Considerations for neuromorphic supercomputing in semiconducting and superconducting optoelectronic hardware

Bryce A. Primavera and Jeffrey M. Shainline

April 2021

Abstract

Any large-scale neuromorphic system striving for complexity at the level of the human brain and beyond will need to be co-optimized for communication and computation. Such reasoning leads to the proposal for optoelectronic neuromorphic platforms that leverage the complementary properties of optics and electronics. Starting from the conjecture that future large-scale neuromorphic systems will utilize integrated photonics and fiber optics for communication in conjunction with analog electronics for computation, we consider two possible paths towards achieving this vision. The first is a semiconductor platform based on analog CMOS circuits and waveguide-integrated photodiodes. The second is a superconducting approach that utilizes Josephson junctions and waveguide-integrated superconducting single-photon detectors. We discuss available devices, assess scaling potential, and provide a list of key metrics and demonstrations for each platform. Ultimately, both platforms hold potential, but their development will diverge in important respects. Semiconductor systems benefit from a robust fabrication ecosystem and can build on extensive progress made in purely electronic neuromorphic computing, but will require III-V light-source integration with electronics at an unprecedented scale, further advances in ultra-low capacitance photodiodes, and success from emerging memory technologies. In contrast, superconducting systems place near theoretically minimum burdens on light-sources (a tremendous boon to one of the most speculative aspects of either platform) and provide new opportunities for integrated, high-endurance synaptic memory. However, superconducting optoelectronic systems will also contend with interfacing low-voltage electronic circuits to semiconductor light-sources, the serial biasing of superconducting devices on an unprecedented scale, a less mature fabrication ecosystem, and cryogenic infrastructure.

1 Introduction

The foundations of cognition remain a great frontier of science, with potentially enormous ramifications for technology and society. A platform capable of neural function with the scale and complexity of the brain or even beyond could be a powerful tool in deciphering this enigma. Achieving such large-scale systems has proven to be non-trivial with established CMOS hardware [1]. A significant challenge will be to enable efficient communication with low-latency amongst billions or trillions of neurons. Optics appears well-matched to the task, as the lack of resistive, capacitive, and inductive parasitics makes optical links more amenable to high fan-out than electrical interconnects [2]. This eliminates the need for time-multiplexing, allowing direct connections between neurons that will not suffer from the traffic induced delays that limit neuromorphic systems using digital communication infrastructures [3], and physically enables larger, faster, and more interconnected networks. However, while the lack of interaction between photons is beneficial for reducing parasitics during communication, it is a detriment to computation. Electronic circuits are thus better suited to implement complex, nonlinear neuronal functions. It is reasonable to anticipate performance gains from optoelectronic neural systems leveraging optics for communication and electronics for computation, provided the hardware can be realized.

Our proposal to fabricate a direct, physical connection between every pair of connected neurons is known as the fully dedicated axon approach to communication [4]. We further specify that all synapses, dendrites, and neurons will utilize fully dedicated electronic circuits as well, so that each element of hardware has a one-to-one correspondence with its information-processing role in the neural system. This stands in contrast to several proposed optoelectronic systems that intend to perform fan-in optically using a shared synapse [5, 6]. This fully dedicated approach is advantageous if one aspires to create a diverse array of synaptic and dendritic behaviors at each neuron, as observed in biological neural systems [7, 8]. For instance, a different time constant or plasticity mechanism could be implemented at every synapse on a single neuron. Perhaps more importantly, fully dedicated components eliminate the auxiliary hardware required to perform multiplexing operations. Further, performing synaptic weighting in the electronic domain allows for binary optical communication, which minimizes the amount of optical energy per spike and reduces noise incurred by communication. The scope of this paper is therefore limited to networks meeting these three conditions:
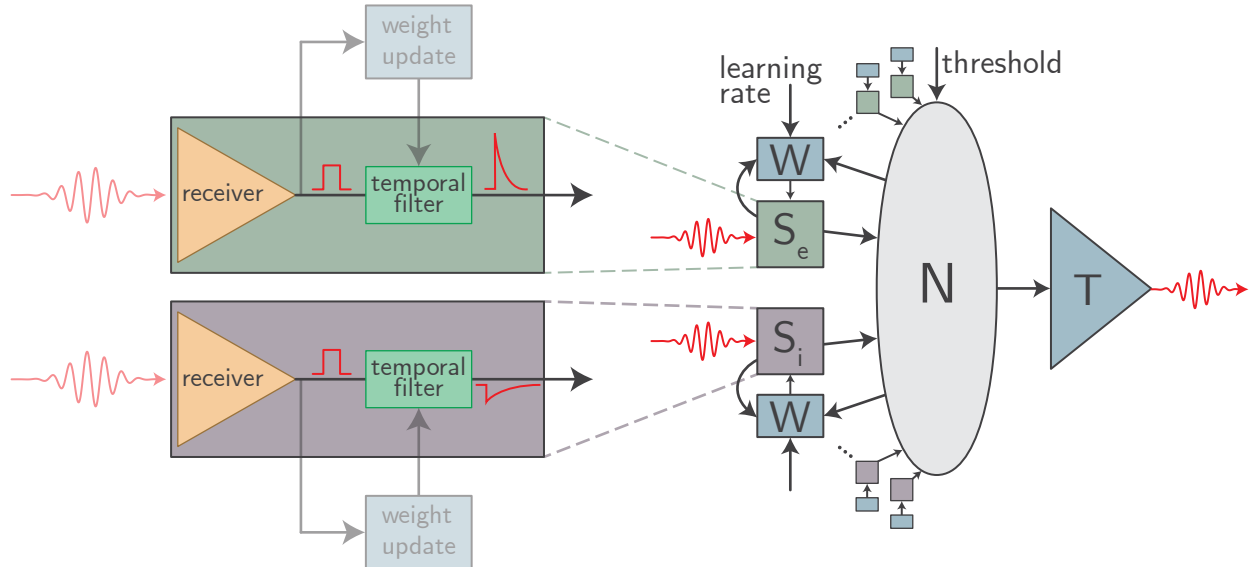
Figure 1: An abstract schematic of the class of optoelectronic neurons meeting our three criteria. Each synapse (S_e and S_i for expiatory and inhibitory synapses respectively) is implemented with a physical circuit block containing a detector and a temporal filter. The detector produces an all-or-nothing electrical pulse upon receipt of an optical spike which is then processed by the filter. The parameters of the filter (time constant, weight, etc.) can be set individually for each synapse. A local weight update circuit (W) implements plasticity mechanisms at each synapse. Synaptic outputs are integrated in the soma (N) which drives an optical transmitter to downstream connections upon reaching threshold.

1. Direct, optical connections are utilized for communication between neurons (fully dedicated axons).

2. Optical communication is binary. The amplitude of the optical signal carries no information.

3. All synaptic, dendritic, and somatic computations are performed by fully dedicated electronic circuits.

With these general axioms established, a picture of the hardware under consideration begins to emerge. There is a single optical transmitter at each neuron. This light emitter produces a short pulse of light each time the neuron spikes. The optical pulse is coupled into a waveguide, and optical power is tapped from the waveguide for each downstream synapse. Each synapse contains a photodetector which registers an all-or-nothing synapse event. From there, all synaptic weighting, spike-train filtering, dendritic processing, signal summation, neuronal thresholding, and plasticity mechanisms are implemented in the electronic domain with tailored integrated circuits. A schematic of this general framework is shown in figure 1.

There are potentially multiple ways to physically implement this model. The remainder of this paper will discuss two possible implementations —a superconducting platform and a room-temperature all-semiconductor system. The superconducting platform, known as SOENs (Superconducting OptoElectronic Networks) is discussed in prior work [2,9–11]. In short, optical links are formed from semi-

conductor light sources and superconducting nanowire single photon detectors (SNSPDs). Computation is performed with analog Josephson junction (JJ) circuits and memory is implemented with persistent current in superconducting loops. The semiconductor implementation is imagined as an exact analogue of the SOENs platform, except without the benefits (or limitations) of cryogenic elements. Traditional photodiodes enable optical communication, analog CMOS circuits provide computation, and emerging memory devices provide synaptic memory.

This paper seeks to analyze the suitability of both platforms for implementing large-scale optoelectronic neuromorphic networks. Despite limiting our discussion only to architectures meeting our three axioms, there remains a vast space of design choices, making it difficult to draw hard-and-fast conclusions. Nevertheless, interesting guidelines can be obtained by analyzing limits of technologies most likely to be used in each platform. Important benchmarks for device performance are also identified, which may be of use in monitoring the development of this field.

2 Communication

2.1 Optical Receivers

We begin analysis of optical interconnects with receivers. There are two ways the receiver influences the power budget of an optical link: (1) The receiver (and the electrical

components it must drive) sets the minimum optical signal that must be produced by the light source, and (2) the receiver may require electrical power of its own to run. It is found that the energy per spike may be quite similar in both platforms once cooling is accounted for in the superconducting case. However, the optical power required by light sources —a speculative element at this scale of integration —is reduced by a factor of 1000 in the superconducting case, at least when compared to the semiconductor receivers of comparable total efficiency, which omit transimpedance amplifiers [12].

2.1.1 SOENs Receivers

As stated previously, the SOENs platform utilizes SNSPDs to detect optical signals as faint as a single photon. Physically, an SNSPD is a superconducting nanowire biased with a current source ($I_{\mathrm{spd}} \approx 10\,\mu\mathrm{A}$). The simple structure makes fabrication and waveguide integration straightforward [13–20]. Photons travelling through a waveguide evanescently couple to a nanowire on the surface of the waveguide. A single photon has enough energy to drive the nanowire from the superconducting phase to a resistive state. In SOENs receivers, this momentarily redirects the bias current along an alternate conduction pathway that activates a JJ circuit to register the synapse event and conduct further synaptic processing (Fig. 2a).

While an SNSPD dissipates zero static power, there is dynamic electrical power consumption associated with detection events. The nanowire has an inductance, L_{spd}, that stores energy from the current bias. During a detection event, this energy is dissipated in the resistor r_{spd}. The electrical energy necessary to detect each photon is then $\frac{1}{2}L_{\mathrm{spd}}I_{\mathrm{spd}}^2$. L_{spd} can be as low as 100 nH, resulting in an electrical energy consumption (E_{spd}) of around 5 aJ/spike.

Since an SNSPD is capable of detecting single photons, it will operate near the quantum limit of optical communication [21]. We assume that the detection of a single photon will be treated as the registering of a spiking event. The probability of a light source producing a spike with a certain number of photons within a fixed time window is given by a Poisson distribution. We will also conservatively assume a detection efficiency η_D of 70% (higher detection efficiency is certainly possible [22, 23]). The probability of measuring zero photons during a spiking event is then given by:

$$P(0) = \sum_{k=0}^{\infty} \frac{N_{ph}^k e^{-N_{ph}}}{k!} (1 - \eta_D)^k = e^{-N_{ph}\eta_D}, \quad (1)$$

where N_{ph} is the average number of photons per spiking event. Neural systems are known for remarkable robustness to and even utilization of noise [24, 25]. Detecting only 99% of spikes may be tolerable. From equation 1, this would correspond to roughly 7 photons (0.9 aJ for $\lambda = 1.5$ um) needed to reach the receiver. The total number of photons produced by the source will need to be higher to account for energy losses in the link. The total optical energy per spike, E_{opt}, will be:

$$E_{\mathrm{opt}} = \frac{N_{\mathrm{ph}}h\nu}{\eta}. \quad (2)$$

$h\nu$ is the energy of a single photon and η is the total energy efficiency of the optical link. η includes all optical losses and the inefficiency of the transmitter. This efficiency factor will be highly dependent on the specifics of the platform, but for now we will leave it as a free variable. The total power consumed by the optical link is the sum of E_{opt} and E_{spd}. Accepting a 1% error rate, these two contributions to the total energy will be roughly equal when $\eta = 20\%$. Such a high efficiency is likely near the limits of physical possibility. For more realistic values of η, E_{opt} will dominate.

Importantly, superconducting electronics come with a cooling overhead (Sec. 5). We conservatively assume that every watt of power produced by superconducting electronics will require about 1 kW of refrigeration power to remove excess heat. System-level effective optical energy per spike for superconducting links will be no less than 1 fJ.

Fabrication of waveguide-integrated SNSPDs has become commonplace in recent years [13–20]. SNSPD materials include NbN, NbTiN, WSi, and MoSi. Superconducting films (3 nm - 10 nm) can be sputtered at room temperature atop many substrates and patterned into wires from 50 nm to 5 µm wide using conventional lithography and etching. Multiple planes of SNSPDs have also been demonstrated [26] —a promising development for future large-scale neuromorphic systems (Sec. 5). Waveguide-integrated NbN SNSPDs can reach photon count rates exceeding 1 GHz [27, 28]. However, slower detectors, such as MoSi and WSi SNSPDs with 20 MHz count rates, have demonstrated the best yields to date (99.7% [29]). Previous statements that SOENs were limited to 20 MHz were motivated by these pragmatic concerns about the current state of fabrication [2].
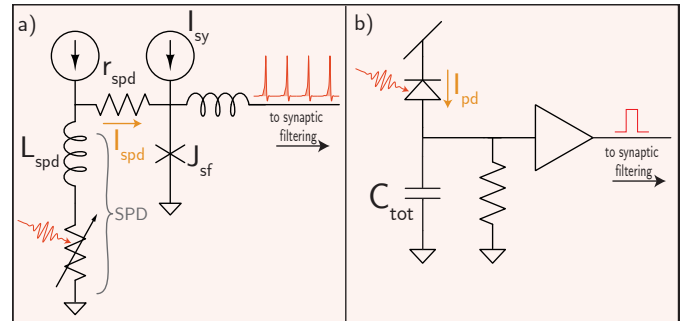

Figure 2: Receivers for the (a) superconducting and (b) semiconducting platforms. Note that synaptic weighting for the semiconductor case is included in the filtering circuitry, shown in figure 4b.

