Making BREAD: Biomimetic Strategies for Artificial Intelligence Now and in the Future

Jeffrey L. Krichmar[1]*, William Severa[2], Muhammad S. Khan[3] and James L. Olds[3]*

[1] Departments of Cognitive Sciences and Computer Science, University of California, Irvine, Irvine, CA, United States, [2] Sandia National Laboratories, Data-Driven and Neural Computing, Albuquerque, NM, United States, [3] Schar School, George Mason University, Arlington, VA, United States

The Artificial Intelligence (AI) revolution foretold of during the 1960s is well underway in the second decade of the twenty first century. Its period of phenomenal growth likely lies ahead. AI-operated machines and technologies will extend the reach of Homo sapiens far beyond the biological constraints imposed by evolution: outwards further into deep space, as well as inwards into the nano-world of DNA sequences and relevant medical applications. And yet, we believe, there are crucial lessons that biology can offer that will enable a prosperous future for AI. For machines in general, and for AI's especially, operating over extended periods or in extreme environments will require energy usage orders of magnitudes more efficient than exists today. In many operational environments, energy sources will be constrained. The AI's design and function may be dependent upon the type of energy source, as well as its availability and accessibility. Any plans for AI devices operating in a challenging environment must begin with the question of how they are powered, where fuel is located, how energy is stored and made available to the machine, and how long the machine can operate on specific energy units. While one of the key advantages of AI use is to reduce the dimensionality of a complex problem, the fact remains that some energy is required for functionality. Hence, the materials and technologies that provide the needed energy represent a critical challenge toward future use scenarios of AI and should be integrated into their design. Here we look to the brain and other aspects of biology as inspiration for Biomimetic Research for Energy-efficient AI Designs (BREAD).

Keywords: AI, biomimetic, energy, edge computing, neurobiology, neuromorphic computing

ARTIFICIAL INTELLIGENCE'S ENERGY REQUIREMENTS

The last few years have seen a rapid expansion of Artificial Intelligence (AI) and Machine Learning (ML) breakthroughs. What were once AI solutions to toy problems have now become human level complex problem-solving. These solutions have moved out of research labs and into commercial applications. However, most AI and ML algorithms for these complex problems are implemented in large data centers housing power hungry clusters of computers and Graphical Processing Units (GPUs). In contrast, natural, biological intelligence is power efficient and self-sufficient. In this article, we argue for Biomimetic Research for Energy-efficient AI Designs (BREAD) as AI moves toward edge computing in remote environments far away from conventional energy sources, and as energy consumption becomes increasingly expensive.

Current Solutions to AI's Energy Requirements

With the growth of the Internet, data traffic (traffic to and from data centers) is escalating exponentially, crossing a zettabyte (1.1 ZB) in 2017 (Andrae and Edler, 2015). **Figure 1** shows this trend. Currently, data centers consume an assessed 200 terawatt hours (TWh) each year equivalent to 1% of global electricity demand (Jones, 2018). A 2017 International Energy Agency (IEA) report noted that with the ongoing explosion of Internet traffic in data centers, electricity demand will likely to increase by 3% (IEA, 2017). While it is difficult to estimate the exact role of AI within data centers, analysis from reports (Andrae and Edler, 2015; Sverdlik, 2016; IEA, 2017; Gagan, 2018; CBInsights, 2019) suggests that it is non-trivial— on the order of 40% (see **Figure 1**). For example, Google projected in 2013 that people searching by voice for three minutes a day using speech recognition deep neural networks would double their datacenters' computation demands, and this was one impetus for developing the Google TPU (Jouppi et al., 2017). Additionally, Facebook has stated that machine learning is "applied pervasively across nearly all sevices" and that "computational requirements are also intense" (Hazelwood et al., 2018).

The rise of highly efficient data factories, known as hyperscale facilities, use an organized uniform computing architecture that scales up to hundreds of thousands of servers. While these hyperscale centers can be optimized for high computing efficiency, there are limits to growth due to a variety of constraints that also affect other electrical grid consumers. However, the shift to hyperscale facilities is a current trend, and if 80% of servers in US conventional data centers were moved over to hyperscale facilities, energy consumption would drop by 25%, according to Lawrence Berkeley National Laboratory report, 2016.

One way the hyperscale centers have cut down their power usage is through efficiencies in cooling. By locating in cooler climates, the data centers can ingest the cool air outside with positive results. Another solution is employing warm water cooling loops, a solution tuned for temperate and warm climates. An innovative solution to address the energy constraints of AI systems is to employ an AI-powered cloud-based control recommendation system. For example, Google employs a cloud-based AI to collect information about the data center cooling system from thousands of physical sensors prior to feeding this information into deep neural networks. The networks then compute predictions for how different combinations of possible activities will affect future energy consumption (Shehabi et al., 2016).

Although hyperscale centers and smart cooling strategies can lower energy consumption, these solutions do not address applications where AI is operating at the edge or when AI is deployed in extreme conditions far away from convenient power supplies. We believe that this is where future AI systems are headed.

Our view is that there is a pressing need to address the energy issue as it applies to the future of AI and ML. While there is a growing research effort toward developing efficient machine learning methods for embedded and neuromorphic processors (Esser et al., 2015; Hunsberger and Eliasmith, 2016; Rastegari et al., 2016; Howard et al., 2017; Yang et al., 2017; Severa et al., 2019), we recognize that these methods do not address the full needs of future applications, despite offering compelling first steps. Generally, current methods modify existing techniques rather than develop de novo algorithms.

In this paper, we emphasize how biology has addressed the power consumption problem, with a particular focus on energy efficiency in the brain. Furthermore, we look at non-neural aspects of biology that also lead to power savings. We suggest that these strategies from biology can be realized in future AI systems.
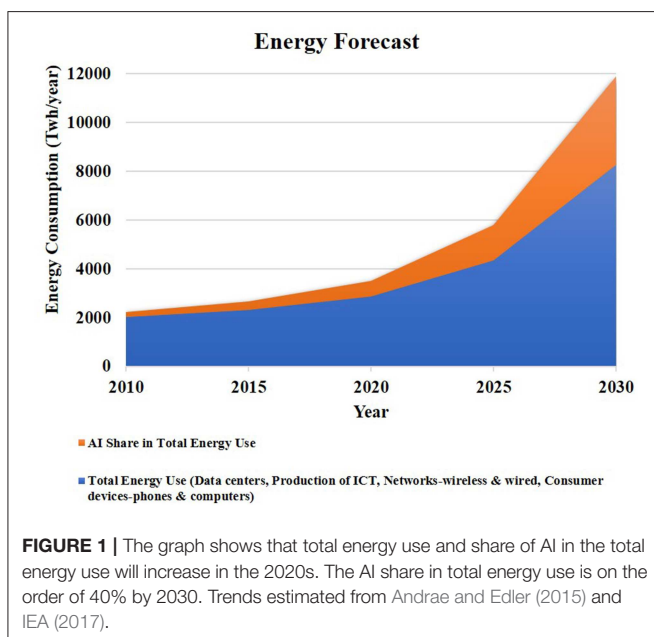
CURRENT STATE OF AI AS IT PERTAINS TO ENERGY CONSUMPTION

Edge Computing in Remote and Hostile Environments

Trends in many human-built systems point to directions where sensing, processing, and actuation is situated on distributed platforms. The emerging Internet-of-Things (IoT) are cyber technologies (Atzori et al., 2010), hardware and software, that interact with physical components in environments which may or may not be populated by humans. IoT devices are often thought of as the "edge" of a large, sophisticated cloud processing infrastructure. Processing data at the "edge," reduces system latency by removing the delays in the aggregation tiers of the information technology infrastructure (Hu et al., 2015; Mao et al., 2016; Shi et al., 2016). In addition to minimizing latency, edge processing increases system security and mitigates privacy concerns when processing data in the cloud. In cases where the data path between the edge and user is very long, such computation can, by feature extraction, reduce the dimensionality and hence the expense of sending

FIGURE 1 | The graph shows that total energy use and share of AI in the total energy use will increase in the 2020s. The AI share in total energy use is on the order of 40% by 2030. Trends estimated from Andrae and Edler (2015) and IEA (2017).

information. However, edge processing may be far away from power sources and may need to operate without intervention over long time periods.

Exploration of remote and hostile environments, such as space and deep ocean, will most likely require AI and ML solutions. These environments are inherently hostile to the circuitry that sub-serves current AI and ML technologies.

