

Optoelectronic Intelligence

Jeffrey M. Shainline

NIST, Boulder, CO, 80305

December 26, 2018

Abstract

We motivate the design of optoelectronic neural systems based on principles of neuroscience and very-large-scale integration. We argue that for large neural systems capable of general intelligence, the strengths of photonics for communication and electronics for computation are indispensable. Based on these considerations, we sketch a concept for optoelectronic hardware, beginning with synaptic circuits and extending to systems at the scale of the human brain and beyond.

1 Introduction

Light is excellent for communication. Fiber optic links carry vast quantities of information across continents and between data centers. An important question in modern computing is: what is the shortest distance over which photonic communication will displace electronic interconnects? Optical links between racks in data centers are becoming common. Major companies are investing seriously in photonics in the package. Monolithic optical links between processor and memory fabricated in a 45-nm CMOS node with no in-line changes have been demonstrated [1], with significant functionality still to be explored in this zero-change approach to optoelectronic integration [2]. A primary challenge affecting further chip-scale electronic-photonic integration is the difficulty of achieving a light source on silicon that is robust, efficient, and economical.

In parallel with hardware considerations affecting optoelectronic integration are questions related to architecture. A prominent theme emerging since clock speed leveled off in the early 2000s is parallelism. Computation is increasingly distributed among more processor cores. Many-core architectures continue to expand into on-chip networks, in some cases resulting in highly distributed, brain-inspired systems [3–8]. As compute grows more distributed, communication across interconnection networks becomes a bottleneck. The demand for energy efficient communication bandwidth has been a major driver of on-chip photonics.

The major drivers for brain-inspired computers fall on a spectrum: energy and algorithmic efficiency for mobile applications reside on one side of the spectrum, and artificial general intelligence (AGI) resides on the other. Knowledge gained from neuroscience informs us that systems with general intelligence will benefit from very large numbers of computational elements as well as extreme communication between them. It is our perspective that hardware

incorporating light for communication between electronic computational elements combined with an architecture of distributed optoelectronic spiking neurons will provide tremendous potential for AGI. While much of the aforementioned computing infrastructure has evolved to implement a von Neumann architecture performing sequential operations in the model of a Turing machine, the functioning of neural systems departs considerably from this model. We think light has even more to offer in a neural computing context, and the hardware requirements are somewhat distinct from integrated photonics for digital logic. Considerations pertinent to the realization of optoelectronic hardware for AGI are the subject of this article.

2 Structure and function across space and time

To guide the design of hardware for AGI, we must simultaneously consider operation across spatial and temporal scales [12]. In Fig. 1, we chart the structures present on various length scales and temporal scales for biological and optoelectronic hardware. The human brain has features spanning roughly eight orders of magnitude in size, from a nanometer to a tenth of a meter, with correlated dynamical activity across four orders of magnitude, from below 1 Hz to around 600 Hz. There are several types of optoelectronic neural systems, and we have proposed a specific approach we see as most conducive to large-scale implementation for AGI [13]. These optoelectronic networks are likely to have features as small as 100 nm and potentially extend up to many kilometers. Dynamical activity may be present from slow temporal scales up to 20 MHz (30,000 times the maximum speed of the brain) for networks measured in kilometers. Note that for the optoelectronic hardware under consideration, the operation of the smallest components (sources and detectors) occurs on a time scale of 50 ns (worst-case scenario), and no aspect of the system limits this speed until the scale of networks 10 m on a side, at which point light-speed communication takes longer than the synaptic or neuronal refractory period. Figure 1 is intended to be viewed qualitatively and emphasizes the point that if communication bottlenecks can be removed, neural systems much larger and much faster than the brain can be achieved. To enable such spatial and temporal scaling, fractal properties must be employed.

Information processing in neural systems employs local clusters of neurons to represent specific features, and the information from these clusters must be shared with other regions of the network to form a multifaceted representation of a complex stimulus. Structurally, this information processing is facilitated by networks with a high clustering coefficient yet also an average path length nearly as short as a random graph [14]. Such graph structures are referred to as small-world networks [15], and are ubiquitous throughout the brain [16]. To achieve small world

