

Atomic Switch Networks for Neuroarchitectonics: Past, Present, Future



R. Aguilera, K. Scharnhorst, S. L. Lilak, C. S. Dunham, M. Aono,
A. Z. Stieg, and J. K. Gimzewski

Abstract Artificial realizations of the mammalian brain alongside their integration into electronic components are explored through neuromorphic architectures, neuroarchitectonics, on CMOS compatible platforms. Exploration of neuromorphic technologies continue to develop as an alternative computational paradigm as both capacity and capability reach their fundamental limits with the end of the transistor-driven industrial phenomenon of Moore's law. Here, we consider the electronic landscape within neuromorphic technologies and the role of the atomic switch as a model device. We report the fabrication of an atomic switch network (ASN) showing critical dynamics and harness criticality to perform benchmark signal classification and Boolean logic tasks. Observed evidence of biomimetic behavior such as synaptic plasticity and fading memory enable the ASN to attain a cognitive capability within the context of artificial neural networks.

R. Aguilera · K. Scharnhorst · S. L. Lilak · C. S. Dunham
Department of Chemistry and Biochemistry, UCLA, Los Angeles, CA, USA

M. Aono
International Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science (NIMS), Tsukuba, Ibaraki, Japan

A. Z. Stieg
WPI Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science (NIMS), Tsukuba, Ibaraki, Japan

California NanoSystems Institute (CNSI), UCLA, Los Angeles, CA, USA

J. K. Gimzewski (✉)
Department of Chemistry and Biochemistry, UCLA, Los Angeles, CA, USA

WPI Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science (NIMS), Tsukuba, Ibaraki, Japan

California NanoSystems Institute (CNSI), UCLA, Los Angeles, CA, USA
e-mail: gimzewski@cnsi.ucla.edu

1 Introduction

In 1965 Gordon Moore, co-founder and chairman of Intel, published a paper on the number of transistors per integrated circuit (microchip) doubling every 2 years. This trend became known as Moore's Law and meant processing speed doubled every 2 years. Recently, Moore's law shows signs of slowing down due to physical and economic constraints. In the next few decades individual elements will approach the scale of a few atoms and in turn the fundamental limits of miniaturization [1]. Feature sizes are constrained by the optical diffraction limit which defines minimum feature size by wavelength divided by two times the numerical aperture of the microscope [2]. Even before physical boundaries are approached, the economical limitations of continuing miniaturization and massive integration will be reached. With extremely small features fabrication costs increase dramatically, making them unsuitable for cost-effective mass production. Thus, as the end of Moore's Law is approached [3] processing power will no longer increase unless alternative individual elements or architectures of integrated circuits are explored.

Modern computers ubiquitously use the Von Neumann architecture first described by John von Neumann in 1945. This architecture separates memory and processing components within integrated circuits [4]. Partitioning data and instruction storage from arithmetic/logic processing leads to a limitation of information transfer known as the Von Neumann bottleneck [5]. When calculating more complex problems requiring mass amounts of data or instructions from memory storage, processors sit idle. In order to maximize information processing and ultimately overcome the Von Neumann bottleneck, significant changes to computer architectures must be made.

Carver Mead, a scientist at Cal Tech that coined the term "Moore's Law" became known for his bio-inspired work in the mid 60s. Particularly his attempts to emulate neural functionality directly into analog hardware implementations and pioneering the "Neuromorphic" computation field. His neuromorphic research was specifically based on analog metal oxide semiconductor (MOS) processors. By 1995 Mead and collaborators demonstrated a single silicon transistor 'synapse' capable of analog learning. However, the ensuing rapid development in digital microprocessors using VLSI superseded early analog computing approaches. In turn, digital neural network software led to today's deep learning and machine learning algorithms.

Concurrently, the growth of the internet-of-things, cloud computing and the rapid explosion of unstructured data has placed new demands on computers in our increasingly interconnected world. Examples of interconnected data are those from satellites, sensors, economic markets, commerce, global climate patterns, social media and consumer habits. This combinatorial complexity challenges the inherently serial Von Neumann architecture of computers which, at their core, scale poorly into supercomputers requiring massive increases in hardware and energy consumption. Currently, China's Tianhe-2 supercomputer uses 18 MW at full power [6] and is still very far from performing many tasks that a human brain can achieve using merely 20 W. In terms of dealing with complex tasks, the scalability of today's computers cannot keep up in a realistic manner with the world in which we live.

In response, we find ourselves in a new era of challenges promulgating research with the aim to develop science and technology that emulate how the brain works and to utilize that knowledge for creating a paradigm shift in computation towards cognition. New technologies such as functional magnetic resonance imaging (fMRI) and advanced image processing have also made major inroads into neuroscience, although the theory of understanding how our brain works is at an early stage. Nevertheless, fundamental scientific experiments to emulate and create brain-like behavioral characteristics are underway.

In this chapter we discuss one such approach using devices, based on the atomic switch's synaptic-like properties, connected in a network called the "Atomic Switch Network" (ASN). The device is inspired, not only by its synaptic function, but also by the brain's inherent network characteristics in the neocortex and its distributed memory.

2 Frameworks for Neuromorphic and Bio-inspired Computing

In order to advance the field of computing, new computational hardware paradigms are required including artificial neural networks, deep learning, reservoir computation, and neuromorphic computing. Additionally, identifying ways to apply these systems for use with established algorithms such as deep learning will play a key role in ushering in the next generation of computing [7].

2.1 Neuromorphic Computing and Artificial Neural Networks

Beautifully designed by the natural world, the human brain is capable of complex multisensory tasks and decision-making using architecture distinct from the hyper-engineered grids of von Neumann computers. Modern neurobiology describes neurons using the formulations of the Hodgkin-Huxley model [8], emulating single neurons as a circuit of capacitors, nonlinear resistors, and a current source. Individual neurons connect to one another through synapses, creating a connectome architecture [9] or network allowing cascading ions to transmit information. Network connectivity determines the efficiency of interneuron communication and has been heuristically observed to follow a small-world network topology [10, 11]. Found throughout nature from galaxy formation to sociological trends, small-world dynamics form the core of understanding complex systems nonlinearizable by current methods. Amazingly, cognitive behavior and memory association practiced by the brain utilize these salient features to be able to perform both computation and information storage within a single synapse [12]. Collectives of interacting neurons

found in the brain act as compartmentalized networks ascribed to specific brain functions, but are capable of growing through heuristic learning. Computing and cognitive capabilities developed through an evolutionary scheme selectively prune weakly correlated neurons and enforce strongly correlated neurons as described in Hebbian learning [13]. These features drastically increase the power of the human brain to complete complicated tasks too computationally expensive for current CMOS technology.

Mead drew upon the natural computing capability of neuron networks to develop the concept of neuromorphic hardware in the mid 1980s. Neuromorphic engineers attempt to emulate neuron functionalities and brain architectures to be capable of performing similar multisensory complex tasks such as associative-dissociative memorization of unstructured data, pattern recognition, and chaotic series prediction, to name a few. Initial attempts to harness brain dynamics for computation relied upon contemporary CMOS technology to construct purpose-built field programmable gate arrays (FPGAs), and supercomputer assisted software within the field of machine learning. Artificial neural networks first developed as the perceptron [14] by Frank Rosenblatt adopted the same connectionist theory as biological neurons. Implemented using camera photocells while contemporary variations are realized in software, the perceptron consisted of a collection of nodes representing neurons with each node owning a number of weighted connections w_i transmitting information x_i . Information propagates from external sensors towards individual nodes. Information traveling through the connections and converge at their respective nodes, activating the node depending on a transfer function or learning rule f_j .

$$y_j(t) = f_j\left(\sum w_i \cdot x_i(t)\right)$$

The overall task is then computed using the sum of all node operations while programmable control is achieved through modification of the weights in a learning procedure [15].

Conceptually, Mead's neuromorphic computing described a system possessing analog circuit elements capable of emulating biological features of the brain. The fundamental element of neuromorphic computing is the spiking neuron, and functions similarly to the logic gates of traditional von Neumann architectures. While traditional logic gates evaluate data as binary states of either 0 or 1, spiking neurons transmit a series of one or more spikes within a fixed period of time where the number of spikes represent the data as discrete continuous values, 0, 1, 2, etc. Thus, neuromorphic computing more closely resembles power-efficient analog systems [16]. A neuromorphic system contains many of these spiking neurons connected in a complex network, which must then be conditioned or trained, using spike-timing-dependent plasticity in order to perform specific tasks such as pattern recognition.

The key advantage of a neuromorphic system is its highly interconnected and parallel architecture providing remarkably reduced power consumption relative to traditional von Neumann architectures. Additionally, neuromorphic systems by design remove the significant bottleneck between memory and processing. This

combination of improved bandwidth between memory and processor, along with parallel processing will theoretically allow neuromorphic computing systems to perform complex operations much faster and more efficiently than conventional systems [17].

Artificial neural networks, or ANNs, are a biologically inspired computing architecture incorporating functional elements designed to mimic the brain's neural networks. The networks form by a collection of artificial neurons. These neurons are interconnected via structures functionally resembling synapses, such that each neuron can transmit a signal it generates or receives to another neuron it is connected with [17]. Each neuron and synapse may possess a different “weight”, dependent on plasticity and history of past signals, which modifies the strength of the transmitted/received signal. The neurons themselves are organized into three layers—input, hidden, and output—and signals traverse these layers one or more times as the network works to solve a given task. Key advantages for this type of architectural framework include fault tolerance characteristics and reliability within hardware [17].

The most basic model of an artificial neural network, the feedforward neural network, only allows signals to propagate in one direction, forward, towards the output layer [17]. An alternative operational mode to the feedforward neural network exists in the form of recurrent neural networks (RNN). Here the connections between units of the network form a feedback loop, which allows the internal memory of the RNN to process arbitrary input sequences [17]. This characteristic widens the RNN's application range to include tasks such as speech recognition. Both feedforward and recurrent artificial neural networks are discussed further in Sect. 3.3.

2.2 *Artificial Neural Networks*

Artificial neural networks (ANNs) have been developed to have brain-like functions using three layers of neurons, input, hidden and output. Hidden layers are unobserved neurons which alter the flow of information between inputs and outputs achieving a level of complexity that prevents outputs being simple functions of inputs. Connections between each layer act as synapses with tunable (weighted) properties. The presence of a feed-back loop also allows for dynamical complexity to dominate the system.

ANNs require training cycles for the tuning of synaptic weights fitted to a desired algorithm. In altering the hysteresis of each synapse, desired outputs can be retrieved from task specific inputs. ANN training techniques can be broken into three primary classes: supervised, unsupervised and reinforcement learning. Supervised learning simultaneously introduces an input vector paired with the desired outputs and adjusts its weights through an implemented learning algorithm (i.e. Manhattan update rule) [18]. Unsupervised learning, also commonly referred to as self-organization, outputs are trained to respond to a cluster of inputs. This results in a system designed to respond to specific input stimuli, and fabricate its own representations of the transmitted information. Reinforcement learning can be thought of as a trial-and-

error system akin to repeating a maze periodically until the optimal route has been determined. The ANN is not given a specific path to take, and must be tuned through numerous training cycles [19].

Recurrent neural networks (RNNs) are a sequence based model of ANNs. RNN's can be separated into two primary classes, the RNN will either be presented with a constant, single input and stabilize to a desired state or presented with variable inputs fluctuating over time with the intention of yielding time-dependent outputs [20]. RNN's are capable of using their internal memory to process input sequences allowing for complex applications including speech and handwriting recognition [20, 21].

Convolutional neural networks (CNNs) function similar to a general ANN, however neurons are now arranged in three dimensions and characterized by height, width and depth. This manipulates the manner in which patterns are constrained and scales more efficiently than a typical ANN for larger pattern recognition systems. CNNs utilize three core layers: the convolutional layer, pooling layer and fully-connected layer (this is present in a typical ANN). The convolutional layer computes output neurons in relation to a dot product between the input region and its corresponding weights. After the convolutional layer, an element based activation function is applied followed by the pooling layer which down samples the width and height dimensions while leaving the CNN volume intact; both of these act as fixed pre-determined functions. Finally, the fully-connected layer will compute class scores which relate the pattern to its desired category. Training through backpropagation is only necessary for the convolutional and fully-connected layers. CNNs core design achieves much more efficient pattern classification than a regular ANN and makes it readily implementable in deep neural networks for image and speech recognition.

Recursive neural networks (rNNs) are another model applicable to deep learning that function by recursively applying an identical set of weights to a given system. This achieves structured predictions based on variable inputs; this architecture also allows for the flow of data in any given direction. While CNNs are good deep learning models for pattern recognition, rNNs offer the ability to predict variations in hierarchical systems, demonstrated in natural language processing, and can be regarded as a linear modification of RNNs though are unable to parse tree-like hierarchies [20].

Multi-layer perceptron's (MLPs) are a class of ANNs which implement feed forward loops. All hidden and output layers in the MLP act as a neuron responding to stimuli (inputs) which rely on a non-linear, typically sigmoidal or hyperbolic tangent, activation function. The activation function acts as a weighted tolerance value governing the output function. MLPs typically utilize backpropagation algorithms for training and the weights can be updated iteratively after each testing phase (often results in chaotic alterations) or in unison after all weights have been analyzed (batch learning, often yields more stable alterations) [22]. By defining classes to analyze, MLPs can be applied to many different tasks including 3D image recognition and handwritten image recognition.

2.3 *Deep Learning*

Modern neuromorphic system design strives to build the same adaptive learning capacity found in biological neural systems into an electronic device. One way to incorporate this system design is through the employment of machine learning algorithms. Machine learning algorithms attempt to emulate neuronal design in software. The learning process can be accomplished in a supervised or unsupervised manner, with the key distinction being whether the data fed into the algorithm is labeled (supervised) or unlabeled (unsupervised) [23]. Presently, supervised learning is the most common form of machine learning, regardless of the algorithm employed.