Unless human beings can be "radiation-hardened," robotic space probes will continue to dominate exploration and exploitation of space in domains ranging from low earth orbit to interstellar exploration. All of these are subject to a variety of hazards which are potentially hazardous to CMOS-based AI. These include collisions with high energy photons (such as gamma rays), micrometeorites, planetary weather, and anthropogenic attacks. Planetary missions such as NASA's Curiosity Mars rover have revealed additional challenges from weather (such as sandstorms) that have put missions at risk. Radioisotope thermal generation (RTG) power was added to the Curiosity Mars rover design to combat the vulnerabilities to solar energy systems on previous missions. While Curiosity's computational systems do not constitute true AI, the power demands of the entire Curiosity rover (including drills and actuators) are of a similar order of magnitude. Recent concerns over limitations on the availability of radioisotopes (Aebersold, 1949) combined with safety concerns (Staff, 1948; Al Kattar et al., 2015) during the launch phase will constrain future deep space missions that might use nuclear power (Billings, 2015; Grush, 2018; Lakdawalla, 2018; Grossman, 2019).

As with Space, in deep ocean environments, power constraints are also a current challenge. Current non-nuclear powered Autonomous Underwater Vehicles (AUVs) have limited capabilities due to restrictions on energy storage and the availability of fuel sources. Furthermore, the extremely high pressures of deep-sea environments offer their own challenges, not only to energy supply for AI but also to the mass and construction of protection containers for the electronics. Ocean glider AUV's use buoyancy engines with fins to convert force in the vertical direction to horizontal motion (Webb et al., 2001; Schofield et al., 2007; Rudnick, 2016; Rudnick et al., 2016, 2018). While very slow, such AUV's are far less energy-constrained than other current technologies. However, the power generated by such engines is not currently suitable for powering AI systems. Batteries are used for such functions and must be recharged at the ocean surface using photo-voltaic cells. There are proposals to use nuclear fission power generation to enable deep-sea battery recharging stations for military AUV's, though these remain at the development stage and have similar safety considerations to those mentioned above for space (Hambling, 2017).

In many of these domains, AI will be the preferred computational modality because of latency issues related to long-distance communication with Earth-based controllers. Operating in such domains will have the additional challenge of energetic constraints because readily available solar power may not always be available in domains such as Earth's moon, solar system planets with weather and deep space (including interstellar). The primary alternative energy source for such domains is nuclear (both fission and fusion-based). Such power sources are in contrast to the current radioisotope thermo-electric technologies used for missions such as the Mars Curiosity Rover. While break-even fusion power has yet to be demonstrated on Earth, the abundance of fusion fuels in the solar system makes such power sources attractive. In all these cases, the nuclear technology must have a similar resiliency to that of the AI in terms of hazards, and it will be optimal to consider such requirements holistically at the design stage.

Existence Proof, Human Brains as Efficient Energy Consumers

The original goal of AI was to extract principles from human intelligence. On the one hand, these principles would allow for a better understanding of intelligence and what makes us human. On the other hand, we could use those principles to build *intelligent* artifacts, such as robots, computers, and machines. In both cases, the goal is to use human intelligence as a use case, which derives from the function of the brain. We believe that there are also important energy efficiency principles that can be extracted from neurobiology and applied to AI. Therefore, the nervous system can provide much inspiration for the construction of low power intelligent systems.

The human nervous system is under tight metabolic constraints. These constraints include the essential role of glucose as fuel under conditions of non-starvation, the continuous demand for approximately 20% of the human body's total energy utilization, and the lack of any effective energy-reserve among others (Sokoloff, 1960). And yet, as is well known, the brain operates on a mere 20 W of power, approximately the same power required for a ceiling fan operating at low speed. While being severely metabolically constrained is at one level a disadvantage, evolution has optimized brains in ways that lead to incredibly efficient representations of important environmental features that stand distinct from those employed in current digital computers.

The human brain utilizes many means to reduce functional metabolic energy utilization. Indeed, one can observe at every level of the nervous system strategies to maintain high performance and information transfer, while minimizing energy expenditure. These range from ion channel distributions, to coding methods, to wiring diagrams (connectomes). Many of these strategies could inspire new methodologies for constructing power efficient artificial intelligent systems.

At the neuronal coding level, the brain uses several strategies to reduce neural activity without sacrificing performance. Neural activity, (i.e., the generation of an action potential, the return to resting state, and synaptic processing) is energetically very costly, and this can drive the minimization of the number of spikes necessary to encode either an engram or the neural representation of a new stimulus (Levy and Baxter, 1996; Lennie, 2003). Such sparse coding strategies appear to be ubiquitous throughout the brain (Olshausen and Field, 1997, 2004; Beyeler et al., 2017).

Furthermore, dimensionality reduction methods from machine learning can explain many neural representations (Beyeler et al., 2016). Because brains face strict

constraints on metabolic cost (Lennie, 2003) and anatomical bottlenecks (Ganguli and Sompolinsky, 2012), which often force the information stored in a large number of neurons to be compressed into an order of magnitude smaller population of downstream neurons (e.g., storing information from 100 million photoreceptors in 1 million optic nerve fibers), reducing the number of variables required to represent a particular stimulus space figures prominently in efficient current coding theories of brain function (Linsker, 1990; Barlow, 2001; Atick, 2011). Such views posit that the brain performs dimensionality reduction by maximizing mutual information between the high-dimensional input and the low-dimensional output of neuronal populations. Although dimensionality reduction is typically used in machine learning to improve generalization, it may have implications for energy efficiency in real and artificial neural networks. By decreasing the number of neurons required to represent stimuli, while adhering to sparsity constraints, energy savings can be achieved without loss of information.

The brain must respond quickly to stimuli and changes in the environment. However, this implies an increase in neural activity, which would be energetically costly. Evidence suggests that this is not the case and that the brain utilizes strategies to maintain a constant rate of activity. For example, the nervous system can respond quickly to perturbations by shifting the specific timing rather than increasing the absolute number of spikes (Malyshev et al., 2013). Moreover, the balance of excitation and inhibition can further maintain a steady rate of neural activity while still being responsive (Sengupta et al., 2013a; Yu et al., 2018). In these ways, the overall energy utilization of the human brain stays relatively constant, while the local rate of energy consumption varies widely and is dependent upon functional neuronal activity and the balance between excitatory and inhibitory neurons (Olds et al., 1994). In a similar way, neural networks and neuromorphic hardware may reduce the activity of unused or unnecessary nodes, when other nodes are highly active. Thus, keeping the overall power budget constant. For example, the SpiNNaker system will turn off turn off cores when they are not needed (Furber et al., 2013). This strategy may be applied dynamically during operation.

At a macroscopic scale, the brain saves energy by minimizing the wiring between neurons and brain regions (i.e., number of axons), yet still communicates information at a high-level of performance (Laughlin and Sejnowski, 2003). Unlike current electronic chips, the brain packs its wiring into a three-dimensional space, which not only reduces the overall volume but also can reduce the energy cost. Energy is further conserved by maintaining high local connectivity with sparse distal connectivity. White matter, which are myelinated axons that transmit information over long distances in the nervous system, make up about half the human brain but use less energy than gray matter (neuronal somata and dendrites) because of the scarcity of ion channels along these axons (Harris and Attwell, 2012). These myelinated axons speed up signal propagation and reduce the volume of matter in the brain. However, information transfer between neurons and brain areas is still preserved by the overall architecture, which essentially is a small world network (Sporns and Zwi, 2004; Sporns, 2006). That is, even though the probability

of any two distal cortical neurons being connected is extremely low, any two neurons are only a few connections away from each other.

The nervous system also optimizes energy consumption at the cellular and sub-cellular levels. Minimizing wiring has energy implications for both hardware and software. In hardware, routing of information within and between processors can have an impact on energy consumption. In software, the handling of synapses, as is in the brain, takes the most processing power. There are typically many more synaptic events to handle than neuron updates. Minimizing the wiring or number of connections could potential yield energy savings.

It has been suggested that the brain strives to minimize its free energy by reducing surprise and predicting outcomes (Friston, 2010). Thus, the brain's efficient power consumption may have a basis in thermodynamics and information theory. That is, the system may adapt to resist a natural tendency toward disorder in an ever-changing environment. Top-down signals from downstream areas (e.g., frontal cortex or parietal cortex) can realize predictive coding (Clark, 2013; Sengupta et al., 2013a,b). In this way organisms minimize the long-term average of surprise, which is the inverse of entropy, by predicting future outcomes. In essence, they minimize the expenditures required to deal with unanticipated events. The idea of minimizing free energy has close ties to many existing brain theories, such as the Bayesian brain, predictive coding, cell assemblies, and Infomax, as well as an evolutionary-inspired theory called Neural Darwinism or neuronal group selection (Friston, 2010). For field robotics, a predictive controller could allow the robot to reduce unplanned actions (e.g., obstacle avoidance) and produce more efficient behaviors (e.g., optimal foraging or route planning). For IoT and edge processing, predictions could reduce communication data. Rather than sending redundant predictable information, it would only need to "wake up" and report when something unexpected occurs.

