Regular Student Paper

Extending neuromorphic engineering beyond electronics

Sunny Bains
Dept. of Electrical and Electronic Engineering
Imperial College London
and
Dept. of Design and Innovation
The Open University
Tel: +44 20 7594-6341
Fax: +44 20 7823-8125 (Notify by e-mail)
E-mail: sunny@sunnybains.com
http://www.sunnybains.com

Topics: analogue electronics, neuromorphic engineering, opto-electronics, optical interconnects, mechanics, embodied artificial intelligence.

ABSTRACT

One of the advantages of the neuromorphic approach is energy efficiency, which comes from the exploitation of the intrinsic physics of electronic devices. Taking the intrinsic efficiency of physics as a guiding principle, we can extend it beyond electronics to other technologies including optical, mechanical, and chemical. In this paper we consider the role that some of these other technologies may has to play in this area, describe some of the work that has already been done, and suggest some advantages of pursuing what we call a physical computational approach to AI.

1. Introduction

This paper considers how the concept of neuromorphic engineering may usefully be broadened to a much wider range of technologies than just electronics. In particular, we discuss how it may be more efficient to allow technologies that are normally considered part of sensing or actuation sub-systems to be considered part of the intelligence of the machine and therefore better integrated.

First we consider why it may make sense to consider this hardware- rather than software-oriented approach to artificial intelligence. In section 3 we consider some lessons from neuromorphic engineering and then the underlying challenge of building brain-like systems in part 4. In sections 5 and 6 we consider two areas where integrating optics with our systems may be particularly useful: increasing synaptic interconnectivity and neural complexity. In part 7 we look at the issue of analogue to digital conversion, the advantages of avoiding it, and discuss robotic implementations where some processing is performed in the body (rather than it all being done in the brain. In addition, we briefly consider the advantage of avilding digitization when it may be necessary to fuse interdependent sensory information. In section 8 we give an example of how optics and electronics can be used more powerfully when properly integrated, and discuss the general implications of this, after which we conclude.

2. Embodied AI as hardware problem

We are concerned with the design of hardware that will implement embodied artificial intelligence, ideally to the point that a machine will be able to react to the environment in ways that make it appear to be as intelligent as a human. By starting with hardware, rather than cognitive models or connectionist networks, we do not intend to imply that software is not important. Rather, we want to show that—with a task as challenging as animal- or human-like AI—hardware may represent a bottleneck that we cannot afford to ignore.

We can break the problem of intelligent interaction down into a process of sensing the outside world, learning and making decisions based on the information acquired, and then actuating to change the world as necessary. If we do this then some of the potential problems become apparent. Are we sensing the world sufficiently (in terms of number, type, and sensitivity of sensors)? Do we have enough computing power to process this information appropriately in order to make decisions about what to do next? Finally, can our actuators implement these decisions sufficiently well? Do they have high enough speed, accuracy, and degrees of freedom to make the appropriate action? Lastly, is the latency of the entire system (from sensor to actuator) sufficiently low that circumstances will not have changed dramatically (at the appropriate macroscopic physical level) by the time actuation takes place.

There are additional engineering constraints: power, size, and heat dissipation among them. We are unlikely to want to build an android that requires its own power station to operate. Or that is the size of a tall building. Or that is too heavy to walk on the floors of our homes and offices. Or that singes our hair if we get within ten feet.

From a hardware point of view we can see that there are sufficient issues to attend to without needing a full knowledge of the algorithms to be run. Further, hardware constraints might lead us to prefer some approaches to AI over others. The application constrains the hardware, then the hardware (may) constrain the software.

3. Neuromorphic engineering

Carver Mead is best known to many of us for pioneering a new way of exploiting analogue electronics. His approach came from an interest in implementing neural circuits in order to process sensory information. In a book that has come to define the field of neuromorphic engineering (Mead,1989), Mead showed that brains are about a billion times more power efficient than conventional computers. Though part of this could be attributed to the inefficiency of the individual transistors, etc. he believed that a performance improvement of six orders of magnitude was possible based on two things: using less wire and more local interactions; and using, not suppressing, the intrinsic physics (i.e. analogue nature) of the device.

The first of these two strategies is an architectural issue that is helpful for any computer whether digital or analogue. In an electronic system, the length of wire between one component and another acts as a power drain. Component A only really wants to charge up component B: that's all that's necessary for the computation. But to do that, it has to charge up all of the wire between the two components as well. The extra power required to do this is fulfilling no computational purpose, it is simply an architectural overhead. Though designers do spend a lot of time trying to keep the amount of wire to a minimum, the kinds of processes that we run on digital computers (and the gate layouts that make them possible) do not lend themselves to the ideal architecture: one in which devices only have to 'talk' to near neighbours. The opposite is true of neural networks which also have major advantages for sensor fusion, see e.g. (Klein,1999)). By changing to this more efficient interconnection scheme, a 100-fold power efficiency gain can be made.