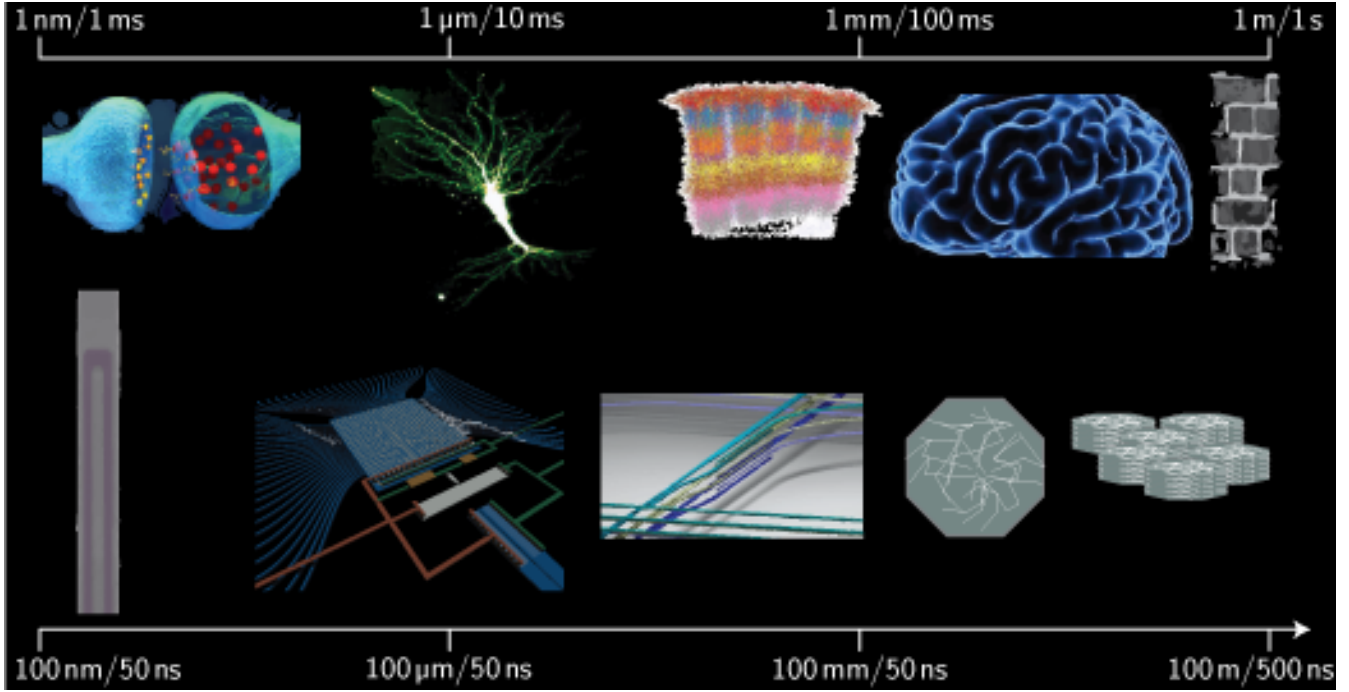


Figure 1: Complexity across scales. Biological systems have functional components on the nanometer scale up to the full brain (roughly 0.3m linear dimension for full human cerebral cortex [9]). The speeds of the various operations are limited by chemical diffusion and axon propagation, which may ultimately limit the size of biological neural systems [10,11]. By contrast, optoelectronic devices rarely have functional components with critical dimension smaller than 100 nm, but the high speed of devices and communication enables optoelectronic systems to extend far beyond the limits imposed by the slow conduction velocity of axons.

networks, long-range connections are necessary. In a random network, near and distant connections are equally probable, so the average path length across the network is small, representing a lower limit on path length for a given number of edges connecting a given number of nodes. In Fig. 2(a) we plot the number of edges required per node to achieve a given average path length as a function of the number of nodes in the network [17]. Consider the case of a network with one million nodes. We see from this plot that if we wish to maintain a path length of two, each node must make, on average, one thousand connections. For the case of a network with 100 million nodes, each node must make 10,000 connections. This is similar to the case of the hippocampus in the human brain, with nearly 100 million neurons, each with 10,000 or more synaptic connections [10]. Maintaining a short average path length across the network is critical to enable efficient information integration, and this appears to be a major factor driving the extensive connectivity of biological neural systems. In the present context, this motivates us to conceive of artificial hardware capable of supporting comparable connectivity, which leads us to consider using light for communication.

In addition to small-world characteristics, networks of the brain demonstrate a hierarchical architecture wherein minicolumns aggregate into columns, columns into modules, and modules into complexes. This fractal property

is necessary to enable networks to continue to scale arbitrarily, with dynamics constrained only by the physical hardware and spatial extent of the system rather than by the ability to communicate across the network [19]. Hierarchical architecture with simple device behavior is sufficient to achieve some of the dynamical features of the brain, such as neuronal avalanches [20,21] that are observed in the resting state [22]. Other dynamical behaviors thought to be necessary for attention, cognition, and learning, such as cross-frequency coupling [23] and synaptic plasticity [24–27], likely require complex capabilities at the device level. Dynamical synapses, dendrites, and neurons allow the same structural network to realize myriad functional networks. On the time scale of neuronal inter-spike intervals, short-term synaptic plasticity enables synapses to rapidly adapt, resulting in temporal filtering of afferent spike trains [28]. Also over short time scales, inhibitory neurons can silence or give voice to specific synapses, clusters of synapses on dendrites, or neurons. Over time scales long compared to an average inter-spike interval, synaptic plasticity is employed to store memories and adapt the network to a small-world architecture [29]. Learning in the presence of a continually changing environment while maintaining long-term memories is achieved through spike-timing-dependent plasticity (STDP) [24,25] and metaplasticity [26,27]. Neurons employing inhibition [30], dendritic processing [31–33], short-