In summary, the brain represents an important existence proof that extraordinarily efficient natural intelligence can compute in very complex ways within harsh, dynamic environments. Beyond an existence proof, brains provide an opportunity for reverse-engineering in the context of machine learning methods and neuromorphic computing.

Energy Efficiency Through Brain-Inspired Computing

A key component in pursuing brain- and neural- inspired computing, coding, and neuromorphic algorithms lies in the currently shifting landscape of computing architectures. Moore's law, which has dictated the development of ever-smaller transistors since the 1960s, has become more and more difficult to follow, leading many to claim its demise (Waldrop, 2016). This has inspired renewed interest in heterogeneous and non-Von Neumann computing platforms (Chung et al., 2010; Shalf and Leland, 2015), which take inspiration from the efficiency of the brain's architecture. Neuromorphic architectures can

offer orders-of-magnitude improvement in performance-per-Watt compared to traditional CPUs and GPUs (Indiveri et al., 2011; Hasler and Marr, 2013; Merolla et al., 2014). However, the benefit of neuromorphic computing can and will depend on the application chosen. For modeling biological neural systems, the performance improvements already may be considerable (e.g., the Neurogrid platform claims 5 orders of magnitude efficiency improvement compared to a personal computer Benjamin et al., 2014). Additionally neural approaches enable IBM's TrueNorth chip to power convolutional neural networks for embedded gesture recognition at less than one Watt (Amir et al., 2017). In another comparison, a collection of image (32 × 32 pixel) benchmark tasks on TrueNorth resulted in approximately 6,000 to 10,000 frames/Second/Watt, whereas the Nvidia Jetson TX1 (an embedded GPU platform) can process between 5 and 200 (ImageNet) frames per second at approximately 10–14 Watts net power consumption (Canziani et al., 2016). Although we note that it is difficult to have fair network and dataset parity across platforms, and that neuromorphic systems supporting even millions of neurons may be too small-scale for application-level machine learning problems.

Neuromorphic architectures refer to a wide variety of computing hardware platforms (Schuman et al., 2017), from sensing (Liu and Delbruck, 2010; Posch et al., 2014) to processing (Indiveri et al., 2011; Merolla et al., 2014), analog (Fieres et al., 2008) to digital (Furber et al., 2013; Merolla et al., 2014); see **Figure 2**. However, in most cases the defining characteristics take inspiration from the brain:

1. Massively parallel, simple integrating processing units (neurons)
2. Sparse and dynamic low-precision communication via "spikes."
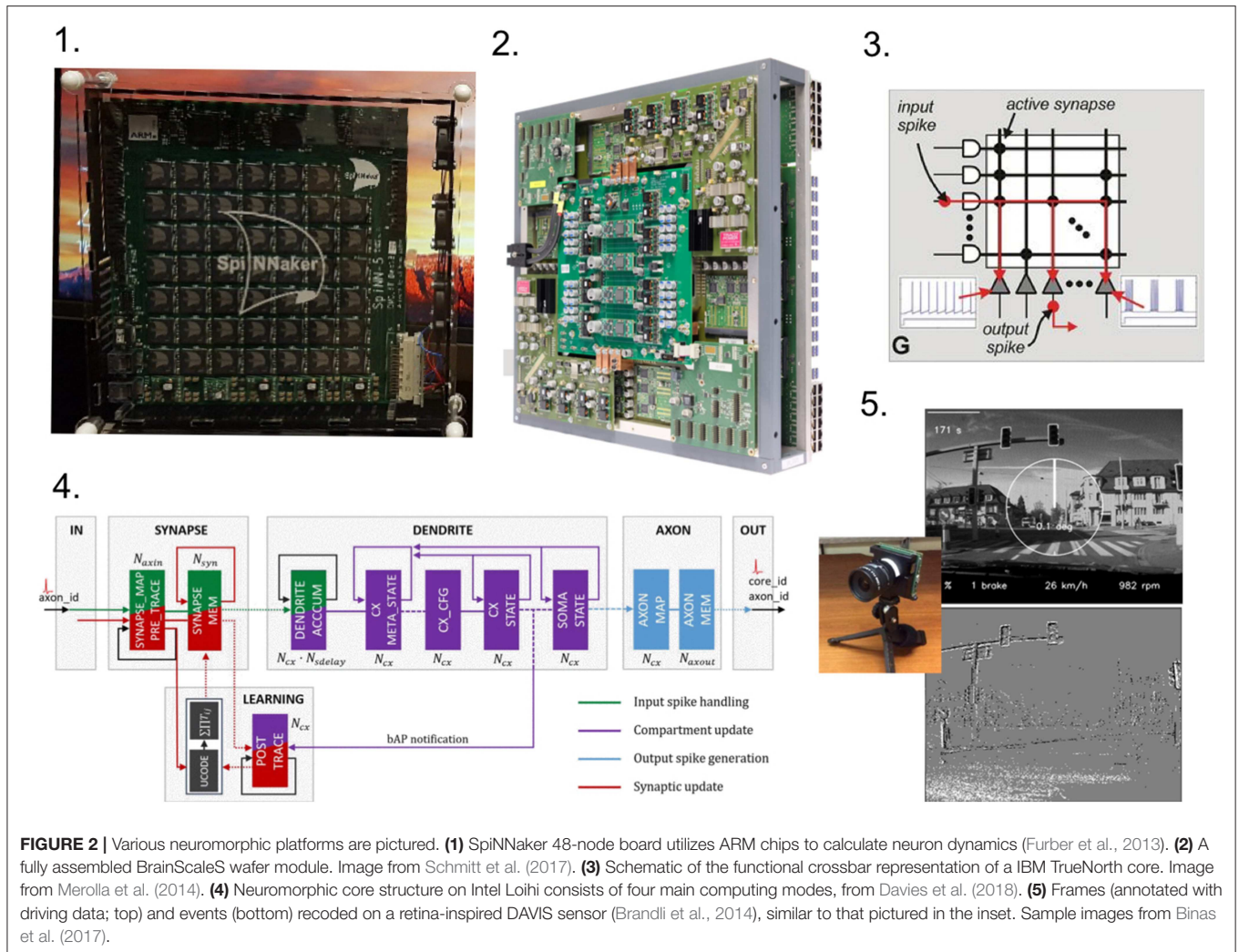3. Event-driven, asynchronous operation.

This event-driven, distributed, processor-in-memory approach provides robust, low-power processing compatible with many neural-inspired machine learning and artificial intelligence applications (Neftci, 2018). Hence, size, weight and power (SWaP) constrained environments, such as edge and IoT devices, can leverage increased effective remote computation capabilities and provide real-time, low-latency intelligent and adaptive behavior. Moreover, the often noisy nature of learned artificial intelligence systems (some incorporate noise by design Srivastava et al., 2014) may lead to more robust computation in extreme environments such as space.

Heterogeneous (spiking and non-spiking) architectures are improving performance and latency, exemplified by a 30–80x improvement on deep learning tasks (Putnam et al., 2014; Jouppi et al., 2017), and new neuromorphic architectures, such as Intel's Loihi, are themselves heterogeneous which improves communication between neural and conventional cores (Davies et al., 2018). Emerging neural computing platforms may benefit traditional large-scale computation both indirectly [e.g., system health (Das et al., 2018), failure prediction (Bouguerra et al., 2013)] and directly (e.g., meshing, surrogate models Melo et al., 2014), and recent work indicates that neuromorphic processors may be useful for direct computation due their

high-communication, highly-parallel nature (Lagorce and Benosman, 2015; Jonke et al., 2016; Aimone et al., 2017; Severa et al., 2018). However, we do remark that currently several challenges exist hindering wide-range adoption of these platforms. Some of the primary difficulties include: (1) Neuromorphic chips are a niche product and difficult to procure at volume; (2) There is insufficient software interfaces for developing applications; (3) Many algorithms are incompatible or may underperform on neuromorphic hardware; (4) Large-scale applications are often too large; (5) Cross-compiling code and I/O require considerable time and bandwidth from a host machine. See Diamond et al. (2016), Severa et al. (2019), Hunsberger and Eliasmith (2016), Disney et al. (2016), Davison et al. (2009), Ehsan et al. (2017), and Wolfe et al. (2018) for more details and possible approaches toward solving these challenges. Moreover, some of the themes from the *Existence proof, human brains as efficient energy consumers* section (e.g., minimizing wiring, keeping firing rates constant, using sparse and reduced representations) could be incorporated into neuromorphic designs.

