Self-assembled networks with neural computing attributes

S. Bandyopadhyay
Virginia Commonwealth University, Richmond, VA

L. Menon
University of Nebraska - Lincoln

N. Kouklin
Virginia Commonwealth University, Richmond, VA

P. F. Williams
University of Nebraska - Lincoln, pfw@moi.unl.edu

Natale J. Ianno
University of Nebraska - Lincoln, nianno1@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/elecengwilliams
Part of the Electrical and Computer Engineering Commons

http://digitalcommons.unl.edu/elecengwilliams/45

This Article is brought to you for free and open access by the Electrical Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in P. F. (Paul Frazer) Williams Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
1. Introduction

The drive to find alternative paradigms for computation, distinct from current silicon CMOS-based circuits, is fueled by a belief in the device research community that the silicon MOSFET will ultimately run up against insurmountable barriers. These barriers could be associated with excessive power dissipation, breakdown of scaling laws, and certain fundamental limits imposed by the laws of quantum mechanics. While the fundamental limits are still a moot issue, excessive power dissipation is universally acknowledged to be a serious problem. The Semiconductor Industry Association's National Technology Roadmap projects that by the year 2007, the dynamic power dissipated in CMOS devices will be 600 nW/logic gate with a gate density of $5 \times 10^7$ cm$^{-2}$, corresponding to a dissipation of 30 W/cm$^{-2}$ of chip area [1]. This figure is not significantly better than today's figure, indicating that there is small chance of improvement in the future. As we make devices more energy efficient, we also keep on adding more devices per unit area so the dissipation per unit area remains approximately constant.

Present day personal computers consume less than 100 W, but this consumption can actually increase in future as clock speed goes up and the memory size increases. The total power consumed by PCs in the US today is about 5% of the total national power generation [2]. Even if future computers do not consume any more power than they do at present, the number of computers in use will probably grow exponentially with time. At the same time, the limited amount of natural fuel reserve, the enormous cost of building a new power plant, the public suspicion of fission energy, and cold fusion being in cold storage (perhaps permanently) mean that power generation is not likely to increase even linearly with time, let alone exponentially. It is therefore likely that computers will soon begin to consume a significant fraction of the national power generation leading to a drain on the energy supply. Therefore, making computing devices more energy efficient (less dissipative) is a primary concern.

In a seminal paper published in 1961 [4], Rolf Landauer addressed the fundamental issue of dissipation and showed that the minimum energy that must be dissipated in a single *logically irreversible* bit operation is $kT \ln 2$ which is about $4 \times 10^{-21}$ J at room temperature. This figure is far smaller than what CMOS or single-electron transistors [5] will dissipate in a logic bit operation by the year 2007, but the very existence of this figure portends a fundamental limit. Assuming that the most advanced devices, constrained only by the Landauer limit, will switch in 1 ps, the power dissipated will be 4 nW/gate. Moreover, assuming that heat sinking technology will allow removal of only
1–10 kW cm\(^{-2}\) [3], the gate density will saturate to 2.5 \times 10^{13} gates cm\(^{-2}\) unless dramatic improvements in heat sinking are achieved. The alternative is to seek ways to circumvent the \(kT \ln 2\) barrier.

Fortunately, Landauer [4] also showed that \(kT \ln 2\) is not an absolute, fundamental limit. Energy dissipation accrues from physical irreversibility which comes about because of logical irreversibility. If a bit operation can be carried out in a logically reversible manner, then the energy dissipation can, in principle, approach zero. Concrete proposals for such “dissipationless” systems were advanced by Bennett [6, 7], Toffoli [8], Fredkin and Toffoli [9], Likharev [10], and Landauer [11] among others. None of these proposals envisioned nanoelectronic implementation. Recently, some nanoelectronic versions have appeared in the literature [12–15]. Ultimately, the burgeoning field of quantum computers may lead to completely dissipationless computing machinery capable of solving classically intractable problems [16, 17]. While considerable progress is being made in these directions, the field is also beset with difficulties accruing mostly from materials and device shortcomings.