2.1.2 Semiconductor Receivers

While semiconductor receivers dominate long-distance optical communication, intra-chip optical receivers deviate significantly from their long-distance counterparts, as traditional transimpedance amplifiers likely consume too much electrical power, despite impressive optical sensitivities. This has led to the proposal of "receiverless" designs that omit amplifiers altogether [12]. Receiverless communication uses a photodetector to directly drive the input of CMOS gates. Photons produce electron-hole pairs in the photodetector, which in turn charge the CMOS gate capcitance up to the switching voltage. A circuit diagram of the scheme is shown in figure 2b in which a photodiode directly drives a CMOS digital buffer. A resistor is also placed in parallel to allow the receiver to reset. In principle the resistor is unnecessary if an optical reset is used as described in [30]. The resistor would increase the minimum optical power necessary to register a spike and limit the bandwidth of the receiver.

With optical link efficiency η, the necessary optical energy required to drive the receiver [12] to a voltage V is:

$$E_{\text{opt}} = \frac{C_{\text{tot}} V}{\eta \mathcal{R}} \qquad (3)$$

$\mathcal{R}$ is the responsivity of the detector, typically of order $1\,\text{A/W}$. C_{tot} includes the photodiode capacitance, the CMOS gate capacitance, and any wiring capacitance. It is reasonable to consider values for C_{tot} at the femto-farad level. For $1.5\,\mu\text{m}$ photons and a required voltage swing of $0.8\,\text{V}$, $E_{\text{opt}} \approx 0.7\,\text{fJ}$ (5000 photons) for unit efficiency. This is very similar to the superconducting case, once cooling is considered. This means that if two optical communications links were identical in all measures (source efficiency, optical losses, etc.) except one was cooled to $4\,\text{K}$ with SNSPDs and the other operated at room-temperature with photodiodes, then communicating a spike would cost nearly the same energy at the system level in each link. The power required for cryogenic cooling pays for itself with reduced light levels in the optical link. For capacitances below $1\,\text{fF}$ (a difficult task), semiconductor receivers could potentially consume even less energy than their superconducting counterparts. Waveguide-integrated femto-farad photodiodes have been demonstrated in both SiGe and Ge [31]. Polysilicon photodiodes are also attractive for increased manufacturability [32]. Most photodiodes have far better speed than required for neuromorphic applications, reaching up to $45\,\text{GHz}$ [31].

Just as with SNSPDs, semiconductor receivers will also require electrical power, even if it is minimized by the receiverless approach. In this case, there will be static power dissipation through the leakage current of the photodiode. Assuming a $1\,\text{V}$ bias, a leakage current on the order of $1\,\text{nA}$ [33], and an optical link efficiency of 1%, this static dissipation would dominate power consumption for average spiking rates below $100\,\text{MHz}$. While very fast neu-romorphic systems are certainly of interest, power limitations may make it difficult for such fast networks to reach the scale of the brain (Sec. 5). The development of low capacitance, zero-bias photodiodes [34] would be a major advantage towards making efficient, low frequency networks. Static power consumption is also a major question for many avalanche photodiode (APD) receivers. Avalanche gain could provide a significant (at least one order of magnitude) reduction in the necessary optical power per spike [12]. While often associated with higher bias voltages, germanium waveguide-integrated avalanche detectors have been demonstrated to provide $10\,\text{dB}$ of gain even at $1.5\,\text{V}$ bias [35]. However, dark current is still typically in the microamp range for such detectors [35, 36]. An intriguing possibility is to reduce static power consumption through cooling, as the dark current could potentially be reduced by orders of magnitude [37]. However, in that case one forfeits a major advantage of the semiconductor approach.

While the receiverless scheme is promising for achieving low energies per spike, it places significant burden on the transmitter side of the link. Neuromorphic applications magnify this burden, as neurons are expected to drive thousands of downstream connections in parallel. Additionally, the receiver capacitance must be charged quickly to maintain high spiking frequencies. The result is that relatively large optical power is required from transmitters. The best case ($\eta = 1$) scenario is shown in figure 3. Semiconductor receivers can be expected to require around one thousand times the optical power of superconducting receivers and the highest spiking frequency of a neuron could very well be limited by the power output of the light source. The ramifications of this result on prospective light sources are discussed in the next section.
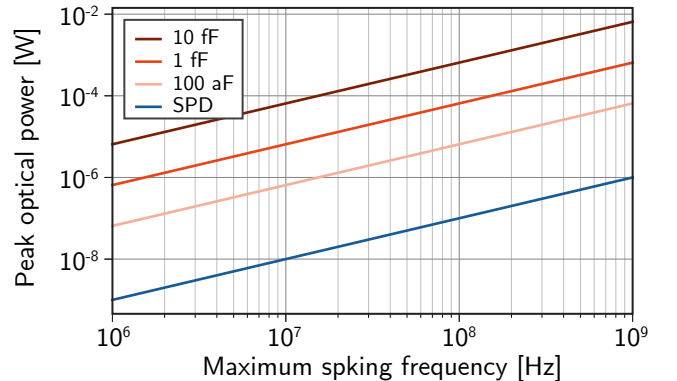

Figure 3: The required optical power in the receiverless scheme to drive 1000 downstream synapses within one inter-spike interval for a given spiking frequency.

2.2 Optical Transmitters

The transmitter is expected to dominate the power budget of optical links for both platforms. Room-temperature, CMOS-integrated light sources have been a holy-grail for decades, but materials integration issues have kept this prized objective out of reach. For superconducting systems, SNSPDs drastically lower the power requirements of light sources, while cryogenic temperatures improve light-source efficiency. Light sources are likely significantly simpler in the superconducting case. However, interfacing low-voltage superconducting electronics with semiconductor light sources presents an obstacle that is entirely absent from the all-semiconductor platform [38].

2.2.1 Integrated Light Sources

Optical coherence is not a requirement for the envisioned system. NanoLEDs are thus an attractive option due to their ease of fabrication, lack of threshold current, and improving efficiency with shrinking scale [39]. However, nanoLEDs struggle to produce optical power significantly greater than $1\,\mu$W [39]. While semiconductor systems targeting spiking frequencies in excess of $1\,$MHz may be forced to turn to lasing, nanoLEDs should be more than sufficient for superconducting platforms. Either way, integrating millions of light sources on a 300 mm wafer remains aspirational. Silicon has excellent electronic properties, but the indirect bandgap drastically reduces light emission. Off-chip light sources are a possibility, but are likely untenable for massive systems, as their high static power consumption is incommensurate with the sparsity of neural activity. Integrated light sources would be a tremendous boon, if not a requirement for the success of large-scale optoelectronic neuromorphic computing. There are two courses of action: (1) force silicon to emit light through either material and/or environmental modifications or (2) integrate direct bandgap materials on silicon.

Many strategies towards silicon light sources have been pursued [40, 41] including quantum confinement in Si-based superlattices [42] and nanocrystals [43], emission from embedded erbium [44, 45], point-defect emitters [46–49], extended defects [50], strain dislocations [51], engineering of the local density of optical states [52], and hexagonal SiGe grown on GaAs [53]. Total efficiency from 0.1% [51] to 1% [52] has been demonstrated at room temperature, but not at powers and areas suitable for the semiconductor receivers introduced in the previous section.

Abandoning silicon as an active optical element, many researchers turned towards integrating III-V materials on silicon substrates. Pending a watershed moment in silicon sources, III-V integration will be required for the semiconductor platform (although not necessarily in the superconductor case, where low-temperature changes the physical context). Epitaxial growth would be the optimal solution [54], but defects due to lattice mismatch have so far prevented this method from large-scale adoption. III-V quantum dots are more robust to such defects and have demonstrated high optical powers with small footprints [54–56], albeit typically grown on offcut Si substrates that are not CMOS compatible or with thick buffer layers that make optoelectronic contact difficult. More work is required to realize scalable, cost-effective integration of III-V quantum dot light sources with CMOS electronics, passive photonic waveguides, and efficient photodetectors. Without epitaxial growth, the semiconductor platform would require more expensive and less scalable bonding technologies. A variety of schemes have been proposed [54, 57], including die-level bonding [58, 59], wafer-level bonding [60–62], transfer printing [63–65], and selective-area epitaxy [66], but these approaches still appear cumbersome when seeking the scale of integration considered here.

The situation is significantly more favorable for cryogenic systems. Low temperature often reduces non-radiative recombination [67–69], improving efficiency for both silicon and III-V light sources alike. This benefit is further compounded by the low optical power requirements of SNSPDs. When integrating III-V light sources with CMOS, the light sources must be integrated on top of the electronics after the high-temperature dopant activation steps have been performed. Superconductor electronics have no such high-temperature processing steps, so the light sources can be produced on a Si wafer before the electronics are realized. Problems related to offcut Si wafers and thick buffer layers are eliminated. Additionally, silicon light sources, with their superior potential for integration, demand exploration with the superconducting platform. Several silicon point defects typically quenched at room-temperature emerge as narrow-linewidth candidates for light sources in the telecommunications band [70–74]. While single-photon emission [75–77] is not the objective in the present context, the narrow linewidth is also attractive for further efficiency gains via the Purcell Effect [78]. LEDs have already been demonstrated with the W-center defect [72, 79], albeit with poor (10^{-6}) efficiencies, limited by electrical injection efficiency rather than emitter lifetime. Photoluminescence studies are promising for orders of magnitude improvement [80], but more work is required to improve emission efficiency in an integrated-circuit context. If cryogenic silicon light sources become viable, the superconducting platform might hold a major scalability advantage over the semiconducting analogue.

2.2.2 Driving Circuitry

Both platforms require neurons to drive semiconductor light sources. The transmitter circuitry is thereby required to produce voltages on the scale of the bandgap of the optical source ($\approx 1\,$V). CMOS circuitry, itself a semiconducting technology, naturally operates on this voltage rendering the driving circuitry essentially a non-issue. Standard MOSFET LED or modulator driving circuits [81, 82] can be straightforwardly adapted for neuromorphic ap-

plications. Superconductors, however, operate in an entirely different regime, with signals usually on the order of the superconducting energy gap ($\approx 1\,\text{mV}$) The optimal method for interfacing superconducting electronics with semiconductor devices is still an area of active research. Recent progress has been made with devices utilizing the massive change in impedance during a phase transition between superconducting and resistive states. In Ref. [38], a resistive element is heated using $50\,\text{mV}$ pulses to thermally trigger a transition in a superconducting meander. The meander transitioned to a state with resistance in excess of $10\,\text{M}\Omega$ and was used to drive a cryogenic silicon light source waveguide-coupled to an SNSPD. While these results are promising, the light source was only pulsed at only $10\,\text{kHz}$ (due to poor source efficiency) and was fabricated on a separate chip. More work is needed to improve the speed, efficiency, to monolithically integrate driving circuitry with LEDs, and to demonstrate pulsed optical links.

3 Electronic Neuronal Computation

Electronic circuitry capable of performing neuronal dynamical operations will also be necessary. Biological neurons are increasingly recognized as sophisticated computational units [83–86]. Emulating such complicated behavior has been the subject of extensive investigation in both semiconducting [87–92] and superconducting platforms [10, 93, 94]. We do not attempt a comprehensive review of circuitry, but rather draw attention to issues specific to optoelectronic networks in both cases.

3.1 Semiconductor Electronics

The maturity of CMOS processing has allowed great strides in neuromorphic computing. While optical communication would likely also be advantageous in digital approaches, we focus on analog CMOS neurons for their perceived efficiency advantages [95, 96]. At a basic level, a neuron must perform three mathematical functions: summation of synaptic inputs, temporal filtering, and threshold detection leading to axon potential generation. Summation can be achieved by exploiting Kirchoff's current law. Filtering can be implemented with elementary resistor-capacitor circuits. Thresholding is a natural function of transistors. Building upon this basic mapping, analog neurons have demonstrated a litany of biologically-inspired models [88, 97].

It was found in the previous section that optical communication requires a minimum of about $1\,\text{fJ}$ of energy to deliver a spiking signal to each synapse. For realistic optical link efficiencies, this value will be at least an order of magnitude larger. Synaptic processing circuits would therefore ideally operate with an energy budget of $10\,\text{fJ}$ - $100\,\text{fJ}$ to process a single spike. Somatic computation could comfortably consume power larger than that of synaptic processing by a factor of the average fan-out (perhaps 1000). Many low-energy neuromorphic demonstrations are promising for reaching these targets. By reducing the membrane capacitance and supply voltage, a neuron capable of $25\,\text{kHz}$ spike rates was demonstrated to consume only $4\,\text{fJ/spike}$ [98]. Many other analog neurons, with energies ranging from fJ to pJ per spike, fall comfortably below the power consumption of optical communication [99]. However, it remains to be seen if more complicated neurons and synapses, implementing a critical subset of behavior necessary for cognition, will be able to maintain such low power operation. In terms of speed, CMOS neurons have demonstrated spike rates in excess of $100\,\text{MHz}$ [100]. Optical communication should face few issues achieving such speeds, *if* sufficiently bright light sources can be efficiently integrated with CMOS circuits.

One challenge for the CMOS approach has been to design compact circuits with long time constants. Long time constants are important for systems targeting biological time scales (upwards of $500\,\text{ms}$) [99] or power law distributions of timescales to implement critical behavior [101]. Subthreshold transistor circuits operating with currents in the femto-picoamp range minimize the size of capacitor needed to implement a specific time constant [88]. The area constraints of this scheme are discussed in Appendix A and compared to the superconducting approach.

For a concrete example, a circuit diagram for a memristor implementation of the popular differential-pair integrator (DPI) synapse is shown in Fig. 4b [102]. The DPI produces a decaying exponential post synaptic signal in response to an input voltage pulse—potentially from an optical receiver. This leaky integrator behavior is characterized by a time constant set by the value of the filtering capacitance and the rate of leakage off the capacitor [103]. The time constant could potentially be programmed using memristors—an advantage over superconducting circuits that have been proposed to date.

3.2 Superconducting Electronics

Superconducting neurons have been studied nearly as long as CMOS implementations, with a mapping between neuronal functions and superconducting electronics identified in the early 1990s [104, 105]. In this case, Faraday's Law, governing the addition of magnetic flux through mutual inductors to superconducting loops provides the necessary synaptic summation function. Filtering is achieved through resistor-inductor blocks (or RC circuits in some cases [93]). Josephson junctions (JJs) provide the requisite nonlinear thresholding element.