However, the advantage of moving from a digital to an analogue neural system is potentially even greater. Mead believes that computation could be performed up to 10,000 times more efficiently (in terms of power) by not throwing away the true functionality of the electronics used. In other words, by using them in a more analogue way. For instance, take a memory cell. Intrinsically, there is nothing about such a circuit that forces it to be either 'full' or 'empty' (to contain logic 1 or logic 0): we simply choose to interpret the information that way when we read it out. Because of this choice, we are forced to use an extra cell for each bit of information we want to store. If we allowed the charge within the cell to vary continuously (or at least to be able to adopt different 'grey levels'), then we could store significantly more information in a much smaller space. In addition, we would need to use less energy to both store and read out the information (because we would only have to go to one location to get it).

The same kind of analysis is true of the information processing side. The electronic circuits we use in computing have interesting, and often useful, responses. For instance, they can perform multiplications or manipulate a signal differently based on its strength or timing. Some of these features are, says Mead, very similar to some of those we see in the brain's processing devices. But we do not really make use of these as computational primitives (low-level processes on which higher ones can be built). Instead, we reduce problems to AND, OR, NOT, etc., and force our hardware to give us the minimalist answer to these questions (1 or 0) we require.

From a power-consumption point of view, it is important to note that the forcing required to achieve this involves driving the electronics in a particularly inefficient mode. More energy is consumed for an incoming signal to trigger a sharp non-linear (thresholded) response, than a sub-threshold (closer to linear) response.

In effect, we are reducing the operations that each component can perform, and increasing the amount of energy it takes. Given the example in the last section, we know why this approach is taken. If we need a precise answer, we would not want to use an amplifier to multiply one number by another. Depending on the noise and the extent to which the device varies from its specifications, we would be likely to get a different answer every time. Thus we string together many gates, each consisting of many different components, to perform a digital multiplication. We get a precise answer, but at a cost in terms of efficiency.

If we have an application where we don't need that kind of precision in the first place, then this approach makes no sense: in this case we are throwing away the useful physics of the devices themselves, recreating the same functions using logic gates, and have nothing to show for it at the end. In such applications, analogue computers can make a real difference in terms of power efficiency without sacrificing functionality.

An illustration of this point is the cellular neural network (CNN) invented by Leon Chua (e.g. (Chua,1998)), and now being used as an image-processing tool in the vision community. The CNN is a device that is digitally programmable but can perform complicated non-linear operations during the analogue transient: the 'switching' time to go from one stable state to the next. It is an apt demonstration of Mead's point. Not only is the device far lower-power than the equivalent image-processor, but it is also up to orders of magnitudes faster (depending on the algorithms implemented).

4. The challenge of building brain-like systems

For many of us, the future of AI will involve, in one way or other, the building of artificial brains. The human brain has of the order of 10^{12} neural processors, linked by as many as 10^{15} synapses (Churchland et al. 1992), all contained in an object about the size and topology of a folded pizza (analogy from Christof Koch, California Institute of Technology). If each of these parallel processors were performing just one logic operation per emission of a neural spike, duplicating this system in digital electronics would be formidable challenge. In fact, according to Yaser Abu-Mustafa, it would be practically impossible today because of the lack of connectivity between neurons (Abu-Mostafa,1988; Abu-Mustafa,1988). In his papers (and echoed in an appendix he wrote for Carver Mead's book (Abu-Mostafa,1989)), he argues that the ability of a local-learning neural network (which would hold for the biological case) is limited by the number of connections between the neurons. This, he says, cannot be made up by simply using a larger network.

If he is right, and biological neural networks do indeed fall into this class, then this could represent a major bottleneck. Either we must find a learning rule that allows us (with a bigger network) to produce the same discrimination ability with a modest level of interconnectivity (of the order of 1-10 connections per neuron rather than the average of 1000 in biological systems), or we find some way to improve connectivity, or we give up on the neuromorphic approach. The first and third of these options may be the same. According to Abu-Mustafa, the local learning rule was considered precisely because it was biologically plausible. Departing from it (necessary to overcome Abu-Mustafa's limit) by definition changes the nature of the system being designed. In any case, connectivity is an issue that has had to be addressed for many connectionist (neural-network base) and distributed computing applications.

5. Connectivity as a hardware problem

David Miller ((Miller,1989, 1997; Miller et al. 1997)) has written extensive critiques on the problems of on-chip electronic interconnections. The problem can be expressed simply: the longer an electronic interconnect is, the higher its capacitance. This means that more energy is required to get a signal from one end to the other and, since circuits have to be designed with the *worst* case scenario in mind, the performance of the entire system ends up being determined by the longest link. As a result, designers try to keep interconnects as short as possible: nearest-neighbour interconnects are ideal. Similarly, the more interconnects a chip has (of whatever length) the more power it must produce to charge them all simultaneously. This is particularly true when *broadcasting* a signal: sending it through all its interconnects at once. Thus the number of interconnects becomes a limiting factor. Since broadcasting does seem to be an important part of some neural functioning (see e.g. (Abu-Mostafa,1988)), these scaling issues must be considered.

When information must be exchanged between one chip and another, there is a further problem: a bottleneck is caused by the fact that electronic die are essentially two-dimensional. If a square chip is length x on a side, then the area (and number of processors accommodated) varies as x^2, while the number of interconnects (pins along the edges of the chips) varies as $4x$. As silicon wafers get larger and feature sizes get smaller, this problem gets worse (even if the size of pins scales down too).

