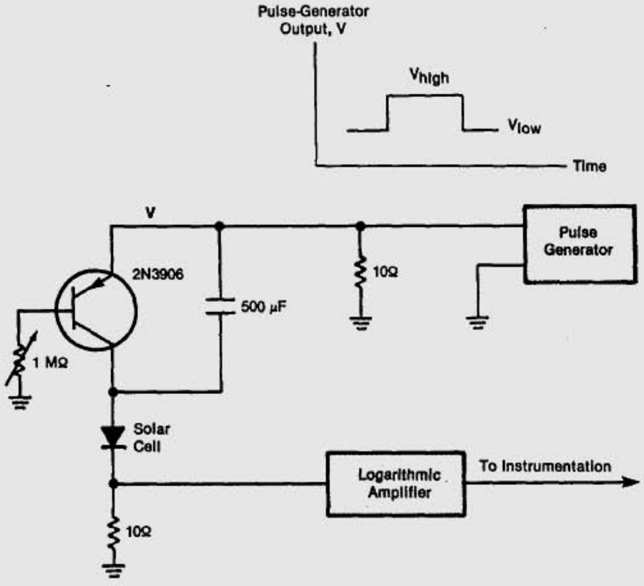

Fig. 1. Switching translator initially applies a forward current to the solar cell during the interval when the pulse-generator output is high. At the transition from V_{high} back to V_{low}, the solar-cell anode is effectively shorted to ground within a nanosecond. The cathode-to-ground voltage is measured and recorded as it changes with time.

continued on page 1963