Historically, machine learning algorithms and techniques struggled to process natural data sets in raw form and only through extensive work by engineers could machine learning be tailored to perform tasks such as speech recognition [23]. This has changed with the development of new methods, feature learning or representation learning, which are capable of identifying the minimum features that represents each class of object for detection and/or classification within a data set. Those methods led to the creation of the well-publicized machine learning technique known as deep-learning that simulates a multi-layered neuron architecture. Its purpose is for organizing unstructured data and applying this learned information to various tasks, such as computer vision, speech recognition and bioinformatics. In the case of supervised learning, the process starts by amassing a large data set where each element of the data set is labeled with its corresponding category. Next, the machine is trained by supplying it with data representative of each category, for which the machine then produces an output in the form of a vector of scores for every category. The goal is to train the machine so that the vector of scores for a particular piece of data has the highest possible score for the category to which it belongs. This is achieved by computing a function that measures the error between the output scores and a desired score pattern, after which the machine modifies its own parameters, or weights, via a gradient vector—a function that adjusts weights depending on whether the error would increase or decrease based on that adjustment—in order to reduce the error. Consequently, training can be an arduous process for which the best results typically require incredibly extensive data dates and hundreds of millions of weights [23].

More advanced machine learning algorithms attempt to completely emulate neuronal design in software, simulating multi-layered neurons similar to the perceptron in an architecture known as ‘deep learning’ [24]. Organized in a hierarchical tree, the output of one perceptron is fed forward to a connected perceptron to compute complex tasks. The interaction between layers increases the difficulty in the learning procedure due to the nonseparability of individual weights from the internal recurrent response. Researchers have developed methods of mitigating the difficulties through feedback learning and error backpropagation schemes that predict network behavior [24, 25].

Contemporary industries alleviate big data demands by relying on these machine learning algorithms distributed across hundreds of servers and graphics processing units (GPUs), which are discussed in a later section, in an attempt to emulate a brain's capability of organizing unstructured data. However, CMOS-based neural networks are trending towards costly power consumption as transfer speeds between buffer memory and logic components as well as miniaturization capabilities are approaching fundamental limits.

2.4 Reservoir Computing

Reservoir computing (RC) is an emerging paradigm that promotes computing using the intrinsic nonlinear dynamics of an excited system called a reservoir. Maass et al. initially proposed a version of RC called the Liquid State Machine (LSM) as a model for cortical microcircuits. Independently, Jaeger introduced a variation of RC called the Echo State Machine (ESM) as an alternative RNN approach for control tasks. Variations of both LSM and ESM have been proposed for many different machine learning and system control tasks. Büssing et al. conducted a comprehensive study of reservoir performance using different metrics as a function of the node connectivity K , the logarithm of the number of states per node m , and the variance of the weights in the reservoir [26].

At its core, the RC paradigm utilizes a reservoir's capability to project an input's information into a mathematical higher representation space, similar to a Fourier transform. A variety of spatially distributed mathematical operations occur throughout the system according to system properties and dimensionality. Computation occurs as a recursive learning algorithm that inscribes a filter on the system such that the projection spans the correct mathematical operations and the desired process is achieved. The reservoir essentially outputs a series of nonlinear transformations of the input which are then trained at the output layer by synaptic weights using linear regression. A system with sufficiently rich dynamics can remember perturbations by an external input over time which compared to other approaches, has many key advantages of using RC such as:

1. Computationally inexpensive training (low programming overhead).
2. Flexibility in the physical reservoir implementation (cost-effective fabrication).
3. A high tolerance to material variation, defects, and faults (robustness).

These factors make RC particularly suitable for emerging unconventional computing paradigms, such as computing using physical phenomena [27] and self-assembled electronic architectures [28]. The high nonlinearity, and thus the high dimensionality, ensures convergence of this algorithm in practical time. Reservoir computing differs from earlier attempts at computing with random assemblies of nano-cells or switches, e.g., Tour et al. [29]. Such systems lacked a formal framework and required complex and time-consuming optimization steps in order to

produce a desired output function. Another problem of these early attempts was the lack of scalability, and thus the difficulty to obtain more complex functions.

3 Hardware Paradigms for Neuromorphic Computing

The previous section described a few framework paradigms for next-generation computing. This section will discuss several examples of the physical implementation of these paradigms, including hardware configurations utilized for deep learning and neuromorphic computing. Throughout the design of a new architecture, its desired neuromorphic properties need to be considered. An array of different metrics can benchmark devices for the implementation of a specific neural network. Considerable factors include efficiency, reconfigurability, and scalability.

3.1 *Neuromorphic Chips*

Neuromorphic refers to any artificial neural system whose core design and architectures are based upon the biological central nervous systems. Traditional computing architecture efficiency is governed by how many floating-point operations per second (FLOPS) per watt can be performed. In developing neuromorphic systems, spiking neurons are fabricated into the hardware; these can be benchmarked by their synaptic operations per second (SOPS) per watt [30]. Floating-point operations are inefficient and slow; even the most powerful super-computers are not capable of obtaining real-time performance on detailed large-scale simulations of neural systems [30]. Synaptic operations offer an advantage in that operations are achieved directly through the hardware; this allows for real-time operation independent of synapse density and/or coupling. This advantage is achieved through circumventing the von Neumann bottleneck; a single component of hardware running specific tasks alleviates the overhead from multiple components communicating with each other achieving real-time analysis.

Silicon neurons are a promising avenue for mimicking biological synaptic/neuronal interactions. Unlike the previously mentioned hardware, a silicon neuron can be broken into computational blocks which can be functionalized for task specific neuronal activity. The synaptic block is capable of carrying out both linear and non-linear input spikes with short and long-term plasticity mechanisms available. Some blocks are a group of sub-blocks designed to computationally represent the theoretical model in which they are based upon. Finally, dendrite and axon circuit blocks account for the spatial structure and interconnectivity of the overall system allowing for an intricate and fully connected array of neurons [31]. Sub-segments of each element can be manipulated for the desired implementation of a specific task or model.

The clear desire for neuromorphic architectures has led to further investigation and developments of different synthetic synapse models. In recent years, atomic switch systems have garnered much interest and are now being modeled and investigated as synthetic synapses with the hopes of scaling them into a connected network yielding synaptic densities and topographies similar to that of the human brain.

ASICs

Carver Mead's work in pioneering analog VLSI implementations for neural systems set the foundation for early neurocomputing architectures, or neuroarchitectures. His work clearly outlined the need for neurons and synapses (weighted connections between neurons) for real-time processing on a single device capable of parallel processing. Mead's investigation of neural networks denoted that various frameworks could be implemented for general purpose neurocomputers, however, the neuroarchitecture will need to be uniquely designed while accommodating the desired neural networks [7]. This presents the potential to develop rigid and highly efficient task specific neurocomputers or, by contrast, versatile and tunable neurocomputers utilizing a hierarchy of neural networks.

Early implementations of neurocomputers were achieved using application-specific integrated circuits (ASICs) fabricated using CMOS VLSI technology. As their name suggests, they were developed for the execution of specific tasks and cannot be reconfigured at a later time by the end user. Each ASIC works as either as a master or slave node interconnected in a ring or bus network which broadcast signals based on their topology. The master node functions to control neurocomputation while the slave nodes enable parallel processing, the backbone of ASIC's efficiency for neuroarchitectures. This system utilizes external memory for the storage of neuron outputs and synaptic weights [32]. Efficiency and speed is ultimately governed by the number of neurons on a chip. An issue with ASIC neuroarchitectures is that they are developed for a specific neural network and are not reconfigurable devices [33] despite this they are notably proficient at specific tasks and highly power efficient.

3.2 FPGAs

Field-programmable gate arrays (FPGAs) while worse in raw performance than ASICs, offer the added benefit of reconfigurability. FPGAs are designed to be manipulated by the end user allowing for a great degree of flexibility in designing FPGA based neuroarchitectures capable of functioning in an array of neural networks for prospective optimization in ASIC implementations. Despite their lower performance, FPGA's re-configurability and capacity to perform parallel processing has overshadowed ASICs in the development of neurocomputers [33].

As FPGA architectures continue to garner interest, programming frameworks have been developed to simplify their manipulation. The Open Computing

Language (OpenCL) platform has been designed to allowing programming through a heterogeneous array of hardware components including FPGAs. Intel has demonstrated a convolutional neural network (CNN) implementation using this platform which currently outperforms traditional CPU based CNNs with a significant increase in efficiency [34]. Advances in software continue to expand the use of FPGAs for a wide array of applications.

Continued developments in FPGA architectures have demonstrated their proficiency as accelerators in deep convolutional neural networks [35], Qiao et al. have developed an FPGA accelerator compatible with Caffe software, which is currently implemented in a wide array of CNNs. A comparison of their FPGA (Xilinx Zynq) accelerator to a traditional CPU (Intel Xeon X5675) and GPU (Nvidia Tesla K20) based CNNs shows FPGAs are superior in versatility and power efficiency to CPUs [34]. While GPUs still offer the capability of much higher performance (486 Gflops to the Zynq's 77.8 Gflops), the large disparity in power consumption (235 W to the Zynq's 14.4 W) implicates FPGAs as the most power efficient device to be implemented in the acceleration of CNNs. Future advances in GPU and FPGA technology may alter this trend.

3.3 Graphics Processing Units (GPUs)

The term graphics processing unit, or GPU, was promulgated by NVIDIA in 1999 to mark the release of the world's first such device, the GeForce 256 [36]. This GPU was touted as an incredible advancement in the world of computer hardware, possessing approximately 23 million transistors (NVIDIA boasted at the time that this was twice as complex as a Pentium III processor) and possessed 50 gigaflops of floating-point calculation capability [36]. Nearly two decades later, GPUs have made enormous strides in power and capability. The most powerful GPU as of 2017 is NVIDIA's Tesla V100, a device possessing 21 billion transistors, over 5000 cores providing 120 teraflops of performance for deep learning applications while drawing only 300 W of power [37]. Thus, these units have increased in processing power by over four orders of magnitude over the last 18 years.

As GPU power increased over time, interest in them from individuals and organizations outside the computer gaming community grew significantly. In the quest to expand upon the capabilities of existing computing architectures and platforms, academic research groups began evaluating GPU's viability for performing certain tasks previously delegated to supercomputers. In 2012, researchers at the University of Toronto demonstrated that GPUs could be used to drastically improve tasks related to computer vision and deep learning, such as image reconstruction, by using the GPUs to run deep neural networks [38]. These GPUs displayed significant performance gains relative to traditional computer processor-based neural networks thanks to their massively parallel architecture involving thousands of individual processor units and exceptional processor-to-memory bandwidth.

Since that time, companies, particularly NVIDIA, have helped lead the way toward an artificial intelligence revolution with the use of GPUs [39]. GPUs are now utilized extensively by companies and universities involved in the fields of big data, machine learning, and genomics, among others [39–41]. GPUs are also seeing increased utilization in areas of academic research as more coding languages and libraries, such as C and FORTRAN, are updated to take advantage of the parallel processing power of this architecture.

3.4 Purpose-Built Chips and History

With an exponentially growing interest in neuromorphic architectures, many researchers began further pursuing their development. In 2005, Fast Analog Computing with Emergent Transient States (FACETS) launched a research initiative funded by the UK to implement brain-like hardware architecture for neuromorphic computing. The project concluded in 2010 with a VLSI implementation of a traditional CMOS fabricated device containing 400 neurons and 100,000 synapses [42]. In 2011, brain-inspired multiscale computation in neuromorphic hybrid systems (BrainScaleS) intended to expand upon FACETS work and concluded in 2015 with an architecture containing 1.6 million neurons and 400 million synapses. In the middle of the BrainScaleS project further interest in neuromorphic architectures arose.

In 2013, the Information and Communication Technologies (ICT) of the EU launched the Human Brain Project. This comprehensive research initiative aimed to advance our understanding of the human brain through numerous fields including neuroscience and computation. In the field of neuromorphic architectures, this flagship intends to refine and expand upon the work of FACETS and BrainScaleS. The same year, the United States launched the Brain Initiative under the Obama administration with similar intent and currently funds numerous agencies including DARPA, NIH and NSF. Through this funding neuromorphic based projects continue to progress.

As these architectures advance, hardware has begun to be designed for task specific applications. IBM has developed TrueNorth, a brain inspired device suitable for complex applications which utilize neural networks. TrueNorth is a 5.4 billion transistor chip containing 4096 neurosynaptic cores interconnected to an intrachip that utilizes 1 million programmable spiking neurons and 256 million configurable synapses; it consumes a mere 70 mW and is able to process 46 billion SOPS, per watt. This greatly exceeds the limit of energy efficient super computers which are only capable of processing 4.5 billion FLOPS per watt. This device has demonstrated high fidelity multi-object recognition in real-time capable of discerning objects based on different classes (i.e. person vs. cyclist). TrueNorth highly excels in task specific applications which do not require reconfigurability similar to ASICs.

The Spiking Neural Network Architecture (SpiNNaker) is an ongoing project which aims to have 1 million ARM processors in parallel. Unlike TrueNorth which

is optimized for specific tasks, SpiNNaker has implemented a versatile, tunable neuromorphic architecture in the same vein as FPGAs at the cost of power efficiency. Both architectures have been designed for scalability and can accommodate cascades of devices allowing for easier implementations of larger systems.

Qualcomm Technologies Zeroth Machine Intelligence platform has also developed a deep learning software development kit (SDK) which has brought the power of deep learning to mobile devices. This breakthrough has removed the need to connect to a cloud server to utilize the benefits of deep learning and gives mobile phones the innate capability to perform numerous complex tasks like facial recognition, object tracking and natural language processing [43].

3.5 *ASNs for Computing*

ASN devices exhibit various properties associated with atomic switches and other memristive systems [44], the latter defined as a system with its internal resistance based on electric flux [45]. Such properties include but are not limited to a requisite forming step and distributed, frequency-dependent hysteretic switching among a collection of dynamically interacting elements. In addition, the functional topology of ASNs has been shown to produce a diversity of complex behaviors, ranging from distributed memory function to emergent critical dynamics similar to those found in both fMRI/EEG of biological brains and multi-electrode array (MEA) studies of neuronal populations [15].

Observations of power-law scaling in various device dynamics and a ‘fading memory’ property of learned states have implicated as an essential component for applications of reservoir computing (RC) using critical states. Initial progress in the use of ASNs as nonlinear reservoirs capable of task performance in the RC paradigm has shown through simulation and experimental implementation of a benchmark task known as waveform generation.

Based on extensive studies of the dynamical response of ASNs (Fig. 4), these devices have been identified as an ideal platform for hardware-based reservoir computation [46]. ASN devices have shown, through both experiment and simulation [47–49], to be a viable platform for hardware-based RC toward applications in pattern recognition, prediction and logic. Based on their ability to integrate, segregate, store and respond to external stimulus, the utility of ASNs as nonlinear reservoirs has been demonstrated through implementation of multiple benchmark tasks including: (1) waveform generation [48] and (2) various logic operations (AND, OR, XOR) (see Sect. 3.2). The speed, density, and [50] scalability of the ASN serve to overcome major hurdles in the RC paradigm.

