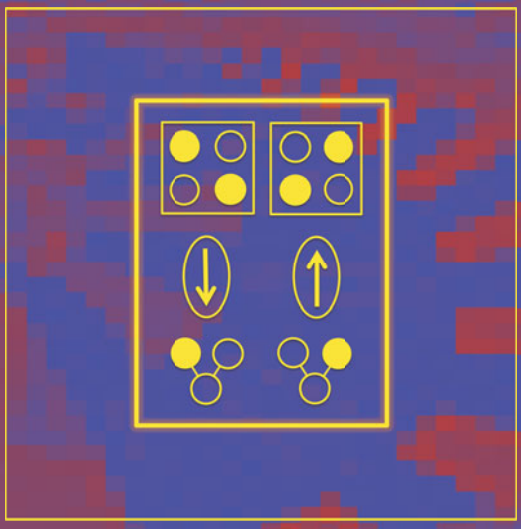
State-of-the-Art Survey

LNCS 8280

Neal G. Anderson
Sanjukta Bhanja (Eds.)

Field-Coupled Nanocomputing

Paradigms, Progress, and Perspectives



Springer

Lecture Notes in Computer Science 8280

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison
Lancaster University, Lancaster, UK
Takeo Kanade
Carnegie Mellon University, Pittsburgh, PA, USA
Josef Kittler
University of Surrey, Guildford, UK
Jon M. Kleinberg
Cornell University, Ithaca, NY, USA
Alfred Kobsa
University of California, Irvine, CA, USA
Friedemann Mattern
ETH Zurich, Zürich, Switzerland
John C. Mitchell
Stanford University, Stanford, CA, USA
Moni Naor
Weizmann Institute of Science, Rehovot, Israel
Oscar Nierstrasz
University of Bern, Bern, Switzerland
C. Pandu Rangan
Indian Institute of Technology, Madras, India
Bernhard Steffen
TU Dortmund University, Dortmund, Germany
Demetri Terzopoulos
University of California, Los Angeles, CA, USA
Doug Tygar
University of California, Berkeley, CA, USA
Gerhard Weikum
Max Planck Institute for Informatics, Saarbruecken, Germany

For further volumes:
http://www.springer.com/series/7407

Neal G. Anderson · Sanjukta Bhanja (Eds.)

Field-Coupled Nanocomputing

Paradigms, Progress, and Perspectives

Springer

Editors
Neal G. Anderson
University of Massachusetts Amherst
Amherst, MA
USA

Sanjukta Bhanja
University of South Florida
Tampa, FL
USA

ISSN 0302-9743 ISSN 1611-3349 (electronic)
ISBN 978-3-662-43721-6 ISBN 978-3-662-43722-3 (eBook)
DOI 10.1007/978-3-662-43722-3
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014940849

LNCS Sublibrary: SL1 – Theoretical Computer Science and General Issues

© Springer-Verlag Berlin Heidelberg 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Field-coupled nanocomputing (FCN) paradigms offer fundamentally new approaches to digital information processing that do not utilize transistors or require charge transport. Information transfer and computation are achieved in FCN via local field interactions between nanoscale building blocks that are organized in patterned arrays. Several FCN paradigms are currently under active investigation, including quantum-dot cellular automata (QDCA), molecular quantum cellular automata (MQCA), nanomagnetic logic (NML), and atomic quantum cellular automata (AQCA). Each of these paradigms has unique features that make it attractive as a candidate for post-CMOS nanocomputing, and each faces critical challenges to realization.

With the hope of bringing the community together to gauge the current status of FCN research and to consider its future directions, we requested support from the National Science Foundation for a dedicated workshop. The result was The 2013 Workshop on Field-Coupled Nanocomputing, which was held at the University of South Florida in February 2013. The international group of participants, which included experienced FCN researchers, postdoctoral scholars, and graduate students, embraced the objectives of this workshop and contributed generously to their achievement. The first objective was to take stock of major milestones that have been achieved in emerging FCN nanocomputing paradigms—at the device, circuit, architecture levels—to provide a snapshot of the current state of research in the field. The second objective was to identify and highlight promising opportunities for FCN and critical challenges facing realization of FCN-based nanocomputers. A panel discussion was dedicated specifically to these concerns, providing all participants—most importantly the graduate student participants—with a variety of perspectives on emerging research priorities and critical next steps.