Neural inspiration has also impacted data collection in the form of spiking neuromorphic sensors which generally follow the same three characteristics as neuromorphic architectures. The two most common categories are silicon cochleas (Watts et al., 1992; Chan et al., 2007) and retina-inspired event-driven cameras (Delbrück et al., 2010; Delbruck et al., 2014), though neuromorphic olfaction is also under active research and development (Vanarse et al., 2016). Neuromorphic sensors can often be thought of as a method for high-speed preprocessing, fundamentally changing the sample space. For example, for imagery this allows for low-bandwidth, high-sampling, and high-dynamic range imagery (Delbrück et al., 2010; Posch et al., 2015). These benefits, in turn, have enabled low-latency, low-power applications such as gesture recognition (Ahn et al., 2011; Amir et al., 2017), robotic control (Conradt et al., 2009; Delbruck and Lang, 2013), and movement determination (Drazen et al., 2011; Haessig et al., 2018). The sparse, spiking representations can pose an algorithmic challenge, at times being incompatible with common processing methods designed around rasterized data. However, spiking sensors are innately compatible with spiking neuromorphic processors, and combining neuromorphic sensors with a neuromorphic processor can avoid the costly conversion between binary data formats and spikes.

Computational requirements of artificial intelligence algorithms limit their remote applications today. Consequently, most current consumer or commercial machine learning technologies are reliant on connections to remote data centers. However, as neuromorphic technologies transition from research platforms to everyday products, learning systems can and will proliferate in capability and scope. Combined with the expected growth of edge and IoT devices, we can expect persistence and pervasive learning devices. These learning devices, extensions of current trends in smart devices (e.g., digital assistants, smart home control, wearables), will be enhanced with personalized online learning and enabled with adaptive, intelligent and context-dependent perception and behaviors. Ultra-low energy neuromorphic chips will carry out computations using milliwatts

FIGURE 2 | Various neuromorphic platforms are pictured. **(1)** SpiNNaker 48-node board utilizes ARM chips to calculate neuron dynamics (Furber et al., 2013). **(2)** A fully assembled BrainScaleS wafer module. Image from Schmitt et al. (2017). **(3)** Schematic of the functional crossbar representation of a IBM TrueNorth core. Image from Merolla et al. (2014). **(4)** Neuromorphic core structure on Intel Loihi consists of four main computing modes, from Davies et al. (2018). **(5)** Frames (annotated with driving data; top) and events (bottom) recoded on a retina-inspired DAVIS sensor (Brandli et al., 2014), similar to that pictured in the inset. Sample images from Binas et al. (2017).

of power. In the industrial, medical and security spaces, the same technologies will provide low-powered sensors capable of extended deployment in a variety of extreme environments.

There is ample precedent for brain-inspired approaches to engineering design that may lead to energy-efficient AI and edge computing. Both designing algorithms to mimic the brain's behavior, and building new computer hardware that mimic neural dynamics can lead to energy efficiency (see **Figure 3**) (Calimera et al., 2013). However, many of the brain's energy efficiency strategies, such as minimizing wiring, maintaining constant activity and prediction outcomes, are not implemented in current neuromorphic architectures and should be explored in the future.

OTHER ENERGY EFFICIENT STRATEGIES IN BIOLOGY

Energy efficiency can also be inspired by observing nature's non-neural solutions. For example, the wing of an aircraft takes inspiration from the wings of flying animals (birds, bats, and insects). The shape of a modern naval submarine has evolved from early boat-like designs prevalent during the First and Second World Wars toward a more streamlined whale-like shape. Even DNA-based computation–by itself incredibly energetically efficient–takes inspiration from the conserved phylogenetic information transfer mechanism of Earth's biosphere. The adaptive immune system, with its sophisticated "learning and memory" through selection also represents a low-energy approach to artificial intelligence that may eventually have applications to AI-enhanced cyber-security applications (Forrest, 1993; Somayaji et al., 1998; Forbes, 2004; Forrest and Beauchemin, 2007; Rice and Martin, 2007; Keller, 2017). This selectionist approach, which was inspired by the immune system, led to an influential brain theory where the synaptic selection took place during neural development and through experiential synaptic plasticity (Edelman, 1987, 1993). Such a Darwinist approach can lead to efficient neural network structures.

As edge computing and mobile sensing devices become ubiquitous, efficient mobility, whether on land, air, or water will become increasingly important. **Figure 4** shows examples
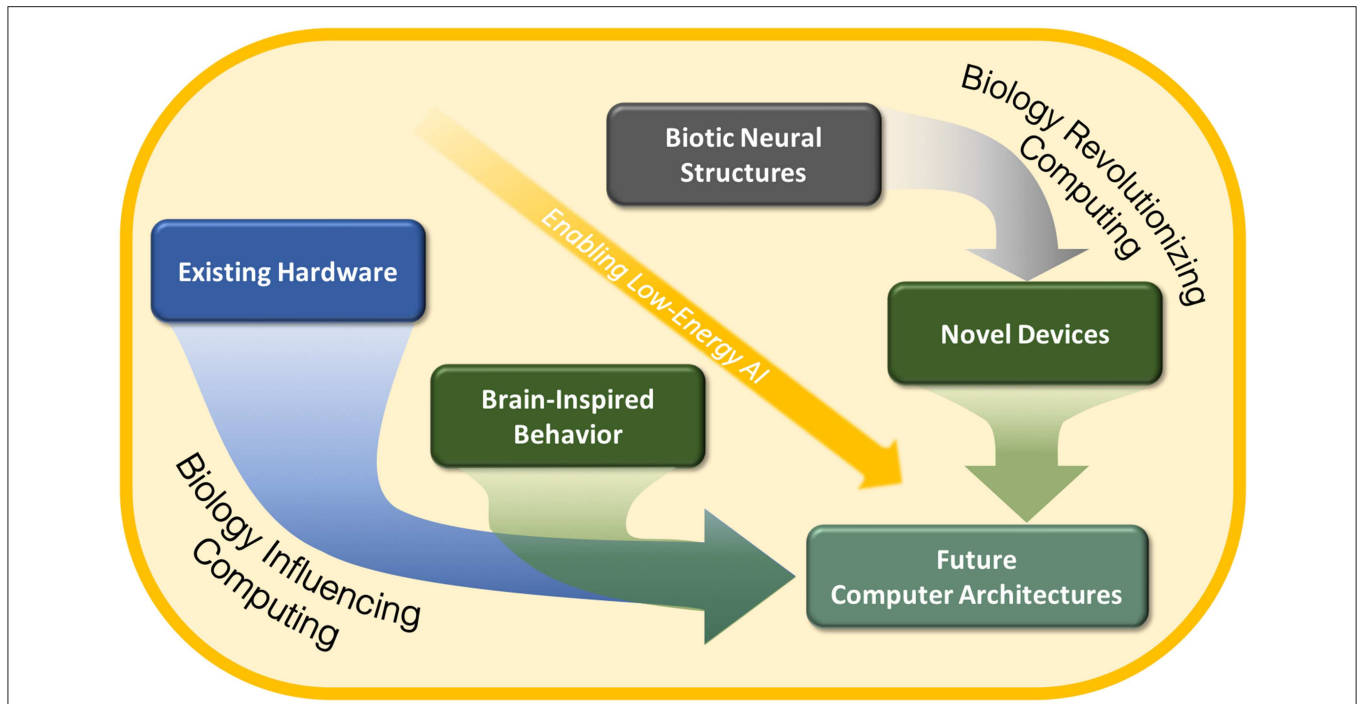
FIGURE 3 | The image shows that in the design of bio-mimetic circuitry, either the existing hardware will be slightly adjusted to copy the behavior of brain parts, or new computing architecture will be designed so that it completely emulate the high energy-efficient biotic neural structures (Image adapted from Calimera et al., 2013).
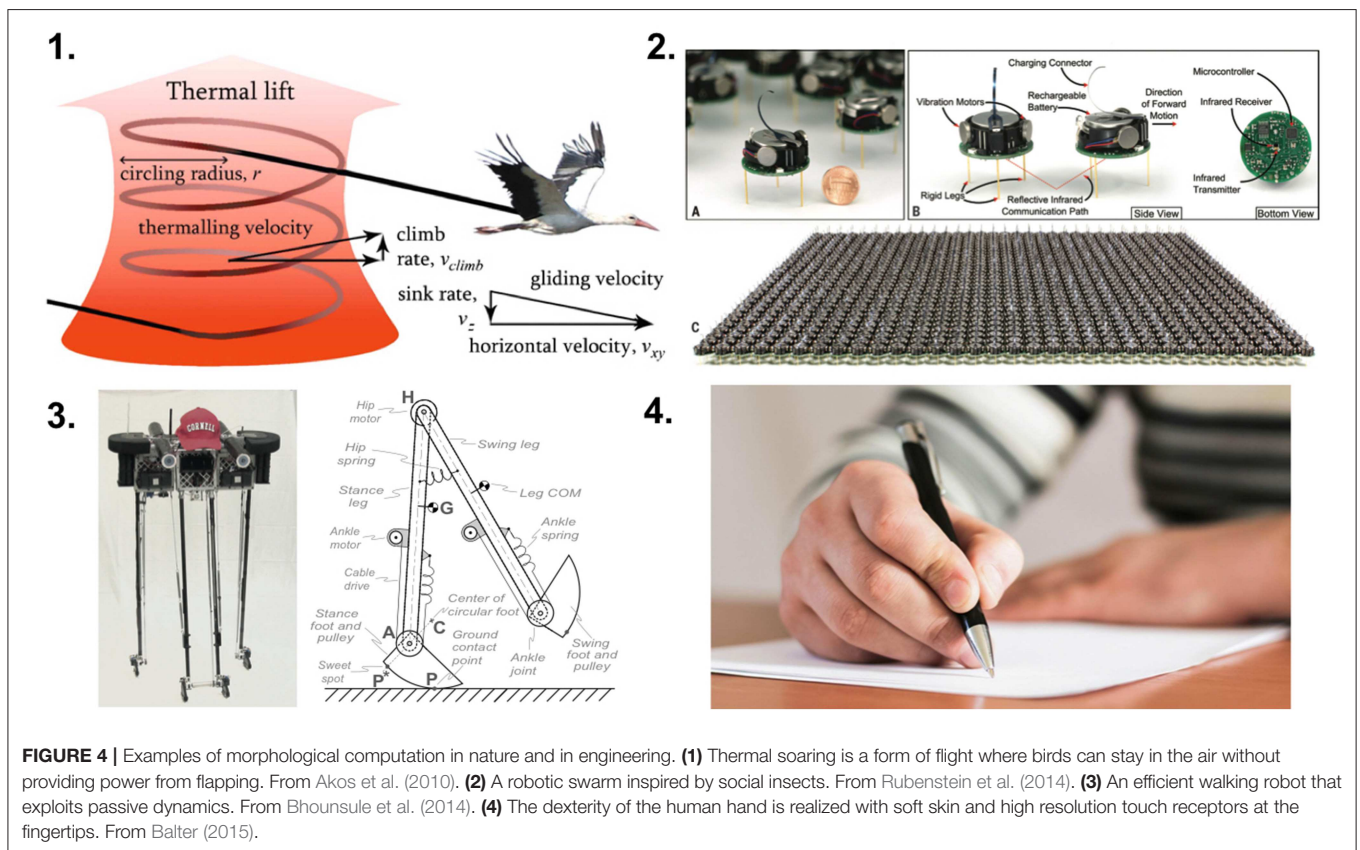