4 Building the Atomic Switch Network

Designing a system capable of neuromorphic computation requires adequate functionality in terms of non-linearity, persistent activity, and recurrent structures. In addition to these metrics, a chip must be power efficient in order to be a viable option for industrial applications. Keeping these requirements in mind our group sought to harness the inherent properties of individual atomic switches (non-linearity, quantized conductance, and memory) and use them as artificial neurons in this device. Through these means, a device was fabricated using both top-down and bottom-up methods. The atomic switch network (ASN) consists of highly recurrent structures that produce dynamic activity through space and time. These networks produce non-linear responses and their emergent behavior is much more complex than that of its individual junctions. This dynamic system's emergent distributed behavior also provides a diverse array of output signals. Configured as neuromorphic chips, ASN devices are capable of performing alternative types of computing.

Atomic Switches as Synthetic Synapses

Establishing specific connections between patterns of electrical activity and brain function is a difficult task that requires studying general features of neuronal structure in order to determine the essential properties required to construct a device capable of learning in a physical sense. These features are believed to include synaptic plasticity, allowing physical reconfiguration of the network to enable functional differentiation and the development of hierarchical structures, which all possess correlated memory distributed throughout the dynamically coupled synapses. Therefore, it can be inferred that learning capacity is connected to dynamic activity within the brain. Specifically, a near-critical or “edge of chaos” operational regime [51] has been associated with the fast, correlated response to stimulation necessary for computation and learning. Though extremely attractive as a construct for developing computational machinery whose operation results from intrinsic critical dynamics, the production of such a device in hardware has proven a daunting task, with ASN devices being one of the few successful demonstrations in the scholarly literature.

The first experiment to measure the transition from an electron quantum tunneling to single point contact regime was reported in 1987 using a scanning tunneling microscope (STM) in ultra-high-vacuum (UHV) on a silver surface [52]. Current-distance characteristics showed that, at sufficiently small tip-surface gaps, an abrupt increase in conductance, G , of $\sim \frac{2e^2}{h} \approx \frac{1}{13} k\Omega$ which is the quantized unit of conductance. Subsequent theoretical analysis verified that at small gap distance the effective tunnel barrier collapses prior to point contact via ballistic electron injection [53]. Later work demonstrated further jumps of $\sim n \frac{2e^2}{h}$, where $n = 1, 2, 3, \dots$ in the conductance occur as the contact area is increased. Such observations were not limited to STM experiments; even two macroscopic wires brought in contact also displayed this effect, albeit in a less controlled manner. Quantized conduction, also

introduces Landauer's concept of transmission where the term t is the transmission [54].

$$G = \frac{2e^2}{h} \sum_n t_n$$

In 2002, experiments by Terabe et. al found that Ag atoms could be transported through an STM tip made of silver coated with silver sulfide and deposited on a surface in a controlled manner [55]. The characteristics of this process also occurred via quantized conduction. However, the mechanism involved ion migration under the influence of an electric field, a process called 'electroionics' meaning that in addition to electron motion, ion motion also occurs simultaneously. Normally ionic diffusion processes on the macro-scale are considered to be slow, but when they are induced on the nanometer scale, they are actually quite fast and can occur on a (sub-) nanosecond time scale depending on the geometry and dimensions of the junction. In 2005, using junctions fabricated using conventional microelectronics, Terabe et al demonstrated atomic switching in silver sulfide junctions with discrete and reversible quantized jumps from $n = 1$ to 10. This was the birth of the "atomic switch". Since that date, a number of researchers have observed quantized conduction in a wide range of materials including sulfide junctions of copper, tungsten sub-oxides as well as various metal-doped polymers.

Aside from the fundamental science of their quantization, interesting electronic features of atomic switches are pinched hysteresis, large ON/OFF conduction ratio, MHz switching speeds and volatility characteristics as well as CMOS compatibility because of their potential in digital electronic memory applications. Indeed, NEC recently have incorporated atomic switch technology into field programmable gate arrays (FPGAs) where a reduction in device foot print, speed, and energy consumption was achieved by replacing certain memory tasks, normally using transistors, into the circuitry [56].

Additional atomic switch functionality was reported in 2011 when studying switching near-threshold conditions [57, 58]. It was found that atomic switches have an on-off memorization property of past switching events. For instance if switching is performed infrequently, the switches remain in the on-state only briefly whereas if frequent switching events are made in rapid succession then the on-state persists for a longer time. A series of careful experiments were able to relate these physical observations to a psychological model of learning call the Atkinson-Schiffin multi-store model. The essence of the model involves sensory memory (SM), short-term memory (STM) and long-term memory (LTM). New information arrives to the brain as sensory memory and that information is passed to short-term memory. In the absence of similar stimulation, information is forgotten. However, if the process is repeated many times the information is moved into long-term memory. Think of learning to play a tune on a piano by diligently practicing repeatedly. In terms of bio-inspiration the operational characteristics of the atomic switch under threshold switching also related to characteristics of biological synapses. The atomic

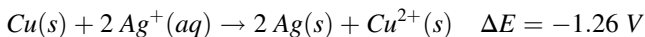
switch therefore has also been called a synthetic synapse where memory is represented by conduction state.

The next step in creating a ‘brain inspired’ device is the fabrication of networks of synthetic synapses (Atomic Switches). Taking the neocortex as a biologically inspiration, self-assembly was used to incorporate atomic switches into a dense dendritic tangle of silver nanowires resulting in a density of $\sim 10^8$ connections/cm². In response to electrical inputs which inject energy into the network, these networks exhibit self-organization, critical power law dynamics and spatio-temporal non-linear outputs at a multiple electrodes. The device is called an Atomic Switch Network (ASN) and is described in detail in the next section.

4.1 Network Fabrication

Several routes to fabricate functionally complex recurrent networks have been experimentally explored, including: seed free networks, random seed networks, and patterned seed networks. The seeds are small areas of deposited copper that react in solution to generate silver wires through electroless deposition. The patterned seed networks proved the most versatile, and utilized a combination of top-down with bottom-up fabrication, a powerful general fabrication approach known as nanoarchitectonics. Initial approaches on implementing atomic switches into a network topology consisted of pipetting 150 μ L of an isopropanol suspension (149.8 mg Ag/L) of monodisperse silver nanowires (120–150 nm \times 20–50 μ m, Aldrich) onto a substrate and allowing to air dry. These devices were then activated using the technique described in Sect. 4.2 and non-linear IV curves were subsequently observed. However, the non-uniformity in the dispersion of nanowires initially caused concern in the area of spatially distributed activity.

With a density controlled network in mind, electrochemistry was used to grow a recurrent silver network via copper seeds. Following the galvanic reaction below, network growth occurs through an electroless deposition (ELD) reaction via individual atom displacement reactions between Ag^+ and Cu^0 based on respective electric potentials. A spontaneous ELD reaction is preferred over an electrically induced reaction to minimize artifacts and maintain the delicate nature of the electrochemical reactions. Here, silver atoms are oxidized while copper is reduced during the galvanic displacement reaction.



Using this ELD reaction, a random seed network was fabricated by pipetting a 1 mL aliquot of copper microspheres 1–10 μ m (99.995% purity, Alfa-Aesar) which was then air dried. Silver nitrate (50 mM) was pipetted (20 μ L) onto the center of the device. Following the ELD described above, silver dendritic structures

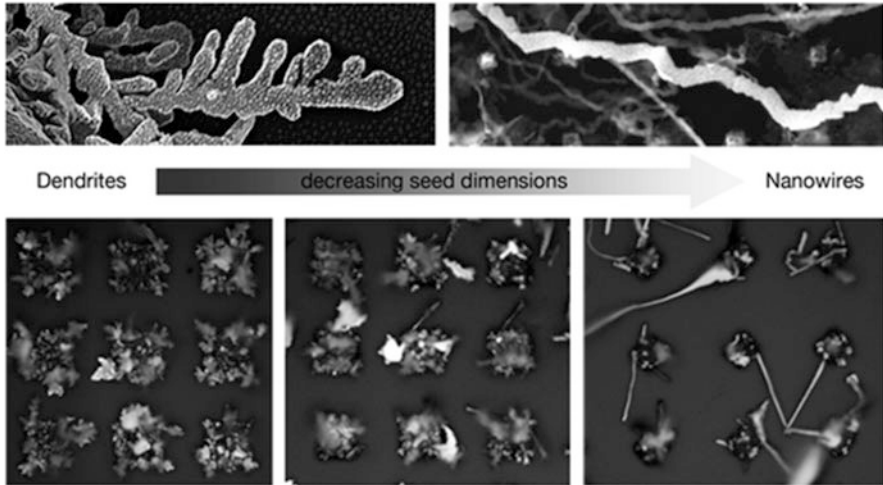


Fig. 1 Above are SEM images depicting the morphological transition seen from changing seed size. Branching dendritic crystals occur above $10\ \mu\text{m}$ posts, while below $3\ \mu\text{m}$ nanowire growth along the (111) lattice is observed. Intermediary post sizes yield a mixture of dendrites and nanowires (Avizienis Crystal Growth & Design)

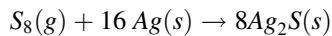
spontaneously formed a complex network. Again, network density exhibited non-uniformity due to the stochasticity associated with drop casting metal suspensions.

Successful implementations of the ELD reaction above allowed us to design a technique using highly patterned top-down photolithography combined with complex spontaneous and self-organized growth. The patterned seed networks consist of a $2\ \mu\text{m}$ layer of AZ nLOF 2020 (a negative photoresist), a soft bake, followed by UV photolithography, and a post-exposure bake. This resist is developed in MF26A, rinsed with isopropanol, and a $300\ \text{nm}$ layer of copper is then deposited and lifted off overnight in acetone. At the end of this process, a patterned grid of copper posts $300\ \text{nm}$ high is left. The size and pitch of these posts were refined over time to give the most desirable silver crystal growth [59].

When first designing a purpose-built device to emulate mammalian brain activity, dendritic silver structures were desired. Experimentally we find that by changing the size of the copper posts, a morphological transition occurred and a seed site of $1 \times 1\ \mu\text{m}$ up to $3 \times 3\ \mu\text{m}$ leads to fine long rhizome-like nanowires. Seeds between $3 \times 3\ \mu\text{m}$ and $10 \times 10\ \mu\text{m}$ yield a mixture of nanowires with branched dendritic structures, while posts larger than $10 \times 10\ \mu\text{m}$ produce predominantly dendrites [59] (Fig. 1).

4.2 Network Functionalization

Devices are functionalized through the deposition of an insulator onto metallic nanowires. In the case of AgI₂SI₂Ag, a temperature controlled sublimation of cyclooctasulfur (S₈) is then directed from a sulfur chamber to another chamber containing the sample using a carrier gas. Sulfur gas is exposed to the silver networks for 5 min and the ASN chips are then removed and stored in vacuum. A slow diffusion reaction governs the movement of sulfur into the silver lattice, which can be assessed by monitoring the electrical this time resistance of the device and if necessary consecutive sulfurizations are performed to obtain the optimum sulfide coverage. The surface chemical reaction can be written as:



Once the desired resistance is reached, the network is initialized in a process called electroforming by sweeping a triangular voltage waveform across the network via the contact electrodes. The triangular sweeps induce a current flow across the device through complex pathways, and with adequate voltage, the device is able to form tiny filaments. A multitude of junctions form creating a conductive pathway to ground, and thus increased current. This increasing current flow through the device can be seen in Fig. 5 and depicts this electroformation of numerous nanofilamentary pathways throughout the atomic switch network.

4.3 Device Fabrication

Over the years there have been multiple generations of ASN devices, starting with simple two electrode devices all the way up to 128 electrodes chips. Devices with 4, 16 and 128 electrodes are shown in Fig. 2. Patterned seed networks consist of a 2 μm layer of AZ nLOF 2020 (a negative photoresist), a soft bake, followed by UV photolithography, and a post-exposure bake. This resist is developed in MF26A, rinsed with isopropanol, and a 300 nm layer of copper is then deposited and lifted off overnight in acetone. Copper lift off occurs leaving a patterned grid of copper posts ~300 nm thick. These patterned seeds enable a level of network density control not previously realized in earlier experiments. The process provides a level of reproducible control over the spontaneous growth of materials that are CMOS compatible and most importantly for alternative computing: structurally complex.

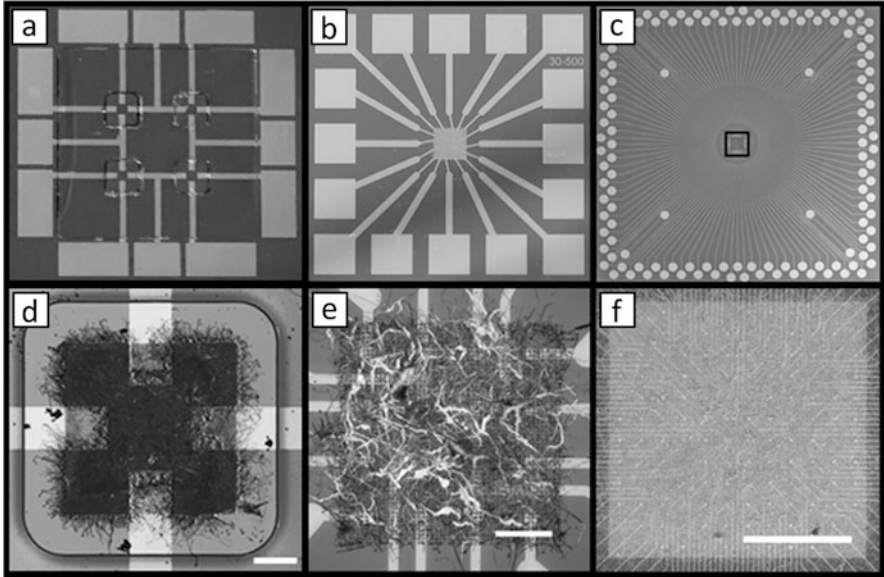


Fig. 2 Different generations of ASN devices are shown above. Chips include (a) 4 electrode device, (b) 16 electrode, and (c) 128 electrode ASN. Inner networks of each device are shown on the bottom row (d–f, scale bars = 100 μm , 1 mm)

4.4 Measurement Platform

The ASN measurement platform was custom designed in CAD and 3D printed (MakerBot Replicator 2.0). ASN chips sit on the bottom of the platform and are attached with gold spring-loaded pins. These pins connect the measurement hardware to the outer electrodes of the ASN chip. Measurement hardware consists of a source measure unit ((National Instruments Model 4141), two data acquisition cards (National Instruments Model 6368), and a switching module (National Instruments Model 2532). These units allow for the ability to define each electrode as an input or output (I/O) signal, specific signals to source, and concurrently record these input signals as well as record spatially-distributed voltage traces that percolate through the network. All measurements on the multielectrode array are performed and recorded simultaneously. Control software for the ASN device was coded in LabView 2012 (National Instruments). Post analysis work was done in MatLab 2010b (MathWorks).