Our third and final objective was to make the workshop proceedings available to a wide readership, and to do so in a way that allowed inclusion of more background, tutorial, and review material than is typically found in conference papers. To this end we invited participants to submit comprehensive, chapter-length expositions of research related to their workshop contributions, and we solicited a few such contributions from researchers who are working on intriguing aspects of FCN-related topics but were not in attendance at the workshop. This invitation was answered with a collection of quality contributions reflecting a remarkably diverse portfolio of current FCN research. These chapters were peer reviewed by referees from pool that included workshop participants and additional FCN experts. We approached Springer about publishing this collection of contributions—together with an edited transcript of the panel discussion—in their well-known *Lecture Notes in Computer Science* (LNCS) series. Alfred Hofmann of Springer was immediately receptive, and suggested publication in the LNCS *State-of-the-Art Survey* series.

The result—this volume—is divided into five topical sections. In the first section (Field-Coupled Nanocomputing Paradigms), pioneering FCN researchers provide

valuable background and perspective on the QDCA, MQCA, NML, and AQCA paradigms and their evolution. The second section (Circuits and Architectures) addresses a wide variety of current research on FCN clocking strategies, logic synthesis, circuit design and test, logic-in-memory, hardware security, and architecture. The third section (Modeling and Simulation) considers the theoretical modeling and computer simulation of large FCN circuits, as well as the use of simulations for gleaning physical insight into elementary FCN building blocks. The fourth section (Irreversibility and Dissipation) considers the dissipative consequences of irreversible information loss in FCN circuits, their quantification, and their connection to circuit structure. The fifth and final section (The Road Ahead: Opportunities and Challenges) includes an edited transcript of the panel discussion that concluded the FCN workshop.

We thank all of the contributors that made this volume possible, the reviewers who enhanced its quality, and the team at Springer—especially Alfred Hofmann, Anna Kramer, and Christine Reiss—who enabled and facilitated its publication. We gratefully acknowledge Dr. Sankar Basu of National Science Foundation for engaging us in the discussions that led to the sponsorship of the FCN workshop—and thus to this volume—and Dr. Robert Trew for attending the workshop as NSF EECS Division Director. Our sincere gratitude to Dinuka Karunaratne, Srinath Rajaram, İlke Ercan, Ravi Panchumarthy, Jayita Das, Drew Burgett, and Kevin Scott and all of the other student volunteers for workshop logistics, technical assistance, and recording of the panel discussion, and to the University of South Florida—especially the USF Student Chapter of IEEE—for local support and arrangements. Finally, we thank Katherine Anderson for assistance in transcribing the panel discussion. We hope that the collective efforts of all involved has yielded an accessible and useful resource for students and researchers who are intrigued by the possibility of future FCN-based nanocomputers and are working toward their realization.

March 2014

Sanjukta Bhanja
Neal G. Anderson

Contents

Modeling and Simulation

Irreversibility and Dissipation

The Road Ahead: Opportunities and Challenges

Field-Coupled Nanocomputing Paradigms

The Development of Quantum-Dot Cellular Automata

Craig S. Lent(✉) and Gregory L. Snider

Department of Electrical Engineering, University of Notre Dame,
Notre Dame, IN 56556, USA
Lent@nd.edu

Abstract. Quantum-dot cellular automata (QCA) is a paradigm for connecting nanoscale bistable devices to accomplish general-purpose computation. The idea has its origins in the technology of quantum dots, Coulomb blockade, and Landauer's observations on digital devices and energy dissipation. We examine the early development of this paradigm and its various implementations.