Finally, there is the problem of crosstalk.

Interconnects must be electrically shielded from each other, otherwise the electric field will create false signals (noise) in neighbouring wires. This leakage also represents power dissipation. The need for shielding represents a limit on the (minimum) volume required for an electrical interconnect.

Inventive ways have been found to get around these problems for experimental systems. The best known electronic method, known as address-event representation (AER, (Boahen,2000; Mahowald,1992)) , essentially gets around the broadcasting problem by using a time-multiplexed, pulsed network. In this asynchronous (analogue) system, individual pixels request access to the bus when active. For example, Eugenio Culurciello ((Culurciello et al. 2001)) built a system where pixels in an imaging array, via artificial neurons, emitted spikes at a frequency proportional to the light intensity. Since AER is a responsive network—access to the communications bus is granted on request—the available bandwidth is allocated according to need. Also, because the spike time is small with reference to the inter-spike-interval time for a given neuron, delay in getting access to the bus does not significantly change the pulse-coded signal.

Unfortunately, though it represents an ingenious and efficient use of infrastructure, AER does not change the fact that there is a serious scaling problem. It was designed to scale well (which it does) for increasing numbers of processing elements on a single array: For N elements in an array to connect to the same number in another array, only $1+log_2N$ wires are needed. If M such arrays have to be connected with each other, then M $(1+log_2N)$ wires would have to be in place. With current technology, only relatively small numbers of chips can be connected in this way. In addition, AER is not designed for on-chip communication, so there is a trade off between making the arrays bigger (for fewer chips in the network) and local connectivity.

There are other techniques under development, some of which are likely to be compatible with AER. Most of these would fall into the broad class of optical interconnects. Unlike wires, optical signals don't have crosstalk (unless handled poorly at the detector), and so can co-exist in the same volume. This comes from the intrinsic physical fact that photons neither repel nor attract each other, nor do they interfere with each other in any way. (Photons, when split, do interfere with *themselves*, but that is another issue, and not relevant here).

The most obvious example of this is the wavelength-division-multiplexed fibre-optical link ((Miki et al. 1978)) where hundreds of closely-spaced wavelength channels can propagate down the same optical waveguide and be differentiated at the other end without their signals becoming mixed. As the light sources, non-linear materials, detectors, etc., for photonic systems become cheaper and more available, increasingly ambitious systems are being built. For example, Nan Jokerst built a substrate-guided or planar-optical system at Georgia Institute of Technology((Jokerst et al. 2000)). Essentially, this gives the light freedom to move in $2^1/_2$ dimensions: the two dimensions of the plane of the interconnect, plus up and down (into devices attached to either side of the plane). By including emitters and detectors on the top of circuits, signals can be either actively or passively routed to other locations (potentially *many* other locations) on the chip or board.

Among many other notable examples is NEC's optical backplane network at its Laboratories in Princeton((Araki et al. 1996)). This was aimed at providing slow reconfigurability for massively parallel computer systems and used free-space optics (optics in air/gas) rather than guided wave optics. Based on an analysis of distributed computing systems, the NEC network was designed to use an electronic crossbar for local interconnections on each board, and had vertical-cavity surface-emitting lasers (VCSELs) emit signals to travel between boards. Each board needed a VCSEL and detector for each other board it has to talk to, and the address is specified by the geometrical location of the VCSEL in the array. This kind of scaling may seem similar to the one that posed such a problem in AER. However, optical paths can cross without affecting each other and, since they do not have problems with heat dissipation, they can be arranged in close-packed arrays, with the area of the array scaling linearly with the total number of interconnected boards.

A more radical solution, using both fibres and free-space, was designed by Ed Frietman at the University of Delft((Frietman,1995)). Rather than producing a network that on some level requires message passing, like the NEC and AER schemes, this system was based on every processor talking to every other processor. This works by each sending out information through an array of light-emitting diodes, one for each bit. This light is then captured by a polymer fibre optic array and connected to a central node known as the Kaleidoscope. The fibres are organized so that their relative positions at the output are the same as at the input: effectively making the data into a 2D image. In the Kaleidoscope, the fibre bundles are tiled to make one large 2D image, the size of which is dependent on the number of bits output by each processor and the number of processors in the array. Using a faceted mirror and lens system, the light from the entire image is broadcast onto separate locations for each processor, coupled into a second fibre bundle, and then received so that the processor can access the data electronically.

It should be noted that not all current strategies to increase connectivity are optical. Irvine Sensors, a company with a history of supporting research into optical interconnects (some of which has recently been published, (Li et al. 2002)) , invented a new mechanism designed to be compatible with its chip-stacking technology((Carson,2000)). The stacking technique involved producing essentially conventional chips and removing the substrate (wafer) through rubbing. Apart from allowing more devices to fit into the same volume, this also gives a new opportunity for interconnection. To capitalize on this they invented the three-dimensional field-effect transistor. This device is constructed as two parts on two separate chips: once they have been bonded together in a chip stack, they operate as a single transistor. This means that the concept of a nearest neighbour is extended to three dimensions rather than two: as it is in the brain.

Though this work is not optical, it is interesting to note that Irvine sensors is has also done significant work in optical interconnection and that, in the company's speculation about building brain-sized systems, they always assumed that it will take many different stacked modules to produce the required power. Interconnection between modules could therefore be optical.