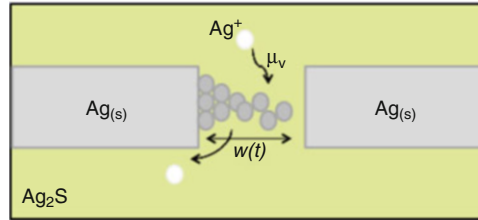


Fig. 3 Atomic switches are comprised of an Ag|Ag₂S|Ag junction. Applied electrical bias causes Ag cation migration to the cathode where it is reduced, forming a stable metallic filament, resulting in resistance change. This migration is modeled by the filament length $w(t)$, Ag cation mobility μ_v and additional stochastic terms (Sillin Nanotechnology 2013)

5 Results: Atomic Switch Network Dynamics

5.1 Operational Characteristics of the Atomic Switch

Experimental studies into the operational characteristics of atomic switches has recently promulgated attention from the perspective of modeling and simulation. Atomic switches are known to operate through two mechanisms: (1) formation/dissolution of conductive filaments, and (2) a phase transition between monoclinic acanthite (α) and body centered cubic argentite (β) within Ag₂S. Application of a bias voltage across the junction has been shown using transmission electron microscopy to induce the formation of nanoscale conducting channels across the Ag₂S interface through a bias-catalyzed phase transition, converting the surrounding α -Ag₂S matrix to the conductive β -Ag₂S phase which exhibits high super ionic mobility (Fig. 3).

Voltage Pulsed STMLTM

In the absence of continued applied bias, the conductive channels eventually return to their stoichiometric, thermodynamically favored equilibrium state, reverting the atomic switch to its initial high resistance. This transition gives rise to a weakly memristive behavior prior to the formation of Ag filaments across the interface. Continued application of bias voltage results in a concurrent increase in current through the device, which then further drives migration of silver cations toward the cathode. At the cathode mobile silver cations are subsequently reduced to metallic Ag⁰, forming a highly conductive Ag nanofilamentary wire. The completion of this filament results in a strong transition to an ON state defined by a significant increase in conductivity with a typical conductance ON/OFF ratio of $\sim 10^5$ [59]. Removal of the applied bias results in filament dissolution as the device again returns its thermodynamic equilibrium state. The completion and dissolution of this filamentary structure characterizes strongly memristive behavior. Continuous application of a bias voltage serves to increase filament thickness as additional silver cations are reduced, causing thickening of the metallic filament. This dynamic process has been shown to alter the dissolution time constant, and can be externally controlled by changing the input bias pattern (e.g. pulse frequency). Such changes in volatility can be interpreted as long-term or short-term memory (LTM and STM) (Fig. 4).

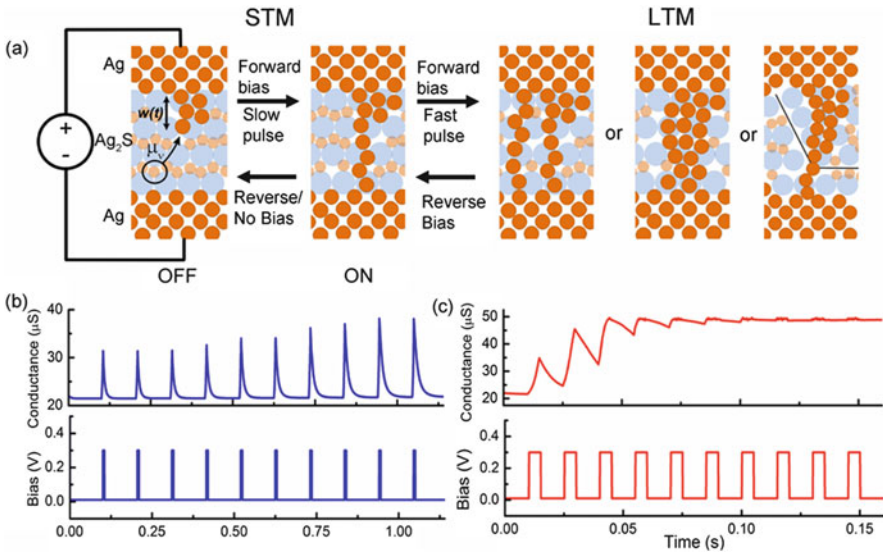


Fig. 4 Resistive switching and long/short term memory effects in at an Ag-Ag₂S-Ag junction arise from (a) increased Ag⁺ mobility in the presence of an externally applied electric field. (b) Short pulses reduce Ag⁺ to form a conductive Ag filament which will quickly re-dissolve in the absence of an applied bias acting as short term memory. (c) Longer pulses of the same amplitude are capable of generating long lasting filaments acting as longer term memory. This likely arises from a combination of multiple filament formations, thicker filaments formed and Ag⁺ ions which have irreversibly crossed a grain boundary until the external bias is removed (AZ Stieg Memristor Networks 2014)

5.2 Device Activation and Switching

The ionic resistive atomic switch has been shown to exhibit fascinating electrical properties analogous to short- and long-term plasticity at single synaptic junctions while operating as a two-terminal device controlled through formation/annihilation of a metal filament within a Metal-Insulator-Metal (MIM) interface. However, the behavior of a collection of atomic switches directly coupled both spatially and electrically is as of yet unknown. In common with the current understanding of switching mechanisms in devices based on Ag₂S, TiO₂, and Ta₂O₅, our network required an initial forming step to create a short-lived high conductivity ‘ON’ state. As measured by current-voltage (I-V) spectroscopy, ASN devices demonstrated non-linear I-V characteristics comprised of a sequential decrease in network resistance with consecutive bias sweeps followed by an abrupt transition to the activated ‘ON’ state. This cascade-type activation required a higher switching voltage (~7 V); however, they exhibit similar time constants (μs) to single atom switches. The switching behavior in the fractal network is attributed to an increased spreading resistance of many individual switching interfaces. In contrast, un-sulfurized control devices comprised of a purely metallic network have substantially lower resistance

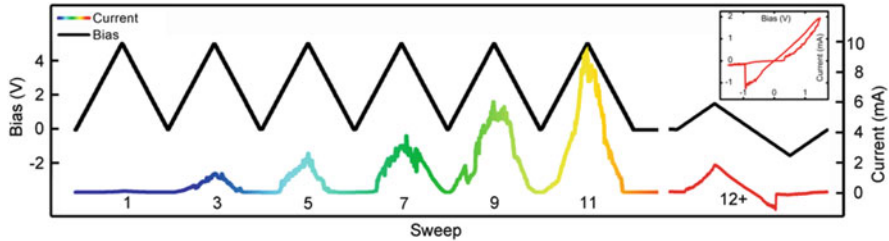


Fig. 5 Activation sweep of an ASN device showing the electrically induced filament formation step. A signal of ± 3 V was input in one corner of the device at a frequency of 10 Hz and current was collected across the network (Avizienis PLoS 2012)

(<100 Ω) and demonstrated repeatable linear, ohmic I-V characteristics at intermediate voltages (± 3 V) followed by irreversible breakdown (melting) at high bias (Fig. 5).

Reproducibility of the switching behavior observed in single atomic switches, i.e. I-V hysteresis, short- and long-term memory, and device activation, were validated using the ASN simulation. In addition, the simulations faithfully reproduced the various emergent properties specific to the ASN architecture. These efforts allow detailed investigation of internal dynamics of the network where it would otherwise have been experimentally impractical. Similar to the electroforming step observed in individual memristive elements, ASNs must undergo an activation process before they display memristive and emergent behaviors [44, 60]. Freshly fabricated ASNs contain Ag_2S interfaces in their low-temperature insulating phase and function as quasi-ohmic resistors. Bias voltage sweeps of the virgin-state network devices exhibited weak memristive/soft switching behavior as silver cations initially migrate into the junctions and are characterized by pinched hysteresis current-voltage curves with a small RON/ROFF ratio and a smooth transition between the two states (Fig. 6a). Continued application of a bias voltage produced an abrupt, nearly discontinuous jump to a state of higher conductance (Fig. 6b). Repeated stimulation with bipolar bias voltage sweeps produced strong memristive/hard switching behavior, typified by abrupt switching between two distinct resistance states (Fig. 6c). While parameters such as threshold voltage and the RON/ROFF are to an extent device specific, the qualitative transition from weak to strong memristive behavior is a general property of the ASNs.

This observed phase transition has been theoretically predicted in simulations of memristor networks [61] and was reproduced in ASN simulation [49]. The transition from soft to hard switching results from the emergence of distinct spatial patterns corresponding to individual hard and soft switching elements (Fig. 6a'). The initial weakly memristive state was characterized by a large fraction of soft switching junctions. As net flux through the network increased, connections became increasingly polarized and conductive. Continued stimulation eventually yields the formation of a percolative pathway comprised of conductive, hard switching elements across the simulated network (Fig. 6b'). Completion of this pathway results in a

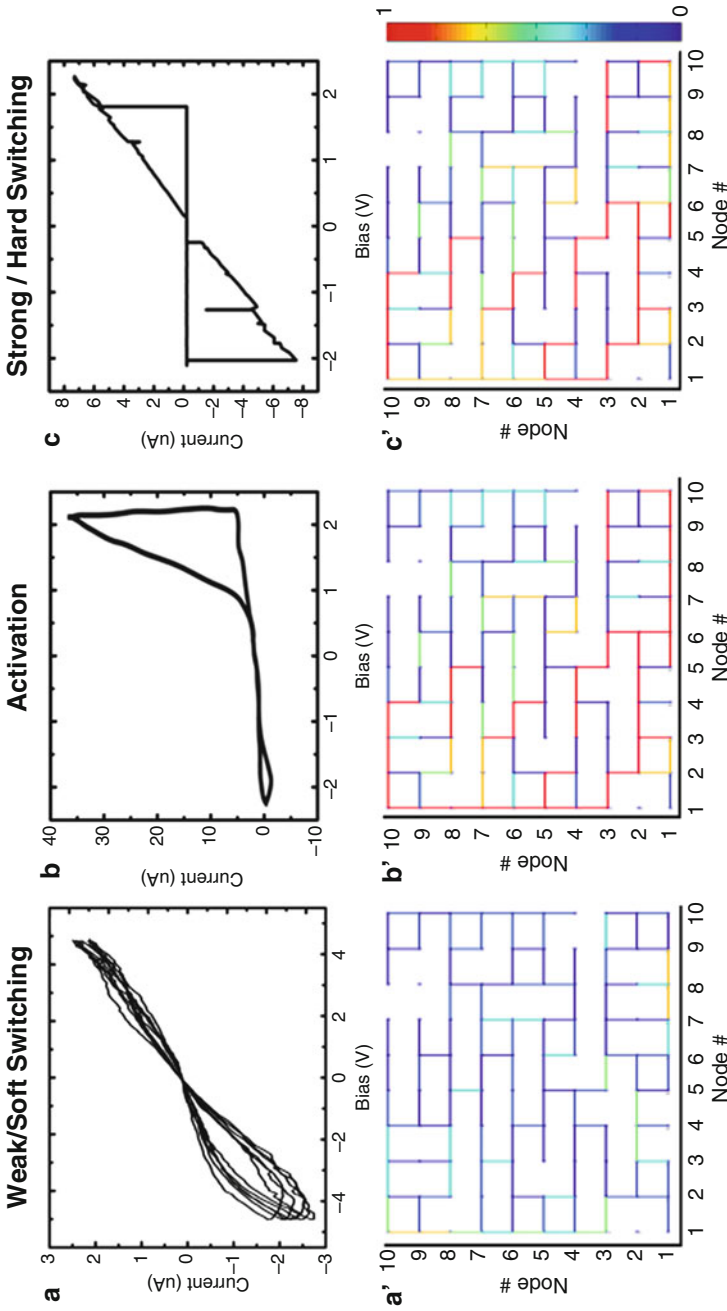


Fig. 6 Simulations of device activation confirmed experimental results where (a) an initial weakly memristive switching ($\alpha = 10\%$) repeats indefinitely until (b) a transition in behavior from weak (blue, $\alpha = 10\%$) to strong (red, $\alpha = 0\%$) switching occurs. Strongly memristive switching (c) persists indefinitely with $\alpha = 0\%$. Internal network connectivity maps (a'–c') reveal conductive pathways in the soft switching, transitional, and hard switching state (Sillin Nanotechnology 2013)

dramatic change in conductance associated with the activated state and a concurrent shift from weak to strong memristive behavior. Subsequent hard switching was observed following the destruction of this highly conductive pathway, as strongly memristive elements were redistributed throughout the network, increasing the probability that connecting a given link would create an equivalent highly conductive path (Fig. 6c'). This is an example of a dynamical self-organization process: different ASNs can have very different initial conditions, yet the basic features of their functional units and network topology cause similar patterns of behavior to emerge during activation and subsequent operation.

Based on experimental and simulation results, a description of physical processes in the ASN was formulated to describe the activation process based on the two mechanisms described in the preceding paragraph: a bias-catalyzed phase transition of Ag_2S , and the subsequent Ag filament formation. A weakly memristive effect is caused primarily by a distribution of phase-transition driven atomic switches, with a small fraction of filamentary driven switches. As overall conduction and the fraction of hard switching elements increases, the electric field intensifies across the remaining soft switching junctions, encouraging further filament formation. Network response changes from weak to strong memristive behavior when a percolative pathway of hard switching junctions forms across the network. Having undergone this transition, the continuously swept network operates as a hard switching memristor, as only a few local switching events are required to reconnect an equivalent path.

5.3 *Coupling and Harmonic Generation*

Individual atomic switches were shown to be directly coupled in configurations using a shared ionic conducting layer, even when separated by large distances. Spatially distributed atomic switch junctions interact through local variations in ionic concentration and electrochemical potential that depend on the combined electrical resistance of the entire network as well as the configuration, or state, of all other electro-ionically interconnected switches.