Keywords: Quantum-dot cellular automata (QCA) · Molecular electronics

1 Introduction

Quantum-dot cellular automata (QCA) is a means of representing binary information in cells, through which no current flows, and achieving device performance by the coupling of those cells through the electromagnetic field. Information is stored in the arrangement of charge (or magnetic dipoles) with in the cell. Importantly, cells have no monopole moment and are designed to be bistable, having two low energy states with different dipole or quadrupole orientation which can encode a binary 1 or 0. For large scale structures it is necessary to guide the switching of the cells with a clocking field that controllably switches the cells between a null state and an active state (either 0 or 1). Clocking provides power gain necessary to restore signal energies which would otherwise decay due to inelastic losses. The interaction energy between two cell, that is the energy difference between neighboring cells holding the same or opposite bits, is termed the kink energy, and determines thermal stability. Raising the kink energy entails moving to smaller geometries, with molecular QCA providing the limit of device density and requiring the ultra-low power dissipation made possible by adiabatic switching of QCA.

Here we sketch the origins of the QCA idea, its early development, and subsequent evolution into several implementations and many research fronts. This is in no way a comprehensive review, but is particularly focused on the perspective of the Notre Dame group which originated the idea, and the basic trajectories that have arisen from the early work. We mean no slight by mentioning only a few of the major subsequent investigators. The elaboration of all the other contributors to this volume is necessary to give a fuller picture.

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 3–20, 2014.
DOI: 10.1007/978-3-662-43722-3_1, © Springer-Verlag Berlin Heidelberg 2014

2 Historical Background

In the 1980's advanced epitaxial growth techniques such as molecular beam epitaxy (MBE) enabled the creation of GaAs-AlGaAs semiconductor heterostructures with very smooth interfaces. This ability to control the composition of crystalline semiconductors with atomic precision made possible the formation of a highly conducting two-dimensional electron gas (2DEG) at the interface between AlGaAs and GaAs. Moreover, the 2DEG, which was essentially a plane of confined electrons, could be further patterned by placing lithographically-defined metal gates on the surface of the semiconductor. A negative potential on the gates would deplete the 2DEG under the gate regions. In 1988 two groups measured quantized conductance through a constriction connecting two 2DEG regions and found that it was quantized [1, 2]. This was convincingly explained by invoking the quantum-mechanical nature of the electrons transiting the constriction. Using the effective mass approximation, one could explain much of the behavior of this layer by solving the Schrödinger equation in two dimensions.

The ability to engineer the effective wavefunction of electrons seemed very promising for potential device applications. Throughout the 1990's (and beyond) many wave-based device designs were proposed which used quantum interference effects as their operating principle, often created in analogy with microwave devices. Truly remarkable experimental demonstrations left little doubt that these quantum mechanical effects were real and could be potentially exploited for device behavior. In addition, it proved possible to create quantum dots by confining the 2DEG in both lateral dimensions (the third dimension was already confined by the heterostructure potential. These quantum dots could be viewed as artificial atoms [3, 4], and also as high-Q resonators for ballistic electron transport [5].

Into the optimism that these new abilities engendered, Rolf Landauer injected a ray of pessimism and realism. In a talk at the first International Symposium on Nanostructure Physics and Fabrication in 1991, Landauer cautioned that interference devices were unlikely to make it in the real world [6]. A "rich" response with many peaks and valleys did not, he argued, make a robust basis for devices, which would have to be tolerant of fabrication variations and environmental perturbations. He argued for devices whose transfer function is nonlinear and saturates at two distinct levels, as does a CMOS inverter. This input signal should be of the same type as the output signal, so that information can be transferred from device to devices.

Another stream of research at the time was the newly emerging and promising phenomenon of Coulomb blockade in small structures. Electrons tunneling onto small metal islands can raise the potential of the island by e2/C, where C is the total capacitance of the island [7]. For very small structures this charging energy could be significant compared to thermal energies. The "orthodox" theories of the Coulomb blockade, even when treating the system quantum mechanically, characterized the island by this macroscopic quantity—the capacitance [8]. While this is quite adequate for metal structures containing very many free electrons, in small semiconductors one should really use a multi-particle approach. In such a model the effective capacitance is a result of the calculation of Coulomb effects, rather than being an input to it.

A study of few-electron systems under bias showed a threshold behavior for single electron transfer that is very nonlinear [9]. The origin of this nonlinearity is fundamentally the quantization of charge. If a region of space is surrounded by barriers that are appropriately high, but still possibly leaky, then the expectation value of the enclosed charge will be very close to an integer multiple of the fundamental charge. When the equilibrium value of the charge changes because of a tunneling event, it will necessarily be a rather abrupt jump between two integers.

