# The Architectural Designs of a Nanoscale Computing Model

**Mary Mehrnoosh Eshaghian-Wilner**
**Department of Electrical Engineering**
**University of California Los Angeles**

**ABSTRACT**

A generic nanoscale computing model is presented in this paper. The model consists of a collection of fully interconnected nanoscale computing modules, where each module is a cube of cells made out of quantum dots, spins, or molecules. The cells dynamically switch between two states by quantum interactions among their neighbors in all three dimensions. This paper includes a brief introduction to the field of nanotechnology from a computing point of view and presents a set of preliminary architectural designs for fabricating the nanoscale model studied.

**Keywords**: Nanoscale Computing Technology, Spintronics, Molecular Electronics, VLSI, MEMS, NEMS, QCA.

## 1. INTRODUCTION

This section is a brief introduction to nanotechnology compiled from the book *Nanotechnology, Basic Science and Emerging Technologies* [1].

Nanotechnology is an anticipated manufacturing technology that allows thorough inexpensive control of the structure of matter by working with atoms. It will allow many things to be manufactured at low cost and with no pollution. It will lead to the production of nanomachines. As explained by Drexler, nanotechnology is the principle of atom-by-atom manipulation through control of the matter structure at the molecular level. It entails the ability to build molecular systems with atom-by-atom precision, yielding a variety of nanomachines. The discovery of nanotechnology in the broadest sense has immediate implications since we can design a whole new range of machines from nanoscale objects, but not necessarily by breaking up matter into individual atoms. Rather, it may be done using bits of crystal or bits of biological materials. The development and use of molecular nanotechnology ñ the building up from atoms ñ will be slower because it will take time to find the exact point where changing only a few atoms in a structure will make a difference. The single electron device (e.g., memory) may be a case where molecular technology research will be commercialized faster.

A consequence of Mooreís law is that the individual feature sizes of electronic components decreases every year despite the continued difficulty in fabricating smaller and smaller electronic components. Following on from Mooreís law, there is a prediction that by the year 2009 the feature sizes of devices will become less than 50 nm, where the electronic properties of the materials will change from obeying the familiar classical physics to the less familiar quantum physics. Transistors will eventually reach a limit of one electron per bit. While quantum effects represent a fundamental limit to the miniaturization that has been one of the key methods of increasing processor performance, a school of thought believes that these effects may be used to our advantage ñ if we knew how to control them. Nanoelectronics is the emerging field of building electronic devices at the atomic level to harness these small-scale ëquantumî properties of nature. The field unites physicists, chemists and biologists in order to understand how nature works at atomic scale and how we can control it.

In the nanoscale regime, electrons in a solid no longer flow through electrical conductors like solid objects, but the electronís quantum mechanical nature also expresses itself as a wave. This wave behavior makes it possible for electrons to do remarkable things, such as instantly tunnel through an insulating layer that normally would have stopped it. To understand how and when quantum effects come into play we must consider what happens to a semiconductor device as it becomes smaller. As we reduce the size, the net electron transit time through the devices is shorter and hence there is an incentive for making electronic devices smaller and smaller. However, there are more fundamental effects, such as the fact that for individual atoms and molecules the electronic states are discrete and quantized [2]. Quantum effects become observable when the separation between these energy levels becomes larger than the thermal energy that allows rapid transitions at operating temperature. As the physical dimensions of the devices are reduced, the separation between the discrete energy levels increases, and quantum effects persist to higher temperatures.

Over the past 40 years scientists have investigated and tried to understand unusual quantum phenomena, but an important question is whether or not it is possible for a new kind of computer to be designed based entirely on quantum principles. The extraordinary power of the quantum computer is a result of a phenomenon called quantum parallelism, a mechanism that enables multiple calculations to be performed simultaneously. This is in contrast to a classical computer, which can only perform operations one at a time, albeit very quickly [3]. The field of quantum computation had remained a largely academic one until the 1990s, when it was shown that for certain key problems quantum computers could, in principle, outperform their classical counterparts. Since then research groups around the world have been racing to pioneer a practical system. However, trying to construct a quantum computer, at the atomic scale, is far from easy since it requires the ability to manipulate and control single atoms. Wiring quantum bits together is a challenging task since it requires the manipulation of electrons and protons within individual atoms without disturbing the particlesí spins. These systems may need to be wired with molecular mimics or even using biological materials.