Like their CMOS counterparts, many superconducting circuits have now been designed to implement sophisticated neuronal dynamics. Superconducting neuromorphic circuits have been designed to implement a variety of bio-inspired neuron models [93, 94, 106], sophisticated den-

dritic processing [10], and have performed image classification in simulation [107]. The natural spiking behavior of JJs may even require a lower device count than analogous CMOS circuits for various leaky-integrate-and-fire models [93]. In short, it does not appear that superconducting circuits are any less capable of complex neuronal computation than CMOS, although experimental demonstrations lag far behind.

Superconducting electronics has long been pursued for gains in energy efficiency and speed. Indeed, superconducting elements dissipate zero static power and spike energies are frequently reported in the sub-femtojoule range, including refrigeration. Optical communication is likely to dominate power consumption for superconducting optoelectronic systems (Appendix B). In terms of speed, fully electronic superconducting neurons may be capable of spike rates up to 100 GHz [106, 107]. However, this is orders of magnitude faster than any SNSPD can respond. This marks a notable difference between the superconducting and semiconducting architectures. While optical communication could be integrated with CMOS neurons with no degradation in speed, optoelectronic superconducting systems will likely be significantly slower than their fully electronic counterparts. This may be the cost of highly connected systems. That said, the extraordinary switching speed of JJs is still leveraged in optoelectronic networks to perform analog computations within synapses, dendrites, and neurons.

The ability of superconducting electronics to go slow might be just as compelling as their ability to go fast. While it can be challenging to implement long, biologically realistic time constants in CMOS neurons, superconducting loops can create time constants orders of magnitude higher than biology by adjusting the L/R ratio in synaptic and neuronal loops (Fig. 2a). See Appendix A for more details. The ability to generate dynamics across many orders of magnitude in time also dovetails nicely with suggestions that critical behavior is important for cognition [108].

Fan-in has traditionally been considered a liability of superconducting electronics. If this were the case, it would clearly be an impediment to mature superconducting neuromorphic systems. For superconducting neurons designed to use single fluxons as synaptic signals, fan-in has recently been analyzed [109], and it has been found that if a single synapse must be able to drive a neuron above threshold, fan-in may be limited to around 100. However, it is often not necessary for each synapse to be able to trigger a neuronal spike event. It has been analyzed elsewhere that if analog signals containing multitudes of fluxons are communicated from synapses to the neuron cell body, fan-in can likely scale to biological levels through the use of mutual inductors [2]. Using more fluxons comes with larger power consumption, but for optoelectronic systems, light production will likely still dominate.

While most diagrams of superconducting circuits (including those here) show many separate biases delivering current to various elements, the ability to construct circuits that can be biased in series will be critical to the scalability of this hardware. A separate bias for every synapse would be untenable in large-scale systems [110]. This mimics the evolution that occurred in superconducting digital electronics, in which the field has turned away from parallel biasing schemes and embraced serially biased platforms [110] and current recycling schemes [111]. SOENs are potenially amenable to serial biasing, but this important point demands further analysis.

A superconducting synaptic filtering circuit is shown in figure 4a. Synaptic weighting is implemented in the receiver circuit (Fig. 2a), so this circuit block is only responsible for converting a train of fluxons into a decaying exponential post-synaptic potential reminiscent of biological and CMOS synapses. A resistor, r_{si} converts a superconducting persistent current loop into a leaky-integrator very similar to the DPI synapse. The time constant is set by L_{si}/r_{si}, and the synaptic current can be added to a neuronal circuit through mutual inductors. Unlike the DPI synapse, this circuit does not have a programmable time constant, but does hold the potential to implement a wide range of different time constants by fabricating different values of L_{si} and r_{si}.
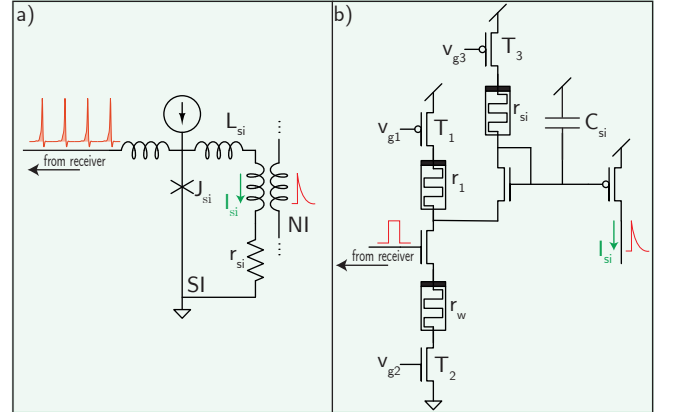

Figure 4: Synaptic filtering circuits for the superconductor (a) and semiconductor (b) cases. Note that weighting in the superconducting case was shown in figure 2. The memristor-integrated DPI circuit pictured here is introduced in [102].

4   Synaptic Memory

It has been apparent to the neuromorphic community for some time that large-scale neural systems will require innovative approaches to synaptic memory. A memory technology intimately integrated with processing units is central to the vision of neuromorphic computing. Important metrics include weight precision, volatility, area, write energy, write speed, and endurance (the effective number of cycles in a device's lifetime). We attempt to provide desired benchmarks for a few of these metrics in the specific

case of optoelectronic networks. For this section, we assume a speedup of about 10^4 over biology, for an average spike rate of 10 kHz and a maximum of 10 MHz. This is commensurate with both the maximum count rates of high-yield SNSPDs and some of the fastest CMOS electronic neuromorphic systems built to-date.

4.1 Memory Benchmarks

4.1.1 Endurance

Large-scale neural systems require significant investments in money and time. Operational lifetimes on the scale of decades (10^9 seconds), if not longer, are therefore essential. Such systems will be expected to learn continually during that lifespan, placing significant requirements on the durability of memory technologies. The number of times a synapse is updated in its lifetime is a function of neuron spiking frequency (f) and the number of synapses that are typically updated after each post-synaptic spike. Neuroscientific evidence has been presented that the number of active presynaptic inputs required to trigger a post-synaptic spike goes as $\sqrt{N}$, where N is the fan-in of the neuron—exceeding 1,000 for brain-like systems [112, 113]. We assume all synapses that contributed to the spiking of the post-synaptic neuron are updated with each spike. We then estimate the number of weight updates ($N_{updates}$) in the synapses's lifetime (L) will be:

$$N_{updates} = \frac{Lf}{\sqrt{N}} \tag{4}$$

For a decades-long lifetime, and a mean spiking frequency of 10 kHz, the total number of weight updates will be 10^{11}. This is a challenging demand for many emerging non-volatile memory technologies.

4.1.2 Update Energy

One would like the power dedicated to weight updates not to exceed the power used for optical communication. Once again invoking the assumption that $\sqrt{N}$ synapses are updated with each postsynaptic spike, we arrive at the following relation between the energy to produce a single spike (E_{opt}) and that to update a single weight (E_{update}):

$$E_{\text{update}} < \sqrt{N} E_{\text{opt}} \tag{5}$$

Using the analysis in Sec. 2, 1 fJ of energy needs to be delivered to the receiver in either platform. Assuming a transmitter efficiency of 1%, this would mean E_{opt} is 100 fJ. Therefore, for a fan-in of 1,000 synapses, E_{update} would ideally be no more than about 3 pJ. This value also includes any energy spent in peripheral circuitry to program the synapse. This efficiency appears to have already been met by several emerging memory technologies [114, 115].

4.1.3 Update Speed

An ideal system would be capable of implementing synaptic updates within the minimum inter-spike interval. While semiconductor optoelectronic systems could potentially produce spike rates in excess of 10 GHz (assuming sufficiently bright, integrated light sources can be achieved), synapses might need to be taken offline during WRITE operations, as it is unlikely that sophisticated plasticity mechanisms can be implemented in under 100 ps. Lower maximum frequencies would allow plasticity to be implemented without ever neglecting a spiking event. For our 10 MHz target, we desire memory updates in under 100 ns. Slower updates may not be completely intolerable, if network dynamics are robust to missed spikes during synaptic updates or to synaptic weights that are in the process of being altered.

4.1.4 Weight Precision

The necessary weight precision will be determined by the specifics of a chosen learning model and the desired application. Weight precision has been the subject of much discussion. It has been suggested that 4-bit precision is sufficient for state-of-the-art mixed signal neuromorphic systems [116]. Deep learning systems have also demonstrated success with 8-bit precision—a significant reduction from 32-bit floating point numbers [117]. Hippocampal synapses in rats have been inferred to allow at least 26 different states (≈ 5 bit), which squares nicely with computer science findings [118]. It has also been argued that metaplasticity mechanisms are more important for lifelong learning than the bit-depth of the synapse [119, 120].

Target values for these key synaptic memory metrics are summarized in Table 4.1.4.

Metric	Goal
Endurance	$> 10^{11}$ updates
Update Energy	< 3 pJ
Update Speed	< 100 ns
Weight Precision	4-8 bits

Table 1: List of desired performance metrics for synaptic memory in a system with average fan-out of 1000, maximum spike rate of 10 MHz, average spike rate of 10 kHz, and spike energy of 100 fJ.

4.1.5 Programming Signals

One important criterion that eludes quantitative benchmarking is the complexity of programming circuitry for synaptic memory. Significant infrastructure for producing programming signals could limit scalability. For example, floating-gate synapses often require programming signals

at significantly higher voltages than are likely to be used in other parts of the network. For large-scale systems, memories with simple programming requirements will be at an advantage. Superconducting loop memory (Sec. 4.2.4) is intriguing from this standpoint, as the plasticity circuits operate with nearly identical signals and circuit blocks as those found in the rest of the network.

4.2 Proposed Technologies

4.2.1 Room-temperature Technologies

Many technologies have been proposed to implement synaptic weighting for room-temperature neuromorphic hardware, each with strengths and weaknesses [121]. The quest to find a suitable device for local synaptic memory dates back to the origins of the field, when Mead and colleagues investigated floating gate transistors [122]. Since then, floating gate synapses have been used to implement STDP [123], are attractive as a mature alternative to emerging devices, and have been proposed for use in large-scale systems [124]. However, there are concerns about high programming voltages, speed, and endurance that may limit floating-gate memories to situations with less-frequent updates. More recently, momentum has shifted to other technologies [114]. Memristive devices [125–127], commonly used in resistive random-access memory (RRAM) have emerged as a popular alternative, with recent demonstrations including monolithic integration with CMOS [128] and unsupervised pattern recognition with a simple network of synapses [129]. Questions remain about high variability (both cycle-to-cycle and device-to-device) [102], linearity, and endurance [114]. Phase Change Memory (PCM) is another option, with its own demonstration of STDP [130]. Thermal management and endurance have been raised as issues [114, 121]. Ferroelectric transistors present another alternative, as they have low variability, good potential for CMOS integration, and linearity [131]. Spin-torque memory, 2D materials, and organic electronics have also been proposed as solutions. Interested readers should consult one of the many review articles on this topic [121, 132, 133]. The field is burgeoning with new devices for synaptic memory, but to-date none has been dominant enough to monopolize research. To our knowledge, no technology has been able to simultaneously meet the targets in table 4.1.4, but progress in this area is encouraging.

4.2.2 Superconducting Technologies

Many of the aforementioned technologies may also apply to superconducting optoelectronic systems, but their cryogenic operation has been scarcely explored. Two other types of memory, only accessible at low temperatures, have received the most attention for superconducting systems: magnetic Josephson junctions (MJJs) and superconducting loop memories. An important distinction from room-temperature technologies is that for superconducting memory to be truly non-volatile, it must retain its state both in the absence of a power supply and upon warming to room-temperature.

4.2.3 Magnetic Josepson junctions

MJJs have been proposed as a (nearly) non-volatile memory technology for superconducting neuromorphic computing. A two-terminal device, the critical current of an MJJ can be programmed by changing the magnetic order of a ferromagnetic material placed in the tunneling barrier of a JJ [115]. MJJs are non-volatile with respect to electrical power, and there is optimism they can be made to retain their memory through a warm-up to room-temperature. Additionally, they provide remarkable performance with respect to the metrics given in Table 4.1.4. The energy per update is on the order of femtojoules (including cooling overhead), switching speeds are commensurate with firing rates exceeding $100\,\mathrm{GHz}$, and devices can be scaled to tens of nanometers. All of these metrics surpass the requirements for optoelectronic networks, and can be exploited in all-electronic superconducting networks as well [106]. More work is needed to analyze the scaling potential of MJJs with respect to yield. The magnetic fields used during programming can be produced with magnetic control lines, but spin-torque mechanisms may provide a more scalable solution. Finding an efficient, scalable solution to programming MJJs in large-scale systems thus remains an area of research that will be critical to their potential for adoption.

4.2.4 Loop Memory

Superconducting loop memories have been in use for decades by the superconducting electronics community [134, 135], but are not ideal for dense memory arrays commonly utilized as RAM in digital computing due to area concerns. In the case of optoelectronic spiking neural systems considered here, the objective is not to produce large RAM arrays, and size as well as addressing challenges do not emerge as significant impediments. Therefore, straightforward extensions of binary loop memories are the synaptic memory technology that appears most promising for the SOENs platform [2, 136]. In these memory cells, circulating current persists indefinitely in a loop of superconducting wire. The current in the loop can be controlled by adding/removing magnetic-flux quanta with standard JJ circuitry. This memory loop is then inductively coupled to a wire supplying a bias current to a Josephson junction at the synapse (J_{sf} in Fig. 2). When the synaptic SNSPD detects a photon, the biased junction will add an integer number of fluxons to another integrating superconductive loop (analogous to the membrane capacitance of a neuron). The number of fluxons added to the integration loop is a function of the bias supplied to the JJ, which is determined by the magnitude of current

circulating in the memory loop. The number of analog memory levels in the memory loop is determined by the inductance of the loop, which is easily set with the length of a wire. High-kinetic-inductance materials [137] enable memory storage loops with over a thousand levels (10 bits) to be fabricated in an area of $5\,\mu m \times 5\,\mu m$.