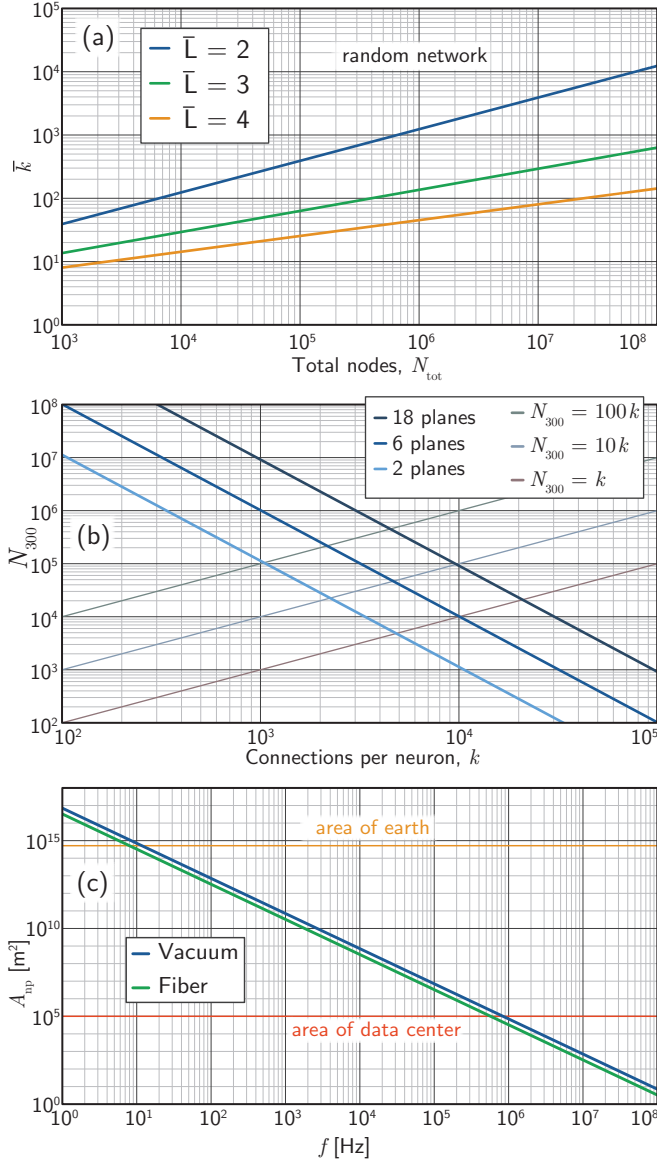


Figure 2: Scaling considerations for optoelectronic neural systems. (a) The average number of connections per node required to maintain a give average path length across a random network as a function of the total number of nodes in the system. (b) The total number of nodes that can fit on a 300 mm wafer as a function of the number of connections per node in the wire-limited regime [18]. (c) The area of the neuronal pool as a function of the frequency of neuronal oscillations assuming light-speed communication [11].

term [28], and long-term synaptic plasticity [24–27] can make efficient use of a structural network by dynamically adapting on time scales from an inter-spike interval to the lifetime of the system. While light is excellent for communication, electrical circuits are better equipped to perform these nonlinear, dynamical functions.

We mention these considerations of network structure and device dynamics to emphasize that neural information processing depends critically on operations best performed by light and well as on operations best performed by electronics. This form of information processing will benefit immensely from optoelectronic integration. We now discuss further considerations regarding optoelectronic implementation of neural systems, beginning with the computational elements.

3 Optoelectronic synapses, dendrites, and neurons

We envision optoelectronic hardware with photonic communication along waveguides playing the role of axons, and computation performed in electronic synapses, dendrites, and neurons. We make two more choices to specify the platform. There are 10^{14} synapses in the human brain. An artificial system must use very low energy per synaptic event if it is going to support activity across this number of synapses with a feasible amount of power. Additionally, if artificial hardware dissipates even 10 nW per synapse to retain the state, the system will require a megawatt just to remember what it learned. Having chosen to communicate synaptic events with light, the quantum limit is a single photon per synaptic connection. We have designed a synapse that detects a single infrared photon and requires no power to retain the synaptic state. The synapses utilizes a superconducting-nanowire single-photon detector (SNSPD), a current-biased strip of superconducting wire roughly 100 nm in width and 5 nm in thickness. In the steady state, the current bias flows straight to ground, and upon detection of a photon, a small section of the wire is driven from the superconducting phase to the normal-metal phase, resulting in a transient resistance that diverts the current bias across a load, usually to produce a voltage for readout.

In the present context, to achieve the desired synaptic operation, a SNSPD can be combined in circuits with Josephson junctions (JJs) and superconducting loops coupled through mutual inductors to achieve the functions we need for neural information processing. Candidate circuit designs have been presented in Refs. [34–37], and much more innovation is certainly possible. In optoelectronic synapses of this design, the current bias across a single JJ establishes the synaptic weight. This current bias can be modified through various photonic and electronic means based on network activity. Inhibition is straightforward with opposing mutual inductors. Dendritic and neuronal nonlinearities are a natural consequence of the fact that

Josephson junctions have a critical current. The extreme nonlinearity of the superconducting phase transition plays a leading role for both for the generation of light during a neuronal firing event and the detection of light at synapses. Due to the prominent role of superconducting current storage loops, we refer to these as loop neurons. In the operation of loop neurons, a single photon triggers a synaptic event, and STDP is induced by two photons—one from each neuron associated with the synapse.

In addition to the choice of SNSPDs as the detectors in the system, we must also select a light source, which must be fabricated across wafers by the millions for economical, brain-scale systems. Because our choice of detectors dictates cryogenic operation, silicon light sources operating at 4 K are an option [38, 39]. This neural system may be one of the few applications where silicon light sources are appropriate and sufficient. The light sources we have in mind are silicon LEDs [40], employing luminescence from defect-based dipole emitters [41, 42]. From the perspective of VLSI, achievement of a silicon light source of sufficient performance would be the greatest contribution to the success of this technology. If cryogenic operation enables both single-photon detectors and silicon light sources, it will almost certainly be worth the effort.