FIGURE 4 | Examples of morphological computation in nature and in engineering. **(1)** Thermal soaring is a form of flight where birds can stay in the air without providing power from flapping. From Akos et al. (2010). **(2)** A robotic swarm inspired by social insects. From Rubenstein et al. (2014). **(3)** An efficient walking robot that exploits passive dynamics. From Bhounsule et al. (2014). **(4)** The dexterity of the human hand is realized with soft skin and high resolution touch receptors at the fingertips. From Balter (2015).

of how biological organisms have evolved to leverage their environment, and this morphological computation can lead to efficient movement and information processing (Pfeifer et al., 2014). For example, swarm intelligence (see **Figure 4**, panel 2), which is inspired by social insects, can solve a number of problems with a collection of low power simple agents. Interesting solutions emerge through the agents' interactions Rubenstein et al. (2014).

Bipedal walking is somewhat of a controlled fall, where energy is conserved by allowing gravity to take over after the swing phase of a step (see **Figure 4**, panel 3). This strategy has been adopted in passive walker robots that utilize orders of magnitude less energy than conventional walking robots (Collins et al., 2005; Bhounsule et al., 2014). Birds of prey and long-range migrating birds take advantage of thermal plumes to reduce energy usage during flight (see **Figure 4**, panel 1) (Akos et al., 2010; Weimerskirch et al., 2016; Bousquet et al., 2017). Gliders have mimicked this strategy in their flight control systems (Allen and Lin, 2007; Edwards, 2008; Reddy et al., 2018). Similar to many fish and marine mammals, oceanographic submersible gliders can harvest energy from the heat flow of thermal gradients (Webb et al., 2001; Schofield et al., 2007; Rudnick, 2016; Rudnick et al., 2016, 2018). These submersible gliders can operate across thousands of kilometers over months to years. Some fish species and flying insects alter their environment (i.e., the water or air vortices) to create additional thrust (Triantafyllou et al., 2002; Sane, 2003). The human hand is a marvel of morphological computing. The shape of the hand naturally and reflexively grasps onto object. The first point of contact is high resolution sensors made of compliant material (i.e., the fingertips). Such structures greatly reduce the neural computing load for complex tasks (see **Figure 4**, panel 4).

Inspiration from biology at the population scale can lead to efficient solutions to problems. Social insects and bacterial colonies have inspired highly distributed robots or computing systems (Rubenstein et al., 2014; Werfel et al., 2014). In these cases, each agent has very low power computation requirements, and no single agent is a point of failure. However, the interactions between these agents can lead to complex problem-solving, which is sometimes referred to as swarm intelligence.

Taken together, future AI systems that take inspiration from biology and other energy harvesting approaches will have a distinct advantage for long-term operation in harsh or remote environments. Following these biological strategies could allow for energy efficient sensor networks at the edge, more efficient manufacturing, and systems that operate over much longer timescales.

CONCLUSIONS

AI is on a trajectory to fundamentally change society in much the same way that the industrial revolution did. Even without the development of General Artificial Intelligence, the trend is toward human-machine partnerships that collectively will have the ability to substantially extend the reach of humans in multiple domains (e.g., space, cyber, deep sea, nano). However, as with many things, there is no free lunch: AI will require energy inputs that we believe must be accounted for at all stages of the AI design process. We believe that such design solutions should leverage the solutions that biology, especially the human brain, has evolved to be energy efficient without sacrificing functionality. These solutions are critical components to what we call *intelligence.*

AI has the potential to change society drastically; the evolving human-machine partnerships will substantially extend the reach of humans in multiple domains. For this to happen, AI will require energy inputs that must be an early component of future integrated AI design processes. A coherent strategy for design solutions should leverage the solutions that biology, especially biological brains, offers in maintaining energy efficiency and preserving functionality. This strategy advances the following recommendations to ensure both private and government support for research and innovation.

Recommendation 1: A Multinational Initiative to Make BREAD

To coordinate investments and channel knowledge from the life sciences to AI energetics into a holistic AI design, we advocate the launch of a global technological innovation initiative, which we call Biomimetic Research for Energy-efficient, AI Designs (BREAD). Ideally, BREAD would be backed by professional societies such as the Association for Computing Machinery (ACM), American Psychological Association (APA), Institute of Electrical and Electronics Engineers (IEEE), and Society for Neuroscience, as well as federal agencies. Energy sustainability would be central in BREAD, but the initiative would encompass all aspects of AI from hardware to sensors.

Recommendation 2: Integrate Biomimetic Energetic Solutions Into Future AI Designs

Future AI development will require an integrated design process where energy supply is not an add-on or assumed. Thermodynamic considerations alone make the energetic considerations important, particularly for those at the Edge, such as IoT and Space environments, which are inherently hostile to CMOS or future successor chip technologies. Current AI, such as deep learning, approaches the problem by situating data centers close to abundant and cheap electric power sources much like what Google, Facebook, and IBM do. We believe that a more fruitful approach for AI design is to leverage the solutions evolved by biology (nervous system, metabolism, morphology) in the future AI design, to what we call 'biomimetic strategies.'

Recommendation 3: An Industry Backed Research Lab or Consortium

Since the initial costs of integrating biomimetic solutions into AI are likely to be front-loaded, we recommend that stakeholder industrial partners with governments establish a pre-competitive research laboratory that preserves intellectual property, similar to IMEC in Belgium. IMEC was created so that CMOS design firms might prototype new chips in a state-of-the-art environment. Alternatively, an industry backed research and development of technology, similar to the Semiconductor Research Consortium

(SRC), would be another model to move forward on energy efficient bio-inspired AI solutions. Such models could catalyze the technological innovations necessary for success.