Quantum computers and dissipationless devices are futuristic constructs. Although they have been demonstrated in superconducting systems, nuclear magnetic resonance devices, ion traps, etc., nanoscale compact solid-state systems have remained elusive. On the other hand, rapid advances are being made in nanoelectronics where there may be immediate opportunities for significant progress in the short term. In particular, there are unexplored vistas in unconventional architectures where nanoelectronics may provide a breakthrough within the next few years.

2. Nanoelectronic architectures

In classical electronic architectures (e.g. MOSFETs, bipolar junction transistors), switching is basically accomplished by moving charges from one region of space to another. In the case of MOSFETs, charge is moved from the source contact into the channel region under the action of a gate potential to switch the transistor “on.” To switch the transistor “off,” charge is moved from the channel into the drain by a change of the gate potential.

It takes energy to move charge around and this energy is ultimately dissipated as heat. The smaller the amount of charge that one has to move (in switching a device on or off), the less the dissipation.

Roughly speaking, the amount of charge \(\Delta Q\) involved in switching a device using a voltage swing \(\Delta V\) is given by

\[
\Delta Q = C \Delta V
\]

where \(C\) is the capacitance associated with the control terminal, namely the terminal where the voltage swing \(\Delta V\) is applied (the “gate” in the case of a MOSFET). The amount of energy dissipated in switching this device is roughly \(C(\Delta V)^2\). It is therefore obvious that decreasing the capacitance \(C\) reduces dissipation. Since \(C\) is related to the area of the device, a small area helps. A good example of a device where this precept is exploited directly is the single-electron transistor that has a very small capacitance and consequently, very little power dissipation [18]. Smaller size promises reduced power dissipation per device. This motivates downscaling of device size. Another motivating factor is the need for faster speed and higher clock frequency. The switching time is basically the time it takes to move charges from one region of space to another, namely the so-called transit time. This time is essentially \(L/v_{sat}\) where \(L\) is the distance over which charge has to be moved (the “channel length” in the case of a MOSFET) and \(v_{sat}\) is the saturated velocity of the charge carriers. Obviously, the shorter the length \(L\), the smaller the switching time and the faster the switching speed. Therefore, “small” also means “fast.”

The association of increased speed and reduced power dissipation (both highly desirable traits) with “smallness” has gradually evolved “microelectronics” to “nanoelectronics.” Industry fabrication lines are at present pursuing 90 nm feature sizes, and floating gate transistors with feature sizes of 10 nm have been demonstrated [19].

2.1. Shortcomings of nanoelectronic devices

While nanometer-sized devices are fast and energy efficient, they also have a few shortcomings. First, the small size of the device makes it difficult to attach very many leads to it. Therefore, random wired architectures, that are used in conventional logic and memory circuits, are inappropriate for nanoelectronics. Second, the small size also precludes large voltage or current swings. Hence the ability to drive several successive stages suffers. Third, the fan-in/fan-out of a nanometer-sized device is small. Take the ultimate case of a device which outputs a single electron charge under a voltage swing. Since the electron cannot be split into halves (or smaller fractions), this device can, at best, drive only one succeeding device. Finally, nanoelectronic devices typically do not have much power gain. Therefore, they are not exactly tailor-made for logic circuits since logic devices require power gain to restore signal levels at logic nodes [20].

It therefore behooves us to look for alternate architectures, very different from today’s Boolean logic-based circuit paradigms, to exploit the full power of nanoelectronics.

2.2. Locally interconnected architectures and edge-driven paradigms

In a series of papers, a group of researchers from Texas Instruments [21] introduced a concept that is suitable for nanoelectronics. This was a generic concept with no particular implementation. The idea had three ingredients:

(i) Every device is connected only to its nearest neighbors (no long-range wiring). Thus, each device has only a few connections. These connections could be of a quantum mechanical nature (such as tunneling) instead of a physical wire.

(ii) All input data are provided to devices on the periphery of the chip. Interior devices are never accessed from external leads since the packing density in the interior of the chip is very dense. All exterior leads access only the peripheral devices.