6. Capturing the complexity of neuron interactions

From our discussion of Carver Mead's work earlier, we know that the replication of simple analogue functions is more efficiently performed using sub-threshold electronics. However, there is much evidence that real neurons are more complex than even Mead's circuits (see e.g. (Churchland et al. 1992) or (Wehr et al. 1996) for a more specific example). Analysis of one of the best-known neural circuits, the Hodgkin/Huxley model of the squid axon((Hodgkin et al. 1952)), suggests that it performs in a mathematically complex way (see e.g. (Arcas et al. 2000)). The excitable membrane of the squid's axon becomes increasingly depolarised due to the incoming signals (*spikes* or *pulses* of ions) until it reaches a threshold potential. At this point it abruptly inverts, producing the neuronal action potential and, thus, the output signal. Afterwards the membrane undergoes a period of no response, followed by another slow build-up. As a result, a pulse arriving immediately after firing will have a completely different effect on the output of the neuron than a pulse arriving immediately before. Thus, spike timing is critical.

Nabil Farhat has done significant work in this area, both from the dynamics perspective ((Farhat et al. 1996)) and the hardware perspective((Farhat et al. 1995)). He claims that the real complexity comes from combining this behaviour with the processing carried out by the neuron's receptors in the dendritic tree. Correlations, caused by synchronicity in the incoming signals, cause a periodic modulation at the excitable membrane. This gives rise to complex ordered patterns of firing that are phase-locked to the periodic modulation, and to disordered (chaotic) firing that depends on the amplitude and frequency of the modulation. In this way the neuron detects coherence or meaning in arriving spike trains and encodes this information in its own output.

Here, an analogy can be made here between spike timing and optical phase. In photography, only the combined intensity of light arriving at the film is recorded: the phase information (encoding distance, direction) is thrown away resulting in a flat image. In holographic imaging, the phases of incoming rays are allowed to interact through interference, and the result of that interaction is recorded((Hecht,1987)). This is what encodes the image's third dimension, depth, and increases the amount of information that can be stored and retrieved((Mok et al. 1992)). In fact, researchers have exploited optical interference to make associative memories (e.g. (Farhat et al. 1985)), based on optical correlators ((Casasent et al. 1976; Juday et al. 1987)) combined with holographic data storage((Psaltis et al. 1990)). Such devices have the advantage of both of high density—through efficient use of material—and content addressibility: possible because the geometrical/topological integrity of images is retained in the way they are stored.

Farhat built both simulations and actual analogue models of neurons that could take timing into account in this way and found that they produced a bifurcated output: they depended on only small changes in input frequency and phase, periodic, m- and quasi-periodic, and chaotic firing patterns were all observed. Further, Farhat designed a system that could take advantage of the high-connectivity available through optical interconnects, the complexity of analogue neurons, and the unusual properties of electron-trapping materials (ETMs): the latter added a further layer of (biologically plausible) nonlinear interaction to the network.

It is worth understanding how the additional complexity is achieved. ETMs are materials that contain two sets of impurities: one with an electron that is easily liberated (Eu2+) and another that provides a trap for it (Sm3+). On illumination with blue light, electrons are excited and either fall back to the Eu2+, producing orange fluorescence, or become trapped. On illumination with infrared, trapped electrons tunnel back to the Eu ions and then fall into its ground state, again producing orange light. Electron beams can also be used instead of the blue wavelength, increasing the complexity of the interaction. Normally these materials are used in a read-write series: where the IR reads out what the blue records. However, when IR and blue light illuminate the ETM simultaneously, the dynamics become complex: especially if one or both beams are changing with time. As a result, Farhat determined that the ETMs could provide input and output (dendritic and synaptic) weights, nonlinearly coupling bifurcation neurons together.

In (Farhat,1998) Farhat showed that, in simulation at least, the bifurcation neurons formed nonlinearly interacting phase-locked *netlets* or neuronal assemblies with external input. The chaotic periods served as noise, and assisted neural functioning through a kind of stimulated annealing process: it would help the netlet to converge by popping it out of local minima. According to Farhat, this behavior was very robust at the netlet level, even if the reactions of particular neurons within it were imprecise. Further, he says that this behaviour seems to mimic that of cortical neurons and will be potentially useful in creating intelligent machines. Experiments with similar networks ((Farhat,1997)) have shown that they can be used to accurately discriminate between objects. Similar conclusions about noise have been reached by researchers at Boston University((Mar et al. 1999)).

The fact that complex (analogue) neural behaviour exists in nature does not prove that it is functional, but Abu-Mostafa argues that it is. In his appendix to Mead's book (Abu-Mostafa,1989), he shows that not only does using analogue neurons with more than two inputs reduce the number of neurons required (to replace N K-input analogue neurons, a minimum of NK^2 binary neurons would be required), but there is no guarantee that such binary network would do the job.