Recommendation 4: A Trainee Pipeline

A critical component of BREAD would be to establish a pool of scientists at the intersection of AI and energy issues who can integrate knowledge from biology, computer science, neuroscience, and engineering. Thus, aligned with BREAD, research institutions should consider new graduate offerings at this nexus. In the US, the Engineering Directorate of the NSF, the DOE or the DOD might support doctoral candidates and post-doctoral trainees in this area through fellowships and scholarships.

In conclusion, we see the future development of AI as requiring new strategies for embedding energy demands of the machine into the overall design strategy. From our standpoint, this must include biomimetic solutions. As indicated above, there is much precedence for this type of engineering in other high aspects of technology, especially those that must operate in challenging environments. Now such engineering must be applied to future AI design so that the technological trajectory of this paradigm-changing technology is secure.

REFERENCES

Aebersold, P. C. (1949). Production and availability of radioisotopes. *J. Clin. Investigat.* 28, 1247–1254. doi: 10.1172/JCI102192

Ahn, E. Y., Lee, J. H., Mullen, T., and Yen, J. (2011). "Dynamic vision sensor camera based bare hand gesture recognition," in *Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 2011 IEEE Symposium on* (Kerkyra: IEEE), 52–59.

Aimone, J. B., Parekh, O., and Severa, W. (2017). "Neural computing for scientific computing applications: more than just machine learning," in *NCS* (Knoxville).

Akos, Z., Nagy, M., Leven, S., and Vicsek, T. (2010). Thermal soaring flight of birds and unmanned aerial vehicles. *Bioinspirat. Biomimet.* 5:045003. doi: 10.1088/1748-3182/5/4/045003

Al Kattar, Z., El Balaa, H., Nassredine, M., and Haydar, P. M. (2015). "Radiation safety issues relevant to radioisotope production medical cyclotron," in *Advances in Biomedical Engineering (ICABME), International Conference on 2015* (Beirut: IEEE), 178–181.

Allen, M., and Lin, V. (2007). "Guidance and control of an autonomous soaring vehicle with flight test results," in *45th AIAA Aerospace Sciences Meeting and Exhibit, Aerospace Sciences Meetings* (Reno, NV: American Institute of Aeronautics and Astronautics), 1–22.

Amir, A., Taba, B., Berg, D. J., Melano, T., McKinstry, J. L., Di Nolfo, C., et al. (2017). "A low power, fully event-based gesture recognition system," in *CVPR* (Honolulu, HI), 7388–7397.

Andrae, S. A., and Edler, T. (2015). On global electricity usage of communication technology: trends to 2030. *Challenges.* 6, 117–157. doi: 10.3390/challe6010117

Atick, J. J. (2011). Could information theory provide an ecological theory of sensory processing? *Network* 22, 4–44. doi: 10.3109/0954898X.2011.638888

Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: a survey. *Comput. Netw.* 54, 2787–2805. doi: 10.1016/j.comnet.2010.05.010

Balter, M. (2015). Humans have more primitive hands than chimpanzees. *Science.* doi: 10.1126/science.aac8845

Barlow, H. (2001). Redundancy reduction revisited. *Network* 12, 241–253. doi: 10.1080/net.12.3.241.253

Benjamin, B. V., Gao, P., McQuinn, E., Choudhary, S., Chandrasekaran, A. R., Bussat, J.-M., et al. (2014). Neurogrid: a mixed-analog-digital multichip system for large-scale neural simulations. *Proc. IEEE* 102, 699–716. doi: 10.1109/JPROC.2014.2313565

Beyeler, M., Dutt, N., and Krichmar, J. L. (2016). 3d visual response properties of mstd emerge from an efficient, sparse population code. *J. Neurosci.* 36, 8399–8415. doi: 10.1523/JNEUROSCI.0396-16.2016

Beyeler, M., Rounds, E., Carlson, K., Dutt, N., and Krichmar, J. L. (2017). Sparse coding and dimensionality reduction in cortex. *bioRxiv.149880.* doi: 10.1101/149880

Bhounsule, P. A., Cortell, J., Grewal, A., Hendriksen, B., Karssen, J. G. D., Paul, C., et al. (2014). Low-bandwidth reflex-based control for lower power walking: 65 km on a single battery charge. *Int. J. Robot. Res.* 33, 1305–1321. doi: 10.1177/0278364914527485

Billings, L. (2015). Nasa struggles over deep-space plutonium power. *Sci. Am.* 313.

Binas, J., Neil, D., Liu, S.-C., and Delbruck, T. (2017). Ddd17: end-to-end davis driving dataset. *arXiv:1711.01458.* doi: 10.5167/uzh-149345

Bouguerra, M. S., Gainaru, A., and Cappello, F. (2013). "Failure prediction: what to do with unpredicted failures," in *28th IEEE International Parallel and Distributed Processing Symposium*, Vol. 2 (Phoenix, AZ).

Bousquet, G. D., Triantafyllou, M. S., and Slotine, J. E. (2017). Optimal dynamic soaring consists of successive shallow arcs. *J. R. Soc. Interface* 14:135. doi: 10.1098/rsif.2017.0496

Brandli, C., Muller, L., and Delbruck, T. (2014). "Real-time, high-speed video decompression using a frame-and event-based davis sensor," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)* (Melbourne, VIC: IEEE), 686–689.

Calimera, A., Macii, E., and Poncino, M. (2013). The human brain project and neuromorphic computing. *Funct. Neurol.* 28, 191–196.

Canziani, A., Paszke, A., and Culurciello, E. (2016). An analysis of deep neural network models for practical applications. *arXiv:1605.07678.*

CBInsights (2019). *The Future of Data Centers*. Research Brief, CBInsights.

Chan, V., Liu, S.-C., and van Schaik, A. (2007). Aer ear: a matched silicon cochlea pair with address event representation interface. *IEEE Trans. Circ. Syst. I* 54, 48–59. doi: 10.1109/TCSI.2006.887979

Chung, E. S., Milder, P. A., Hoe, J. C., and Mai, K. (2010). "Single-chip heterogeneous computing: does the future include custom logic, fpgas, and gpgpus?," in *Microarchitecture (MICRO), 2010 43rd Annual IEEE/ACM International Symposium* (Atlanta, GA: IEEE), 225–236.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Collins, S., Ruina, A., Tedrake, R., and Wisse, M. (2005). Efficient bipedal robots based on passive-dynamic walkers. *Science* 307, 1082–1085. doi: 10.1126/science.1107799

Conradt, J., Berner, R., Cook, M., and Delbruck, T. (2009). "An embedded aer dynamic vision sensor for low-latency pole balancing," in *Computer Vision Workshops (ICCV Workshops), IEEE 12th International Conference on 2009* (Kyoto: IEEE), 780–785.

Das, A., Mueller, F., Siegel, C., and Vishnu, A. (2018). "Desh: deep learning for system health prediction of lead times to failure in hpc," in *Proceedings of the 27th International Symposium on High-Performance Parallel and Distributed Computing* (Tempe, AZ: ACM), 40–51.

Davies, M., Srinivasa, N., Lin, T.-H., Chinya, G., Cao, Y., Choday, S. H., et al. (2018). Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro* 38, 82–99. doi: 10.1109/MM.2018.112130359

Davison, A. P., Brüderle, D., Eppler, J. M., Kremkow, J., Muller, E., Pecevski, D., et al. (2009). Pynn: a common interface for neuronal network simulators. *Front. Neuroinformatics* 2:11. doi: 10.3389/neuro.11.011.2008

Delbruck, T., and Lang, M. (2013). Robotic goalie with 3 ms reaction time at 4% cpu load using event-based dynamic vision sensor. *Front. Neurosci.* 7:223. doi: 10.3389/fnins.2013.00223

Delbrück, T., Linares-Barranco, B., Culurciello, E., and Posch, C. (2010). "Activity-driven, event-based vision sensors," in *Circuits and Systems (ISCAS), Proceedings of IEEE International Symposium on 2010* (Paris: IEEE), 2426–2429.

Delbruck, T., Villanueva, V., and Longinotti, L. (2014). "Integration of dynamic vision sensor with inertial measurement unit for electronically stabilized event-based vision," in *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on* (IEEE), 2636–2639.