While 'weak' and 'strong' behavior can be exhibited by single elements, the most interesting features of this complex atomic switch device are its network-specific properties. Infrared imaging was used track Joule heating from current flow during DC bias sweeps in order to confirm distributed network conductance. The IR images (Fig. 7) show power dissipation occurring across the network, indicating that the phase change in network I-V behavior was not attributable to percolation but rather due to the sum of parallel current flows, meaning that network structure and connectivity actively influence device function. Additional evidence for the distribution of switch function stemmed from the analysis of the device's frequency response. Theoretical simulations indicate that second harmonic generation will occur under an applied sinusoidal voltage in networks whose percentage of hard switching junctions exceeds the percolation threshold. Furthermore, the relative

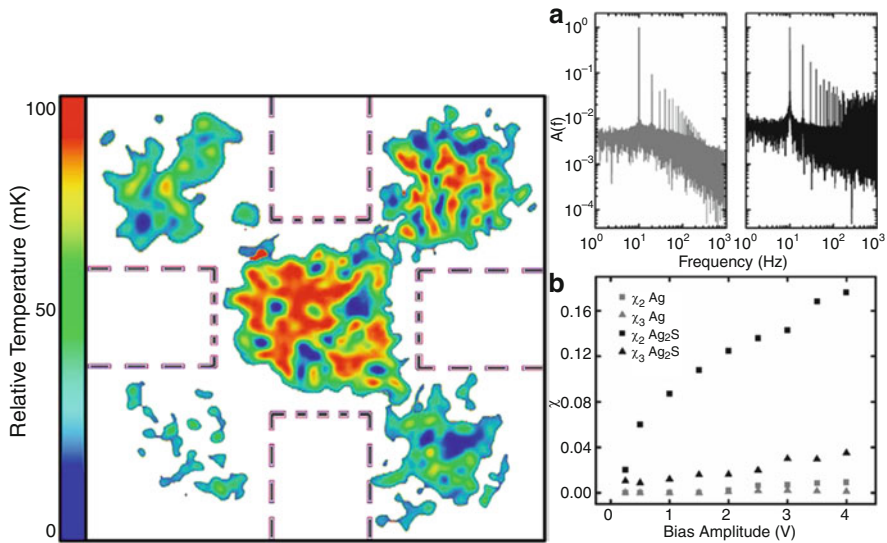


Fig. 7 Network-specific behaviors. (Left) Representative IR image (sensitivity <20 mK) of Joule heating in atomic switch network during bias sweeps indicating current flow distributed throughout the device. Electrode positions are indicated by dashed lines. The image was taken from data integrated during <1 min of bias. (Right) Frequency Response (a) Fourier transforms of a functional device's current response (black) to a 2 V, 10 Hz sinusoidal input signal shows enhanced overtones of the input signal with respect to a control device (gray). (b) Plot of normalized amplitudes (χ) of 2nd and 3rd harmonic generation for varying sinusoidal signal voltages in both functional (black) and control (gray) networks. Sulfurized networks generate higher harmonics, as was theoretically predicted for random memristor/resistor networks with 80% or more strongly memristive (switching) elements

magnitude of higher harmonics increases with the relative number of hard switching junctions. Device response to a 10 Hz sinusoidal voltage signal revealed a large increase in higher frequency components after functionalization (Fig. 7). The proportion of higher harmonics generated increases with signal amplitude, indicating that the network contains a distribution of switching voltage thresholds. As the bias voltage increases, so does the number of memristive junctions operating in the hard switching regime, producing a larger degree of higher harmonic generation. This confirms the IR observation of distributed activity throughout the network, with different regions activating at different voltages.

5.4 Memory and Plasticity

A general objective in designing a functional device platform included a direct interface between memory/logic elements embedded in ASN architecture and

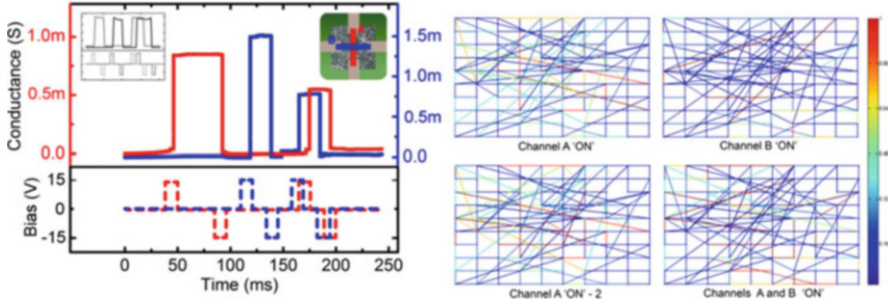


Fig. 8 (Left) Simulation of spatially overlapping channels being modified independently by write/rewrite pulses, emulating the 2-bit switching functionality of actual device behavior (inset). (Right) Simulated internal network configurations ($N = 219$) at different ON/OFF configurations describing the formation of feedforward assemblies. In ON states of the network, conductances do not distribute uniformly. In fact, the simulation shows that several different configurations may correspond to the same ON/OFF channel configuration depending on the history of channel switching. For example, the internal configurations responsible for the ON of channel A at the two time points when it is activated before/after the activation/deactivation of channel B (blue), is shown

externally controlled input/output contact electrodes allowing for data processing. In the current device configuration, ASNs can only be electrically probed using the macroscopic interface electrodes. It was therefore essential to confirm that these electrodes could be effectively coupled to local ensembles of atomic switches within a particular spatial region of the network. Experimental observations of network plasticity [44] as a mechanism for the formation of feedforward pathways within ASNs was addressed through simulation [49] as seen in Fig. 8. Monitoring the conductance of all electrode combinations throughout the stimulation regimen revealed dynamic patterns of activity in regions free from intentional manipulation. The coexistence of localized changes in network connectivity alongside complex system-wide correlations suggest a capability for autonomous, higher-dimensional information processing through formation of specialized functional regions.

5.5 *Fluctuations, Correlations and Power Laws*

Our group examined the ASN device for emergent properties considered fundamental to brain function, which are not observed for individual atomic switches operating in simpler geometries, namely recurrent dynamics and the activation of feedforward subnetworks [44]. The presence of recurrent loops and dynamics within the ASN devices were demonstrated by applying a constant DC bias (Fig. 9a) across a particular region of the network. This produced persistent, bidirectional fluctuations—both increases and decreases—in network conductivity.

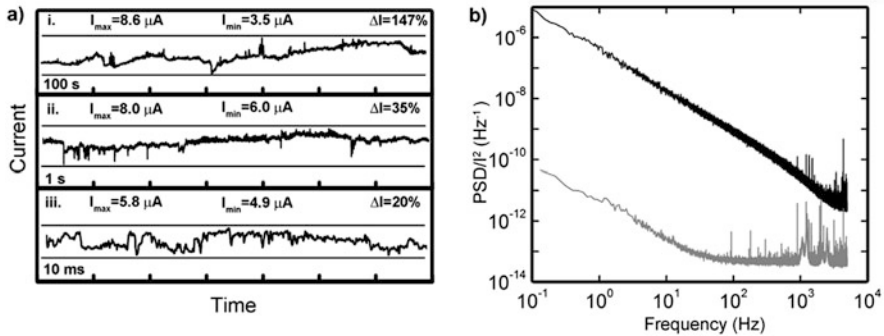


Fig. 9 DC response (a) time traces of current response to 2 V DC bias show current increases and decreases at all time scales around a mean of $5.81 \mu A$ (standard deviation $0.88 \mu A$), behavior specific to recurrent AS networks. (b) Fourier transforms of DC bias response for Ag control (grey) and functionalized Ag-Ag₂S (black) networks. The power spectrum of the functionalized network displays $1/f$ power law scaling, indicating a high level of temporal correlation and memory (Avizienis PLoS 2012)

In the absence of recurrent structures within the network, conductivity would increase monotonically under constant DC bias, as in the case of a single atomic switch. However, bidirectional fluctuations in the current response persisted for several days under constant applied voltage, demonstrating that the complex network connectivity inherently resists localized positive feedback that would lead to the serial formation of a single, dominant high conductivity pathway between electrodes. Previously unreported current fluctuations of this kind are ascribed to recurrent loops in the network that create complex couplings between switches, resulting in network dynamics that do not converge to a steady state even under constant bias. A single switch turning ON does not simply lead to an increased potential drop across the next junction in a serial chain, but redistributes voltage across many recurrent connections that can ultimately produce a net decrease in network conductivity. These fluctuations are not attributable to noise, as shown by comparing the Fourier transformed current responses (Fig. 9b) of the devices to constant voltage before and after functionalization. The formation of atomic switch junctions expands the degree of correlation in current fluctuations, producing $1/f$ -like behavior across the entire sampled range. This behavior is distinct from that of control devices (unsulfurized silver network, grey line in Fig. 9b), which flattens to white noise and some high energy, high frequency fluctuations attributed to arcing between neighboring wires.

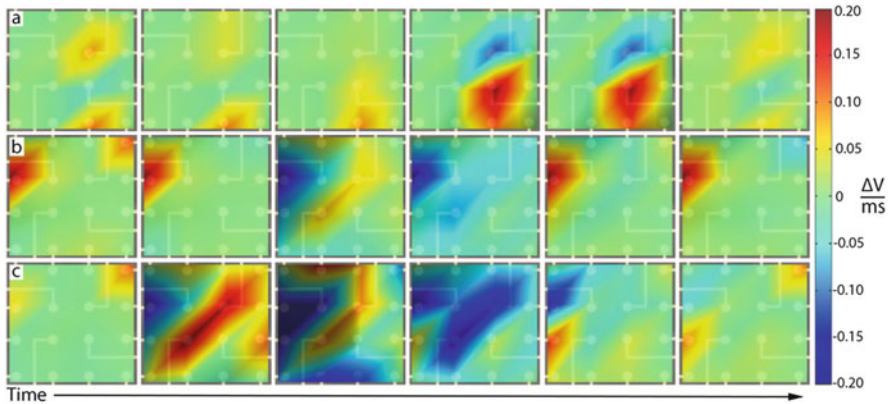


Fig. 10 Spatial temporal switching activity in the ASN is seen in rows (a)–(c). Each row represents 3 ms with a 0.5 ms frame rate while a 5 V DC bias was applied at the upper right electrode and grounded at the lower left electrode. Both (a, b) localized switching, and (c) distributed switching affecting the entire network are observed (Demis Nanotech. 2015)

5.6 Distributed Switching/Correlations

Further examples of correlations within the network exist as local and global switching events are monitored in the form of voltage fluctuations at each electrode. With an applied DC bias, the network moves through different patterns of activity. These patterns include local perturbations seen in rows (a) and (b) of Fig. 10, as well as large cascading switching events distributed throughout the whole network (Fig. 10c). Potential changes vary through time from low to high as current flows throughout the complex architecture. Electrodes many orders of magnitude larger than the individual atomic switches capture a general potential map along with the separability of outputs. Thus, the highly recurrent and interconnected coupling of individual switches leads to the emergence of distributed activity, which enables for a wide array of outputs and in turn potential applications in alternative computing paradigms.

5.7 Temporal Metastability and Criticality

To our knowledge, the atomic switch network represents a unique implementation of a purpose-built self-assembled network composed of coupled non-linear elements that clearly demonstrate the essential characteristics of criticality, specifically power-law scaling of: $1/f$ fluctuations, energetic avalanches, as well as temporal metastability. The emergent complex behaviors observed in their temporally metastability indicate a capacity for memory and learning via persistent critical states with

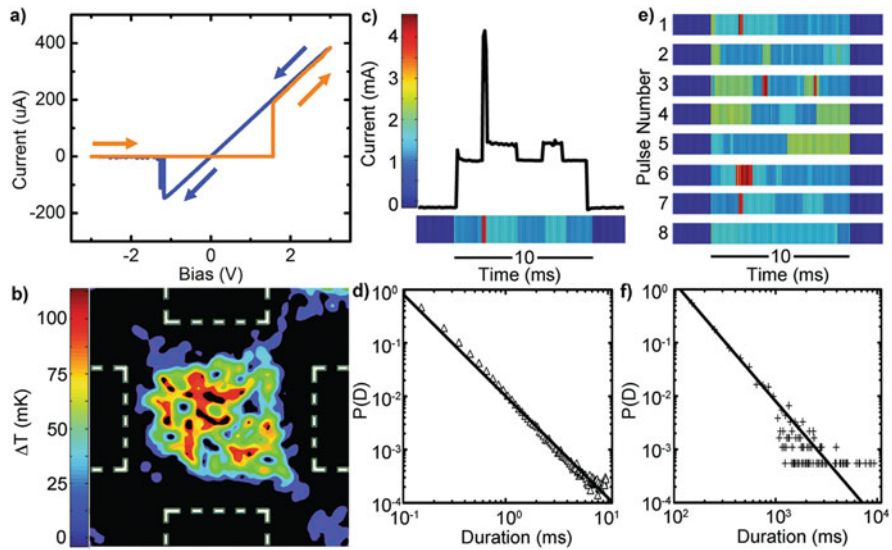


Fig. 11 Electrical characteristics of complex nanoelectronic networks. **(a)** Experimental I-V curve demonstrating pinched hysteresis; $\text{RON} = 8 \text{ K}\Omega$, $\text{ROFF} > 10 \text{ M}\Omega$. **(b)** Ultrasensitive IR image of a distributed device conductance under external bias at 300 K. **(c, e)** Representative network current response to a 2 V pulse showing switching between discrete, metastable conductance states. **(d, f)** Temporal correlation of metastable states observed during pulsed stimulation demonstrated power law scaling for probability, $P(D)$, of duration. Power law scaling existed for residence time both **(d)** within a single 10 ms pulse and **(f)** over 2.5 s during extended periods of pulsed stimulation (Stieg Adv. Mat 2012)

potential utility for the creation of physically intelligent machines capable of evolution and learning. In addition to the exciting behaviors described in Sect. 5.6, complex networks of coupled nonlinear elements such as this commonly manifest non-trivial spatiotemporal evolution through dynamic system reconfigurations. Such reconfigurations, readily described by critical dynamics, enable enhanced maintenance of system correlations and more effective signal propagation. Indicators of criticality typically include power-law scaling of $1/f$ fluctuations and temporal metastability. Analysis of the power spectral density of network conductivity in the activated state revealed $1/f$ power law scaling over five orders of magnitude with $f \approx 1.4$ (data not shown). Temporal metastability was observed during sub-threshold pulsed voltage stimulation, analogous to methods employed in neuroscience to probe cortical cultures. Under typical conditions, the current response fluctuated over a wide range of metastable conductance states associated with discrete network configurations (Fig. 11c-f), as classified by residence times in a given state ranging from milliseconds (within a single stimulation pulse) to several seconds (across hundreds of pulses). Comparing the probability of state duration with its time duration indicated a power law distribution with ≈ 1.8 (Fig. 11e-f), indicating a

diverging temporal correlation length. Observation of both increased and decreased conductivity during stimulation, similar to fluctuations under DC stimulation described above, were again attributed to recurrent network dynamics.