This approach has several advantages. The memory is nearly analog and updates are nearly linear. Memory is updated by the switching of a JJ, which involves only a change of the phase of the superconducting wave function. This phase can switch 10^{11} times in a second, so the endurance metric defined in the previous section is not an issue. This stands in contrast to room-temperature memories requiring material changes (filament formation, phase changes, etc.) which are often associated with degradation over time. Loop memory is also attractive from a fabrication perspective as it requires no additional materials or devices. The simplicity of the memory lends itself favorably to 3D integration, provided cross-talk from nearby loops can be mitigated. Plasticity circuits based on loop memories will also operate at the energy scale of single photons and flux quanta ($10^{-19}\,J$), which is commensurate with the rest of the circuitry in the network. This allows weight updates to be performed with the spikes the network produces in standard operation, reducing peripheral circuitry. There is no need to engineer differently shaped pulses for READ and WRITE operations, and the synapse does not need to be taken offline during programming. Simulations have demonstrated STDP learning with circuits containing only four additional Josephson junctions [2].

Two aspects of loop memory are concerning. First, loop memory is not strictly non-volatile. While circulating current can persist in a superconducting loop without any power supply, superconductivity must be maintained. If the temperature of the system is raised above the critical temperature of the superconducting material, the memory will be lost. Mechanisms for transferring weights stored in current loops to non-volatile solutions will need to be developed if the system's state is to be persevered upon reaching room-temperature (i.e. for maintenance or during a power interruption). The second weakness of loop memory is the size. The employed superconducting loops, as well as the transformers that couple them, will be large compared to all of the other solutions discussed. The consequences of these large-area components must be considered in the context of the entire system, which we discuss next.

5 System Level Considerations

Here we consider aspects concerning the integration of the components previously discussed and how systems may reach the scale of the brain. Basic graph theory metrics and the assumption of 300-mm fabrication processes allow us to assess area constraints and the benefits of 3D integra-

tion. It is found that at least 4 planes of photonic routing will be required in either platform to achieve brain-scale systems. Prospects for 3D integration of active elements are addressed. It also must be stressed that an optoelectronic system of the complexity of the human brain will be abjectly impossible on a single 300-mm wafer in either case. A possible vision for connecting many wafers is discussed. Finally, we analyze cooling and power concerns, finding that neither should preclude the development of brain-scale systems in either platform.

5.1 Considerations from graph theory

Neurons in brain regions active in cognition, such as the cerebral cortex and hippocampus, are characterized by a high degree of connectivity—often in excess of ten thousand connections per neuron [138,139]. These connections often extend across appreciable spatial distances. Creating and maintaining these connections comes with high metabolic and spatial costs. The severely constrained biological brain would not support such expenditures if they were not advantageous to cognition [140].

One reason why such high connectivity is necessary relates to efficient communication across the network. Rapid communication can only be achieved if the average path length across the network is small. In the language of graph theory, a network is a collection of nodes connected by edges. To calculate the shortest average path length across the network, one calculates the number of edges that must be traversed to travel from each node to every other node in the network. One takes the mean of this quantity over all pairs of nodes. The shortest average path length ($\bar{L}$) is a global metric that offers a glimpse at the efficiency with which information can be communicated across space.

Equation 6 provides the relationship between $\bar{L}$ and the number of edges connected to a node, or in our case, the number of synapses ($\bar{k}$) per neuron for a random network. In a random network, nearby and distant connections are equally probable. Specifically, the equation holds for Erdös-Rényi random graphs of networks with N_{tot} neurons [141]:

$$\bar{k} = \exp\left[\frac{\ln(N_{\text{tot}}) - \gamma}{\bar{L} - 1/2}\right], \tag{6}$$

where $\gamma \approx 0.5772$ is Euler's constant. For a network with 10^6 neurons, each neuron must make nearly 10,000 connections to support an average path length of two, and 200 synapses must be formed to support a path length of three. For a network with 10^8 neurons, more than 100,000 synapses are required for a path length of two, and more than 1,000 for a path length of three. The human hippocampus is a module with roughly 10^8 neurons, each with 10,000-50,000 nearly spatially random connections. The objective of achieving an average path length between two and three may be an important reason why the hippocampus prioritizes this exceptional degree of connectivity [139]. The cerebral cortex in the human brain contains

more than 10^{10} neurons, each with roughly 10,000 connections. This analysis indicates that a path length between two and three cannot be achieved across the entire cortex, and accordingly the cortex is constructed with a hierarchical, modular architecture [142, 143] with high connectivity and efficient communication within smaller modules, and more sparse connectivity between modules separated by larger distances [143–146].

While more sophisticated graph metrics can further elucidate the network concepts underlying cognition [147], the simple, global metric of average shortest path length can help inform scaling analysis of artificial cognitive hardware. We next consider the constraints $\bar{L}$ puts on the size of synaptic circuits.

5.2 Generic spatial constraints

Based on the significance of the interplay between the hippocampus and cerebral cortex in cognition [148], we assume hardware for artificial neural systems will make use of similar architectural principles. Here we assume optoelectronic circuits will be fabricated using the conventional sequential, planar processing techniques of the silicon microelectronics industry. We further specify to consideration of 300-mm wafers and seek a relationship between the network path length and the size of components on the wafer.

The area of a neuron occupied by its photonic waveguides can be approximated in a similar manner to the wires for electronic circuits [149]. This gives the following expression for the area of passive photonic circuitry:

$$A_p = \left(\frac{k w_{\mathrm{wg}}}{p_p} \right)^2 \tag{7}$$

p_p is the number of photonic waveguide planes, k is the degree of each neuron (assumed identical), and w_{wg} is the pitch of waveguides. The area of a neuron due to electronic circuits is given by

$$A_e = \frac{k w_{\mathrm{sy}}^2}{p_e} \tag{8}$$

w_{sy} is the width of a synapse and p_e is the number of planes of electronic circuits. Both w_{wg} and w_{sy} are subject to the area constraint of a 300 mm wafer. We use these relations to plot the number of planes (electronic and photonic) that will be required to maintain a path length of 2.5 across a network of a given size (Fig. 5). See appendix C for analysis of path length dependence. We note that more than ten million neurons (less than a mouse brain) appears out of reach for any platform.

5.3 Fabrication Processes

We assume 300-mm silicon wafer processing. Wafer-scale integration has already been demonstrated for electronic neuromorphic systems [150]. Still, even at this scale,
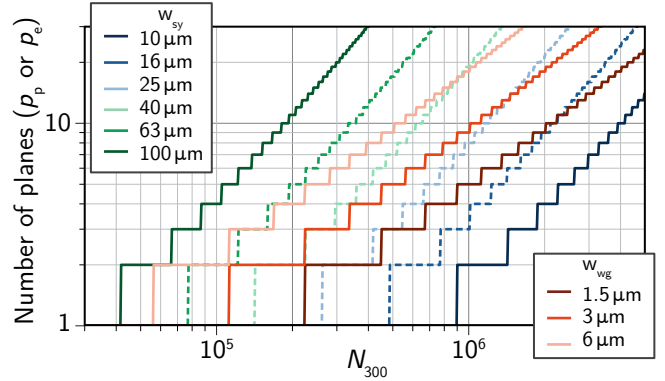

Figure 5: Number of planes of active synaptic circuits (p_e) and passive photonic waveguides (p_e) required to maintain a path length of 2.5 as a function of the number of neurons on a 300-mm wafer (N_{300}).

reaching 10^6 optoelectronic neurons per wafer is a tall order for either platform (Fig. 5). We choose this integration metric somewhat arbitrarily; 10^6 neurons per wafer corresponds to 10^4 wafers for a human cortex scale system. This is roughly the same order as the number of processing units in modern supercomputers. If this target is to be reached, 3D integration at some level will be necessary. From figure 5, it is clear that either platform will require a minimum of 4 photonic planes. Fortunately, photonic planes are quite amenable to 3D integration. Common waveguide materials include amorphous silicon (aSi), silicon nitride (SiN_x) and silicon oxynitride (SiO_xN_y). These dielectric materials can be deposited at low temperature, enabling several multi-planar demonstrations [151–154]. Additionally, low-temperature deposition makes such processes compatible with back-end CMOS fabrication. It should be noted that four photonic planes represents a best-case scenario, as wider waveguides have lower loss and only minimal reduction in average path length (Appendix C).

3D integration of active electronics is less straightforward, particularly for the semiconductor approach. 3D CMOS integration has been the subject of decades of research [155–162] and still faces uncertainty. Required high-temperature processing steps for dopant activation and contact anneals typically have a degrading effect on previous layers. Much of 3D integration of silicon microelectronics takes place at the die scale [161], which is incommensurate with the scale of systems under consideration. For the semiconductor path, the best course of action may be to abandon 3D active electronics altogether in favor of simply reducing the footprint (w_{sy}) of synapses. We see again from figure 5 that nearly 10^6 neurons can be integrated on a single plane if each synapse is on the order of $10\,\mu\mathrm{m} \times 10\,\mu\mathrm{m}$. This may be a challenging benchmark to reach with high-functionality synapses implementing complex plasticity and dynamics. Subthreshold circuits that have embraced larger CMOS nodes for decreased variability may need to adjust to more modern nodes, of which

there is some precedent [163]. Additionally, photodetectors will be on the micron scale and long time-constant capacitors can require significant area (Appendix A) [99]. Both of these elements would however be fabricated on separate planes from MOSFETs.

Superconducting platforms would likely take the opposite approach, embracing 3D integration in the face of necessarily large device areas. Superconducting electronics, including active JJs, are routinely deposited at low temperatures ($< 180 \, °C$). Integrated circuits with two stacked planes of JJs have been demonstrated by two research laboratories [164, 165], along with multiple of planes of SNSPDs [26]. This is particularly important, as superconducting systems will not be able to reach 10^6 neurons per wafer without 3D integration. A reasonable estimate for a superconducting synapse may be $30 \, \mu m$ on a side (Appendix B). Such a size would require eight electronic planes.

We note that even if $p_p = p_e = 1$, it is still possible to fabricate wafers with 10^6 neurons, provided $\bar{k} = 100$, giving $\bar{L} = 3.5$ (figures 9 and 10 in Appendix C). While this does not match the short path lengths of cognitive circuits in the brain, such a network is likely to have significant technological and scientific utility while offering an intermediate-term practical objective.

5.4 Constructing Multi-Wafer Systems

Given that neither system will scale to billions of neurons on a single wafer, many wafers ($\approx$10,000) will need to be connected together to support human-brain-scale computing. A vision for a multi-wafer system is discussed in reference [166] for the SOENs platform. Briefly, wafers are stacked to form highly inter-connected columns mimicking the modular structure of biological circuits [143–146, 167]. Columns are coupled to each other with inter-plane and free-space connections (both water and liquid helium are transparent at the wavelengths under consideration), but such connectivity is more sparse than that within a column. Optical fibers provide low-loss communication over long distances.

Achieving systems of this scale requires advances, particularly in wafer-scale circuit integration and system-level construction. A phenomenon akin to Moore's law, with ever-decreasing feature sizes enabling ever-higher integration density is unlikely to carry this concept forward, as many device sizes are limited by other physical considerations. Metrics related to number of synapses on a wafer, number of planes of integrated circuits, and number of wafers in a system may be more relevant to chart progress in neuromorphic supercomputing. Gradual progress may be possible by consistently scaling up, but it is difficult to envision this sustained trend without a powerful economic drive.

5.5 Power consumption and cooling

5.5.1 Cooling Systems

Cooling systems will be a key component to either platform. For superconducting electronics, the system will fail completely if the temperature rises above the critical temperature (T_c). Superconducting neuromorphic systems will rely on niobium ($T_c = 9.3$K) or a material with a similarly low T_c. Liquid helium (4.2 K) is the cryogen of choice for such temperatures. Cooling systems will add significantly to the power consumption of superconducting electronics. The power efficiency of a refrigeration system is measured by its specific power [168]. The specific power gives the number of watts consumed by the refrigeration system for every watt of heat removed. The theoretical limit for specific power, given by the Carnot limit, is $\frac{T_H - T_C}{T_C}$. For liquid helium temperature (4.2 K), the Carnot limit demands that at least 74 Watts of refrigeration power are required to remove every watt of heat produced on-chip if the system is operated in a 300 K ambient. State-of-the-art systems have reached specific powers below 400 W/W. Auspiciously, the most efficient refrigeration systems also tend to have the highest heat loads. Heat loads as high as 10 kW at 4.4K have already been demonstrated by commercially available systems [169]. Note that throughout this paper, we assume a more conservative specific power of 1000 W/W, representative of the smaller scale cryogenic systems used in most laboratories today. Overall, it does not appear that cryogenic capability will be an insurmountable obstacle towards large-scale superconducting neural systems.

5.5.2 Power Limitations

Modern supercomputers typically consume megawatts of power. Tianhe 2, for instance, requires 17.8 MW for operation (and another 6.4 MW for cooling) [110]. If we thus assume a total power budget on the order of 10 MW, we can analyze the trade-off between average firing rate and the number of neurons. We assume 1 fJ of optical energy is required to initiate a firing event at each synapse and plot the maximum average frequency spiking frequency for several different optical link efficiencies in figure 6.

Power does not appear to be a limiting factor in achieving brain-scale and brain-speed optoelectronic networks. If the power resources of modern supercomputers were dedicated to a brain-scale optoelectronic neuromorphic system, average spiking rates on the order of 10 kHz (10^4 speedup over biology) appear feasible even with relatively inefficient optical links. Such a system may enable brain-scale computation with time accelerated by four orders of magnitude.