That said, there is no reason *in principle* to argue that neuron complexity cannot be traded-off against speed, network size, etc.. However, *in practice* (and this chapter is concerned with the engineering aspects of systems), though imperfect, biology is generally very efficient. For example, the use of ATP (adenosine triphosphate) as a fuel for motor proteins (such as kinesin and myosin) is highly efficient((Howard,2001)). It has almost no thermal by-product, turning chemical energy almost entirely into mechanical energy. By over-simplifying neural processes and ignoring the complexity of biological exemplars, we risk turning human-like embodied AI from a problem that could be solved practically (implemented in a machine the size of a man) into one that cannot (where the machine would have to be the size of a planet).

7. Hybrid system design and A/D conversion

Most work in embodied artificial intelligence (see e.g. (Johnson et al. 1995)) has a conventional structure: sensors receive analogue information from the outside world, turn it into bits, and send it to be processed by a digital computer. Once the processor has decided what to do, the decision is passed to a controller that turns this information into the appropriate (often analogue) driving signal for the actuators. Such machines can be defined as *hybrid*—both analogue and digital—and their structure as *conventional*: the A/D and D/A conversion stages necessitate the isolation of each of the sensor, processor, and actuation stages from each other.

It is not always necessary to design systems in this way: instead, all three stages can be merged into a single physical system. Such an approach has two main advantages. The most obvious is in terms of speed and power consumption. This fact is increasingly being recognised by roboticists. For instance Matthew Williamson, who worked on Rodney Brooks *Cog* project, was charged with engineering the robot's limbs in such a way as to ease both the information-processing and energy burden they represented for the machine as a whole. A mechanical engineer by training, he designed Cog's arms and wrists so that they were compliant ((Williamson,1995)) and could respond *mechanically* to changes in the environment rather than purely through conventional sensor-processor-actuator loops. The result was not only a more mechanically-efficient and natural-looking movement, but considerably lower computational overheads.

Others interested in exploiting a balance between information processing and a machine's physical dynamics (in this case, mechanical) computation include Lungarella and Berthouze((Lungarella et al. 2002)). They have looked at how temporarily restricting the degrees of freedom of a mechanical system can improve a robot's ability to learn how to manipulate it. In another recent paper, Pfeiffer gives numerous examples of the importance of mechanical design to machine intelligence((Pfeifer,2002)). Going further back, biologically-inspired roboticists have looked at how cats survive falls from tall buildings through a mechanical process of turning, and through using their legs as a buffer between critical systems (the body and brain) and the ground((Cameron et al. 1991)).

Another advantage of avoiding the conventional setup is that it is not necessary to determine what resolution of A/D/A conversion is necessary. Conventionally—in the case of machine vision, for instance—an engineer has consider how many grey levels are necessary to implement a particular task (such as floor-following (Horswill,1993), defect detection (Davies,1990), or number recognition((Hinton,2000))) in a particular set of conditions (eg. over a given range of light levels, object orientations, distances, etc.) in order to specify the hardware. Though machine vision is still at a primitive stage, the reverse-engineering process for choosing sensor sensitivity, resolution, and dynamic range is generally well understood and is used for most applications.

However, because one of the prerequisites for such reverse engineering—a clear specification—is unavailable for our application, the interface between analogue and digital layers becomes a problem. This fact is intrinsic to point S3 in our definition of human-like AI given earlier: the machine must be free to adapt to its environment based what's important to its survival. This suggests that the machine should be free to use the dynamic range of its sensors as it sees fit, without an artificial A/D conversion limit. If the maximum bit resolution is specified in advance, then information that came in through the sensors may be withheld from the processor. As this information is unavailable for scrutiny then the embodied agent will never be able to learn whether it was important or not.

In humans and animals, sensors output their information as asynchronous spike trains: the timing between spikes can vary continuously and is not regulated by a clock. This freedom from being forced into discrete values can be useful. For some problems, non-linear greyscales (such as logarithmic, see Weber-Fechner law, e.g. (Walker,1995)) are preferable to linear: particular light levels may more important than others and require more discrimination, others less. Further, if the scale can be changed for different applications or environmental conditions, then the whatever information is available may be exploited to it's maximum potential.

Perhaps the clearest example of where A/D/A conversion can be a problem is in systems where feedback is important. In electronics, positive and negative feedback are used to either minimize or maximize small changes in an input, thus making a circuit either stable to small changes or extremely sensitive to them((Young,1988)). A nice example can be drawn from opto-electronics ((Dupertuis et al. 2000)) where a semiconductor optical amplifier (SOA) can be made several times faster without reducing gain or increasing current through the injection of a continuous-wave light beam. The technique exploits the transparency point of a photonic device—where a wavelength is absorbed and emitted equally—using the incoming light to boost the production of carriers when they are most needed. The problem is that, when the incoming signal at one wavelength is amplified by the SOA, there is a decrease in the number of charge carriers available to produce gain. Electronically it takes a long time to replenish these. However, the decrease in charge carriers has a secondary effect: it shifts the transparency point of the material. The continuous-wave beam, which is at a different wavelength, is at this transparency point, having almost no effect on the device before the signal arrives. While the signal is present and amplification takes place, however, the resulting drop in charge carriers pushes the injection wavelength into the device's absorbing region, and the absorption, in turn, quickly produces the needed charge carriers.