To achieve complex neural circuits, we must combine light sources, detectors, and various other circuit elements. We have demonstrated waveguide coupling of light from these micron-scale light sources to integrated SNSPDs on a silicon photonic chip [40]. The performance achieved in the first iteration of these optical links was not particularly impressive (10^6 photons per second generated with 5×10^{-7} efficiency), yet the simplicity of both the source and detector made the demonstration of a monolithic optical link far easier than if room-temperature operation were required. It is important to realize that for this application, high-performance light sources are not required. The light sources are only required to produce incoherent pulses of 10,000 photons (≈ 1 fJ) at 20 MHz when operating at 4 K. If silicon light sources can meet these performance specifications, the hardware stands a chance of enabling brain-scale systems with 30,000 times the speed.

4 Communication with guided light

The central premise of this work is that photons are superior to electrons for communication across large-scale neural systems. To explain why we place this conjecture at the center of hardware development, we briefly summarize the physical limitations of electrical interconnection networks [43]. It is impracticable in silicon electronics for a single device to source current to many other devices. A shared communication network must be employed. In contemporary computing, switched media networks are used for this purpose. Each device must then only communicate to the nearest switch in the network. Because the

communication infrastructure is shared, devices must request access to the switch network to transmit messages. When multiple devices request access simultaneously, arbitration must be performed, wherein devices are queued and sequentially granted access. This approach to communication between electronic devices leverages the speed of electronic circuits to compensate for the difficulties in communication. The limitations are reached when many devices need to communicate with many other devices with a high frequency of communication events. Unfortunately, this is exactly the situation encountered in neural information processing. During a neuronal avalanche, many neurons may need to communicate simultaneously across the network. As more neurons, each with many synapses, are added to the network, the average frequency of neuronal firing events must decrease, and integration of information across the network is limited by the communication infrastructure.

The physics of light is complimentary to that of electrons. Photons, being uncharged bosons, interact very weakly with each other. In the linear regime in which this few-photon technology will operate, photons can co-propagate on waveguides independently of one another without wiring capacitance. This enables a pulse of photons to fan out to many destinations without a charging penalty due to wiring. This is not to say photonic communication can address an arbitrarily large number of recipients without consequence. For each new recipient, the number of photons in the initial pulse must increase, and as destinations get further away, more energy is dissipated to propagation loss. These realities notwithstanding, it appears feasible for devices communicating with photons to make direct connections to thousands of destinations, thereby eliminating the need for the shared communication infrastructure that is the primary impediment to achieving AGI with electrical interconnections.

Having made this claim, the burden is upon us to provide evidence of the feasibility of photonic communication in large-scale neural systems. Much like electrical interconnection networks utilize different technologies at the scales of chips versus data centers, photonic interconnection technologies will vary at different scales to enable communication in chip-scale systems as well as across networks the size of a data center. The small-world architectures that are suitable for neural information processing require neurons to communicate seamlessly across all scales of network hierarchy. The large wavelength of light relative to the size of electronic devices (and relative to the size of devices in the brain) cause concern for the size of optoelectronic brain-scale networks. To build confidence for the feasibility of the endeavor, we sketch a vision of how a general optoelectronic neural system may be constructed.

A successful neural technology must leverage the fabrication infrastructure of silicon electronics. We conjecture that silicon photonics technology will be utilized to move light between neurons. At the wafer scale, light will be

guided in dielectric waveguides, as shown schematically in Fig. 3(a). Silicon photonics provides three primary dielectric materials that can be used for these passive waveguides: Si, SiN, and SiO₂. The indices of refraction of these materials are 3.5, 2.0, and 1.5, respectively, for λ close to 1550 nm. Perhaps not coincidentally, these are the three primary dielectrics used in CMOS technology as well. Achieving the dense routing required to connect large numbers of neurons on a wafer will require multiple planes of waveguides, just as integrated electronics requires multiple wiring layers. We anticipate optoelectronic neural systems will utilize dielectric waveguide layers deposited in the back-end-of-line in the fabrication process, with lower layers having higher index and being utilized for local connections, and higher layers having gradually lower index with lower propagation loss used for more distant connections.

We wish to approximate the area of such photonic interconnection networks. Following Keyes [18], we approximate the area required for the waveguides entering a neuron as $A_n = (n_{\text{in}}w/k)^2$, where n_{in} is the number of waveguides entering the neuron (in-directed synaptic connections), w is the waveguide pitch, and k is the number of planes of waveguides. For tiling multi-wafer assemblies, wafers diced into octagons may be advantageous, so we take the area of a wafer to be $A_8 = 2\sqrt{2}r^2$ with $r = 150$ mm. The number of neurons that can be supported on a 300-mm wafer is given by the ratio

$$N_8 = \frac{A_8}{A_n} = 2\sqrt{2}r^2 \left(\frac{k}{wn_{\text{in}}} \right)^2. \quad (1)$$

This expression is plotted in Fig. 2(b). This estimate informs us that a 300 mm wafer with six waveguide planes can support roughly one million neurons if they each have one thousand connections. More involved analysis finds a slightly smaller number [44]. As a point of comparison to electrical neural systems, Ref. [45] finds that through multi-layer, wafer-scale integration of logic and memory, 250 million electrical neurons could fit on a 300 mm wafer. The trade-off is speed, as the shared communication network would limit the electrical neurons studied in Ref. [45] to 10 Hz operation when 1000 synaptic connections are made per neuron. Nevertheless, the message of Fig. 2(b) is that photonic routing results in large area consumption. The human cortex contains over 10 billion neurons. An optoelectronic brain larger than that of a bumble bee will not fit on a wafer.