Another factor to consider is power density. There is a maximum power density that can be handled by heat removal systems for both the semiconducting and superconducting case. In the semiconductor case, high-performance computing routinely generates power den-
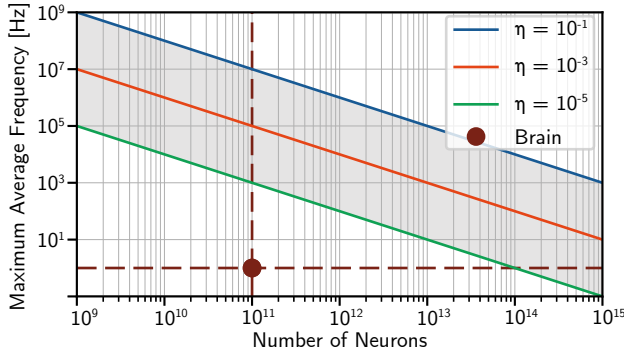
Figure 6: Tradeoff between size and average spiking frequency for a population of optoelectronic neurons with a power budget of 10 MW. Fan-out is 10^3 and the optical energy needed at each synapse is assumed to be 1 fJ (accounting for cooling in superconductor case). This likely would correspond to the limits of either superconductor or semiconductor neurons.

sities of hundreds of watts per square centimeter [110]. A theoretical limit of around $1\,\mathrm{kW/cm^2}$ is postulated in ref [170]. In contrast, superconducting systems will be required to operate at significantly lower power densities. Roughly $1\,\mathrm{W/cm^2}$ is a conservative limit for on-chip power density that can be cooled with liquid helium [110]. Interestingly, superconducting optical links appear to be capable of dissipating about three orders of magnitude less energy per bit, approximately cancelling out the limited power density requirements of superconducting systems in comparison with semiconductors. In practice, it might well be the case that mature, sophisticated synapses and neurons will occupy so much area that these power density limitations will be of no consequence. For instance, even with link efficiencies of 1×10^{-4}, a synapse would require a lateral dimension of less than $30\,\mu\mathrm{m}$ for power density considerations to limit spiking to less than $1\,\mathrm{GHz}$. Section 5 argued that superconducting synapses are not likely to be smaller than this. $10\,\mu\mathrm{m}$ semiconducting synapses could reach $1\,\mathrm{GHz}$ with 1×10^{-3} efficiency. However, it should also be noted that optoelectronic systems will have extremely nonuniform power dissipation across the chip/wafer, with most of the power being dissipated at the light sources. A more in-depth analysis is required to see if heat removal will be an issue near the light-sources in particular, but for the superconducting case it is convenient that the light sources themselves are not superconducting, and can afford to be raised to higher temperatures without failure. Concerns about local heating may be assuaged with layouts that sufficiently shield and/or separate thermally sensitive devices from the light sources.

6   Conclusion

The prospects of neuromorphic systems at the scale of the brain and beyond are tantalizing. The fan-out capability of optical communication coupled with the computational power of electronic circuitry makes optoelectronic systems a promising framework for realizing these high ambitions. However, there is no technology platform that is ready to support optoelectronic spiking networks of the scale and sophistication of the human brain. Making this vision a reality will require breakthroughs at the device level, no matter which path is chosen, particularly with regard to integrated light sources. Beyond that, several different classes of devices must be integrated alongside each other, which further reduces the likelihood for success. Efficient, densely integrated light sources, waveguide-coupled detectors, local memory devices, and capable neuronal circuitry all must be consolidated onto a single platform. Candidates for all requisite devices can be proposed for either semiconducting or superconducting platforms, and the two systems may be capable of similar performance. However, the technological paths toward achieving brain-scale systems with the two platforms diverge in important respects.

Semiconductor platforms hold advantages in technological maturity, room-temperature operation, and perhaps speed. Spike rates in excess of $10\,\mathrm{GHz}$ may be feasible, but only for systems significantly smaller than the human brain. Semiconductor receivers can potentially operate with extremely low energies per spiking event (sub femto-joule), making them a worthy competitor of superconducting single photon detectors. However, these low energy receivers require significant optical power from integrated light sources. To achieve biological-scale fan-out, either very bright light sources, repeatering schemes (costing area and yield), or additional gain stages (costing power) will need to be included. In terms of neuronal computation, semiconductor neurons have already demonstrated impressive functionality and low-power operation that should be capable of integration with optical communication infrastructure, provided the long-standing challenges with CMOS-integrated III-V light sources can be overcome. Synaptic memory is a major open question, but a variety of non-volatile memory solutions have seen extensive investigation, and time will tell if one technology can meet the requirements we have laid out for brain-scale optoelectronic systems. 3D integration of transistors, photodetectors, and memory may not be a feasible solution, meaning aggressive scaling of synaptic circuits while maintaining complex functionality is perhaps a better strategy. The fabrication processes for mature semiconductor neural systems may prove to be prohibitively complicated and heterogeneous, perhaps requiring different processing strategies for sources, detectors, and memories. If wafer-scale monolithic integration of these components cannot be achieved, and chip-scale die-stacking techniques are required, the outlook for achieving brain-scale systems is limited.

Both Platforms
Monolithic integration of light sources, detectors, memory devices, and electronics
At least 4 planes of photonic planes
Wafer-scale processing
Inter-wafer optical links
Demonstrate memory meeting requirements in section 4.1
Semiconductor Platform
Demonstrate femtojoule optical receivers, ideally with low static power dissipation
III-V integration with electronics (1 million light sources/wafer)
Synapses and local plasticity mechanisms reach 10µm x 10µm
Superconductor Platform
Either III-V integration *or* cryogenic silicon light sources (1 million light sources/wafer)
Interface superconducting electronics with semiconductor light sources
Serial biasing or current-recycling schemes for current biasing of synapses and neurons
Demonstrate 8 planes of Josephson junctions per wafer

Figure 7: Summary of targets for each platform

Superconducting optoelectronic neural systems suffer from a comparatively primitive fabrication ecosystem, but the incorporation of superconducting devices provides several intriguing properties. SNSPD receivers place nearly the theoretical minimum burden on integrated light sources. This attribute compounds positively with the improvements in efficiency for light sources operating at cryogenic temperatures. Integration of light sources with superconducting electronics does not appear to have the same materials integration challenges as integration with CMOS, but this state of affairs may be due to the lack of attention the effort has received. These factors make the large-scale integration of light sources appear more tractable than in the semiconductor case—perhaps even opening the door to silicon as an active optical material. Driving these light sources with superconducting electronics, however, has yet to demonstrate the performance required for this application. The implementation of a high-impedance pulse-and-reset circuit remains an open challenge. For computation, superconducting neuronal circuits appear just as capable of implementing complex neuronal and synaptic behaviors as their CMOS counterparts, but will need to be designed with serial biasing in order to scale. Additionally, some speed advantages present in superconducting electronics will be negated by the response time of SNSPDs ($<1\,$GHz). Of course, even if maximum spike rates are limited to 20 MHz, this would still represent a speed-up of about 4 orders of magnitude over biological systems. Memory seems to be a strength for the superconducting platform, as superconductivity provides new avenues of storing synaptic weights. Loop memory in particular may be capable of implementing plasticity mechanisms that operate with only the signals produced through normal network activity. Caution is in order here, however, as superconducting synaptic plasticity mechanisms have scarcely been explored. 3D integration may yield more readily in the superconductor platform. The inconvenience of cryogenic cooling is a serious consideration, but power and heat removal estimations indicate this is unlikely to be a limiting factor for brain-scale systems. *If* all these issues can be resolved, superconducting optoelectronic systems may require simpler manufacturing processes than the semiconductor approach, as the material ecosystem could potentially be parsimonious. Of course, superconducting foundries are far less developed than their semiconductor counterparts, which may negate these advantages in the near-term.

We would be remiss to paint the quest for neuromorphic supercomputing as only a question of hardware. The inner workings of the brain are the subject of intense investigation, and the emergent phenomena of cognition and consciousness remain taunting, increasingly lonely enigmas entrenched in the netherworld between philosophy and science. Watershed breakthroughs in neuroscience and algorithmic development will be required for the discussed hardware platforms to have practical applications, although they may be of use in helping to unravel some of these mysteries. The question of whether it is imprudent to develop hardware before algorithms has pestered the field of neuromorphic computing since its inception. In this case, we believe that the length of development, rich opportunities for spin-off technologies, and inestimable potential make such hardware development well-worth pursuing even at this incipient stage.

Acknowledgements

A   Implementing Long Time Constants

For the DPI synapse (Sec. 3) the time constant is given by [103]:

$$\tau = \frac{C_{si}V_{th}}{\kappa I_\tau} \tag{9}$$

Where I_τ is the current leaking off of the capacitor, set with transistor T_3, V_{th} is the thermal voltage, and κ is the sub-threshold slope factor (typically order 1). Operating in the subthreshold regime allows I_τ to be reduced to femtoamps [171]. Metal Insulator Metal (MIM) capacitors ultizing high-k dielectrics can reach capacitance densities around $20\,\text{fF}/\mu\text{m}^2$ [172]. The maximum achievable time constant as a function of synapse width is shown in figure 8 for $I_\tau = 10\,\text{fA}$, $\kappa = 1$, and $V_{th} = 25\,\text{mV}$. Since MIM capacitors can be fabricated on a separate layer from transistors, the entire $10\,\mu\text{m} \times 10\,\mu\text{m}$ per synapse area target identified in Sec. 5 could be dedicated to capacitor area.

For SOENs synapses, L_{si}/r_{si} sets the time constant. Inductors and resistors will most likely be fabricated on separate layers, again conserving space. A meandering geometry gives the maximum inductance L_{si}, that can be fabricated in an area w_{sy}^2 as:

$$L_{si} = \frac{w_{sy}^2 l}{w_{wire}(w_{wire} + w_{gap})} \tag{10}$$

Where l is the inductance per square of the material, and w_{wire} and w_{gap} are the minimum feature sizes. Small value resistors are fabricated by putting many wide resistances in parallel. The smallest (nonzero of course) resistor that can be fabricated in an area of w_{sy}^2 is then:

$$r_{si} = \frac{R_s w_{gap}(w_{wire} + w_{gap})}{w_{sy}^2} \tag{11}$$

Where R_s is the sheet resistance. The maximum time constant (τ_{max}) in area w_{sy}^2 is given by L_{si}/r_{si}:

$$\tau_{max} = \frac{w_{sy}^4 l}{R_s w_{wire} w_{gap}(w_{wire} + w_{gap})^2} \tag{12}$$

The maximum achievable time constants as a function of synapse width is plotted in figure 8 for $R_s = .001\,\Omega/\square$ (corresponds to a 200 nm gold layer with resistivity of $2 \times 10^{-10}\,\Omega\text{m}$ at $4\,\text{K}$ [173]), $w_{wire} = w_{gap} = 100\,\text{nm}$, and $l = 160\,\text{pH}/\square$, corresponding to MoSi. For the large

synaptic areas expected to be available to superconducting synapses via 3D integration, the superconducting approach can support significantly larger time constants than the semiconducting case. In ref [99], it is suggested that the human brain itself is limited by the maximum achievable time constant, and relies on network dynamics for certain types of long-term memory. The ramifications of a hardware with time-constants far greater than biology are intriguing.
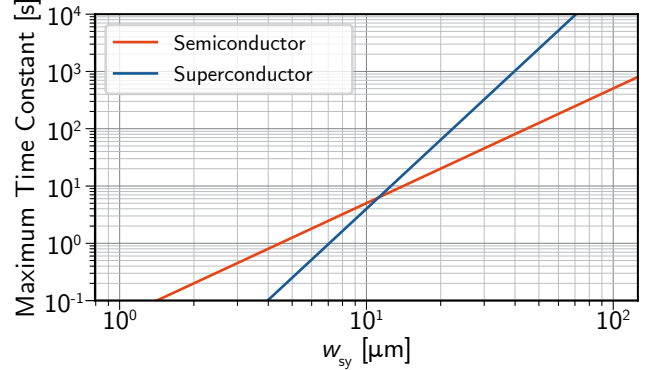

Figure 8: Maximum achievable time constant as a function of synapse width.

B   Area and Energy in Superconducting Synapses

JJs will never be as small as MOSFETs [110], but the area of synaptic circuits is not limited by JJs. Large inductors or transformers that couple flux from storage loops to receiving SQUIDs are likely to be the components that consume the most area. The sizes of these components is determined by the critical current of the junctions used in the SQUIDs (I_c), the permeability of free space (μ_0), SQUID inductance (L), and the magnetic flux quantum (Φ_0). Based on the typical SQUID design criterion we expect $2LI_c/\Phi_0 = 1$ [174]. For the usual washer-type inductor geometry, $L \approx 1.25\mu_0 w_{sq}$, where w_{sq} is the inner dimension of the hole [175]. The energy, E_{sq} to produce two fluxons is approximately $2I_c\Phi_0$ for an appropriately biased junction. There is thus a trade-off between the size of a SQUID and its energy:

$$w_{sq} \approx \frac{\Phi_0^2}{1.25\mu_0 E_{sq}} \tag{13}$$

Additionally, I_c is constrained by noise and the ability to interface with SNSPDs. $I_c = 60\,\mu\text{A}$ is reasonable and would correspond to $w_{sq} \approx 11\,\mu\text{m}$ and $E_{sq} \approx 0.2\,\text{aJ}$. If the optical energy per synapse is around 100 aJ for a 1% efficient link, about 1000 fluxons can be produced per synapse event without dominating the power budget. In terms of size, a synapse would likely be composed of

three or four SQUIDs—one for synaptic integration, one for loop memory, and perhaps two more for various plasticity functions. A reasonable estimate for synaptic size is then about $30\,\mu m \times 30\,\mu m$.

C Further Scaling Analysis

Figures 9 and 10 provide insight into how network connectivity constrains hardware for any planar fully-dedicated system. Figure 9 plots equation 6, giving the average node degree (number of synapses per neuron) necessary to maintain a given path length as a function of network size.
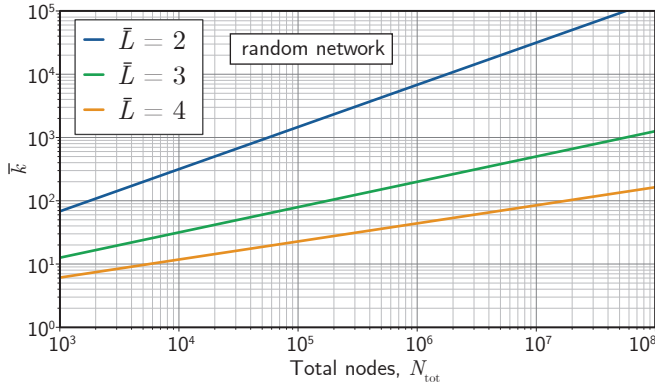
Figure 9: Average node degree as a function of network size for different path lengths.

Figure 10 plots the minimum achievable path length as function of synaptic width and waveguide width (equations 8 and 7). We note that path length is relatively insensitive to waveguide pitch, suggesting that wider, and therefore lower loss waveguides may be beneficial. In contrast, we see that the synaptic size can be a major impediment to achieving low path lengths.