Finally, opportunities for teaching and learning through utilizing metastable critical states were initiated through close collaboration with our theoretical team. This process required: the development of an analytical model for the complex behavior observed in such devices, the design and implementation of teaching algorithms, and exploration of means to interact with the network.

5.8 Altered Critical Power-Law Dynamics

ASN devices have also demonstrated altered power spectral density, or PSD, slopes. As seen in Figs. 1 and 2, the mean current of a given state showed a marked dependence on both the length of the state and the probability ($P(D)$) of its occurrence where longer state durations and increased activity, characterized by $1/f$ alpha power-law scaling of PSD, were observed for intermediate mean current values. In an effort to control these properties, a current-controlled feedback loop was implemented in a similar fashion to that shown in Fig. 3. Real-time maintenance of a defined current set-point between two arbitrarily defined electrodes was readily achieved through application of an applied bias voltage (Fig. 14). While persistent fluctuations in ASN conductance are known to exhibit non-trivial spatiotemporal correlations characterized by power-law scaling of PSD [44, 46], procedures to control such correlations have yet to be reported. Utilization of the current-control approach provides a direct method to tune network dynamics as seen in Fig. 12, where higher current setpoints generated larger, more rapid network reconfigurations in the form of resistance switching as indicated by steeper PSD slopes (α) for both current and local voltage. Reliable transitions between resistance states, and thus

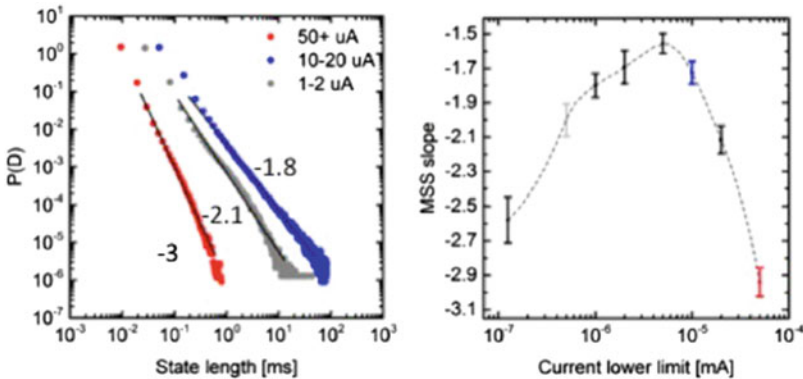


Fig. 12 Representative probability distribution $P(D)$ of metastable state duration (left) obeyed power law scaling with exponents dependent on the mean current during a given state (right)

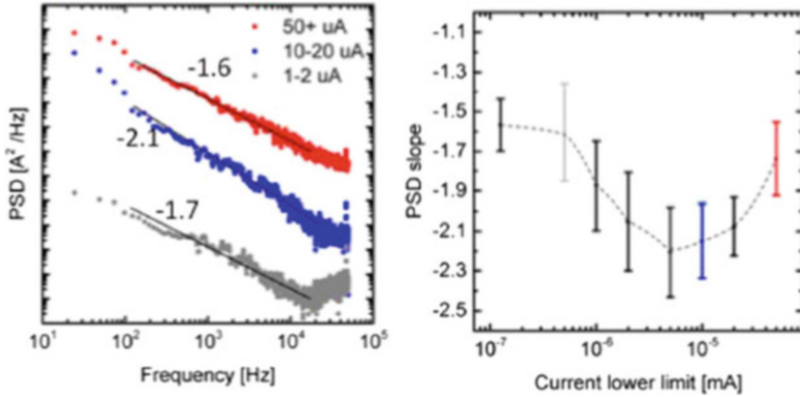


Fig. 13 Representative example of power law ($1/f^\alpha$) scaling (left) of the PSD slope (α) which was observed to depend on the mean value of the current output (right)

regimes of operational dynamics, was achieved as seen in Fig. 13. Finally, cross-correlation analysis of spatiotemporal correlations in device activity at various points throughout the network provided further support of the current set-point as a control parameter for network dynamics. During periods of limited activity, observed correlations were attributed to shared background noise. Periods of activity, at higher current set-points, resulted in a diversity of voltage recordings throughout the network characterized by broadening of correlation coefficients.

6 Computing with the Atomic Switch Network

6.1 Theoretical Constructs

The ASN is one of a limited number of CMOS compatible platforms capable of performing RC [62]. Within the context of the RC formalism, each atomic switch is a functional node in the reservoir and the connective weights between each node are mediated by the silver nanowires. The multi-electrode array on which the network is grown can be adjusted to control input and readout functionality of the electrodes to measure all nodes in 10–50 μm regions of the network. All necessary criteria such as short-term memory, increased fault tolerance, and an arbitrarily scalable number of higher-dimensional outputs are fulfilled by the ASN. The memristive behavior of individual atomic switches bestows the ASN with a fading memory characteristic. This ensures that previous inputs to the network do not exert considerable influence over the current state. The power-law dynamics (Fig. 12) indicate that the system has a scale-free topology that allows it to operate at the “edge-of-chaos,” a dynamical regime providing a balance between memory and instability. The non-linear transformations Fig. 14 are an intrinsic behavior of the system that can be harnessed to

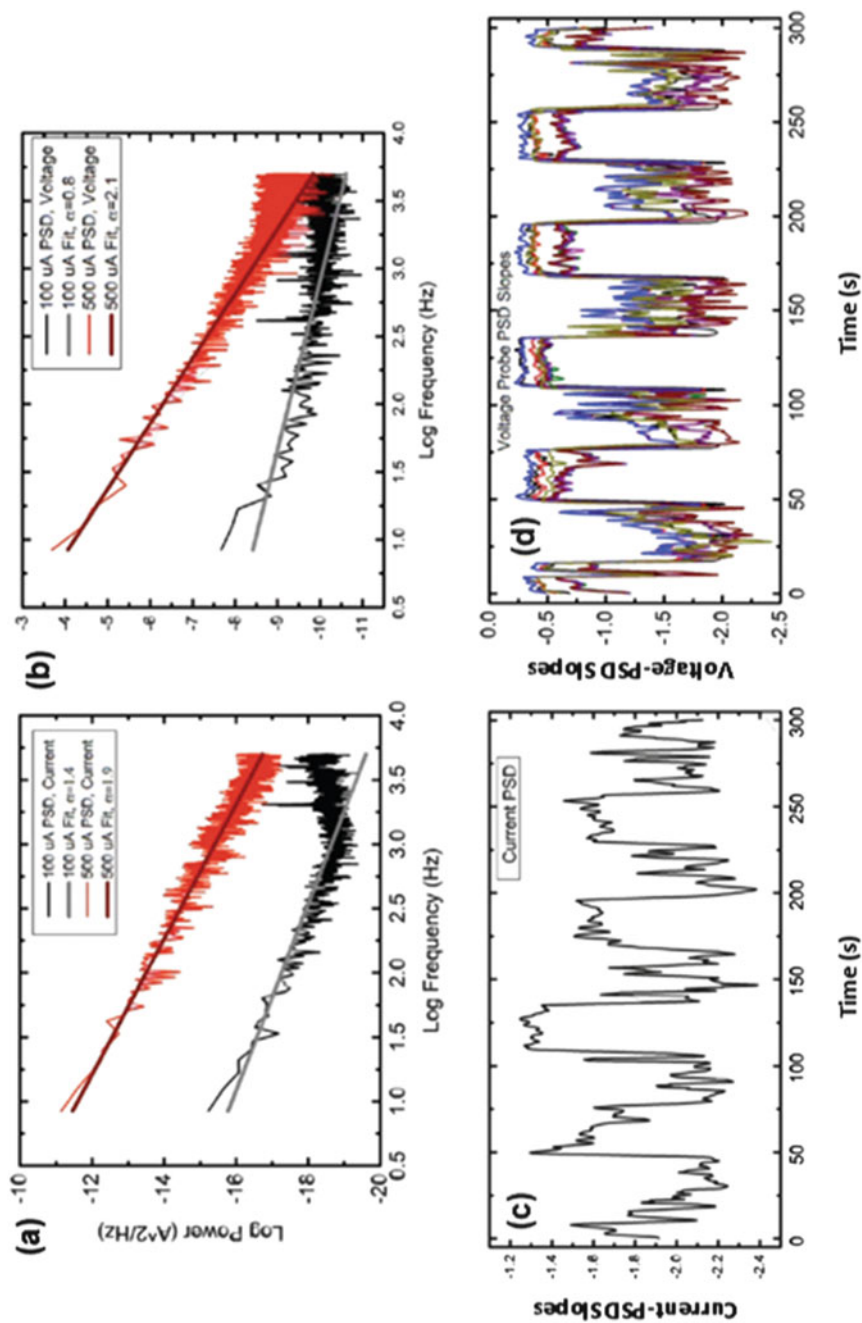


Fig. 14 Dependence of power law scaling (PSD α value) on current setpoint using a current-controlled feedback loop operating in real-time through control of the applied bias voltage. PSD slope (α) of the current output (left) and simultaneously acquired voltage (right). Repeated cycling of current setpoint demonstrated an ability to tune the PSD slope (α), and hence temporal dynamics, of the current output (top) and local voltage throughout the ASN device (bottom)

increase performance. These transformations manifested as separable unique voltage signals and are a consequence of the input signal entering a higher representational space that is then used in reconstructing the desired target waveform. Unlike current computational models of explicitly programmable algorithms, RC relies on systems operating in a regime where they are able to ‘learn’ through experience, circumventing the need for intelligent programming.

Generally, RC utilizes a randomly connected network, dubbed the ‘reservoir,’ composed of interacting elements called neurons with a topology based on mammalian brain neural networks. Information propagates through network connections via electrical or electrochemical signals which are preferentially directed towards neurons with the greatest connective strength. Neurons may be activated by the incoming signal based on a set of learning rules causing them to undergo internal modifications and excitatory transformation of the incoming signal. This stimulated response of the neuron is then communicated through the neuron’s outgoing connections, thereby allowing the signal to percolate throughout the network. Computation is achieved by recording outputs from a few neurons assigned during initialization and these neurons are trained to achieve the desired functionality. Training resembles the evolutionary phenomenon observed in biological and natural systems adapting due to environmental changes. External stimulation of the reservoir allows the system to independently evolve into a number of diverse neurons due to signal propagation and local interactions which activate and modify neuron properties. Clusters of neurons are capable of displaying distinct activity and emergent behaviors due to local interactions interplaying with signal percolation. Training selects a number of clusters desirable for computation and reinforces these neurons for specialization into relevant mathematical processors. After the system is properly trained using sample tasks, the neurons are passivated such that the system retains the knowledge and experience for future computation. RC is both a simple and elegant construction that avoids the need for absolute control over programmable elements while capable of producing powerful processors due to evolutionary concepts. Performance is controlled by neuron connectivity and distribution of strong and weak connections. Operational utility is thus dependent on the reservoir’s statistical characteristics and global parameters, focusing on emergent qualities of the network instead of individual elements.

The non-linear dynamics exhibited by the ASN device is uniquely positioned to be exploited using the paradigm of reservoir computing [49, 63]. Interactions between atomic switches in the network produce non-linear transformations of input signals not present in Ohmic resistors or single switches. The additional non-linear transformations of the input signal increases the diversity of the output signals and thereby increase output separability and potential for computational capability. In the reservoir computing approach, the input signal undergoes a projection into a higher dimensional representation space, which can be thought of as an expansion of the output into a sum of mathematical elements [64, 65]. Having a rich collection of elements provides a large repertoire of possible transformations of the input via mathematical representation vectors. The separability and utility of multiple transformations enable the construction of mathematical algorithms by

integrating desired representations. Within this representation space, reservoir computing is able to identify useful elements of the output function to attempt to solve a specific task. Depending on the degree of non-linearity, which in turn determines the size and resolution of the reservoir, a task may be optimally solved in this higher dimensional representation space. In the ASN device, variations in connectivity, atomic switch density, and junctions contribute to the degree of non-linearity by creating a larger range of functional elements that have different activation and operational voltages.

6.2 Implementations

6.2.1 Waveform Regression

Simulations of ASNs have indicated that the system has the fundamental capacity to perform waveform regression [66]. From simulation results, performance depended on the level of higher harmonics produced and the harmonic distortion (Fig. 16) required for the specific task. For example, the cosine task only requires a shift in its periodicity and, therefore, does not require extensive higher harmonic generation. Conversely, the square wave task requires infinitely distributed harmonics to produce a straight line through wave interference. Further, voltage dependent simulations showed that increasing device activation controlled these harmonic generations. Here, the device was expected to perform in a similar way with task difficulty increasing from cosine, triangle, sawtooth, and square due to increasing harmonic requirements. Device initialization and activation to achieve the best performance is described in the previous section.

Experimental performance of various waveform regression tasks using ASN devices are presented in Fig. 15. To implement waveform regression the ASN was stimulated with a bipolar sinusoidal voltage, inducing switching activity and placing the network in an active state. The output potentials measured at each electrode were then combined using the Moore-Penrose linear regression and optimized during a training period [67–69]. Two-second epochs of data were used to evaluate the ASN's computational capability, where 1 s of data was allocated for both training and testing. Performance was measured during a 1 s period after training where the ASN accomplished various tasks (Fig. 15). The performance of the ASN was quantified by calculating the normalized mean squared error between the target and generated waveforms [70]. Here, the difference between error and unity was used to calculate accuracy.

