Neuromorphic sensory systems
Shih-Chii Liu and Tobi Delbruck

Biologists provide examples of efficient machines which greatly outperform conventional technology. Designers in neuromorphic engineering aim to construct electronic systems with the same efficient style of computation. This task requires a melding of novel engineering principles with knowledge gleaned from neuroscience. We discuss recent progress in realizing neuromorphic sensory systems which mimic the biological retina and cochlea, and subsequent sensor processing. The main trends are the increasing number of sensors and sensory systems that communicate through asynchronous digital signals analogous to neural spikes; the improved performance and usability of these sensors; and novel sensor processing methods which capitalize on the timing of spikes from these sensors. Experiments using these sensors can impact how we think the brain processes sensory information.

Address
Institute of Neuroinformatics, University of Zurich and ETH Zurich, Winterthurerstrasse 190, CH-8057, Zurich, Switzerland

Corresponding author: Liu, Shih-Chii (shih@ini.phys.ethz.ch) and Delbruck, Tobi (tobi@ini.phys.ethz.ch)

Introduction
Nature offers many examples of compact, energy-efficient, adaptable, and intelligent systems. Even a simple animal like the bee displays exquisite flight motor skills and cognitive behaviors, with a body weight of less than a gram and a brain that dissipates only on the order of 10 μW. Yet we are not able to construct an artificial system that displays even a subset of this creature’s abilities using the latest technology; see for example, the performance of the state-of-art robotic vehicles in the recent DARPA Desert and Urban Challenges which relied heavily on absolute positioning via GPS and carried roughly 1 kW of computing power [1].

To identify the factors contributing to nature’s computational efficiency in dealing with its uncontrolled environment, many researchers rely on simulations of brain networks using supercomputers [2] which are constructed from fast, high-precision digital hardware with high power dissipation. However, brains are composed of slow asynchronous neural components which use a combination of both analog and digital signal representations. Although computers are valuable for detailed simulation of biophysics, they are not physical emulations of biological neural computation.

A small number of labs in the field of neuromorphic engineering are actively pursuing a different approach towards understanding the architecture of brains. By emulating the neuronal organization and function of nervous systems in electronic devices, neuromorphic engineers hope to harness the brain’s efficient and powerful style of physical computation for future artificial systems [3–5]. The fruits of this effort are illustrated in neuromorphic microchips that emulate biological sensing and spike-based neuronal processing.

Neuromorphic sensors and sensory systems have made the greatest strides in recent years with many designs using a new form of asynchronous output representation which carries timing information similar to spikes in the nervous systems. This representation has led to novel systems where computing occurs in an event-driven manner similar to that of nervous systems [6].

Artificial technology: similarities between biology and silicon
Neuromorphic electronic devices are constructed primarily on silicon, the same technology used for fabricating the majority of analog and digital computing chips. The primary silicon primitive is the transistor where current flows through this device as a function of the voltage at its four terminals. It shares much of the same physics of neurons thus making it an ideal device for emulating the circuitry of nervous systems [4,7]. In the “subthreshold” region of operation, the transistor current is exponentially dependent on its terminal voltages analogous to the exponential dependence of active populations of voltage-sensitive ionic channels as a function of the potential across the membrane of a neuron. This similarity allows us to construct using a small number of transistors, compact circuits that implement electronic models of voltage-controlled conductance-based neurons and conductance-based synapses [8,9] and useful computational biological primitives such as phototransduction, logarithmic functions, amplification, multiplication, inhibition, correlation, thresholding, and winner-take-all selection [7].
Communicating spikes by using the address-event representation

Building large arrays of interconnected neurons on a silicon chip can pose a challenge. Unlike the 3D world of neurons, transistors are permanently patterned on a 2D substrate with only a few layers of wiring. Point-to-point dedicated wiring is expensive and is restricted to local neighborhoods. Wiring and transistor circuits on a chip are also fixed after chip fabrication, restricting the range of possible neuronal architectures that can be expressed on one chip. However, the mobility of current flow (electrons) in a transistor is $10^7$ times higher than the mobility of ions in biology. This huge difference in speed is used to make up for the fixed circuitry, as we will now discuss.