Another example, particularly relevant for our application, is local inhibition in the retina (for a brief review and recent developments, see (Roska et al. 2000)). Here, small differences in the intensity of light received are amplified so that only the brightest pixel amongst nearest neighbors fires. This principle has been incorporated into neuromorphic systems called *winner-take-all* networks ((Lazzaro et al. 1989)) and these have been used in navigation and sensing (e.g. (Indiveri et al. 1996)). Part of their advantage comes from their analogue nature, which means it is almost impossible that two incoming signals can appear to have the same intensity. Even the smallest difference can be leveraged to produce a clear difference. If digitized, this would not be the case.

Another area in which resolution issues can become important is sensor fusion: this is particularly because of the inter-dependence of sensor systems within a complex embodied intelligence such as a human. In essence, the sensors perform an application that is more than implementing a particular sense. Specifically, the eyes must not only supply sufficient information to allow, for instance, visual pattern recognition to take place, but it must supply sufficient information to guide and supplement (for example) locomotion or audition. Further, it must supply this information in an appropriate form.

To give a more specific example, in the spinal chords of many vertebrates (Cohen et al. 1992), sensory information is directly incorporated into the gait of the animal through a process of entrainment. Cohen and colleagues have modelled the interaction with visual feedback both theoretically and experimentally. The spinal chord has its own driving frequency, the source of which has been modelled as a coupling between adjacent oscillator sections (such as the vertebrae). The coupling takes place through the neural integration of spike trains, and so can be manipulated by the addition of extra spikes.

The word 'entrainment' refers to the modulation of one signal by another (similar) oscillation, with the output both frequency- and phase-locked to the latter. This kind of entrainment does occur with some sensory input, particularly tactile sensors directly connected to the locomoting limbs. However, in the visual case, the signals are not of similar frequency: instead the integrating neuron reaches its threshold condition earlier when additional excitatory spikes are added, more slowly when the extra signal is inhibitory. In this way, the analogue signals from one sensory modality can directly interact with the behaviour of another system: without any explicit processing.

This is a crucial point. Studies of the brain in many different animals (see examples for human, monkey, and cat in (Foxe et al. 2000), (Duhamel et al. 1998), (Wallace et al. 1997), respectively) have shown that visual and auditory signals are not just processed by their own processor (or cortex), but that there are connections that leave each of these sensory systems very early in the chain of processing. Thus, the resolution required for visual applications is not relevant in providing the correct determination for those systems that share visual information. As in the spinal chord example, such systems are not necessarily making use of the results of conventional visual processes (like pattern recognition, object tracking etc.). Rather, they are processing the visual signals in their own ways for their own ends.

Because of the fact that the sharing systems—those directly exploiting sensory signals that are nevertheless outside their primary modality—are working with relatively raw (unprocessed) signals, they can potentially make use of any information that such signals contain. If the signal is digitised then, again, the information available has been artificially restricted. Once again we return to the problem that, in order to be able to set a resolution that is indistinguishable from the original analogue, then it is necessary to fully understand both the neural processes and the applications they are serving. Such an expectation would not be permitted given our definition of an embodied human-like artificial intelligence given previously.

8. An integrated approach

An excellent example of the advantages to be derived from taking a more integrated approach—in this case integrating optics and electronics—to designing intelligent systems is illustrated by CDM Optics (Boulder, CO) wavefront coding. This approach attempts to match the optical design to the detector electronics and application. In it, unexploited resolution is traded for increased depth of field and a slight increase in noise. The resolution may be unused, for instance, because the pixel separation is significantly greater than the focal-spot size, and can be exchanged for either a deeper image, or for one with a more relaxed focusing tolerance (see figure). One application of this is in color imaging: a single lens can be designed that focuses light over a broad bandwidth—like white light—without chromatic dispersion. The technique works by designing optics that have a large—but uniformly distributed—point-spread function (PSF). Images produced by such optics appear blurry but, because the blur is uniform, it can easily be extracted by image processing: the PSF can be used as a kernel filter.

Lenses designed to produce the wavefront to be decoded generally look very different from conventional optics. In the case of the conformal IR-imaging system, the imaging side is conventional but the detector side has a cosine-form surface—one with three teardrop-shaped peaks circled around, and pointing toward, the center, [Kubala 2003]. This design not only takes account of the imaging characteristics that are required, but also of the image-processing overhead. Because increasing the kernel-filter size rapidly increases the processing time, the size had to be limited to 10Å~ 10 pixels. The ability of wavefront coding to accommodate such constraints is one of the things that makes it so powerful.

Examples such as this one—as well as many of the others in this paper—show how important it is that sensing, processing, and actuation as not treated as different design problems but part of the a single physical problem. We have developed a model of physical computation [Bains 2003] that seeks to show why this is the case, and have begun to assess the implications of this fact for engineering. One of the most basic results of this work is that information processing is just a slice of what any given physical object (even one that we built entirely for information processing) can do. Understanding the relationship between the virtual machine and physical machine is, we believe, the first step towards a more integrated approach to co-designing the brains and bodies of robots.

Conclusion

In this paper we have described a number of different projects all of which, we believe, point to the fact designing the most efficient embodied intelligent systems requires an integrated approach to robotics. Such an approach, which is consistent with neuromorphic engineering, moves beyond the consideration of brain (electronics) and body (optics, electronics, mechanics) as separate sub-systems. Instead, it seeks to combine them and integrate them fully, exploiting the processing ability of all system components regardless of their primary purpose.