The ASN was capable in achieving up to ~90% accuracy using 62 of the 64 measurement electrodes for each task. Task complexity increased from cosine to square wave due to the increasing mismatch between the sinusoidal input and the target waveform. In the case of cosine generation, the overall waveform of the input is preserved save for a shift in its periodicity. The cosine generation was the simplest task where the ASN performed with the highest accuracy, ~90%. Note that the

Waveform Generation

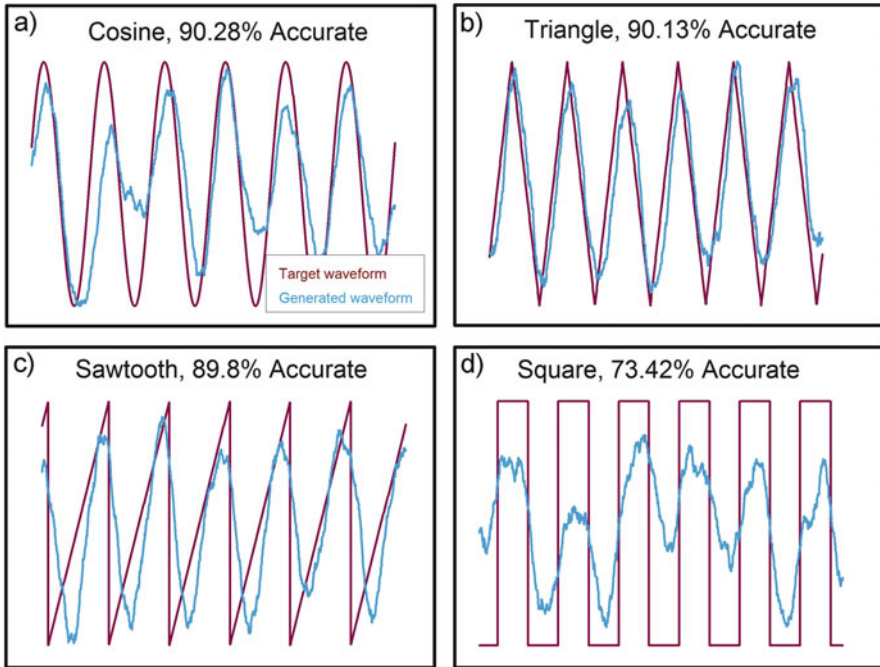


Fig. 15 Computation of a sinusoidal wave into various waveforms. The above figure shows several waveforms (sawtooth, square, triangle, and cosine) produced using the ASN as a computational device using the setup in Fig. 4. Each plot contains the desired signal (red) and the computed signal (blue) with their accuracy w.r.t. the desired signal shown above the curves. All tasks share an 11 Hz frequency for their waveforms and share the same dataset with only differences in the target task. The dataset was approximately 1 min long, divided into 2 s epochs, and 1 s within each interval was allocated for training and testing. A 1 s excerpt which best represents device behavior during testing are shown above (Sillin Nanotechnology 2013)

cosine regression shown in Fig. 15a would not be possible using a grid of regular resistors due to their intrinsic linear response. Since individual atomic switches have a non-linear memristive response, it is possible to harness that state function into the highly recurrent structure of the ASN. The highly recurrent structure allowed higher levels of coupled interactions that cannot be captured by a single atomic switch, resulting in emergent behaviors. Particularly, the network was capable of producing delayed responses and enabled the network to shift the phase of the input signal by a half-wavelength, producing a cosine.

Figure 15b shows hardly any mismatch in the triangle generation, achieving a similar $\sim 90\%$ accuracy and visually validates the performance metric used throughout our analysis. A similar argument is used to explain the high performance of the triangle wave when compared to the cosine task. The determining factor for reservoir performance is the level of similarity between the target and input signal, where the

reservoir acts as a transformational operator to minimize dissimilarities. In both cases, the target waveform is aesthetically similar to a sinusoidal wave and maintains the overall shape of the input signal. Despite steeper edges in the triangle task, the algorithm is able to correct any differences by selectively combining different representations produced by the ASN.

The ASN generated sawtooth (Fig. 15c) waveforms with similar accuracy to previously reported simulations of memristive networks at roughly 90% accuracy [71]. Despite the requirement to produce an instantaneous drop, the ASN delivered the sawtooth waveform with astounding accuracy. Figure 15c illustrates significant mismatch between the target and generated waveform at the turning point leading to a minor drop in accuracy. Basic visual inspection shows the sawtooth task retains the overall shape of the sinusoidal input while the square wave task requires complete transformation of the input signal into a two-valued function.

Figure 15d, on the other hand, shows significant mismatch throughout the series. The square wave generation was carried out with roughly a 78% accuracy, which was much lower than the accuracy of the other tasks. To recreate a straight horizontal line, an infinite series of higher harmonics is necessary in order to satisfy the spectral theorem in the algorithm [49]. Fourier analysis showed that the square wave task was relatively selective in utilizing higher harmonics to construct the waveform. While the sawtooth and square wave both require an infinite series of sinusoidal harmonics, the square wave requires continuous constructive interference patterns to produce a horizontal line, which limits it to odd or even harmonics and drastically diminishes the regression algorithm. In this case, the ASN was only capable of producing a finite number of higher harmonics. However, further post-processing such as setting a threshold on the voltage to binarize the data can be performed to expand the device's response to a square wave input, a necessity for reliable Boolean logic computing [72].

It was found that the ASN was capable of replicating computing performances typical of reservoirs with 10^3 output signals [73]. Theoretical studies predicts the performance to scale with an increasing number of output signals due to the dependence on the regression algorithm [74]. However, how can a reservoir with much fewer output signals outperform reservoirs with output signals orders of magnitude higher than the ASN? Further inspection of the mathematical formalism [74] show that performance is additionally characterized by the uniqueness of each output signal. Obtaining a set of unique signals allows us to linearly combine the output signals into a number of unique solutions, where the number of unique solutions scales with the number of unique output signals. The larger set of solutions increases the size of the "net" we cast which increases the probability and approximation of producing the correct solution (Fig. 16).

6.2.2 Logic

Expanded efforts to assess their performance in Boolean logic operations using non-temporal inputs based on randomized Boolean input streams. Zero and one

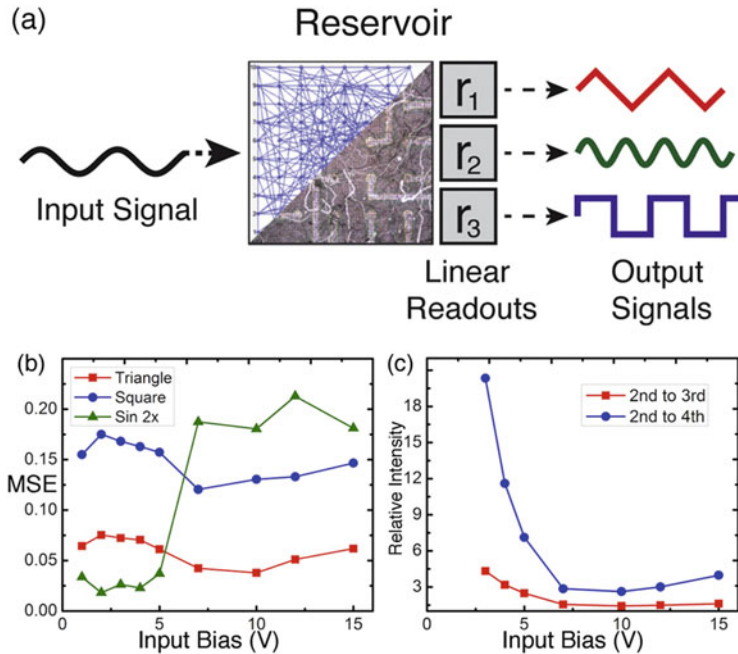


Fig. 16 (a) Schematic of network simulation used in the waveform generation RC task, with specific electrodes chosen as inputs/outputs (16 output electrodes). RC was implemented using a 10×10 node network with a 5 V, 10 Hz sinusoidal input signal and tasked to produce 10 Hz triangle/square and 20 Hz sinusoidal waveforms. (b) Mean-squared error (MSE) for each task with respect to driving amplitude showed minimal error in triangle/square waveform generation task at 10 V, corresponding to the onset of higher harmonic generation (see red curve of Fig. 6b). Performance in the 20 Hz sinusoidal waveform generation task decreased when (c) the relative amplitude of the average 2nd harmonic intensities of the readouts becomes increasingly diminutive. These results correspond to a strong dependence on the 2nd harmonic for 20 Hz sine generation and the need for HHG in triangle/square generation as expected by Fourier analysis (Sillin Nanotechnology 2013)

bits were converted to negative and positive DC voltage pulses, respectively. Next, a linear readout layer was applied to an array of voltage outputs from the device to reconstruct target output signals for the given task. ASNs produced nearly perfect results at low voltages for AND, OR, and NAND with more than 95% confidence. XOR, which requires non-linearity to solve, was able to be partially solved at high voltages with more than 95% confidence owed to stable, non-temporal, non-linear behaviors in the device under optimized operational conditions. As opposed to previous works which have investigated temporal computation in ASNs, this work was the first to demonstrate semi-predictable, non-temporal, non-linear behavior within the device. These results demonstrate that the device connectivity is complete enough to perform complex computations. With a more comprehensive view of ASN behavior, these devices will be capable of performing functions currently implemented in CMOS while occupying less area and processing more inputs simultaneously (Fig. 17).

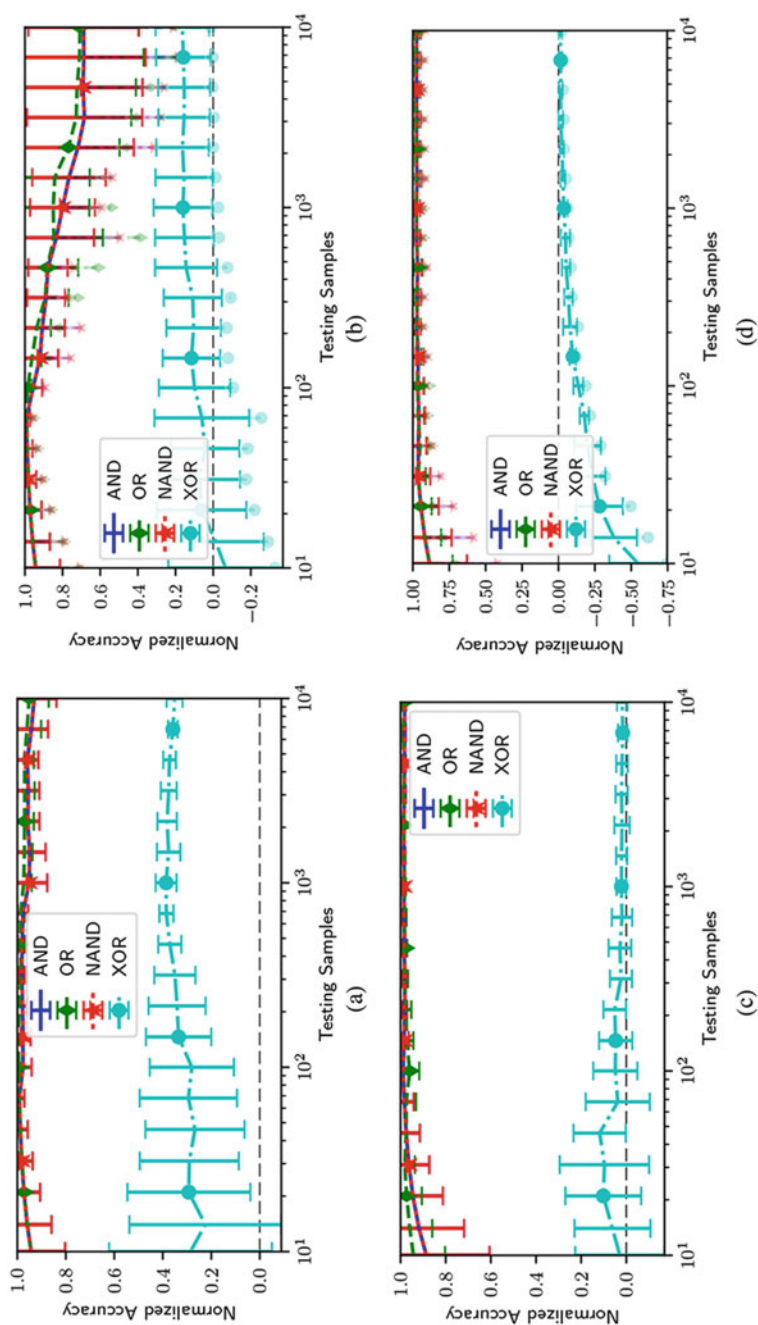


Fig. 17 Accuracy on all Boolean logic functions learned using 1000 training samples. Figure 4a, b used 3 V inputs; Fig. 4c, d used 0.01 V. The left two plots represent normalized accuracy of a single electrode combination. Error bars represent the standard deviations of the electrode combination's performance at different points in time. The right two plots are the 95% confidence accuracies across all electrode combinations; that is, we are 95% sure that a random electrode combination, at any point in time, will produce a regression that will outperform the mean line shown. Faded markers below lines indicate the worst-performing electrode combination's 95% confidence accuracy

7 Outlook

If we consider the future of A.I. certainly we are at a turning point in history where increasingly we see the practically limitless applications in diverse areas such as driverless vehicles, medical diagnosis—healthcare, climate prediction, even social comfort. Our digital computer systems of today are nevertheless being pushed to the ultimate limits of fabrication with the end of Moore’s Law in clear sight. In other words, it will become prohibitively expensive to maintain the advances required for A.I. to create a post-human world. The poor scaling of computer hardware and software required as combinatorial complexity increases will not be overcome in future von Neumann machines.

Predictions of increasing computational capacity in comparison to that of the human brain or the “singularity”, as coined by Ray Kurtzweil are unlikely using digital approaches. The scenario of computational equivalence to a human was proposed to give rise to a sudden and massive increase in dominantly non-biological intelligence, and we propose it can best be approached by using biological inspiration in making a system that has physical operational characteristics closer to ourselves. The Atom Switch and networks of them have a potential role to play in Hybrid-CMOS morphic systems where methods such as Reservoir Computation in physical analogue hardware can be integrated into a morphic system that utilizes the optimum performances of CMOS digital with the ASN approach. Already Atomic Switches have successfully integrated into FPGA devices by NEC reducing energy consumption, footprint, size and transistor count [56]. The Atomic Switch technologies are more robust in terms of sensitivity to electromagnetic noise and radiation than Flash making them candidates for robotic and space satellite applications. The key differences of the ASN approach to conventional computation are in the elimination of programming and error correcting each step of a calculation with the RC paradigm. Likewise, ASN devices use distributed fading memory not RAM similar to living systems. The ability to handle multiple tasks in parallel is another advantage of such an approach. Although accurate calculations of arithmetic operations will always be superior in digital systems, analog systems such as the ASN excel in decision making, or noisy and error prone data that have no precise solution but rather a range of outcomes with a best guess of the outcome probabilities a bit like Newtonian vs. Quantum mechanics where deterministic solutions are replaced by probabilities. The potential impact of A.I. in society has become quite heated in terms of the dangers it poses to our society and discussions of government regulation (Elon Musk) and possibly imposing taxes on A.I. robots and systems have even been proposed (Bill Gates). However, in many advanced countries there is a future need for such technology as baby-boomers retire and the population of a young work force declines A.I. will be essential in healthcare, welfare, national security, and in many other areas of societal enhancements.