We use the higher speed of electronics to our advantage by devising a different signal transmission scheme than that of real neurons. Connections in real neurons are built by synapses which connect an axon to a dendrite. In electronics, we simply confer an address to every neuron and synapse on-chip. We transmit spikes off-chip as a spike address that carries the “what” (which neuron) and “when” (when it spiked) information. These addresses are transmitted within a few nanoseconds to another external device. For example, one 16-bit address bus can carry spikes from up to $2^{16} = 65,536$ neurons. Since the on-chip neurons are designed to have low firing rates of about 100 Hz or less, multiple spike addresses can be transmitted off-chip on the shared address bus before a neuron makes the next spike. The availability of “what” and “when” information allows great flexibility for subsequent processing. For example, digital memory chips can store a routing table which describes how a spike address is routed to other neurons through their pre-specified spike addresses, thus emulating a soft configurable connectivity scheme for multiple neuron on-chip. This scheme is called the address-event representation (AER) and is the de facto signal transmission scheme used in most recent neuromorphic sensors and multi-chip systems [10**].

Event-driven computing with spike-timing

What are the advantages of using AER? Conventional sensors and AER systems differ in who is the master of the communication. In conventional sensors, an external process regularly polls the sensor array at sufficiently high rate to capture all frequencies of interest. AER sensors, like their biological counterparts, make the sensor the master: Each retina pixel or cochlea channel autonomously decides when it should transmit events that carry useful information about the world. This form of event coding has two advantages: first, these sensors transmit only informative non-redundant events, thus reducing power dissipation both for communication and for subsequent processing. Second, because the pixels initiate their own communication, output events are transmitted with very short latencies. These spikes initiate computation in the post-processing network, which we call event-driven computation. A recipient network can, for example, use the timing information of the events to configure the connections between neurons [3,11**], to initiate synaptic plasticity in a network [12,13], or to drive the motor output of a visual-motor system with low-latency [14**,15].

Emulation of biological sensing

Neuromorphic engineers have focused significant effort on mimicking the retina and cochlea. The primary reasons for this focus are that these sensors have immediate applications in artificial vision and audition, and they allow the exploration of neural-like architectures for sensory processing in the brain.

Silicon retinas

Conventional imagers like our digital cameras create sequences of highly redundant pictures. However, retinas serve a different function; they asynchronously transmit relevant information about the world, thus reducing redundancy and increasing efficiency in power dissipation and information transfer [16]. Neuromorphic designers have made significant strides in copying the functionality of the retina based on the wealth of knowledge from neuroscientists [17]. The earliest silicon retina by Mahowald and Mead [18] illustrated the functionality of major retinal cells and demonstrated useful properties of biological retinas, for example, the local gain control of photoreceptors over many decades of background intensity, and the reduction of redundant spatial and temporal information. The majority of recent silicon retina designs transmit their outputs as AER spikes. Improvement in the silicon analog circuit design techniques has led to recent useable designs with varieties of cell types and functionality [19,20**–25,26**] (see Figure 1).

Unlike conventional imagers that generate framesampled intensity values, silicon retinas emulate the local processing, local gain control, and asynchronous spike transmission properties of biological retinas. From an engineering perspective, silicon retinas offer advantages of increased dynamic range in illumination (>10^5) and higher effective sampling rates (several kHz) over conventional cameras, which have typical dynamic range of a factor of 300 in illumination and sampling rates of under 60 Hz. The cost of processing the sparse output of silicon retinas can be reduced by a factor of 100 compared to the cost of processing the outputs of conventional cameras [26**].

Silicon cochleas

Silicon cochleas model the functionality of the biological counterpart, in particular the traveling wave responses of the basilar membrane (BM) as a result of incoming sound waves. The transport of energy along the BM is modeled by an electronic filter bank consisting of multiple filters...
Figure 1