While we employ quantum and tunneling effects due to the below 50nm level of integration, the style of the computation we use is classical rather than quantum. The main objective of this paper is to study the challenges in design and fabrication of nanoscale chips for computing using spins and quantum dots.

Quantum dots (QD) are nano-sized deposits of one semiconductor embedded in another semiconductor. Since the dot material has an energy band gap that is smaller than that of the surrounding material, it can trap charge carriers. While quantum dots are particles made up of hundreds to thousands of atoms, in many of their characteristics they behave like a single gigantic atom. The optical and transport properties of quantum dots ñ particularly the ease of customizing those properties by adjusting the size or composition of the dots ñ make them very suitable for molecular electronics. In the category of QDs there are individual dots (a.k.a. ì artificial atomsî ), as well as coupled dots (ì quantum-dot moleculesî ), and a composite device of four or five QDs called a ì QD cell.î The integration of these into various architectures is shown later in this paper.

In 1990, Supriyo Datta and Biswajit A. Das, then at Purdue University, proposed a design for a spin-polarized field-effect transistor, or spin FET. The Datta-Das spin FET has a ferromagnetic source and drain so that the current flowing into the channel is spin-polarized. When a voltage is applied to the gate, the spins rotate as they pass through the channel and the drain rejects these antialigned electrons. Macroscopic spin transport was first demonstrated in n-doped gallium arsenide. Recent experiments have successfully driven coherent spins across complex interfaces between semiconductor crystals of different composition. For more information, refer to the overview article cited [4].

Another alternative for designing nanoscale computing chips would be to employ molecular electronics. It is possible to build molecular electronic switches. The key issue in designing a molecular switch as compared to a large-scale switch such as a transistor is the ability to control the flow of electrons. Using a device that moves, such as rotaxanes or catenanes, is one method, provided there is some way of recording the movement. Another way to do this in a molecule is to control the overlap of electronic orbitals. For example, with the right overlap it may be possible for electrons to flow, but if we disturb the overlap it may be possible to block the flow. The task of fabricating and testing such tiny molecular devices is possible by the use of a scanning tunneling microscope (STM). STMs use a sharpened, conducting tip with a bias voltage applied between the tip and the sample. When the tip is brought within about 1 nm of the sample, electrons from the sample begin to pass through the 1nm gap into the tip, or vice versa, depending upon the sign of the bias voltage. The process involves the wave properties of an electron to move across an energy barrier at lower energy than if it were a particle.

The rest of the paper is organized as follows. In the next section, we present a generic model for nanoscale computing. In Section 3, we will introduce two implementations of an architecture based on the nanoscale computing model introduced. The first implementation, called the H3D-QCN architecture, employs quantum dots, and the second implementation, called the H3D-SPN architecture, uses spins for switching. Concluding remarks are presented in Section 4.

## 2. A NANOSCALE MODEL OF COMPUTATION

One of the steps towards designing architectures and algorithms that involve a new technology is to design an abstract model of computation. Based on this model, units of computation can be defined in terms of time and space. This model, which could represent trade-offs between space and time, should reflect a generic mode of computation.

For example, the Thompson VLSI model of computation in the late 1970s illustrated that all VLSI chips have similar properties with respect to limitations on the number of layers, and therefore VLSI is a planar technology. Based on this, the VLSI space-time trade-offs were obtained, which gave an accurate measure of how much speed-up can be expected if a chip with a larger VLSI area is used for a particular application. This model gave a concrete tool for design and analysis of VLSI architectures and algorithms. In the late 1980s new computational models were introduced to represent VLSI architectures with optical interconnects. Among them are the VLSIO model by Barakat and Reif, and the OMC by Eshaghian. The volume-time trade-offs of both these models are superior to the VLSI model of computation that reflects the computation power gained through the third dimension of connectivity [5].