References

[1] Steve Furber. Large-scale neuromorphic computing systems. Journal of neural engineering, 13(5):051001, 2016.

[2] Jeffrey M Shainline, Sonia M Buckley, Adam N McCaughan, Jeffrey T Chiles, Amir Jafari Salim, Manuel Castellanos-Beltran, Christine A Donnelly, Michael L Schneider, Richard P Mirin, and Sae Woo Nam. Superconducting optoelectronic loop neurons. Journal of Applied Physics, 126(4):044902, 2019.

[3] Aaron R Young, Mark E Dean, James S Plank, and Garrett S Rose. A review of spiking neuromorphic hardware communication systems. IEEE Access, 7:135606–135620, 2019.
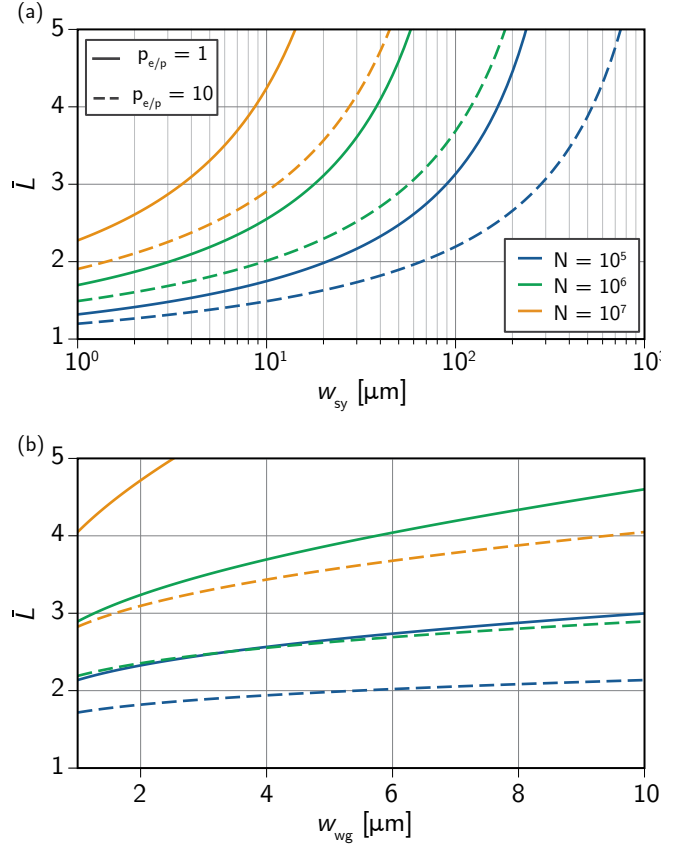
Figure 10: Path length versus feature size. (a) Path length versus width of electronic synapses. (b) Path length versus waveguide pitch. In both figure the solid lines correspond to a single plane of synapses or waveguides, and the dashed lines correspond to 10 planes of synapses or waveguides. The blue, green, and yellow traces correspond to the cases of 10^5, 10^6, and 10^7 neurons per 300-mm wafer, respectively.

[4] C. Segal, A. Dalakoti, M. Miller, and F. Brewer. Connectivity Effects on Energy and Area for Neuromorphic System with High Speed Asynchronous Pulse Mode Links. SLIP, 16, 2016.

[5] Mohammadamin Nazirzadeh, Mohammadsadegh Shamsabardeh, and SJ Ben Yoo. Energy-efficient and high-throughput nanophotonic neuromorphic computing. In CLEO: Applications and Technology, pages ATh3Q–2. Optical Society of America, 2018.

[6] Bhavin J Shastri, Alexander N Tait, T Ferreira de Lima, Wolfram HP Pernice, Harish Bhaskaran, C David Wright, and Paul R Prucnal. Photonics for artificial intelligence and neuromorphic computing. Nature Photonics, 15(2):102–114, 2021.

[7] E Marder. Neurotransmitters and neuromodulators. In The crustacean stomatogastric system, pages 263–306. Springer, 1987.

[8] Thomas Euler and Winfried Denk. Dendritic processing. Current opinion in neurobiology, 11(4):415–422, 2001.

[9] Jeffrey M Shainline, Sonia M Buckley, Richard P Mirin, and Sae Woo Nam. Superconducting optoelectronic circuits for neuromorphic computing. Physical Review Applied, 7(3):034013, 2017.

[10] Jeffrey M Shainline. Fluxonic processing of photonic synapse events. IEEE Journal of Selected Topics in Quantum Electronics, 26(1):1–15, 2019.

[11] J.M. Shainline. Optoelectronic Intelligence. Appl. Phys. Lett., 118:160501, 2021.

[12] David AB Miller. Attojoule optoelectronics for low-energy information processing and communications. Journal of Lightwave Technology, 35(3):346–396, 2017.

[13] J.P. Sprengers, A. Gaggero, D. Sahin, S. Jahanmirinejad, G. Frucci, F. Mattioli, R. Leoni, J. Beetz, M. Lermer, M. Kamp, S. Höfling, R. Sanjines, and A. Fiore. Waveguide superconducting single-photon detectors for integrated quantum photonic circuits. Appl. Phys. Lett., 99:181110, 2011.

[14] W. Pernice, C. Schuck, O. Minaeva, M. Li, G. Goltsman, A. Sergienko, and H. Tang. High speed travelling wave single-photon detectors with near-unity quantum efficiency. Nat. Comm., 3:1325, 2012.

[15] Mohsen K Akhlaghi, Ellen Schelew, and Jeff F Young. Waveguide integrated superconducting single-photon detectors implemented as near-perfect absorbers of coherent radiation. Nature communications, 6(1):1–8, 2015.

[16] S. Ferrari, O. Kahl, V. Kovalyuk, G.N. Goltsman, A. Korneev, and W.H.P. Pernice. Waveguide-integrated single- and multi-photon detecton at telecom wavelengths using superconducting nanowires. Appl. Phys. Lett., 106:151101, 2015.

[17] D. Sahin, A. Gaggero, J.-W. Weber, I. Agafonov, M.A. Verheijen, F. Mattioli, J. Beetz, M. Kamp, S. Höfling, M.C.M. van de Sanden, R. Leoni, , and A. Fiore. Waveguide nanowire superconducting single-photon detectors fabricated on gaas and the study of their optical properties. IEEE J. Sel. Top. Quant. Electron., 21:3800210, 2015.

[18] J.M. Shainline, S.M. Buckley, N. Nader, C.M. Gentry, K.C. Cossel, J.W. Cleary, M. Popović, N.R. Newbury, S.W. Nam, and R.P. Mirin. Room-temperature-deposited dielectrics and superconductors for integrated photonics. Opt. Express, 25:10322, 2017.

[19] Simone Ferrari, Carsten Schuck, and Wolfram Pernice. Waveguide-integrated superconducting nanowire single-photon detectors. Nanophotonics, 7(11):1725–1758, 2018.

[20] S.M. Buckley, A.N. Tait, J. Chiles, A.N. McCaughan, S. Khan, R.P. Mirin, S.W. Nam, and J.M. Shainline. Integrated-Photonic Characterization of Single-Photon Detectors for Use in Neuromorphic Synapses. Phys. Rev. Applied, page 054008, 2020.

[21] Behzad Razavi. Design of integrated circuits for optical communications. John Wiley & Sons, 2012.

[22] 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.

[23] D.V. Reddy, R.R. Nerem, S.W. Nam, R.P. Mirin, and V.B. Verma. Superconducting nanowire single-photon detectors with 98% system detection efficiency at 1550 nm. Optica, 7:1649, 2020.

[24] Richard B Stein, E Roderich Gossen, and Kelvin E Jones. Neuronal variability: noise or part of the signal? Nature Reviews Neuroscience, 6(5):389–397, 2005.

[25] Mark D McDonnell and Lawrence M Ward. The benefits of noise in neural systems: bridging theory and experiment. Nature Reviews Neuroscience, 12(7):415–425, 2011.

[26] V.B. Verma, F. Marsili, S. Harrington, A.E. Lita, R.P. Mirin, and S.W. Nam. A three-dimensional, polarization-insensitive superconducting nanowire avalanche photodetector. Appl. Phys. Lett., 101:251114, 2012.

[27] D Rosenberg, AJ Kerman, RJ Molnar, and EA Dauler. High-speed and high-efficiency superconducting nanowire single photon detector array. Optics express, 21(2):1440–1447, 2013.

[28] Andreas Vetter, Simone Ferrari, Patrik Rath, Rasoul Alaee, Oliver Kahl, Vadim Kovalyuk, Silvia Diewald, Gregory N Goltsman, Alexander Korneev, Carsten Rockstuhl, et al. Cavity-enhanced and ultrafast superconducting single-photon detectors. Nano letters, 16(11):7085–7092, 2016.

[29] E.E. Wollman, V.B. Verma, A.E. Lita, W.H. Farr, M.D. Shaw, R.P. Mirin, and S.W. Nam. Kilopixel array of superconducting nanowire single-photon detectors. Opt. Express, 27:35279, 2019.

[30] Christof Debaes, Aparna Bhatnagar, Diwakar Agarwal, Ray Chen, Gordon A Keeler, Noah C Helman,

Hugo Thienpont, and David AB Miller. Receiverless optical clock injection for clock distribution networks. IEEE journal of selected topics in quantum electronics, 9(2):400–409, 2003.

[31] Christopher T DeRose, Douglas C Trotter, William A Zortman, Andrew L Starbuck, Moz Fisher, Michael R Watts, and Paul S Davids. Ultra compact 45 ghz cmos compatible germanium waveguide photodiode with low dark current. Optics express, 19(25):24897–24904, 2011.

[32] K.K. Mehta, J.S. Orcutt, J.M. Shainline, O. Tehar-Zahav, Z. Sternberg, R. Meade, M.A. Popović, and Rajeev J. Ram. Polycrystalline silicon ring resonator photodiodes in a bulk complementary metal-oxide-semiconductor process. Opt. Lett., 39:1061, 2014.

[33] Yu Zhang, Anirban Samanta, Kuanping Shang, and SJ Ben Yoo. Scalable 3d silicon photonic electronic integrated circuits and their applications. IEEE Journal of Selected Topics in Quantum Electronics, 26(2):1–10, 2020.

[34] Kengo Nozaki, Shinji Matsuo, Takuro Fujii, Koji Takeda, Akihiko Shinya, Eiichi Kuramochi, and Masaya Notomi. Forward-biased nanophotonic detector for ultralow-energy dissipation receiver. APL Photonics, 3(4):046101, 2018.

[35] Solomon Assefa, Fengnian Xia, and Yurii A Vlasov. Reinventing germanium avalanche photodetector for nanophotonic on-chip optical interconnects. Nature, 464(7285):80–84, 2010.

[36] Léopold Virot, Paul Crozat, Jean-Marc Fédéli, Jean-Michel Hartmann, Delphine Marris-Morini, Eric Cassan, Frédéric Boeuf, and Laurent Vivien. Germanium avalanche receiver for low power interconnects. Nature communications, 5(1):1–6, 2014.

[37] Andrea Pizzone, Srinivasan Ashwyn Srinivasan, Peter Verheyen, Guy Lepage, Sadhishkumar Balakrishnan, and Joris Van Campenhout. Analysis of dark current in ge-on-si photodiodes at cryogenic temperatures. In 2020 IEEE Photonics Conference (IPC), pages 1–2. IEEE.

[38] Adam N McCaughan, Varun B Verma, Sonia M Buckley, JP Allmaras, AG Kozorezov, AN Tait, SW Nam, and JM Shainline. A superconducting thermal switch with ultrahigh impedance for interfacing superconductors to semiconductors. Nature electronics, 2(10):451–456, 2019.

[39] Bruno Romeira and Andrea Fiore. Physical limits of nanoleds and nanolasers for optical communications. Proceedings of the IEEE, 108(5):735–748, 2019.

[40] S. S. Iyer and Y.-H. Xie. Light emission from silicon. Science, 260:40, 1993.

[41] J.M. Shainline and J. Xu. Silicon as an emissive optical medium. Laser and Photonics Reviews, 1:334, 2007.

[42] J Warga, R Li, SN Basu, and L Dal Negro. Electroluminescence from silicon-rich nitride/silicon superlattice structures. Applied Physics Letters, 93(15):151116, 2008.

[43] Robert J. Walters, George I. Bourianoff, and Harry A. Atwater. Field-effect electroluminescence in silicon nanocrystals. Nature Materials, 4:143–146, 2005.

[44] H. Ennen, G. Pomrenke, A. Axmann, K. Eisele, W. Haydl, and J. Schneider. 1.54 μ m electroluminescence of erbium-doped silicon grown by molecular beam epitaxy. Appl. Phys. Lett., 46:381, 1985.

[45] J. Palm, F. Gan, B. Zheng, J. Michel, and L.C. Kimerling. Electroluminescence of erbium-doped silicon. Phys. Rev. B, 54:17603, 1996.

[46] T.G. Brown and D.G. Hall. Observation of electroluminescence from excitons bound to isoelectronic impurities in crystalline silicon. J. Appl. Phys., 59:1399, 1986.

[47] P.L. Bradfield, T.G. Brown, and D.G. Hall. Electroluminescence from sulfur impurities in an p-n junction formed in epitaxial silicon. Appl. Phys. Lett., 55:100, 1989.

[48] E. Rotem, J.M. Shainline, and J.M. Xu. Electroluminescence of nanopatterned silicon with carbon implantation and solid phase epitaxial regrowth. Opt. Express, 15:14099, 2007.

[49] Jiming Bao, Malek Tabbael, Taegon Kim, Supakit Charnvanichborikarn, James S. Williams, Michael J. Aziz, and Federico Capasso. Point defect engineered si sub-band gap light-emitting diode. Opt. Express, 15:6727, 2007.

[50] Wai Lek Ng, MA Lourenco, RM Gwilliam, S Ledain, Guosheng Shao, and KP Homewood. An efficient room-temperature silicon-based light-emitting diode. Nature, 410(6825):192–194, 2001.

[51] V. Kveder, M. Badylevich, E. Steinman, A. Izotov, M. Seibt, and W. Schröter. Room-temperature silicon light-emitting diodes based on dislocation luminescence. Appl. Phys. Lett., 84:2106, 2004.