References

Abu-Mostafa, Yaser S. (1988), *Connectivity versus entropy,* in **Neural Information Processing Systems,** American Institute of Physics.

Abu-Mostafa, Yaser S. (1989), *Complexity in Neural Systems,* in **Analog VLSI and Neural Systems,** Addison Wesley.

Abu-Mustafa, Yaser S. (1988), *Lower bound for connectivity in local-learning neural networks,* **Journal of Complexity 4**, pp. 246–255.

Araki, S., M. Kajita, K] Kasahara, K. Kubota, K. Kurihara, I. Redmond, E. Schenfeld, and T. Suzaki (1996), *Experimental free-space optical network for massively parallel computers,* **Applied Optics 35** (8), pp. 1269–1281.

Arcas, Blaise Agüera y, Adrienne L. Fairhall, and William Bialek (2000), *What Can a Single Neuron Compute?*, Paper presented at **Neural Information Processing Systems**.

Bains, Sunny (2003), *Intelligence as physical computation and embodied artificial intelligence,* **AISB J. 1** (3), July 2003.

Boahen, Kwabena A. (2000), *Point-to-point connectivity between neuromorphic chips using address-events,* **IEEE Transactions on Circuits and Systems II 47** (5), pp. 416–434.

Cameron, Jonathan and Ronald C. Arkin (1991), *Survival of falling robots,* **Proceedings of the SPIE 1613**, pp. 91–102.

Carson, John C. (2000), *3D Silicon and Recognition-Based Logic -- enabling the road to HAL,* **Proceedings of the SPIE 4109**, pp. 264–270.

Casasent, David and Demetri Psaltis (1976), *Position, rotation, and scale invariant optical correlation,* **Applied Optics 15** (7), pp. 1795–1799.

Chua, Leon O. (1998), **CNN: A Paradigm for Complexity,** World Scientific.

Churchland, P. S. and T. J. Sejnowski (1992), **The Computational Brain,** MIT Press.

Cohen, Avis H., G. B. Ermentrout, T. Kiemel, N. Kopell, K. Sigvardt, and T. Williams (1992), *Modelling of intersegmental coordination in the lamprey central pattern generator for locomotion,* **Trends in Neurosciences 15**, pp. 434–438.

Culurciello, Eugenio, Ralph Etienne-Cummings, and Kwabena A. Boahen (2001), *Arbitrated address-event representation digital image sensor,* **Electronics Letters 37** (24), pp. 1443–1445.

Davies, E. R. (1990), **Machine Vision**, pp. 374–377, Academic Press.

Duhamel, Jean-Rene, Carol L. Colby, and Michael E. Goldberg (1998), *Ventral Intraparietal Area of the Macaque: Congruent Visual and Somatic Response Properties,* **Journal of Neurophysiology 79**, pp. 126–136.

Dupertuis, M. A., J. L. Pleumeekers, T. P. Hessler, P. E. Selbmann, B. Deveaud, B. Dagens, and J. Y. Emery (2000), *Extremely fast high-gain and low-current SOA by optical speed-up at transparency,* **IEEE Photonics Technology Letters 12** (11), pp. 1453–1455.

Farhat, Nabil H. (1997), *Cognitive networks for ATR: The roles of synchronicity, bifurcation and chaos*, Final Report: Office of Naval Research.

Farhat, Nabil H. (1998), *Biomorphic Dynamical Networks for Cognition and Control,* **Journal of Intelligent and Robotic Systems 21**, pp. 167–177.

Farhat, Nabil H., Demetri Psaltis, Aluizio Prata, and Eung Paek (1985), *Optical implementation of the Hopfield model,* **Applied Optics 24** (10), pp. 1469–1475.

Farhat, Nabil H. and Z. Wen (1995), *Large-scale photonic neural networks with biology-like processing elements: the role of electron trapping materials,* **Proceedings of the SPIE 2565**, pp. 2–11.

Farhat, Nabil H. and E. del Moral Hernandez (1996), *Recurrent networks with recursive processing elements: paradigm for dynamical computing,* **Proceedings of the SPIE 2824**, pp. 158–170.

Foxe, John J., Istvan A. Morocz, Micah M. Murray, Beth A. Higgins, Daniel C. Javitt, and Charles E. Schroeder (2000), *Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping,* **Cognitive Brain Research 10**, pp. 77–83.

Frietman, E. E. E. (1995), **Opto-Electronic Processing & Networking: A Design Study,** Ph.D. Thesis, Delft University of Technology.

Hecht, Eugene (1987), **Optics,** Addison Wesley.

Hinton, Geoffrey E. (2000), *Training Products of Experts by Minimizing Contrastive Divergence*, Technical Report GCNU TR 2000-004, Gatsby Computational Neuroscience Unit, University College London.

Hodgkin, A. L. and A. F. Huxley (1952), *Currents carried by sodium and potassium ions through the membrane of the giant axon of* Loligo, **Journal of Physiology 116**, pp. 449.