The goal here is to understand what kind of trade-offs are obtained in using the nanoscale technology for computing when the level of integration is small enough to be subject to quantum effects. To answer this question from a computer science point of view, we would first need to design a new model of computation representing the type of architectures such as the ones presented here in this paper. Towards that goal, we present the following preliminary model that is intended to capture the nature of this technology. As shown in the figure below, the model is a three-dimensional box that consists of a collection of nanoscale-size cells that are small enough to be subject to quantum effects. They do computations through tunneling effects. Whether the cells are implemented using quantum dots, spins and/or molecules is just an implementation issue. Similarly, it is not necessary to specify the type of connectivity placed above or among this model. Based on this, it is easy to see that the space-time trade-offs of this model are similar to those for three-dimensional VLSI which is $VT^{3/2}=\Omega(I)^{3/2}$. The difference here is that the three-dimensional model was not implementable due to fabrication limitations in the number of layers. This lowerbound matches the lowerbounds for the VLSIO model of Barakat and Rief [6].

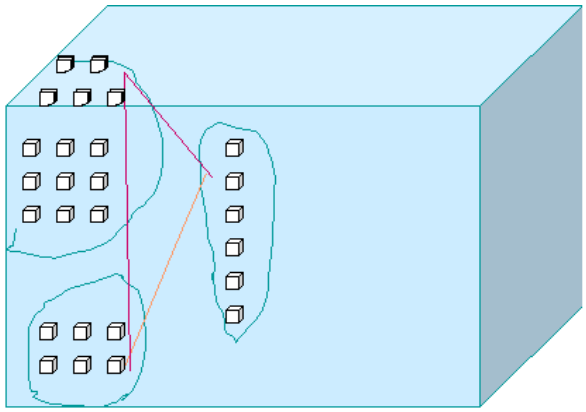

**Fig. 1.** The Nanoscale Model

## 3 TWO ARCHICTECTURAL IMPLEMENTATIONS

Two implementations of an architecture based on the model presented in the previous section is described here. As shown in Figure 2, this architecture is essentially a NEMS (Nano-Electro-Mechanical Systems) design [7,8] in which there are two layers: the processing layer below and the deflecting layer on top. The processors can intercommunicate using a standard reconfigurable mesh through the local switchable connections, and also using the reconfigurable micro-electromechanical mirrors with free-space optical interconnects. Each of the processors contains some local memory and is attached to a nanoscale computing cube. In each cube there are nanoscale cells laid out in three-dimensional format as shown.

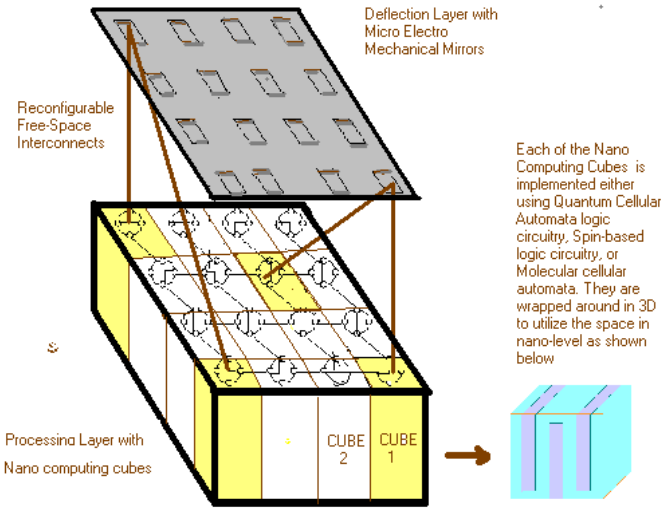


**Fig. 2.** The Architecture of H3D-QCN and H3D-SPN

Below, we explain two possible implementations of the above NEMS architecture. The first one is called H3D-QCN [9], in which each cube is a three-dimensional Quantum Cellular Array (QCA [10,11]). The second one is called H3D-SPN, where each of its cubes is a spin-based computational module.

### H3D-QCN

The Quantum Cellular Automata (QCA) have been extensively studied by a group of researchers at the University of Notre Dame for several years [10,11]. The basic idea behind QCA is that when the level of integration is very small, cells interact with each other through quantum effects and tunneling. Utilizing quantum dots, the size of an elementary cell can be shrunk down to hundreds or tens of nanometers and the inter-cell interaction can be realized via quantum tunneling without wires. Moreover, the product of energy of switching, E, and of switching time, $\tau$, may approach a fundamental limit. Using this concept, simple cells have been developed mainly using five quantum dots called a quantum dot molecule. The five dots are close enough to enable electrons to tunnel between the dots. The barriers between cells are assumed to be sufficient to completely suppress intercellular tunneling. Two electrons occupy each cell. The occupancy can be stabilized because of the large energy splitting between different charge states of the cell. The Coulomb interaction between electrons in a cell acts to produce two distinct cell states with different charge configurations. If the barriers to tunneling are sufficiently high, the two-electron ground-state wave function in the cell will localize the two electrons on antipodal sites. This localization is due to Coulomb exclusion, a phenomenon closely related to the well-known Coulomb blockade of current, and results in nearly exact quantization of charge in each dot.