Diamond, A., Nowotny, T., and Schmuker, M. (2016). Comparing neuromorphic solutions in action: implementing a bio-inspired solution to a benchmark classification task on three parallel-computing platforms. *Front. Neurosci.* 9:491. doi: 10.3389/fnins.2015.00491

Disney, A., Reynolds, J., Schuman, C. D., Klibisz, A., Young, A., and Plank, J. S. (2016). Danna: a neuromorphic software ecosystem. *Biol. Inspir. Cogn. Architect.* 17, 49–56. doi: 10.1016/j.bica.2016.07.007

Drazen, D., Lichtsteiner, P., Häfliger, P., Delbrück, T., and Jensen, A. (2011). Toward real-time particle tracking using an event-based dynamic vision sensor. *Exp. Fluids* 51:1465. doi: 10.1007/s00348-011-1207-y

Edelman, G. M. (1987). *Neural Darwinism: The Theory of Neuronal Group Selection.* New York, NY: Basic Books.

Edelman, G. M. (1993). Neural darwinism: selection and reentrant signaling in higher brain function. *Neuron* 10, 115–125. doi: 10.1016/0896-6273(93)90304-A

Edwards, D. (2008). "Implementation details and flight test results of an autonomous soaring controller," in *Guidance, Navigation, and Control and Co-located Conferences. American Institute of Aeronautics and Astronautics.* (Honolulu, HI).

Ehsan, M. A., Zhou, Z., and Yi, Y. (2017). "Neuromorphic 3d integrated circuit: a hybrid, reliable and energy efficient approach for next generation computing," in *Proceedings of the on Great Lakes Symposium on VLSI 2017* (Alberta: ACM), 221–226.

Esser, S. K., Appuswamy, R., Merolla, P., Arthur, J. V., and Modha, D. S. (2015). "Backpropagation for energy-efficient neuromorphic computing," in *Advances in Neural Information Processing Systems* (Montreal, QC), 1117–1125.

Fieres, J., Schemmel, J., and Meier, K. (2008). "Realizing biological spiking network models in a configurable wafer-scale hardware system," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence).* (Hong Kong: IEEE), 969–976.

Forbes, N. (2004). *Imitation of Life: How Biology Is Inspiring Computing.* Cambridge: MIT Press.

Forrest, S. (1993). Genetic algorithms: principles of natural selection applied to computation. *Science* 261, 872–878. doi: 10.1126/science.8346439

Forrest, S., and Beauchemin, C. (2007). Computer immunology. *Immunol. Rev.* 216, 176–197. doi: 10.1111/j.1600-065X.2007.00499.x

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11:127. doi: 10.1038/nrn2787

Furber, S. B., Lester, D. R., Plana, L. A., Garside, J. D., Painkras, E., Temple, S., et al. (2013). Overview of the spinnaker system architecture. *IEEE Trans. Comput.* 62, 2454–2467. doi: 10.1109/TC.2012.142

Gagan, O. (2018). "Here's how AI fits into the future of energy," in *World Economic Forum.*

Ganguli, S., and Sompolinsky, H. (2012). Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu. Rev. Neurosci.* 35, 485–508. doi: 10.1146/annurev-neuro-062111-150410

Grossman, D. (2019). Scientists find a new way to create the plutonium that powers deep space missions. *Popular Mech* (accessed January 08, 2018).

Grush, L. (2018). Ideas for new nasa mission can now include spacecraft powered by plutonium. *The Verge.* (accessed March 19, 2018).

Haessig, G., Cassidy, A., Alvarez, R., Benosman, R., and Orchard, G. (2018). Spiking optical flow for event-based sensors using ibm's truenorth neurosynaptic system. *IEEE Trans. Biomed. Circ. Syst.* 12, 1–11. doi: 10.1109/TBCAS.2018.2834558

Hambling, D. (2017). Why russia is sending robotic submarines to the arctic. *BBC Future.* (accessed November 21, 2017).

Harris, J. J., and Attwell, D. (2012). The energetics of cns white matter. *J. Neurosci.* 32, 356–371. doi: 10.1523/JNEUROSCI.3430-11.2012

Hasler, J., and Marr, H. B. (2013). Finding a roadmap to achieve large neuromorphic hardware systems. *Front. Neurosci.* 7:118. doi: 10.3389/fnins.2013.00118

Hazelwood, K., Bird, S., Brooks, D., Chintala, S., Diril, U., Dzhulgakov, D., et al. (2018). "Applied machine learning at facebook: a datacenter infrastructure perspective," in *IEEE International Symposium on High Performance Computer Architecture (HPCA) 2018* (Vienna: IEEE), 620–629.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv: 1704.04861.*

Hu, Y. C., Patel, M., Sabella, D., Sprecher, N., and Young, V. (2015). *Mobile Edge Computing-a Key Technology Towards 5G.* European Telecommunications Standards Institute.

Hunsberger, E., and Eliasmith, C. (2016). Training spiking deep networks for neuromorphic hardware. *arXiv: 1611.05141.*

IEA (2017). *Digitalization: A New Era in Energy*? Report, International Energy Agency.

Indiveri, G., Linares-Barranco, B., Hamilton, T. J., Van Schaik, A., Etienne-Cummings, R., Delbruck, T., et al. (2011). Neuromorphic silicon neuron circuits. *Front. Neurosci.* 5:73. doi: 10.3389/fnins.2011.00073

Jones, N. (2018). The information factories. *Nature* 561, 163–166. doi: 10.1038/d41586-018-06610-y

Jonke, Z., Habenschuss, S., and Maass, W. (2016). Solving constraint satisfaction problems with networks of spiking neurons. *Front. Neurosci.* 10:118. doi: 10.3389/fnins.2016.00118

Jouppi, N. P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., et al. (2017). "In-datacenter performance analysis of a tensor processing unit," in *Computer Architecture (ISCA), ACM/IEEE 44th Annual International Symposium on 2017* (Toronto, ON: IEEE), 1–12.

Keller, K. L. (2017). Leveraging biologically inspired models for cyber–physical systems analysis. *IEEE Syst. J.* 12, 1–11.

Lagorce, X., and Benosman, R. (2015). Stick: spike time interval computational kernel, a framework for general purpose computation using neurons, precise timing, delays, and synchrony. *Neural Comput.* 27, 2261–2317. doi: 10.1162/NECO_a_00783

Lakdawalla, E. (2018). *The Design and Engineering of Curiosity: How the Mars Rover Performs Its Job.* Springer Praxis Books.

Laughlin, S. B., and Sejnowski, T. J. (2003). Communication in neuronal networks. *Science* 301, 1870–1874. doi: 10.1126/science.1089662

Lennie, P. (2003). The cost of cortical computation. *Curr. Biol.* 13, 493–497. doi: 10.1016/S0960-9822(03)00135-0

Levy, W. B., and Baxter, R. A. (1996). Energy efficient neural codes. *Neural Comput.* 8, 531–543. doi: 10.1162/neco.1996.8.3.531

Linsker, R. (1990). Perceptual neural organization: some approaches based on network models and information theory. *Annu. Rev. Neurosci.* 13, 257–281. doi: 10.1146/annurev.ne.13.030190.001353

Liu, S. C., and Delbruck, T. (2010). Neuromorphic sensory systems. *Curr. Opin. Neurobiol.* 20, 288–295. doi: 10.1016/j.conb.2010.03.007

Malyshev, A., Tchumatchenko, T., Volgushev, S., and Volgushev, M. (2013). Energy-efficient encoding by shifting spikes in neocortical neurons. *Eur. J. Neurosci.* 38, 3181–3188. doi: 10.1111/ejn.12338

Mao, Y. Y., Zhang, J., and Letaief, K. B. (2016). Dynamic computation offloading for mobile-edge computing with energy harvesting devices. *IEEE J. Select. Areas Commun.* 34, 3590–3605. doi: 10.1109/JSAC.2016.2611964

Melo, A., Cóstola, D., Lamberts, R., and Hensen, J. (2014). Development of surrogate models using artificial neural network for building shell energy labelling. *Energy Policy* 69, 457–466. doi: 10.1016/j.enpol.2014.02.001

Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., et al. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* 345, 668–673. doi: 10.1126/science.1254642

Neftci, E. (2018). Data and power efficient intelligence with neuromorphic learning machines. *iScience.* 5, 52–68 doi: 10.1016/j.isci.2018.06.010

Olds, J., Frey, K., and Agranoff, B. (1994). Sequential double-label deoxyglucose autoradiography for determining cerebral metabolic change: origins of variability within a single brain. *Neuroprotocols* 5, 12–24.