(iii) The inputs provided to the peripheral devices are communicated to the interior devices via the connections between the peripheral and interior devices. The interior devices then perform the signal processing or
computation in response to the input and convey the results to other peripheral devices which act as output ports. The external leads connect to these output ports to access the results. This scheme is pictorially depicted in figure 1. This architecture is synergistic with nano-electronic devices.

2.3. A neuromorphic implementation

Our past work has involved a specific implementation of a locally interconnected edge-driven architecture that performs neuromorphic functions. It can be adapted to Boolean logic as well. Most importantly, it was inspired by recent advances in chemical self-assembly and is therefore synergistic with an inexpensive and versatile production method.

This architecture, exhibiting the attributes of universal computing machinery, was proposed by Roychowdhury et al. [22–25]. The basic system is shown in figure 2. It consists of a two-dimensional periodic array of nanometer-sized metallic islands with nearest-neighbor electrical interconnections, self-assembled on a substrate whose current–voltage characteristic has a non-monotonic non-linearity. The simplest choice for the substrate is a resonant tunneling diode (RTD). Other choices, such as an Esaki tunnel diode, are also possible and may be preferable for silicon-based implementations.

The system in figure 2(a) can realize logic circuits, associative memory, signal processors, and combinatorial optimizers which solve such problems as the traveling salesman problem by mapping it onto the charging dynamics of the network. The details have been given in a number of publications such as [22–26]. The reader is referred to these references for more detail. In particular, the last article is a review article summarizing the salient features of this paradigm.

3. Self-assembled networks for computing architectures

In this paper, we will not repeat the theoretical foundations of the architecture which have been dealt with in detail in the above-cited references. Instead, we will focus on experimental progress towards implementing the critical components of the circuitry.

The architecture in question has been developed with the generic features of self-assembly synthesis in mind. We designed the system such that it must not only be compatible but also synergistic with self-assembly. Our own experience with self-assembly gives us enough reasons to adopt this philosophy.

Traditionally, nanoscale patterns have been delineated by direct-write fine-line lithography. Direct writing of highly complex and dense integrated circuits is extremely time consuming and can run into several hours. This, coupled with the fact that direct writing is a serial technique whereby each wafer is patterned one at a time, can lead to an unacceptably slow throughput.

Chemical self-assembly techniques, on the other hand, are parallel in nature, i.e. several wafers can be processed simultaneously. They are well suited for realizing highly uniform sheets of organized nanostructures, e.g. molecules on surfaces (self-assembled monolayers or SAMs), clusters/nanoparticles/wires of controlled dimensions at the nanometer scale, and uniform two-dimensional and/or three-dimensional arrays of structures such as clusters, islands, lines, and nanopores on surfaces. These techniques are useful for defining various nanometer-scale ultradense arrays with relatively low cost and high throughput. Of course, uniform arrays by themselves are usually not sufficient for realizing useful computational or signal processing circuits. It is likely that some form of lithography, at a scale larger than the minimum element size, will be required in order to connect and isolate blocks of devices. This is not a serious drawback since the finest features will still be produced by self-assembly.

A key component required to realize circuits of the type shown in figure 2 is the formation of ordered metallic dot arrays on a substrate. Each dot must be coupled to the substrate via a non-linear conductor exhibiting a non-monotonic current–voltage characteristic. The non-linear element of choice is an Esaki tunnel diode that exhibits a negative differential resistance. Thus, each dot must be used as an etch mask to electrically isolate a columnar Esaki diode structure underneath. Finally, nearest-neighbor electrical connections must be established between these dots.

Self-assembly synthesis techniques can provide highly ordered arrays of metallic islands at the nanometer scale on arbitrary substrates. These islands can be used as a natural mask to mesa-isolate structures with desired transport characteristics in the underlying substrate. Finally, a resistive film can be evaporated on the surface to complete electrical connections between nearest-neighbor dots. This is the most straightforward, but perhaps not the most elegant, approach.