Example of silicon retina system [24]. (a) The sensor emulates a part of the retina pathway that is not well served by conventional image sensors. We made the analogy between the sustained pathway of the biological retina and conventional cameras, as opposed to the transient pathway of the biological retina, which we model with the so-called Dynamic Vision Sensor (DVS). (b) The DVS pixel models a simplified 3-layer retina, where individual pixels are decoupled spatially but still use a temporal "surround" computed by each pixel independently. (c) The DVS pixel forms an abstraction of the photoreceptor-bipolar-ganglion cell information flow. It consists of three parts: a logarithmic photoreceptor, a differencing amplifier (bipolar cells), and two decision units (ganglion cells). The pixel output consists of asynchronous ON and OFF address-events that signal scene reflectance changes. (d) The spike events are computed by the pixel as illustrated. The continuous-time photoreceptor output, which encodes intensity logarithmically, is constantly monitored for changes since the last event was emitted by the pixel. A detected change in log intensity which exceeds a threshold value results in the emission of an ON or OFF event. The threshold is typically set to about 10% contrast. Communication of the event to the periphery resets the pixel, which causes the pixel to memorize the new log intensity value. (e) The pixels are arranged in an array and fabricated in a standard chip-making process. AER circuits along the periphery of the chip handle the access to the shared AER bus and ensure that all events are transmitted, even if there are collisions. Colliding pixels must wait for their turn for access to the AER bus. (f) The chips are integrated into a camera, interfaced either to a computer by USB, directly to a microcontroller, or to another neuromorphic chip via its AER interface. (g) Data collected from the DVS shows its characteristics: the events can be histogrammed in 2D-space over a certain time window to form an image which displays either the ON and OFF events as contrast (Faces), or as a gray scale showing the relative event time (Juggling event time), or they can be viewed in space-time to see the spatiotemporal structure (space-time spike events).
which are arranged to model the physical properties of the BM. Initial prototypes used a 1D cascaded filter bank to model the response of the BM [27] (Figure 2b). Subsequent designs use filter banks with a cascaded or parallel architecture; or include circuits modeling the role of the fluid coupling between cochlea stages [28,29,30*]; and the local gain control in the biological cochlea [30*,31–33]. The latest designs also include circuits after the BM that emulate the hair cells and spiral ganglion cells [30*,34–36] and transmit asynchronous AER spikes like the silicon retina (see Figure 2). The timing information from the AER cochleas can be used for inferring the location of an acoustic source, for example, from the interaural time differences between sounds arriving to the two ears [37] or by emulating the echolocalization mechanism of bats [38]. The spike outputs can also be used to extract higher level auditory features suitable for tasks such as speaker identification [39]. The commonality of output representation in the retina and cochlea should open up a fruitful area in sensory fusion which exploits timing information for a coherent bimodal representation.

**Neuromorphic sensory systems**

In neuromorphic multi-chip sensory systems, AER sensor spikes travel in real-time to multi-neuron processors implementing neuron and synaptic circuit models of varying details [40,41,42*,43*]. Through the AER infrastructure, configuration of the network connectivity...
on-chip and across chips allows the expression of various network and system architectures. Multi-neuron chip systems are increasing in scale [42,44] with as many as 65,000 neurons on a single chip (unpublished working system). Multi-chip sensory systems can demonstrate cortical visual properties such as orientation selectivity [45–47], stereopsis [48], and motion sensing with tracking [49**]. Cortical-like multi-layered AER systems which can implement neuroscience models of visual processing are also being realized [47,49**] along with sensory-motor AER systems [14**,49**] (see Figure 3 for two examples).

**Challenges to application**

Like every area in electronics, neuromorphic chip designers ultimately face the challenge of validating their approach by implementing commercially viable silicon designs. Transistor current variability is one of the main challenges because it severely limits precision, especially across an array of nominally identical circuits [50]. Silicon circuits cannot easily imitate the complex, rich molecular machinery of cells which enable the latter to continuously adapt their operating parameters to changes in their environment. Synchronous digital logic circuits are par...
6 Sensory systems

tially so successful because they deal with this transistor variability by simply restoring every logic circuit output to one of two possible voltage values. However, clever circuit and system design can be used to mimic biology’s adaptation schemes and to reduce unwanted response variances across pixels in the chips. This understanding has led to recent neuromorphic sensors with performance that can exceed conventional technology, especially in applications like low power, high speed, or wide dynamic range vision [20–25,26†], and neural prosthetics [51*,52†]. A small number of studies have exploited transistor variability for computation. For example, an orientation-selective network refines its local orientation preference starting from the intrinsic preference embodied by the built-in variability of transistors [53] and a network of spiking neurons refines its synchronicity using relative spike-timing [12].

Future neuromorphic computation

For much of its history, neuromorphic engineers have aimed to build “pure” neuromorphic AER systems consisting of multiple interconnected, mostly feed-forward neuromorphic chips. These hardware systems have proved to be more difficult to configure than software, in large part because of the general lack of understanding of how we can compose real-time behaving systems out of independent modules. Although we understand something about building hierarchical networks for feature extraction and classification (e.g. [54,55]), little is known about building systems that express a suite of behaviors using these generic modules.