Finally, the QCA architecture was inspired by classical cellular automata (CA) architectures [10], of the type popularized by Conway's Game of Life. These are mathematical models of evolution that proceed from discrete generation to generation according to specified rules. The state of each cell is determined by the state of the neighboring cells in a previous generation. The neighbor-to-neighbor coupling is a natural match for nano-devices, since one expects that one very small device may influence its neighbors, but not distant devices. CA's represent a means of computation that departs from the current-switch paradigm of transistors. But cellular automata are mathematical models that can operate with any set of evolution rules. The question for device architecture was not simply what local CA rules will produce computational behavior, but what rules does the actual physics of cellular interaction support.

The original QCA idea was the result of the confluence of these four ideas: (1) the ability to create configurations of quantum dots which localize charge, (2) the convincing argument by Landauer that any practical device would need bistable saturation in the information transfer function, (3) the nonlinearity of charge tunneling between such dots because of charge quantization, and (4) the notion of a locally-coupled architecture in analogy to cellular automata.

It is worth noting that the connection to cellular automata is by analogy. Classical CA's are almost always regular one or two-dimensional arrays of cells. The physics of the interaction between QCA cells does not yield very interesting results for regular arrays. QCA circuits look more like wires connecting devices; highly non-regular layouts of cells provide the specific function. Mathematical model CA's evolve in discrete generations, but physical systems interact continuously.

3 Developments

The first QCA paper demonstrated the bistability of a QCA cell using a multi-electron Hamiltonian and a direct solution of the Schrödinger equation [11]. This direct approach avoided the problems of sorting out exchange and correlation effects; within the site model it was exact. This bistability remains a key feature of QCA. Though it is somewhat appealing to explore a multi-state QCA cell and multi-state logic, only a bistable system can truly saturate in both logic states. An intermediate state is always subject to drifting off from stage to stage.

It was soon realized that a line of QCA cells acted like a binary wire and a junction of two or three wires could form a logic gate [12]. The first proposal was to have a special cell at the junction which could be internally biased to the 1 or 0 state and thereby act as an OR gate or an AND gate [13]. It was soon clear that the bias could be

applied by another wire, forming a three-input majority gate [14]. Inverters could be formed from diagonal interactions between cells. With these basic elements, any logical or arithmetic function can be formed.

With a year of the first QCA publication, other implementations were suggested. Small metal islands could serve as dots and form QCA cells if they were coupled by tunnel junctions [15]. One advantage of metal-dot QCA is that the electric field lines from the dot can be guided by the conductors to influence the neighboring dots; in the semiconductor depletion dots, the field spreads out in all directions. It was also natural to envision molecular versions of QCA where the role of the quantum dot was played by a part of the molecule that could localize charge [14]. A magnetic model of QCA was constructed from three-inch magnets held in Lucite blocks which rotated on low-friction jeweled pivots. These magnetic cells were used during talks and lectures by both of the present authors to demonstrate QCA wires and gates. They prefigured (at enormous scale) the nanomagnetic QCA under active research today and discussed in other contributions to this volume.

A detailed examination of QCA dynamics and the development of several levels of quantum description of QCA arrays [16], were prompted by another observation of Landauer [17]. While encouraging QCA exploration, he expressed concern that a weak link in a QCA wire would cause a switching error because the incorrect "old" state downstream would have more influence than the upstream cells with the new state. By treating the whole wire quantum mechanically it could be shown that this would only be a temporary problem. But the exercise focused attention on the nature of computation in QCA systems, which were designed from the beginning to map the ground state onto the computationally correct state. This mapping can be robust, while the details of the transient response of the system are inherently more fragile. We wanted to avoid computing with the transient. A byproduct of these calculations was the development of several approximate treatments for both equilibrium and dynamic calculations. The mapping between QCA and the Ising model in a transverse field was also made precise.