There are two possible configurations with the electrons on opposite corners of the dots, as shown below. The polarization of the states is defined as +1 and ñ1. Binary information can be encoded using the cell polarization. A cell polarization of +1 corresponds to a bit value of 1; a cell polarization of ñ1 corresponds to a bit value of 0. The Coulomb interaction between cells causes the state of one cell to affect the state of a neighboring cell. Even a slight polarization in a neighboring cell induces essentially complete polarization in the target cell. This means that at every stage the signal level is restored. This will enable a line of QCA cells to act as a robust binary wire. Similarly, a series of logic gates can be built using a specific arrangement of such cells. Therefore, it is possible to implement logic circuits in QCA. A schematic for a Full adder is shown below.

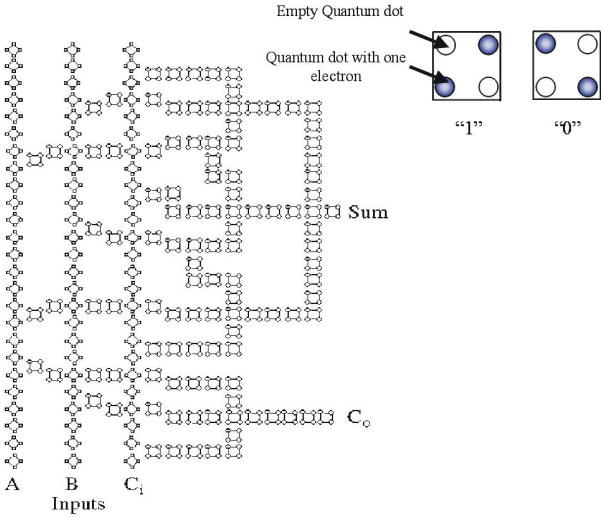


**Fig. 3.** The Implementation of Full Adder on QCA [10,11]

In the H3D-QCN, the computations within each QCA cube is done in a similar fashion as a standard QCA except that the two-dimensional QCA logic circuits are laid out in three dimensions, as shown the bottom right of Figure 2. In other words, the QCA blocks can be used to compute millions of logic operations locally by techniques already developed for QCA. The computations are done as the neighboring cells interact with each other through quantum tunneling effects. Once the local computations within each cube are completed, the results are forwarded to their corresponding processing units. The processors can then store the data in their local memory and/or intercommunicate with other processing units using the electronically reconfigured mesh and/or the micro-electro-mechanical mirrors.

The implementation of the H3D-QCN has the low temperature operation limitation. A solution to this could be to implement the cells using molecules as described in the previous section. Using molecular magnetic switches, it could be possible to simulate the QCA that operates at room temperature.

**H3D-SPN**

It is possible to replace the QCA cubes with spin-based computational cubes as shown below. The overall operation at the architectural level is still the same. Computations are done within the cubes using quantum effects but based on spins instead of based on the polarities of the quantum cells. Once the cubes complete the computations, they send their results to their cube-designated processor, which will intercommunicate with other cubes using electro-optical interconnectivity. Once the results of each cube are obtained, the operation among the MEMS level processors proceeds via the electro-optical connectivity available.