References

1. Waldrop, M.M.: The chips are down for Moore's law. *Nature*. **530**, 144–147 (2016)
2. Abbe, E.: *Contributions to the Theory of the Microscope and the Microscopic Perception*. Springer (1873)
3. International Technology Roadmap for Semiconductors (2015)
4. Neumann, J.V.: First draft of a report on the EDVAC. *IEEE Ann. Hist. Comput.* **15**, 27–75 (1993)
5. Backus, J.W.: Can programming be liberated from the von Neumann style? A functional style and its algebra of programs. *Comm. ACM.* **21**, 613–641 (1978)
6. Dongarra, J.: Visit to the National University for Defense Technology Changsha, China. University of Tennessee (2013)
7. Mead, C.: Neuromorphic electronic systems. In: *Proceedings of the IEEE*. IEEE (1990)
8. Hodgkin, A.L., Huxley, A.F.: A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500–544 (1952)
9. Haimovici, A., Tagliazucchi, E., Balenzuela, P., Chialvo, D.R.: Brain organization into resting state networks emerges at criticality on a model of the human connectome. *Phys. Rev. Lett.* **110** (17), 178101 (2013)
10. Wang, X.F., Chen, G.: Complex networks: small-world, scale-free and beyond. *IEEE Circ. Syst. Mag.* **3**, 6–20 (2003)
11. Sporns, O.: Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *Biosystems.* **85**, 55–64 (2006)
12. Abbott, L.F., Nelson, S.B.: Synaptic plasticity: taming the beast. *Nat. Neurosci.* **3**, 1178 (2000)
13. Hebb, D.O.: *Organization of Behavior*. Wiley, New York (1950)
14. Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386–408 (1958)
15. Hopfield, J.J.: Artificial neural networks. *IEEE Circ. Dev. Mag.* **4**, 3–10 (1988)
16. Gomes, L.: Neuromorphic chips are destined for deep learning—or obscurity. *IEEE Spectrum* (2017)
17. Schuman, C.D., Potok, T.E., Patton, R.M., Douglas Birdwell, J., Dean, M.E., Rose, G.S., Plank, J.S.: A survey of neuromorphic computing and neural networks in hardware. *arXiv* (2017)
18. Christie, P., Stroobandt, D.: The interpretation and application of Rent's rule. *IEEE Trans. VLSI Syst.* **8**, 639–648 (2000)
19. Abraham, A.: Artificial neural networks. In: Sydenham, P.H., Thorn, R. (eds.) *Handbook of Measuring System Design*. Wiley, New York (2005)
20. Medsker, L., Jain, L.C.: *Recurrent Neural Networks: Design and Applications*. CRC, Boca Raton, FL (2001)
21. Graves, A., Mohamed, A., Hinton, G.: Speech recognition with deep recurrent neural networks. *ArXiv* (2013)
22. Khotanzad, A., Chung, C.: Application of multi-layer perceptron neural networks to vision problems. *Neural Comput. Appl.* **7**, 249–259 (1998)
23. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature*. **521**, 436–444 (2015)
24. LeCun, Y., Boser, B.E., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W.E., Jackel, L. D.: Handwritten digit recognition with a back-propagation network. In: Touretsky, D.S. (ed.) *Advances in Neural Information Processing Systems*. Morgan Kaufmann, San Mateo (1990)
25. Hinton, G.E.: Learning multiple layers of representation. *Trends Cogn. Sci.* **11**(10), 428–434 (2007)
26. Büssing, L., Schrauwen, B., Legenstein, R.: Connectivity, dynamics, and memory in reservoir computing with binary and analog neurons. *Neural Comput.* **22**, 1272–1311 (2010)
27. Sojakka, C. F.: Pattern recognition in a bucket. In: Wolfgang Banzhaf, J. Z. (ed.) *European Conference on Artificial Life: Advances in Artificial Life* (2003)
28. Goudarzi, A., Teuscher, C., Gulbahce, N., Rohlf, T.: Emergent criticality through adaptive information processing in Boolean networks. *Phys. Rev. Lett.* **108**, 128702 (2012)

29. Tour, J.M., Cheng, L., Nackashi, D.P., Yao, Y., Flatt, A.K., Angelo, S.K.S., Mallouk, T.E., Franzon, P.D.: Nanocell electronic memories. *J. Am. Chem. Soc.* **125**, 13279–13283 (2003)
30. Merolla, P.A., Arthur, J.V., Alvarez-Icaza, R., Cassidy, A.S., Sawada, J., et al.: A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*. 345, 668–673 (2014)
31. Indiveri, G., Linares-Barranco, B., Hamilton, T., van Schaik, A., Etienne-Cummings, R., et al.: Neuromorphic silicon neuron circuits. *Front. Neurosci.* **5**, 1–23 (2011)
32. Shimokawa, Y., Fuwa, Y., Aramaki N. A parallel ASIC VLSI neurocomputer for a large number of neurons and billion connections per second speed. In: *IEEE International Joint Conference on Neural Networks*. Singapore (1991)
33. Omondi, A.R., Rajapakse, J.C.: *FPGA Implementations of Neural Networks*. Springer, Dordrecht (2006)
34. Nurvitadhi, E., Sheffield, D., Sim, J., Mishra, A., Venkatesh, G., Marr, D. Accelerating binarized neural networks: comparison of FPGA, CPU, GPU, and ASIC. In: *International Conference on Field-Programmable Technology (FPT)*, IEEE (2016)
35. Qiao, Y., Shen, J., Xiao, T., Yang, Q., Wen, M., Zhang, C.: FPGA-accelerated deep convolutional neural networks for high throughput and energy efficiency. *Concurr. Comput. Pract. Exp.* **29** (2016)
36. NVIDIA launches the world’s first graphics processing unit: GeForce 256 [Online]. http://www.nvidia.com/object/IO_20020111_5424.html (1999)
37. Jager, C.: Nvidia unveils Volta: the most powerful GPU ever [online]. <https://www.lifehacker.com.au/2017/05/nvidias-unveils-volta-gv100-the-most-powerful-gpu-ever/> (2017)
38. Krizhevsky, A., Sutskever, I., Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems* (2012)
39. Saxena, A.: Deep learning pioneers boost research at NVIDIA AI labs around the world [online]. <https://blogs.nvidia.com/blog/2017/07/11/deep-learning-pioneers-boost-research-at-nvidia-ai-labs-around-the-world/> (2017)
40. Romero, A., et al.: Diet networks: thin parameters for fat genomics. *ArXiv* (2017)
41. Finn, C., Levine, S.: Deep visual foresight for planning robot motion. *ArXiv* (2017)
42. Meier, K.: The FACETS project. Available https://facets.kip.uni-heidelberg.de/images/4/48/Public%2D%2DFACETS_15879_Summary-flyer.pdf (2010)
43. Qualcomm helps make your mobile devices smarter with new Snapdragon machine learning software development kit. <https://www.qualcomm.com/news/releases/2016/05/02/qualcomm-helps-make-your-mobile-devices-smarter-new-snapdragon-machine> (2016)
44. Avizienis, A.V., Sillin, H.O., Martin-Olmos, C., Shieh, H.H., Aono, M., Stieg, A.Z., Gimzewski, J.K.: Neuromorphic atomic switch networks. *PLoS One*. **7**(8), e42772 (2012)
45. Yang, J.J., Strukov, D.B., Stewart, D.R.: Memristive devices for computing. *Nat. Nanotechnol.* **8**, 13–24 (2013)
46. Stieg, A.Z., et al.: Self-organization and emergence of dynamical structures in neuromorphic atomic switch networks. In: Adamatzky, A., Chua, L. (eds.) *Memristor Networks*. Springer, Cham (2014)
47. Demis, E.C., Aguilera, R., Sillin, H.O., Scharnhorst, K., Sandouk, E.J., Aono, M., Stieg, A.Z., Gimzewski, J.K.: Atomic switch networks nanoarchitectonic design of a complex system for natural computing. *Nanotechnology*. **26**, 204003 (2015)
48. Demis, E.C., Aguilera, R., Scharnhorst, K., Aono, M., Stieg, A.Z., Gimzewski, J.K.: Nanoarchitectonic atomic switch networks for unconventional computing. *Jpn. J. Appl. Phys.* **55**, 1102B2 (2016)
49. Sillin, H.O., Aguilera, R., Shieh, H.H., Avizienis, A.V., Aono, M., Stieg, A.Z., Gimzewski, J. K.: A theoretical and experimental study of neuromorphic atomic switch networks for reservoir computing. *Nanotechnology*. **24**, 384004 (2013)
50. Scharnhorst, K.S., Carbajal, J.P., Aguilera, R.C., Sandouk, E.J., Aono, M., Stieg, A.Z., Gimzewski, J.K.: Atomic switch networks as complex adaptive systems. *Jpn. J. Appl. Phys.* **57**, 03ED02 (2018)

51. Langton, C.G.: Computation at the edge of chaos: phase transitions and emergent computation. *Phys. D.* **42**, 12–37 (1990)
52. Gimzewski, J.K., Möller, R.: Transition from the tunneling regime to point contact studied using scanning tunneling microscopy. *Phys. Rev. B.* **36**(2), 1284–1287 (1987)
53. Lang, N.D.: Theory of single-atom imaging in the scanning tunneling microscope. *Phys. Rev. Lett.* **56**, 1164–1167 (1986)
54. van Houton, H., Beenakker, C.: Quantum point contacts. *Phys. Today.* **49**(7), 22–27 (1996)
55. Terabe, K., Nakayama, T., Hasegawa, T., Aono, M.: Formation and disappearance of a nanoscale silver cluster realized by solid electrochemical reaction. *J. Appl. Phys.* **91**, 10110–10114 (2002)
56. NEC. NEC integrates nanobridge in the Cu interconnects of Si LSI. <https://phys.org/news/2009-12-nec-nanobridge-cu-interconnects-si.html> (2009)
57. Ohno, T., Hasegawa, T., Tsuruoka, T., Terabe, K., Gimzewski, J.K., Aono, M.: Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nat. Mater.* **10**, 591–595 (2011)
58. Hasegawa, T., Nayak, A., Ohno, T., Terabe, K., Tsuruoka, T., Gimzewski, J.K., Aono, M.: Memristive operations demonstrated by gap-type atomic switches. *Appl. Phys. A.* **102**, 811–815 (2011)
59. Avizienis, A.V., Martin-Olmos, C., Sillin, H.O., Aono, M., Gimzewski, J.K., Stieg, A.Z.: Morphological transitions from dendrites to nanowires in the electroless deposition of silver. *Cryst. Growth Des.* **13**(2), 465–469 (2013)
60. Stieg, A.Z., Avizienis, A.V., Sillin, H.O., Martin-Olmos, C., Aono, M., Gimzewski, J.K.: Emergent criticality in complex turing B-type atomic switch networks. *Adv. Mater.* **24**, 286–293 (2011)
61. Oskoe, E.N., Sahimi, M.: Electric currents in networks of interconnected memristors. *Phys. Rev. E.* **83**, 031105 (2011)
62. Goudarzi, A., Lakin, M.R., Stefanovic, D., Teuscher, C.: A model for variation-and fault-tolerant digital logic using self-assembled nanowire architectures. In: *IEEE/ACM International Symposium on Nanoscale Architectures*. ACM, pp. 116–121 (2014)
63. Verstraeten, D.: Reservoir computing: computation with dynamical systems. PhD thesis, Ghent University (2009)
64. Legenstein, R., Maass, W.: What makes a dynamical system computationally powerful? In: Haykin, S., Principe, J.C., Sejnowski, T.J., McWhirter, J. (eds.) *New Directions in Statistical Signal Processing: From Systems to Brain*. MIT Press, Cambridge, MA (2005)
65. Lukoševičius, M., Jaeger, H.: Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* **3**, 127–149 (2009)
66. Wyffels, F., Schrauwen, B.: A comparative study of reservoir computing strategies for monthly time series prediction. *Neurocomputing.* **73**, 1958–1964 (2010)
67. Castro, L.N.D.: Fundamentals of natural computing: an overview. *Phys. Life Rev.* **4**, 1–36 (2007)
68. Modha, D.S., Ananthanarayanan, R., Esser, S.K., Ndirango, A., Sherbondy, A., Singh, R.: Cognitive computing. *Commun. ACM.* **54**, 62–71 (2011)
69. Yu, S., Kuzum, K., Philip Wong, H. S.: Design considerations of synaptic device for neuromorphic computing. In: *IEEE International Symposium on Circuits and Systems*, Melbourne, VIC. IEEE, pp 1062–1065 (2014)
70. Schrauwen, B., Verstraeten, D., Van Campenhout, J.: An overview of reservoir computing: theory, applications and implementations. In: *15th European Symposium on Artificial Neural Networks*, pp. 471–482 (2007)
71. Bürger, J., Goudarzi, A., Stefanovic, D., Teuscher, C.: Hierarchical composition of memristive networks for real-time computing. In: *IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*. IEEE (2015)

72. Gacem, K., Retrouvey, J.M., Chabi, D., Filoramo, A., Zhao, W., Klein, J.O., Derycke, V.: Neuromorphic function learning with carbon nanotube based synapses. *Nanotechnology*. **24**, 384013 (2013)
73. Snyder, D., Goudarzi, A., Teuscher, C.: Computational capabilities of random automata networks for reservoir computing. *Phys. Rev. E*. **87**, 042808 (2013)
74. Carbajal, J.P., Dambre, J., Hermans, M., Schrauwen, B.: Memristor models for machine learning. *Neural Comput.* **27**, 725–747 (2015)