Olshausen, B. A., and Field, D. J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by v1? *Vis. Res.* 37, 3311–3325. doi: 10.1016/S0042-6989(97)00169-7

Olshausen, B. A., and Field, D. J. (2004). Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* 14, 481–487. doi: 10.1016/j.conb.2004.07.007

Pfeifer, R., Iida, F., and Lungarella, M. (2014). Cognition from the bottom up: on biological inspiration, body morphology, and soft materials. *Trends Cogn. Sci.* 18, 404–413. doi: 10.1016/j.tics.2014.04.004

Posch, C., Benosman, R., and Etienne-Cummings, R. (2015). Giving machines humanlike eyes. *IEEE Spectrum* 52, 44–49. doi: 10.1109/MSPEC.2015.7335800

Posch, C., Serrano-Gotarredona, T., Linares-Barranco, B., and Delbruck, T. (2014). Retinomorphic event-based vision sensors: bioinspired cameras with spiking output. *Proc. IEEE* 102, 1470–1484. doi: 10.1109/JPROC.2014.2346153

Putnam, A., Caulfield, A. M., Chung, E. S., Chiou, D., Constantinides, K., Demme, J., et al. (2014). A reconfigurable fabric for accelerating large-scale datacenter services. *ACM SIGARCH Comput. Architect. News* 42, 13–24. doi: 10.1145/2678373.2665678

Rastegari, M., Ordonez, V., Redmon, J., and Farhadi, A. (2016). "Xnor-net: imagenet classification using binary convolutional neural networks," in *European Conference on Computer Vision* (Amsterdam: Springer), 525–542.

Reddy, G., Wong-Ng, J., Celani, A., Sejnowski, T. J., and Vergassola, M. (2018). Glider soaring via reinforcement learning in the field. *Nature.* 562, 236–239. doi: 10.1038/s41586-018-0533-0

Rice, J., and Martin, N. (2007). Using biological models to improve innovation systems: the case of computer anti-viral software. *Eur. J. Innovat. Manag.* 10, 201–214. doi: 10.1108/14601060710745251

Rubenstein, M., Cornejo, A., and Nagpal, R. (2014). Robotics. Programmable self-assembly in a thousand-robot swarm. *Science* 345, 795–799. doi: 10.1126/science.1254295

Rudnick, D. L. (2016). Ocean research enabled by underwater gliders. *Annu. Rev. Mar. Sci.* 8, 519–541. doi: 10.1146/annurev-marine-122414-033913

Rudnick, D. L., Davis, R. E., and Sherman, J. T. (2016). Spray underwater glider operations. *J. Atmospher. Ocean. Technol.* 33, 1113–1122. doi: 10.1175/JTECH-D-15-0252.1

Rudnick, D. L., Sherman, J. T., and Wu, A. P. (2018). Depth-average velocity from spray underwater gliders. *J. Atmospher. Ocean. Technol.* 35, 1665–1673. doi: 10.1175/JTECH-D-17-0200.1

Sane, S. P. (2003). The aerodynamics of insect flight. *J. Exp. Biol.* 206:4191. doi: 10.1242/jeb.00663

Schmitt, S., Klähn, J., Bellec, G., Grübl, A., Güttler, M., Hartel, A., et al. (2017). "Neuromorphic hardware in the loop: training a deep spiking network on the brainscales wafer-scale system," in *2017 International Joint Conference on Neural Networks* (Anchorage, AK, IJCNN), 2227–2234.

Schofield, O., Kohut, J., Aragon, D., Creed, L., Graver, J., Haldeman, C., et al. (2007). Slocum gliders: robust and ready. *J. Field Robot.* 24, 473–485. doi: 10.1002/rob.20200

Schuman, C. D., Potok, T. E., Patton, R. M., Birdwell, J. D., Dean, M. E., Rose, G. S., et al. (2017). A survey of neuromorphic computing and neural networks in hardware. *arXiv: 1705.06963*.

Sengupta, B., Laughlin, S. B., and Niven, J. E. (2013a). Balanced excitatory and inhibitory synaptic currents promote efficient coding and metabolic efficiency. *PLoS Comput. Biol.* 9:e1003263. doi: 10.1371/journal.pcbi.1003263

Sengupta, B., Stemmler, M. B., and Friston, K. J. (2013b). Information and efficiency in the nervous system-a synthesis. *PLoS Comput. Biol.* 9:e1003157. doi: 10.1371/journal.pcbi.1003157

Severa, W., Lehoucq, R., Parekh, O., and Aimone, J. B. (2018). "Spiking neural algorithms for markov process random walk," in *International Joint Conference on Neural Networks 2018* (Rio De Janeiro: IEEE).

Severa, W., Vineyard, C. M., Dellana, R., Verzi, S. J., and Aimone, J. B. (2019). Training deep neural networks for binary communication with the whetstone method. *Nat. Mach. Intell.* 1, 86–94. doi: 10.1038/s42256-018-0015-y

Shalf, J. M., and Leland, R. (2015). Computing beyond moore's law. *Computer* 48, 14–23. doi: 10.1109/MC.2015.374

Shehabi, A., Smith, S. J., Sartor, D. A., Brown, R. E., Herrlin, M., Koomey, J. G., et al. (2016). *United States Data Center Energy Usage Report*. Technical report, LBNL.

Shi, W. S., Cao, J., Zhang, Q., Li, Y. H. Z., and Xu, L. Y. (2016). Edge computing: vision and challenges. *IEEE Int. Things J.* 3, 637–646. doi: 10.1109/JIOT.2016.2579198

Sokoloff, L. (1960). The metabolism of the central nervous system *in vivo*. *Handb. Physiol. I Neurophysiol.* 3, 1843–1864.

Somayaji, A., Hofmeyr, S., and Forrest, S. (1998). "Principles of a computer immune system," in *Proceedings of the 1997 Workshop on New Security Paradigms* (Langdale: ACM), 75–82.

Sporns, O. (2006). Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *Biosystems* 85, 55–64. doi: 10.1016/j.biosystems.2006.02.008

Sporns, O. and Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics* 2, 145–162. doi: 10.1385/NI:2:2:145

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.

Staff (1948). Availability of radioisotopes. *Chem. Eng. News Arch.* 26:2260. doi: 10.1021/cen-v026n031.p2260

Sverdlik, Y. (2016). *Here's How Much Energy All United States Data Centers Consume*. Report, Data Center Knowledge.

Triantafyllou, M. S., Techet, A. H., Zhu, Q., Beal, D. N., Hover, F. S., and Yue, D. K. (2002). Vorticity control in fish-like propulsion and maneuvering. *Integr. Comp. Biol.* 42, 1026–1031. doi: 10.1093/icb/42.5.1026

Vanarse, A., Osseiran, A., and Rassau, A. (2016). A review of current neuromorphic approaches for vision, auditory, and olfactory sensors. *Front. Neurosci.* 10:115. doi: 10.3389/fnins.2016.00115

Waldrop, M. M. (2016). The chips are down for moore's law. *Nat. News* 530:144. doi: 10.1038/530144a

Watts, L., Kerns, D. A., Lyon, R. F., and Mead, C. A. (1992). Improved implementation of the silicon cochlea. *IEEE J. Solid State Circ.* 27, 692–700. doi: 10.1109/4.133156

Webb, D. C., Simonetti, P. J., and Jones, C. P. (2001). Slocum: an underwater glider propelled by environmental energy. *IEEE J. Ocean. Eng.* 26, 447–452. doi: 10.1109/48.972077

Weimerskirch, H., Bishop, C., Jeanniard-du Dot, T., Prudor, A., and Sachs, G. (2016). Frigate birds track atmospheric conditions over months-long transoceanic flights. *Science* 353:74. doi: 10.1126/science.aaf4374

Werfel, J., Petersen, K., and Nagpal, R. (2014). Designing collective behavior in a termite-inspired robot construction team. *Science* 343, 754–758. doi: 10.1126/science.1245842

Wolfe, N., Plagge, M., Carothers, C. D., Mubarak, M., and Ross, R. B. (2018). "Evaluating the impact of spiking neural network traffic on extreme-scale hybrid systems," in *2018 IEEE/ACM Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS)* (Dallas, TX: IEEE), 108–120.

Yang, T.-J., Chen, Y.-H., and Sze, V. (2017). "Designing energy-efficient convolutional neural networks using energy-aware pruning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Honolulu, HI), 5687–5695.

Yu, L., Shen, Z., Wang, C., and Yu, Y. (2018). Efficient coding and energy efficiency are promoted by balanced excitatory and inhibitory synaptic currents in neuronal network. *Front. Cell. Neurosci.* 12:123. doi: 10.3389/fncel.2018.00123