Optoelectronic intelligence will require communication between wafers. Wafers can be stacked vertically, and free-space optical links can send photons from a source on one wafer to a detector on a wafer above or below, as illustrated in Fig. 3(b). Such 3D wafer-stacking techniques are being developed for electronics [46,47], but the ability of light to propagate through free space (or liquid helium) and the ten-micron alignment tolerances enabled by wide-area photodetectors [48] make such 3D integration

promising for photonic communication as well. Assuming SNSPDs receiving vertical communication have a pitch of 25 μm , a 300-mm octagon could support 10^8 vertical communication links between two wafers. Assuming half of this area is for feed-forward communication from the lower wafer to the upper wafer, and half is for feed-back from the upper wafer to the lower wafer, this would result in 5×10^7 synaptic connections originating from neurons on one wafer and terminating in neurons on a vertically adjacent wafer. If each wafer had one million neurons with one thousand connections per neuron within the wafer, the total number of intra-wafer synaptic connections would be 10^9 . Therefore, the number of synapses present in a layer of this network that originated on a previous layer would be 5%, similar to the fraction observed in the laminar structure of biological cortex (Ref. [10], pg. 286).

In addition to free-space vertical coupling, inter-wafer communication can be achieved at wafer edges with in-plane waveguide couplers, as shown in Fig. 3(c). In the octagonal (truncated square) tiling used here for illustration, each wafer makes such connections to neighbors in the cardinal directions. With a 10 μm pitch, 11,500 wafer-edge couplers could be supported in each of the cardinal directions with 46,000 total in-plane, lateral connections. Such a system would demonstrate strong connectivity within the vertical stack of the wafers, and weaker lateral connectivity between wafers in the same horizontal plane. Such an architecture resembles the columnar organization of cortex.

The wafer tiling we have just described leads to a picture of optoelectronic networks with vertically stacked columns of wafers with horizontal connectivity emanating from the perimeter of each wafer. To achieve communication from within these columns to other (perhaps distant) regions of the network, optical fibers are ideal. Within the truncated square tiling under consideration, the square areas at diagonals between wafers can support fiber-optic bundles. These optical fiber tracts are analogous to white matter in the brain. One such region could house a million single-mode fibers of 125 μm diameter. These fibers will emanate from all wafers within the column, so the number of outputs available to each wafer will depend on how many vertically integrated wafers are utilized in a column. If six wafers are stacked in a column, each wafer would have roughly 167,000 output fibers to carry information to distant regions of the network. With one million neurons on a wafer, not every neuron would be able to couple to a fiber for long-distance communication. This again is consistent with brain organization wherein the number of long-distance axons emanating from a region is smaller than the number of neurons within the region. Note, however, that each of these fibers can branch as it extends through the white matter, so a neuron with access to a single wafer-edge fiber could establish multiple long-range synaptic connections.

On a wafer, photonic fan-out across dielectric waveguides enables neurons to make thousands of direct con-