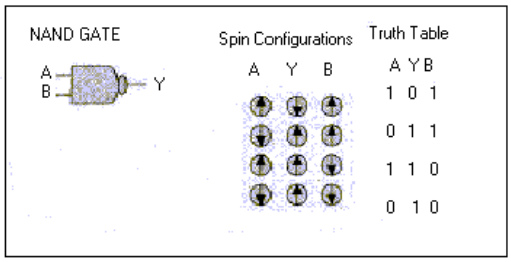


**Fig. 4.** Spin-based Logic NAND Gate [2,12,13].

The basic logical operations in both H3D-QCA and H3D-SPN can be performed within the cells that dynamically switch between two states by quantum interactions among their neighbors in all three dimensions. However, currently all the implementations based on bistable devices without directionality, such as the QCA and spins, suffer from the failure to ensure propagation of the logic signals from the input to the output. That is, if two bistable devices are connected together in series, then there must be some isolation between the input and output so that the input drives the output and not the reverse. Coulomb interactions between two identical charge polarizations are reciprocal so that it is impossible to distinguish the input polarization from the output polarization. In other words, the output influences the input just as much as the input influences the output. Consequently, logic signals can not propagate unidirectionally from the input to the output, from one stage to the next, leading to an operational failure. This occurs because the input can not uniquely and predictably determine the output. This problem is pathological in many proposed schemes of nano-electronic architecture. One needs to ensure that the signals propagate from the input to the outputs and the whole system does not get stuck in metastable states. For more details see the article by Anantrand and Roychowdhury [12].

## 4. CONCLUSION

Two high-level implementations of a hierarchical architecture based on a nanoscale model of computation presented were discussed in this paper. The architecture consists of a collection of fully interconnected nanoscale three-dimensional nanoscale computing cubes, where each cube can be implemented using quantum dots (in the QCA-based version called H3D-QCN), or using spins (in the spin-based version called H3D-SPN). The fabrication of these architectures currently faces a number of challenges. While nanoscale and molecular computing is still at an infancy stage, it is a promising alternative to todayís CMOS technology [13].


## 5. ACKNOWLEDGMENTS

This work was supported in part by a fellowship at the NSF Institute of Pure and Applied Mathematics at the University of California Los Angeles (UCLA) as part of the Mathematics in Nanoscale Science and Engineering Program. The author is especially thankful for the valuable feedback and experience gained through collaboration with several members of the California Nanosystems Institute at UCLA.


## 6. REFERENCES


[1] M. Wilson, K. Kannangara, G. Smith, M. Simmons, B. Raguse, ìNanotechnology, Basic Science and Emerging Technologies,î Chapman & Hall/CRC, (2000).

[2] P. Balasingam and V. P. Roychowdhury, ìNanoelectronic Functional Devices,î Purdue University Technical Report: TR-EE 94-24, (1994).

[3] A. Y. Kitaev, A.H. Shen, M. N. Vyalyi, ìClassical and Quantum Computation, ìAmerican Mathematical Society, (2002).

[4] D.D. Awschalom, M.E. FlattÈ and N. Samarth, ìSpintronics,î Scientific American May (2002).

[5] M. M. Eshaghian, ìParallel Algorithms for Image Processing on OMC,î IEEE Transactions on Computers, Volume 40, Number 7, Pages 827-833, (1991).

[6] R. Barakat and J. Reif, ìLowerbounds on the Computational Efficiency of Optical Computing Systems,î Applied Optics, (1987).

[7] M. M. Eshaghian, and Lili Hai, ìEfficient Parallel Computing with an Optically Interconnected Reconfigurable Mesh,î Journal of Parallel and Distributed Computing, Volume 61, pages 737-747, (2001).

[8] S. E. Lyshevski, ìMEMS and NEMS, Systems, Devices, and Structures,î CRC Press, (2002).

[9] M. M. Eshaghian, ìNanoscale Computing Structures,î Proceedings of the 7th World Multi-conference on Systemics, Cybernetics, and Informatics, SCI2003, Florida, July 2003.

[10] C.S. Lent, P.D. Togaw, and W. Porod, ìBistable Saturation in Quantum Dots for Quantum Cellular Automata,î Applied Physics Letters 62, 714 (1993).

[11] G. Snider, A. Orlov, I. Amlani, X. Zuo, G. B. Stein, C. Lent, J. Mez, and W. Porod, ìQuantum-dot Cellular Automata,î Journal of Applied Physics (1999).

[12] M.P. Anantram and V.P. Roychowdhury, ìMetastable States and Information Propagation in a One-dimensional Array of Locally Coupled Bistable Cells,î Journal of Applied Physics 85, (1999).

[13] M. M. Eshaghian-Wilner, A H. Flood, A. Khitun, V. Roychowdhury, J. F. Stoddart, and K. Wang, ìMolecular and Nanoscale Computing and Technology,î, to appear as a chapter in the edited volume by Albert Zomaya, entitled ìHandbook on Innovative Computing,î to be published by Springer-Verlag (USA), 2005.