[52] Martin A. Green, Jianhua Zaho, Aihua Wang, Peter J. Reese, and Michael Gal. Efficient silicon light-emitting diodes. Nature, 412:805, 2001.

[53] Elham M.T. Fadaly, Alain Dijkstra, Jens Renè Suckert, Dorian Ziss, Marvin A.J. van Tilburg, Chenyang

Mao, Yizhen Ren, Victor T. van Lange, Ksenia Korzun, Sebastian Kölling, Marcel A. Verheijen, David Busse, Claudia Rödl, Jürgen Furthmüller, Friedhelm Bechstedt, Julian Stangl, Jonathan J. Finley, Silvana Botti, Jos E.M. Haverkort, and Erik P.A.M. Bakkers. Direct-bandgap emission from hexagonal Ge and SiGe alloys. Nature, 580(7802):205–209, 2020.

[54] Justin C Norman, Daehwan Jung, Yating Wan, and John E Bowers. Perspective: The future of quantum dot photonic integrated circuits. APL Photonics, 3(3):030901, 2018.

[55] S. Chen, W. Li, J. Wu, Q. Jiang, M. Tang, S. Shutts, S.N. Elliott, A. Sobiesierski, A.J. Seeds, I. Ross, P.M. Smowton, and H. Liu. Electrically pumped continuous-wave III-V quantum dot lasers on silicon. Nat. Photonics, 10:307, 2016.

[56] Daehwan Jung, Justin Norman, MJ Kennedy, Chen Shang, Bongki Shin, Yating Wan, Arthur C Gossard, and John E Bowers. High efficiency low threshold current 1.3 μ m inas quantum dot lasers on on-axis (001) gap/si. Applied Physics Letters, 111(12):122107, 2017.

[57] Mingchu Tang, Jae-Seong Park, Zhechao Wang, Siming Chen, Pamela Jurczak, Alwyn Seeds, and Huiyun Liu. Integration of iii-v lasers on si for si photonics. Progress in Quantum Electronics, 66:1–18, 2019.

[58] B. Song, C. Stagarescu, S. Ristic, A. Behfar, and J. Klamkin. 3D integrated hybrid silicon laser. Opt. Express, 24:10435, 2016.

[59] G. Crosnier, D. Sanchez, S. Bouchoule, P. Monnier, G. Beaudoin, I. Sagnes, R. Raj, and F. Raineri. Hybrid indium phosphide-on-silicon nanolaser diode. Nat. Photon., 11:297, 2017.

[60] Y. Hu, D. Liang, K. Mikherjee, Y. Li, C. Zhang, G. Kurcveil, X. Huang, and R.G. Beausoleil. III/V-on-Si MQW lasers by using a novel photonic integration method of regrowth on a bonding template. Light: Science & Applications, 8:93, 2019.

[61] B. Szelag, K. Hassan, L. Adelmini, E. Ghegin, P. Rodriguez, F. Nemouchi, P. Brianceau, E. Vermande, A. Schembri, D. Carrara, P. Cavalié, F. Franchin, M.-C. Roure, L. Sanchez, C. Jany, and S. Olivier. Hybrid iii-V/Silicon Technology for Laser Integration on a 200-mm Fully CMOS-Compatible Silicon Photonics Platform. IEEE J. Sel. Top. Quant. Electron., 25:8201210, 2019.

[62] Y. Jiao, J. van der Tol, V. Pogoretskii, J. van Engelen, A.A. Kashi, S. Reniers, Y. Wang, X. Zhao, and W. Yao et al. Indium Phosphide Membrane Nanophotonic Integrated Circuits on Silicon. Physica Status Solidi, 217:1900606, 2020.

[63] J. Justice, C. Bower, M. Meitl, M.B. Mooney, M.A. Gubbins, and B. Corbett. Wafer-scale integration of group III-V lasers on silicon using transfer printing of epitaxial layers. Nat. Photonics, 6:610, 2012.

[64] J. Zhang, B. Haq, J. O'Callaghan, A. Gocalinska, E. Pelucchi, A.J. Trindade, B. Corbett, G. Morthier, and G. Roelkens. Transfer-printing-based integration of a III-V-on-silicon distributed feedback laser. Opt. Express, 26:8821, 2018.

[65] Jing Zhang, Grigorij Muliuk, Joan Juvert, Sulakshna Kumari, Jeroen Goyvaerts, Bahawal Haq, Camiel Op de Beeck, Bart Kuyken, Geert Morthier, Dries Van Thourhout, et al. Iii-v-on-si photonic integrated circuits realized using micro-transferprinting. APL photonics, 4(11):110803, 2019.

[66] Y. Han, Y. Xue, Z. Yan, and K.M. Lau. Selectively Grown III-V Lasers for Integrated Si-Photonics. J. Lightwave Technol., 39:940, 2021.

[67] M Gurioli, A Vinattieri, M Colocci, C Deparis, J Massies, G Neu, A Bosacchi, and S Franchi. Temperature dependence of the radiative and nonradiative recombination time in gaas/al x ga 1- x as quantum-well structures. Physical Review B, 44(7):3115, 1991.

[68] V Dolores-Calzadilla, B Romeira, F Pagliano, S Birindelli, A Higuera-Rodriguez, PJ Van Veldhoven, MK Smit, A Fiore, and D Heiss. Waveguide-coupled nanopillar metal-cavity lightemitting diodes on silicon. Nature communications, 8(1):1–8, 2017.

[69] D.J. Sandiford. Temperature Dependence of Carrier Lifetime in Silicon. Proc. Phys. Soc., 71:1002, 1958.

[70] Gordon Davies. The optical properties of luminescence centres in silicon. Physics reports, 176(3-4):83–188, 1989.

[71] H. Sumikura, E. Kuramochi, H. Taniyama, and M. Notomi. Ultrafast spontaneous emission of copper-doped silicon enhanced by an optical nanocavity. Sci. Rep., 4:5040, 2014.

[72] Sonia Buckley, Jeffrey Chiles, Adam N McCaughan, Galan Moody, Kevin L Silverman, Martin J Stevens, Richard P Mirin, Sae Woo Nam, and Jeffrey M Shainline. All-silicon light-emitting diodes waveguide-integrated with superconducting single-photon detectors. Applied Physics Letters, 111(14):141101, 2017.

[73] C. Beaufils, W. Redjem, E. Rousseau, V. Jacques, A.Y. Kuznetsov, C. Raynaud, C. Voisin, A. Benali, T. Herzig, S. Pezzagna, J. Meijer, M. Abbarchi, and G. cassabois. Optical properties of an ensembe of G-centers in silicon. Phys. Rev. B, 97:035303, 2018.

[74] C. Chartrand, L. Bergeron, K.J. Morse, H. Riemann, N.V. Abrosimov, P. Becker, H.-J. Pohl, S. Simmons, and M.L.W. Thewalt. Highly enriched ^{2}8Si reveals remarkable optical linewidths and fine structure for well-known damage centers. Phys. Rev. B, 98:195201, 2018.

[75] M. Hollenback, Y. Berencén, U. Kentsch, M. Helm, and G.V. Astakhov. Engineering telecom single-photon emitters in silicon for scalable quantum photonics. Opt. Express, 28:26111, 2020.

[76] W. Redjem, A. Durand, T. Herzig, A. Benali, S. Pezzagna, J. Meijer, A. Yu. Kuznetsov, H.S. Nguyen, S. Cueff, J.-M. Gérard, I. Robert-Philip, B. Gil, D. Caliste, P. Pochet, M. Abbarchi, V. Jacques, A. Dréau, and G. Cassabois. Single artificial atoms in silicon emitting at telecom wavelengths. Nat. Electronics, 3:738, 2020.

[77] L. Bergeron, C. Chartrand, A.T.K. Kurkjian, K.J. Morse, H. Riemann, N.V. Abrosimov, P. Becker, H.-J. Pohl, M.L.W. Thewalt, and S. Simmons. Silicon-Integrated Telecommunications Photon-Spin Interface. PRX Quantum, 1:020301, 2020.

[78] Bruno Romeira and Andrea Fiore. Purcell effect in the stimulated and spontaneous emission rates of nanoscale semiconductor lasers. IEEE Journal of Quantum Electronics, 54(2):1–12, 2018.

[79] Jiming Bao, Malek Tabbal, Taegon Kim, Supakit Charnvanichborikarn, James S Williams, Michael J Aziz, and Federico Capasso. Point defect engineered si sub-bandgap light-emitting diode. Optics Express, 15(11):6727–6733, 2007.

[80] Sonia M Buckley, Alexander N Tait, Galan Moody, Bryce Primavera, Stephen Olson, Joshua Herman, Kevin L Silverman, Satyavolu Papa Rao, Sae Woo Nam, Richard P Mirin, et al. Optimization of photoluminescence from w centers in a silicon-on-insulator. Optics express, 28(11):16057–16072, 2020.

[81] John E Bowers, Tin Komljenovic, Michael Davenport, Jared Hulme, Alan Y Liu, Christos T Santis, Alexander Spott, Sudharsanan Srinivasan, Eric J Stanton, and Chong Zhang. Recent advances in silicon photonic integrated circuits. In Next-Generation Optical Communication: Components, Sub-Systems, and Systems V, volume 9774, page 977402. International Society for Optics and Photonics, 2016.

[82] Hubert Halbritter, Claus Jäger, Rolf Weber, Michael Schwind, and Frank Möllmer. High-speed led driver for ns-pulse switching of high-current leds. IEEE Photonics Technology Letters, 26(18):1871–1873, 2014.

[83] C. Koch and I Segev. The Role of Single Neurons in Information Processing. Nature Neuroscience, 3:1171, 2000.

[84] G.J. Stuart and N. Spruston. Dendritic integration: 60 years of progress. Nature Neuroscience, 18:1713, 2015.

[85] 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.

[86] 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.

[87] R.J. Vogelstein, U. Mallik, J.T. Vogelstein, and G. Cauwenberghs. Dynamically reconfigurable silicon array of spiking neurons with conductance-based synapses. IEEE Trans. Neural Networks, 18:253, 2007.

[88] Giacomo Indiveri, Bernabé Linares-Barranco, Tara Julia Hamilton, André Van Schaik, Ralph Etienne-Cummings, Tobi Delbruck, Shih-Chii Liu, Piotr Dudek, Philipp Häfliger, Sylvie Renaud, et al. Neuromorphic silicon neuron circuits. Frontiers in neuroscience, 5:73, 2011.

[89] T. Pfeil, A. Grubl, S. Jeltsch, E. Müller, M.A. Metrovici, M. Schmuker, D. Brüderle, J. Schemmel, and K. Meier. Six networks on a universal neuromorphic computing substrate. Frontiers in Neuroscience, 7:1, 2013.

[90] S. Brink, S. Nease, and P. Hasler. Computing with networks of spiking neurons on a biophysically motivated floating-gate based neuromorphic integrated circuit. Neural Networks, 45:39, 2013.

[91] B.V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A.R. Chandresekaran, J.-M. Bussat, R. Alvarez-Icaza, J.V. Arthur, P.A. Merolla, and K. Boahen. Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. Proceedings of the IEEE, 102:699, 2014.

[92] K. Abu-Hassan, J.D. Taylor, P.G. Morris, E. Donati, Z.A. Bortolotto, G. Indiveri, J.F.R. Paton, and A. Nogaret. Optimal solid state neurons. Nat. Comm., 10:5309, 2019.

[93] Patrick Crotty, Dan Schult, and Ken Segall. Josephson junction simulation of neurons. Physical Review E, 82(1):011914, 2010.

[94] Emily Toomey, Ken Segall, and Karl K Berggren. Design of a power efficient artificial neuron using superconducting nanowires. Frontiers in neuroscience, 13:933, 2019.

[95] Carver Mead. Neuromorphic electronic systems. Proceedings of the IEEE, 78(10):1629–1636, 1990.

[96] Bipin Rajendran, Yong Liu, Jae-sun Seo, Kailash Gopalakrishnan, Leland Chang, Daniel J Friedman, and Mark B Ritter. Specifications of nanoscale devices and circuits for neuromorphic computational systems. IEEE Transactions on Electron Devices, 60(1):246–253, 2012.

[97] S.-C. Liu, T. Delbruck, G. Indiveri, A. Whatley, and R. Douglas, editors. Event-based neuromorphic systems. John Wiley and Sons, 2015.

[98] Ilias Sourikopoulos, Sara Hedayat, Christophe Loyez, François Danneville, Virginie Hoel, Eric Mercier, and Alain Cappy. A 4-fj/spike artificial neuron in 65 nm cmos technology. Frontiers in neuroscience, 11:123, 2017.

[99] Giacomo Indiveri and Yulia Sandamirskaya. The importance of space and time for signal processing in neuromorphic agents: the challenge of developing low-power, autonomous agents that interact with the environment. IEEE Signal Processing Magazine, 36(6):16–28, 2019.

[100] Johannes Schemmel, Laura Kriener, Paul Müller, and Karlheinz Meier. An accelerated analog neuromorphic hardware system emulating nmda- and calcium-based non-linear dendrites. In 2017 International Joint Conference on Neural Networks (IJCNN), pages 2217–2226. IEEE, 2017.

[101] J.M. Beggs. The criticality hypothesis: how local cortical networks might optimize information processing. Philosophical transactions of the Royal Society A, 366:329, 2007.

[102] Thomas Dalgaty, Melika Payvand, Filippo Moro, Denys RB Ly, Florian Pebay-Peyroula, Jerome Casas, Giacomo Indiveri, and Elisa Vianello. Hybrid neuromorphic circuits exploiting non-conventional properties of rram for massively parallel local plasticity mechanisms. APL Materials, 7(8):081125, 2019.

[103] Elisabetta Chicca, Fabio Stefanini, Chiara Bartolozzi, and Giacomo Indiveri. Neuromorphic electronic circuits for building autonomous cognitive systems. Proceedings of the IEEE, 102(9):1367–1388, 2014.

[104] Y. Harada and E. Goto. Artificial neural network circuits with josephson devices. IEEE Trans. Magnetics, 27:2863, 1991.

[105] M. Hidaka and L.A. Akers. An artificial neural cell implemented with superconducting circuits. Supercond. Sci. Technol., 4:654, 1991.