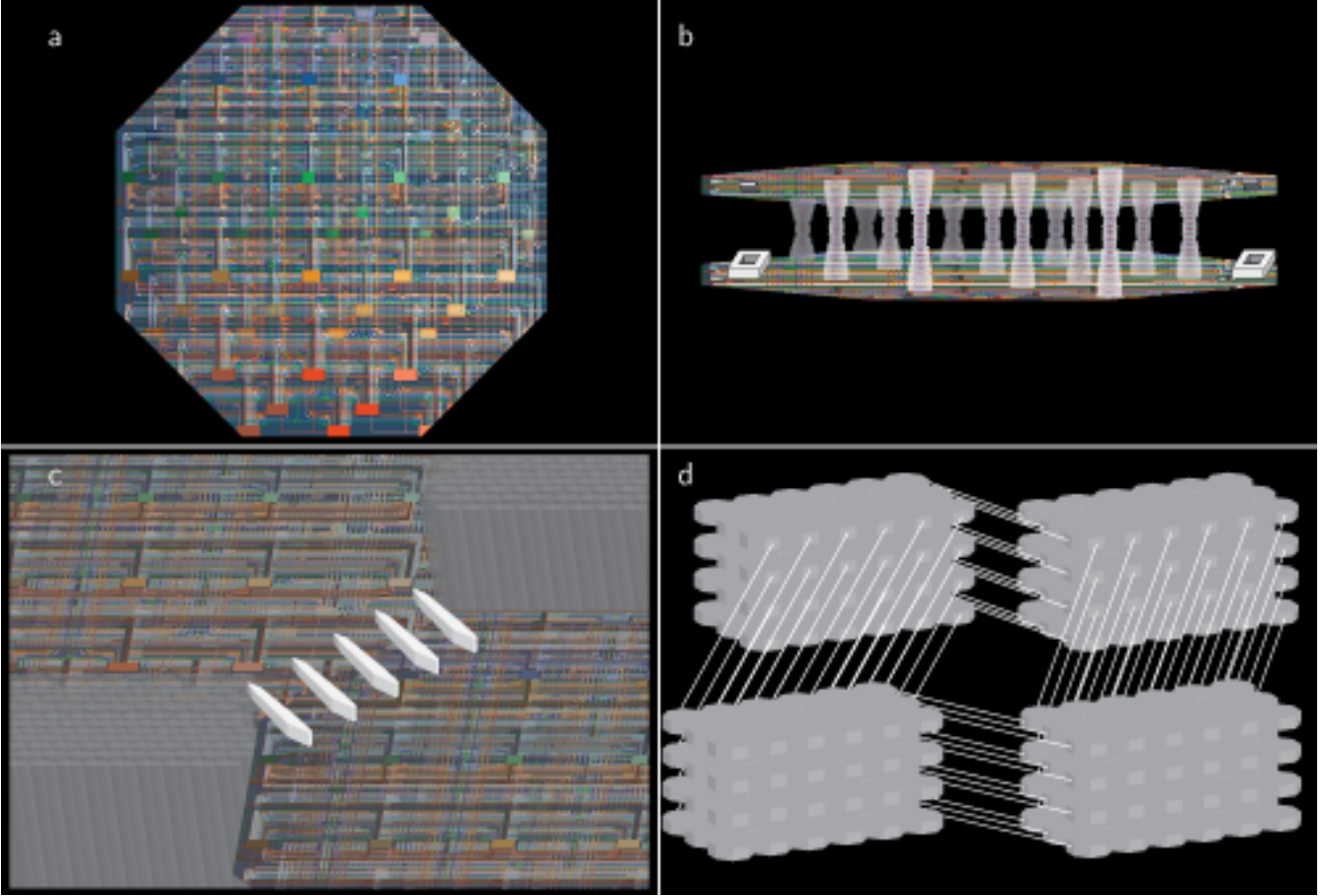


Figure 3: Schematic illustration of communication across scales in optoelectronic neural systems. (a) Schematic of a neural system implemented on a 300 mm wafer, cut into an octagon for tiling. (b) Vertical photonic links between two stacked wafers. (c) Lateral wafer-edge links with fiber optic bundles for long-range communication. (d) A large neural system with fiber-optic connectivity between massive modules, each containing hundred to thousands of wafers.

nections without the limits of a shared switching network. Free-space and wafer-edge couplers enable significant inter-wafer communication conducive to columnar information processing. Such columns can communicate to one another locally and globally over fiber optic links. With this configuration in mind, we can assess the feasibility of constructing systems on the scale of the human cerebral cortex, with 10 billion neurons, each with thousands of synaptic connections. If a wafer holds a million neurons, a brain-scale assembly requires 10,000 wafers and would fit in a volume two meters on a side—the size of a few server racks in a closet.

We are optimistic that this approach to neural information processing will be successful for physical and practical reasons. Physically, due to photonic signaling, it is possible to achieve efficient communication across the network for systems with orders of magnitude more than the 10,000 wafers comprising a brain-scale system. As we see in Fig. 2(c), networks with activity at 20 MHz (a safe estimate for the maximum firing rate of loop neurons) can span an area hundreds of meters on a side before communication delays limit the speed of network activity. Due to superconducting electronics, the power density of such systems will be low enough for liquid helium cooling. On the practical side, fabrication of loop neurons at industrial scale appears feasible. All the proposed circuits can be created on 300 mm wafers with existing infrastructure, such as a 45 nm CMOS node. Ten thousand wafers move through such a foundry every day. If dedicated to fabrication of optoelectronic intelligence, a single foundry may be able to produce multiple brain-scale systems per year. Assembly of the wafers into a functional system would probably not be more difficult than the construction of a contemporary supercomputer. The requirement of liquid-helium cooling is not a major impediment.

5 Discussion

At present, the challenge of creating an artificial intelligence comparable to a human appears formidable with the use of silicon electronics alone. The primary challenge arises because shared communication infrastructure is required, resulting in a connectivity/speed tradeoff. The use of photonic communication will mitigate this tradeoff, despite the increased size of photonic interconnection networks. Photonic fanout enables direct connections between large numbers of neurons, and the velocity of light enables communication across ten-meter systems before communication limits network speed below the 20 MHz where loop neurons are limited by electronic reset times.

Light is excellent for communication, while electronics excel at computation. Artificial neural hardware should be designed and constructed to leverage photonic communication while performing synaptic, dendritic, and neuronal functions with electronic circuits. Superconducting optoelectronic circuits naturally implement these functions, in

part because of the utility of Josephson nonlinearities for neural computation. This hardware differs in important ways from the silicon transistors that implement Boolean logic in a von Neumann architecture. Such digital computers emerged to perform arithmetic calculations in a manner based on a Turing machine, while neural information processing departs markedly from the sequential operation of a Turing apparatus. We should not be surprised that optimal hardware may differ when the objective is so different. Yet for the superconducting optoelectronic hardware discussed here, the same fabrication infrastructure as CMOS can be employed. To achieve neural information processing, light sources and detectors can and must be densely integrated for communication between primitive compute nodes (neurons). For distant communication across the network, the low loss of fiber-optic waveguides is required. The system leverages the strengths of integrated silicon photonics as well as fiber-optic networks.