Horswill, I. (1993), *Polly: A Vision-Based Artificial Agent,* in **Proc. AAAI-93.**

Howard, Jonathon (2001), **Mechanics of Motor Proteins and the Cytoskeleton,** Sinauer.

Indiveri, G., J. Kramer, and Christof Koch (1996), *Neuromorphic Vision Chips: intelligent sensors for industrial applications*, Paper presented at **Advanced Microsystems for Automotive Applications**, Dec.

Johnson, Jeffrey and Philip Picton (1995), **Designing Intelligent Machines: Concepts in Artificial Intelligence,** The Open University; Butterworth-Heinemann.

Jokerst, N. M., Martin A. Brooke, Joy Laskar, D. Scott Wills, April S. Brown, Olivier Vendier, Steven W. Bond, Jeffrey B. Cross, Michael Vrazel, Mikkel Thomas, Myunghee Lee, Sungyong Jung, YoongJoon Joo, and Jae J. Chang (2000), *Smart Photonics: Optoelectronics Integrated with SI CMOS VLSI Circuits,* **Proceedings of the SPIE 4109**, pp. 241–251.

Juday, Richard D., Stanley E. Jr Monroe, and Don A. Gregory (1987), *Optical Correlation with Phase Encoding and Phase Filtering,* **Proceedings of the SPIE 825**, pp. 149–155.

Klein, Lawrence A. (1999), **Sensor and Data Fusion Concepts and Applications,** SPIE Press.

Kenneth Kubala et al. (2003), **Optics Express 11**(18), p. 2102, 8 September.

Lazzaro, John, S. Ryckebusch, Misha A. Mahowald, and Carver A. Mead (1989), *Winner-Take-All Networks Of O(N) Complexity,* in **Advances in Neural Information Processing Systems,** Morgan Kaufman.

Li, Guoqiang, Dawei Huang, Emel Yuceturk, Philippe J. Marchand, Sadik C. Esener, Volkan H. Ozguz, and Yue Liu (2002), *Three-dimensional optoelectronic stacked processor by use of free-space optical interconnection and three-dimensional VLSI chip stacks,* **Applied Optics 41** (2), pp. 348.

Lungarella, Max and Luc Berthouze (2002), *Adaptivity through physical immaturity*, Oral presented at **2nd International Workshop on Epigenetic Robotics**.

Mahowald, Misha A. (1992), **VLSI Analogs of Neuronal Visual Processing: A Synthesis of Form and Function,** Ph.D. Thesis, California Institute of Technology.

Mar, D. J., C. C. Chow, W. Gerstner, R. W. Adams, and J. J. Adams (1999), *Noise shaping in populations of coupled model neurons,* **Proceedings of the National Academy of Sciences USA 96**, pp. 10450–10455.

Mead, Carver A. (1989), **Analog VLSI and Neural Systems,** Addison Wesley.

Miki, Tetsuya and Hideki Ishio (1978), *Viabilities of the Wavelength-Division-Multiplexing Transmission System Over an Optical Fiber Cable,* **IEEE Transactions on Communications Com-26** (7), pp. 1082–1087.

Miller, D. A. B. (1989), *Optics for low-energy communication inside digital processors: quantum detectors, sources and modulators as efficient impedance converters,* **Optics Letters 14** (2), pp. 146–148.

Miller, D. A. B. (1997), *Physical Reasons for Optical Interconnection,* **International Journal on Optoelectronics 11** (3), pp. 155–168.

Miller, D. A. B. and H. M. Özaktas (1997), *Limit to the Bit-Rate Capacity of Electrical Interconnects from the Aspect Ratio of the System Architecture,* **Journal of Parallel and Distributed Computing 41**, pp. 42–52.

Mok, Fai, Demetri Psaltis, and Geoffry Burr (1992), *Spatially- and Angle-multiplexed Holographic Random Access Memory,* **Proceedings of the SPIE 1773**, pp. 334–345.

Pfeifer, Rolf (2002), *On the role of embodiment in the emergence of cognition: Grey Walter's turtles and beyond*, Invited Speaker presented at **Biologically-Inspired Robotics: The Legacy of W. Grey Walter**.

Psaltis, Demetri, Mark A. Neifeld, Alan Yamamura, and Seiji Kobayashi (1990), *Optical memory disks in optical information processing,* **Applied Optics 29** (14), pp. 2038–2055.

Roska, Botond, Erik Nemeth, Laszlo Orzo, and Frank S. Werblin (2000), *Three Levels of Lateral Inhibition: A Space–Time Study of the Retina of the Tiger Salamander,* **The Journal of Neuroscience 20** (5), pp. 1941–1951.

Walker, Peter M. B. (1995), **Wordsworth Science and Technology Dictionary,** Wordsworth Reference.

Wallace, Mark T. and Barry E. Stein (1997), *Development of Multisensory Neurons and Multisensory Integration in Cat Superior Colliculus,* **Journal of Neuroscience 17** (7), pp. 2429–2444.

Wehr, M. and G. Laurent (1996), *Odour encoding by temporal sequences of firing in oscillating neural assemblies,* **Nature 384**, pp. 162–166, 14 Nov.

Williamson, Matthew M. (1995), **Series Elastic Actuators,** MIT AI Lab.

Young, E. C. (1988), **Dictionary of Electronics,** Penguin.