However, we have started to make some inroads into event-driven sensory and sensory-motor algorithms by combining AER sensors with conventional computers. An open-source software project (jaer.wiki.sourceforge.net) provides the possibility of processing AER sensor events using software algorithms. The software infrastructure is optimized for using the event timing. For example, local features like orientation and motion can be extracted using temporal coherence and relative timing of events [56]. The main practical application area so far has been fast object tracking using the temporal clustering of events [57] (see Figure 3b for an example). These developments show advantages gained from the reduction in latency and processor load with event-driven sensor representations [14,15,57].

Just as large microprocessor companies support open-source computer vision and machine learning software projects to understand the required features of future generations of processors, software processing of AER data will suggest neuromorphic hardware that needs to be developed. Future advances in neuromorphic sensory systems will also depend on ongoing work of how we can systematically combine modules to produce a desired function [58], on understanding probabilistic inference methods in nervous systems, and on the evolution of algorithms into hardware.

Acknowledgements

We thank members and former members of the Institute of Neuroinformatics; in particular RJ Douglas and K Martin, and the hardware group; the following readers for useful comments on the articles, U Rutishauser, A Benecchi, N. da Costa, P. Baker; and the NSF Telluride Neuromorphic Cognitive Engineering Workshop.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest
•• of outstanding interest


These cars were utterly dependent on Global Positioning System (GPS) for navigation, hardly used vision, and carried on the order of 1 kW of conventional computing power.


This “cat scale” brain simulation requires a 1.4 MW supercomputer with 147,000 processors which weighs 28,000 kg, requires 500 tons of air conditioning, and runs simplified point neurons about 700 times slower than real time.


The authors describe an AER infrastructure for connecting multiple chips in a daisy-chain which allows for system scalability. Systems that consist of multiple chips spiking the same AER language can then be composed out of single chip modules almost like LEGO blocks. Because of the similarity with biology of the spike transmission process in the neuromorphic circuits, neuromorphic engineering is in a good position to look at event-driven computation in natural systems and event-driven coding schemes. Continual discussions of rate coding and temporal coding schemes pervade the neuroscience field and similar questions can be asked of the neuromorphic hardware.


This work shows how the timing of spikes can be used to develop topographical connections between layers when the cells in both input and output layers have only local knowledge of spike activity.


Please cite this article in press as: Liu S-C, Delbruck T. Neuromorphic sensory systems, Curr Opin Neurobiol (2010), doi:10.1016/j.conb.2010.03.007.
Neuromorphic sensory systems

Liu and Delbruck


By using the output from two AER silicon retinas, this pencil balancing robot can be built using only simple microcontrollers burning a total of less than a watt of power. Using conventional high-frame-rate cameras for the same robot would require an entire state-of-the-art PC which burns more than a hundred times the power.


A high-performance sustained-pathway retina system which uses a conventional frame-sampled image sensor pixel which is then coupled to a lateral resistive network which models the horizontal cell network. The difference between the local intensity value and the spatially low-pass filtered image is output from the sensor.


Highlights state-of-the-art work in developing high-density arrays of biophysical models of neurons and synapses. The latest system has 61 graded and 18 binary programmable parameters allowing Neurocore to model a wide variety of spiking and connectivity patterns. Each neuron has a soma, a dendrite, four synapse populations, and four ion-channel populations. Circuits implement pairs of gating variables to model ion-channel populations. The synaptic circuits that consist of low-pass filters can be configured to drive the soma, dendrite, or to see an ion-channel population’s maximum conductance. The same type of gating-variable circuits are used to model different synapses and receptors and cholinergic modulation.


This review focuses on silicon neurons, synapses, and communication technology.


8 Sensory systems


The largest scale multi-chip AER sensory processing system to date. This project integrated neuromorphic AER chips designed by four groups from different countries.


Although it would seem that neuromorphic sensors would serve as ideal front ends for neural prosthetics, in practice developers of prosthetics are concerned mostly with issues of neural interfacing, long-term stability, power supply, device control, and clinical trials. As a result, recent advances in vision prosthetics such as this one focus on device fabrication and neural stimulation.


This paper describes state-of-the-art work in preprocessing audio for implants using a bio-inspired approach. State-of-the-art hearing aids burn only about 1mW and do a substantial amount of digital signal processing. Cochlear implants put most of the computation external to the body; the implant itself is a stimulator which provides pulses to stimulate the spiral ganglion cells. For this auditory prosthetics, power consumption is dominated by stimulation and communication, so there has not been a strong industrial drive for the reduced power consumption that a neuromorphic auditory front end could offer.