4. Self-assembled template-based synthesis

In this section, we describe our simple approach to self-assembling ordered arrays of nanometer-sized dots on a p–
n+ wafer that acts as an Esaki tunnel diode structure. This is the most critical step in the realization of the neuromorphic architecture.

Anodic alumina films containing ordered arrays of nanopores are widely used for self-assembling semiconductor quantum dots and wires of uniform diameter. We will adapt that technique to create a regimented array of metal nanodots on a p+-n+ wafer.

Electrochemical self-assembly of the anodic alumina film consists of the following steps. A 99.999% pure 100 µm aluminum foil is first degreased in trichloro-ethylene, washed in distilled water, and electropolished at 30 V for 60 s in LECO-1 solution consisting of perchloric acid, ethanol, butyl cellusolve, and water. The foil is then dc anodized in either sulfuric or oxalic acid at room temperature using a current density of 25 mA cm\(^{-2}\). This results in the formation of a porous alumina film on the surface with a pore diameter of about 8 nm and an areal pore density of \(10^{11}\) cm\(^{-2}\) for sulfuric acid anodization. For oxalic acid anodization, the pore diameter is 50 nm and the pore density is \(10^{10}\) cm\(^{-2}\).

An atomic force micrograph of a porous film produced by anodization in oxalic acid is shown in figure 3.

The porous film can be used to create an ordered array of metallic nanodots following a technique proposed by Masuda and Satoh [27]. The porous film is coated on the surface with an organic binder (which provides mechanical support during later processing steps). Next, the aluminum backing is dissolved in HgCl\(_2\). Then, the alumina barrier layer is removed from the bottom of the film (for a definition of “barrier layer,” see figure 4) by etching in phosphoric acid at 100 °C. Thereafter, the organic binder is dissolved in acetone to create an ultrathin alumina film with “see-through” pores that floats up to the surface of the acetone. This ultrathin film is then captured on a silicon p+-n+ substrate. Finally, gold is evaporated on the surface using e-beam evaporation. The Au atoms travel through the pore openings and lodge themselves on the surface of the silicon wafer. The alumina template is then removed in phosphoric acid, leaving behind a regimented array of gold nanodots on the surface. These steps are shown in figure 4. An atomic force micrograph of 50 nm diameter Au nanodots produced by this technique is shown in figure 5.

The next step in the process is to use the Au dots as an etch mask to reactive-ion-etch mesas into the underlying
Self-assembled networks with neural computing attributes

This is a challenging task, but the recent creation of pillars using 9 nm dots as an etch mask [28] holds out significant promise in this direction. Our task however extends beyond the mere creation of pillars by reactive ion etching. We need to ensure that these pillars are electrically conducting (not pinched off by surface damage and Fermi level pinning caused by the etching step) and that the conduction characteristic retains the non-monotonic non-linearity. Our current efforts are directed towards this goal.

5. Conclusions

In this paper, we have described the experimental strategy for realizing a self-assembled neuromorphic circuit and demonstrated initial success towards this goal. Future work is geared towards extending this technique to create functional Esaki diode pillars of nanometer-sized diameter.

Acknowledgment

Various facets of this ongoing work are currently supported by the US Office of Naval Research under grant N00014-01-10742 and by the US National Science Foundation under grant ECS-0089893.

References


Figure 4. Steps involved in the creation of an ordered array of metallic dots on a semiconductor surface. (a) Anodization of aluminum to form porous alumina film on the surface, (b) depositing an organic binder on top for mechanical strength, (c) removing the aluminum substrate in HgCl₂, (d) etching the alumina barrier layer away in phosphoric acid or NaOH, (e) dissolving the organic binder in acetone to create “see-through” pores, (f) capturing the see-through porous film on a substrate, (g) evaporating gold through the pores, (h) removing the alumina template in phosphoric acid to create an ordered array of metal nanodots on the substrate.

Figure 5. An atomic force micrograph of 50 nm Au dots produced on a silicon surface by following the steps outlined in figure 4.