What are the next steps to realize this technology? Low-cost source-detector integration at the wafer scale is required. These active devices must be augmented with improvements in deposited dielectrics for photonic routing to enable more planes with lower loss. For system scaling, improved fiber-to-waveguide coupling and multi-wafer modules must be demonstrated. All the hardware improvements will not lead to AGI without further theoretical analysis at device, circuit, and system levels. Further understanding the principles of network information processing and designing the architecture to achieve general intelligence are likely to be much more challenging than understanding the operation of a Turing machine. Theoretical progress is required achieve high-performance neural systems, train them, and make them intelligent.

References

- [1] C. Sun, M.T. Wade, Y. Lee, J.S. Orcutt, L. Alloatti, M.S. Georgas, A.S. Waterman, J.M. Shainline, R.R. Avizienis, S. Lin, B.R. Moss, R. Kumar, F. Pavanello, A.H. Atabaki, H.M. Cook, A.J. Ou, J.C. Leu, Y.-H. Chen, K. Asanović, R.J. Ram, M.A. Popović, and V.M. Stojanović. Single-chip microprocessor that communicates directly using light. *Nature*, 528:534, 2015.
- [2] V. Stojanovic, R.J. Ram, M. Popovic, S. Lin, S. Moazeni, M. Wade, C. Sun, L. Alloatti, A. Atabaki, F. Pavanello, N. Mehta, and P. Bhargava. Monolithic silicon-photonics platforms in state-of-the-art cmos soi processes. *Opt. Express*, 26:13106, 2018.
- [3] K.A. Boahen. Point-to-point connectivity between neuromorphic chips using address events. *IEEE Tran. Circ. Sys. II*, 47:416, 2000.
- [4] T. Pfeil, A. Grubl, and K. Meier. Six networks on a universal neuromorphic computing substrate. *Frontiers in Neuroscience*, 7:11, 2013.

- [5] P.A. Merolla, J.V. Arthur, R. Alvarez-Icaza, A.S. Cassidy, J. Sawada, F. Akopyan, B.L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S.K. Esser, R. Appuswamy, B. Taba, A. Amir, M.D. Flickner, W.P. Risk, R. Manohar, and D.S. Modha. A million spiking-neuron integrated circuit with scalable communication network and interface. *Science*, 345:668, 2014.
- [6] S.B. Furber, F. Galluppi, S. Temple, and L.A. Plana. The spinnaker project. *Proceedings of the IEEE*, 102:652, 2014.
- [7] J. Park, T. Yu, S. Joshi, C. Maier, and G. Cauwenberghs. Hierarchical address event routing for reconfigurable large-scale neuromorphic systems. *IEEE Trans. Neural Networks and Learning Systems*, 28:2408, 2017.
- [8] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S.H. Choday, and G. Dimou. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 1:82, 2018.
- [9] H.G. Schnack, N.E.M. van Haren, R.M. Brouwer, A. Evans, S. Durston, D.I. Boomsma, R.S. Kahn, and H.E. Hulshoff Pol. Changes in thickness and surface area of the human cortex and their relationship with intelligence. *Cerebral Cortex*, 25:1608, 2014.
- [10] G. Buzsaki. *Rhythms of the brain*. Oxford University Press, 2006.
- [11] J.M. Shainline. The largest cognitive systems will be optoelectronic. In *IEEE International Conference on Rebooting Computing*. IEEE, Nov 2018.
- [12] R.F. Betzel and D.S. Bassett. Multi-scale brain networks. *NeuroImage*, 160:73, 2016.
- [13] J.M. Shainline, S.M. Buckley, A.N. McCaughan, J. Chiles, R.P. Mirin, and S.W. Nam. Superconducting optoelectronic neurons I: General principles. *arXiv*, page 1805.01929, 2018.
- [14] E. Estrada and P.A. Knight. *A First Course in Network Theory*. Oxford, Oxford, United Kingdom, first edition, 2015.
- [15] D.J. Watts and S.H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440, 1998.
- [16] O. Sporns. *Networks of the Brain*. The MIT Press, Cambridge, Massachusetts, first edition, 2010.
- [17] A. Fronczak, P. Fronczak, and J.A. Holyst. Average path length in random networks. *Phys. Rev. E*, 70:056110, 2004.
- [18] R.W. Keyes. The wire-limited logic chip. *IEEE J. Sol.-Sta. Circuits*, SC-17:1232, 1982.
- [19] D. Plenz and T.C. Thiagarajan. The organizing principles of neuronal avalanches: cell assemblies in the cortex? *Trends in Neuroscience*, 30:101, 2006.
- [20] J.M. Beggs. The criticality hypothesis: how local cortical networks might optimize information processing. *Philosophical transactions of the Royal Society A*, 366:329, 2007.
- [21] E.J. Friedman and A.S. Landsberg. Hierarchical networks, power laws, and neuronal avalanches. *Chaos*, 23:013135, 2013.
- [22] T. Petermann, T.C. Thiagarajan, M.A. Lebedev, M.A.L. Nicolelis, D.R. Chialvo, and D. Plenz. Spontaneous cortical activity in awake monkeys composed of neuronal avalanches. *PNAS*, 106:15921, 2009.
- [23] B.J. He, J.M. Zempel, A.Z. Snyder, and M.E. Raichle. The temporal structures and functional significance of scale-free brain activity. *Neuron*, 66:353, 2010.
- [24] S. Song, K.D. Miller, and L.F. Abbott. Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, 3:919, 2000.
- [25] H. Markram, W. Gerstner, and P.J. Sjöström. Spike-timing-dependent plasticity: a comprehensive overview. *Frontiers in Synaptic Neuroscience*, 4:2, 2012.
- [26] W.C. Abraham. Metaplasticity: tuning synapses and networks for plasticity. *Nature Neuroscience*, 9:387, 2008.
- [27] S. Fusi, P.J. Drew, and L.F. Abbott. Cascade models of synaptically stored memories. *Neuron*, 45:599, 2005.
- [28] L.F. Abbott and W.G. Regehr. Synaptic computation. *Nature Reviews*, 431:796, 2004.
- [29] C.-W. Shin and S. Kim. Self-organized criticality and scale-free properties in emergent functional neural networks. *Phys. Rev. E*, 74:045101(R), 2006.
- [30] G. Buzsaki and A. Draguhn. Neuronal oscillations in cortical networks. *Science*, 304:1926, 2004.
- [31] J. Hawkins and S. Ahmad. Why neurons have thousands of synapses, a theory of sequence memory in neocortex. *Frontiers in Neural Circuits*, 10:23, 2016.
- [32] G.J. Stuart and N. Spruston. Dendritic integration: 60 years of progress. *Nature Neuroscience*, 18:1713, 2015.
- [33] S. Sardi, R. Vardi, A. Sheinin, A. Goldental, and I. Kanter. New types of experiments reveal that a neuron functions as multiple independent threshold units. *Scientific reports*, 7:18036, 2017.