[106] Michael L Schneider, Christine A Donnelly, and Stephen E Russek. Tutorial: High-speed low-power neuromorphic systems based on magnetic josephson junctions. Journal of Applied Physics, 124(16):161102, 2018.

[107] Michael L Schneider, Christine A Donnelly, Stephen E Russek, Burm Baek, Matthew R Pufall, Peter F Hopkins, and William H Rippard. Energy-efficient single-flux-quantum based neuromorphic computing. In 2017 IEEE International Conference on Rebooting Computing (ICRC), pages 1–4. IEEE, 2017.

[108] Luca Cocchi, Leonardo L Gollo, Andrew Zalesky, and Michael Breakspear. Criticality in the brain: A synthesis of neurobiology, models and cognition. Progress in neurobiology, 158:132–152, 2017.

[109] ML Schneider and K Segall. Fan-out and fan-in properties of superconducting neuromorphic circuits. Journal of Applied Physics, 128(21):214903, 2020.

[110] Sergey K Tolpygo. Superconductor digital electronics: Scalability and energy efficiency issues. Low Temperature Physics, 42(5):361–379, 2016.

[111] D.E. Kirichenko, S. Sarwana, and A.F. Kirichenko. Zero static power dissipation biasing of rsfq circuits. IEEE Trans. Appl. Supercond., 21:776, 2011.

[112] C. van Vreeswijk and H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science, 274:1724, 1996.

[113] T.P. Vogels, K. Rajan, and L.F. Abbott. Neural network dynamics. Annu. Rev. Neurosci., 28:357, 2005.

[114] Furqan Zahoor, Tun Zainal Azni Zulkifli, and Farooq Ahmad Khanday. Resistive random access memory (rram): an overview of materials, switching mechanism, performance, multilevel cell (mlc) storage, modeling, and applications. Nanoscale research letters, 15:1–26, 2020.

[115] Michael L Schneider, Christine A Donnelly, Stephen E Russek, Burm Baek, Matthew R Pufall, Peter F Hopkins, Paul D Dresselhaus, Samuel P Benz, and William H Rippard. Ultralow power artificial synapses using nanotextured magnetic josephson junctions. Science advances, 4(1):e1701329, 2018.

[116] Thomas Pfeil, Tobias C Potjans, Sven Schrader, Wiebke Potjans, Johannes Schemmel, Markus Diesmann, and Karlheinz Meier. Is a 4-bit synaptic weight resolution enough?–constraints on enabling spike-timing dependent plasticity in neuromorphic hardware. Frontiers in neuroscience, 6:90, 2012.

[117] Naigang Wang, Jungwook Choi, Daniel Brand, Chia-Yu Chen, and Kailash Gopalakrishnan. Training deep neural networks with 8-bit floating point numbers. arXiv preprint arXiv:1812.08011, 2018.

[118] Thomas M Bartol Jr, Cailey Bromer, Justin Kinney, Michael A Chirillo, Jennifer N Bourne, Kristen M Harris, and Terrence J Sejnowski. Nanoconnectomic upper bound on the variability of synaptic plasticity. Elife, 4:e10778, 2015.

[119] S. Fusi, P.J. Drew, and L.F. Abbott. Casdcade models of synaptically stored memories. Neuron, 45:599, 2005.

[120] S. Fusi and L.F. Abbott. Limits on the memory storage capacity of bounded synapses. Nature Neuroscience, 10:485, 2007.

[121] Navnidhi K Upadhyay, Hao Jiang, Zhongrui Wang, Shiva Asapu, Qiangfei Xia, and J Joshua Yang. Emerging memory devices for neuromorphic computing. Advanced Materials Technologies, 4(4):1800589, 2019.

[122] Chris Diorio, Paul Hasler, Bradley A Minch, and Carver Mead. Floating-gate mos synapse transistors. In Neuromorphic Systems Engineering, pages 315–337. Springer, 1998.

[123] Shubha Ramakrishnan, Paul E Hasler, and Christal Gordon. Floating gate synapses with spike-time-dependent plasticity. IEEE Transactions on Biomedical Circuits and Systems, 5(3):244–252, 2011.

[124] Jennifer Hasler and Harry Bo Marr. Finding a roadmap to achieve large neuromorphic hardware systems. Frontiers in neuroscience, 7:118, 2013.

[125] D.B. Strukov, G.S. Snider, D.R. Stewart, and R.S. Williams. The missing memristor found. Nature, 453:80, 2008.

[126] J.J. Yang, D.B. Strukov, and D.R. Stewart. Memristive devices for computing. Nat. Nanotech., 8:13, 2012.

[127] I. Abraham. The case for rejecting the memristor as a fundamental circuit element. Nature, 8:10972, 2018.

[128] Shihui Yin, Yulhwa Kim, Xu Han, Hugh Barnaby, Shimeng Yu, Yandong Luo, Wangxin He, Xiaoyu Sun, Jae-Joon Kim, and Jae-sun Seo. Monolithically integrated rram-and cmos-based in-memory computing optimizations for efficient deep learning. IEEE Micro, 39(6):54–63, 2019.

[129] Daniele Ielmini. Brain-inspired computing with resistive switching memory (rram): Devices, synapses and neural networks. Microelectronic Engineering, 190:44–53, 2018.

[130] Stefano Ambrogio, Nicola Ciocchini, Mario Laudato, Valerio Milo, Agostino Pirovano, Paolo Fantini, and Daniele Ielmini. Unsupervised learning by spike timing dependent plasticity in phase change memory (pcm) synapses. Frontiers in neuroscience, 10:56, 2016.

[131] Min-Kyu Kim and Jang-Sik Lee. Ferroelectric analog synaptic transistors. Nano letters, 19(3):2044–2050, 2019.

[132] Sun Gil Kim, Ji Su Han, Hyojung Kim, Soo Young Kim, and Ho Won Jang. Recent advances in memristive materials for artificial synapses. Advanced Materials Technologies, 3(12):1800457, 2018.

[133] Yang Zhang, Zhongrui Wang, Jiadi Zhu, Yuchao Yang, Mingyi Rao, Wenhao Song, Ye Zhuo, Xumeng Zhang, Menglin Cui, Linlin Shen, et al. Brain-inspired computing with memristors: Challenges in devices, circuits, and systems. Applied Physics Reviews, 7(1):011308, 2020.

[134] T. Van Duzer and C.W. Turner. Principles of superconductive devices and circuits. Prentice Hall, USA, second edition, 1998.

[135] Alan M. Kadin. Introduction to superconducting circuits. John Wiley and Sons, USA, first edition, 1999.

[136] 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.

[137] S.K. Tolpygo, V. Bolkhovsky, D.E. Oates, R. Rastogi, S. Zarr, A.L. Day, T.J. Weir, A. Wynn, and L.M. Johnson. Superconductor Electronics Fabrication Process with MoN_x Kinetic Inductors and Self-Shunted Josephson Junctions. IEEE Transactions on Applied Superconductivity, 28(4):1100212, 2018.

[138] V. Braitenberg and A. Schuz. Cortex: statistics and geometry of neuronal connectivity. Springer, Berlin, Germany, 1998.

[139] G. Buzsáki. Rhythms of the Brain. Oxford University Press, 2006.

[140] E. Bullmore and O. Sporns. The economy of brain network organization. Nature Reviews Neuroscience, 13:336, 2012.

[141] A. Fronczak, P. Fronczak, and J.A. Holyst. Average path length in random networks. Phys. Rev. E, 70:056110, 2004.

[142] H.A. Simon. The Architecture of Complexity. Proc. Amer. Phil. Soc., 106:467, 1962.

[143] D. Meunier, R. Lambiotte, and E.T. Bullmore. Modular and hierarchically modular organization of brain networks. Frontiers in Neuroscience, 4(200):1, 2010.

[144] V.B. Mountcastle. The columnar organization of the neocortex. Brain, 120:701, 1997.

[145] M. Bota, O. Sporns, and L.W. Swanson. Architecture of the cerebral cortical association connectome underlying cognition. PNAS, page E2093, 2015.

[146] R.F. Betzel and D.S. Bassett. Multi-scale brain networks. NeuroImage, 160:73, 2017.

[147] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience, 10:186, 2009.

[148] K. Friston and G. Buzsáki. The Functional Anatomy of Time: What and When in the Brain. Trends in Cognitive Sciences, 20(7):500, 2016.

[149] R.W. Keyes. The wire-limited logic chip. IEEE J. Sol.-Sta. Circuits, SC-17:1232, 1982.

[150] Johannes Schemmel, Daniel Brüderle, Andreas Grübl, Matthias Hock, Karlheinz Meier, and Sebastian Millner. A wafer-scale neuromorphic hardware system for large-scale neural modeling. In 2010 IEEE International Symposium on Circuits and Systems (ISCAS), pages 1947–1950. IEEE, 2010.

[151] K. Shang, S. Pathak, B. Guan, G. Liu, and S.J.B. Yoo. Low-loss compact multilayer silicon nitride platform for 3D photonic integrated circuits. Opt. Express, 23:21334, 2015.

[152] W.D. Sacher, Y. Huang, G.-Q. Lo, and J.K.S. Poon. Multilayer silicon nitride-on-silicon integrated photonic platforms and devices. J. Lightwave Tech., 33:901, 2015.

[153] J. Chiles, S. Buckley, N. Nader, S.W. Nam, R.P. Mirin, and J.M. Shainline. Multi-planar amorphous silicon photonics with compact interplanar couplers, cross talk mitigation, and low crossing loss. APL Photonics, 2:116101, 2017.

[154] Y. Zhang, Y. Ling, Y. Zhang, K. Shang, and S.J.B. Yoo. High-density wafer-scale 3-d silicon-photonic integrated circuits. IEEE J. Sel. Topics Quantum Electron., 24:8200510, 2018.

[155] A.L. Rosenberg. Three-Dimensional VLSI: A Case Study. J. Assoc. Computing Machinery, 30(3):397, 1983.

[156] J U Knickerbocker, P S Andry, B Dang, R R Horton, M J Interrante, C S Patel, R J Polastre, K Sakuma, R Sirdeshmukh, E J Sprogis, A M Stephens, A W Topol, C K Tsang, B C Webb, and S L Wright. Three- dimensional silicon integration. IBM J. Res. & Dev., 52(6):553–569, 2008.

[157] K. Sakuma, P.S. Andry, C.K. Tsang, S.L. Wright, B. Dang, C.S. Patel, B.C. Webb, J. Maria, E.J. Sprogis, S.K. Kang, R.J. Polastre, R.R. Horton, and J.U. Knickerbocker. 3D chip-stacking technology with through-silicon vias and low-volume lead-free interconnections. IBM J. Res. & Dev., 52(6):611, 2008.

[158] M. Vinet et al. 3D monolithic integration: Technological challenges and electrical results. Microelectronic Engineering, 88:331, 2011.

[159] J. Zhao, Y. Xie, and Q. Zou. Overview of 3-D Architecture Design Opportunities and Techniques. IEEE Design & Test, page 60, 2015.

[160] S.K. Lim. Design for High Performance, Low Power, and Reliable 3D Integrated Circuits. Springer, 2013.

[161] 3D Stacked Chips. Springer, 2016.

[162] M. Li, J. Shi, M. Rahman, S. Khasanvis, S. Bhat, and C.A. Moritz. Skybridge-3D-CMOS: A Fine-Grained 3D CMOS Integrated Circuit Technology. IEEE Trans. Nanotech., 16(4):639, 2017.

[163] Arianna Rubino, Melika Payvand, and Giacomo Indiveri. 2019 26th IEEE International Conference on Electronics, Circuits and Systems, ICECS 2019. 2019 26th IEEE International Conference on Electronics, Circuits and Systems, ICECS 2019, pages 458–461, 2019.

[164] S.K. Tolpygo, V. Bolkhovsky, R. Rastogi, S. Zarr, A.L. Day, E. Golden, T.J. Weir, A. Wynn, and L.M. Johnson. Planarized Fabrication Process With Two Layers of SIS Josephson Junctions and Integration of SIS and SFS π-Junctions. IEEE Transactions on Applied Superconductivity, 29(5):1101208, 2019.

[165] T. Ando, S. Nagasawa, N. Takeuchi, N. Tsuji, F. China, M. Hidaka, Y. Yamanashi, and N. Yoshikawa. Three-dimensional adiabatic quantum-flux-parametron fabricated using a double-active-layered niobium process. Supercond. Sci. Technol., 30:075003, 2017.

[166] Jeffrey M Shainline. Optoelectronic intelligence. arXiv preprint arXiv:2010.08690, 2020.

[167] V.B. Mountcastle. An Organizing Principle for Cerebral Function: The Unit Module and the Distributed System. The MIT Press, 1978.

[168] A Alekseev. Basics of low-temperature refrigeration. arXiv preprint arXiv:1501.07392, 2015.

[169] D Scott Holmes, Andrew L Ripple, and Marc A Manheimer. Energy-efficient superconducting computing—power budgets and requirements. IEEE Transactions on Applied Superconductivity, 23(3):1701610–1701610, 2013.

[170] Victor V Zhirnov, Ralph K Cavin, James A Hutchby, and George I Bourianoff. Limits to binary logic switch scaling-a gedanken model. Proceedings of the IEEE, 91(11):1934–1939, 2003.

[171] Bernabé Linares-Barranco and Teresa Serrano-Gotarredona. On the design and characterization of femtoampere current-mode circuits. ieee journal of solid-state circuits, 38(8):1353–1363, 2003.

[172] Yung-Hsien Wu, Bo-Yu Chen, Lun-Lun Chen, Jia-Rong Wu, and Min-Lin Wu. Metal-insulator-metal capacitor with high capacitance density and low leakage current using zrtio 4 film. Applied Physics Letters, 95(11):113502, 2009.

[173] Richard Allen Matula. Electrical resistivity of copper, gold, palladium, and silver. Journal of Physical and Chemical Reference Data, 8(4):1147–1298, 1979.

[174] J. Clarke and A.I. Braginski, editors. The SQUID handbook. Wiley-VCH.

[175] J Jaycox and M Ketchen. Planar coupling scheme for ultra low noise dc squids. IEEE Transactions on Magnetics, 17(1):400–403, 1981.