- [34] J.M. Shainline, S.M. Buckley, A.N. McCaughan, M. Castellanos-Beltran, C.A. Donnelly, M.L. Schneider, R.P. Mirin, and S.W. Nam. Superconducting optoelectronic neurons II: Receiver circuits. *arXiv*, page 1805.02599, 2018.
- [35] J.M. Shainline, A.N. McCaughan, S.M. Buckley, C.A. Donnelly, M. Castellanos-Beltran, M.L. Schneider, R.P. Mirin, and S.W. Nam. Superconducting optoelectronic neurons III: Synaptic plasticity. *arXiv*, page 1805.01937, 2018.
- [36] J.M. Shainline, A.N. McCaughan, A. Jafari-Salim, S.M. Buckley, R.P. Mirin, and S.W. Nam. Superconducting optoelectronic neurons IV: Transmitter circuits. *arXiv*, page 1805.01941, 2018.
- [37] J.M. Shainline, S.M. Buckley, A.N. McCaughan, J. Chiles, A. Jafari-Salim, R.P. Mirin, and S.W. Nam. Circuit designs for superconducting optoelectronic loop neurons. *J. Appl. Phys.*, page 152130, 2018.
- [38] Gordon Davies. The optical properties of luminescence centres in silicon. *Physics Reports*, 176:83–188, 1989.
- [39] J.M. Shainline and J. Xu. Silicon as an emissive optical medium. *Laser and Photonics Reviews*, 1:334, 2007.
- [40] S. Buckley, J. Chiles, A.N. McCaughan, G. Moody, K.L. Silverman, M.J. Stevens, R.P. Mirin, S.W. Nam, and J.M. Shainline. All-silicon light-emitting diodes waveguide-integrated with superconducting single-photon detectors. *Appl. Phys. Lett.*, 111:141101, 2017.
- [41] G. Davies, E.C. Lightowers, and Z.E. Ciechanowaska. The 1018 meV (π of π) vibronic band in silicon. *J. Phys. C*, 20:191, 1987.
- [42] M. Aboy, I. Santos, P. Lopez, L.A. Marques, and L. Pelaz. W and x photoluminescence centers in crystalline Si: chasing candidates at atomic level through multiscale simulations. *J. Electronic Materials*, 47:5045, 2018.
- [43] J.L. Hennessy and D.A. Patterson. *Computer Architecture*. Elsevier. Appendix F.
- [44] J.M. Shainline, J. Chiles, S.M. Buckley, A.N. McCaughan, R.P. Mirin, and S.W. Nam. Superconducting optoelectronic neurons V: Networks and scaling. *arXiv*, page 1805.01942, 2018.
- [45] A. Kumar, Z. Wan, W.W. Wilcke, and S.S. Iyer. Toward human-scale brain computing using 3D wafer scale integration. *ACM Journal on Emerging Technologies in Computing Systems*, 13:45, 2017.
- [46] M.M. Shulaer, G. Hills, R.S. Park, R.T. Howe, K. Saraswat, H.-S. Philip Wong, and S. Mitra. Three-dimensional integration of nanotechnologies for computing and data storage on a single chip. *Nature*, 547:74, 2017.
- [47] S. Salahuddin, K. Ni, and S. Datta. The era of hyper-scaling in electronics. *Nature Electronics*, 1:442, 2018.
- [48] F. Marsili, V.B. Verma, J.A. Stern, S. Harrington, A.E. Lita, T. Gerrits, I. Vayshnker, B. Baek, M.D. Shaw, R.P. Mirin, and S.W. Nam. Detecting single infrared photons with 93% system efficiency. *Nat. Photon.*, 7:210, 2013.