Clocking of QCA arrays arose out of the detailed consideration of switching dynamics and the desire to retain the robustness of the mapping between the ground state and the computationally correct state for large systems. Clocking QCA entails gradually moving cells between a neutral state and an active state with a clocking signal. The active state can be either a binary 0 or 1; the neutral state is usually denoted as a "null" state that carries no information. The first version of clocking, proposed the year following the initial QCA papers, contemplated raising and lowering the inter-dot tunneling barriers [18, 19]. This would gradually (adiabatically) transition the cell between a delocalized electron configuration (null) and the localized configuration of the active state. Koroktov and Likharev subsequently suggested a version of metal-dot QCA called the single electron parametron which used a complicated rotating electric field as a clock [20]. This had several drawbacks (e.g., information could only move in one direction in an array), but the idea of using as the null state a localized state on an intermediate dot was adopted for clocking QCA, particularly for molecular implementations. It is much easier to change the potential on the intermediate dot than to directly influence the tunnel barriers between dots, and the effect is the same. Adiabatic clocking QCA [21] solved the problem of switching

dynamics getting caught, even temporarily, in a metastable state; this was the heart of Landauer's objection. It retained the advantages of (local) ground-state mapping. During switching each cell is always very close to its instantaneous ground state (the definition of adiabaticity). Though we did not fully appreciate it at the time, this essentially turned QCA into a concrete implementation of the gedanken experiments which had led Landauer to conclude that there was no fundamental lower bound to the energy that must be dissipated to compute a bit [22]. Clocking further allowed much larger computational structures to be envisioned, combining memory and processing.

3.1 Semiconductor QCA

The Cavendish group of Smith et al. demonstrated QCA operation in GaAs/AlGaAs heterostructures with confining top-gate electrodes [23–25], as originally envisioned in the earliest QCA publications. The group of Kern et al. demonstrated a QCA cell in silicon, using an etching technique to form the dots [26–29]. The group of Mitic et al. used a novel method to form dots from small clusters of phosphorus donors in silicon [30]. They succeeded in demonstrating QCA operation in that system. Interestingly, their long-term goal is coherent quantum computing and they conceive QCA devices as providing an ultra-low-power layer of interface electronics to connect a cryogenic quantum computer to standard CMOS electronics [31].

The challenges of all semiconductor implementations have been two-fold. Firstly, the lithographically accessible sizes for quantum dots are large enough that kink energies are low and cryogenic operation is required. More importantly, the perfection of the interface and electronic environment becomes an issue. While dots with tens of electrons effectively screen small amounts of impurity and imperfections, in the limit of single occupancy, semiconductor dots become very sensitive to the details of the environment. Even MBE-grown samples have enough random imperfections that the dot is often not exactly where one expects it to be based on lithography [32].

3.2 Metal-Dot QCA

Although electronic QCA has been demonstrated in a number of material systems, metal dot implementations have proven to be the most successful, so far, building on the fabrication techniques developed for single-electron transistors [33, 34]. The advantages of metal dots are that the fabrication yield is relatively high, and they are electrically well-behaved, meaning that energy required to add each additional electron to the dot typically remains constant over the addition of many electrons. This makes it easier to load the QCA cell with the proper number of electrons and to bias the cell so that the two polarizations are energetically degenerate. However, in semiconductor dots [33, 34] the fabrication yield is low and the addition energy typically differs for each additional electron, and the electrical behavior of the dot can change from run to run, making it difficult to prepare the cell for proper operation. This makes the metal dot an attractive option for QCA experiments. The main disadvantages of metal dots are background charge fluctuations [35], and low operating temperature. Background charge fluctuations are caused by the random arrangement

of stray charge in the vicinity of the QCA cell, which affects the bias point and polarization degeneracy of the cell. The arrangement of this charge changes with time, and the gate biases applied to each dot in the cell must be adjusted to keep the cell operational. The low operating temperature of metal dot QCA cell is due to the size of the lithographically defined cell.

The metal dot QCA are composed of aluminum islands separated by tunnel junctions. Fabrication of the cell is done by electron-beam lithography using the Dolan bridge technique [7] where the tunnel junctions are formed by evaporation of aluminum from two angles, with an intervening oxidation step. The resulting tunnel junctions are composed of two layers of aluminum separated by a thin layer, 1–2 nm, of aluminum oxide. The area of the overlap between the two layers of aluminum determines the capacitance of the junction, and since it is typically the dominant capacitance of the dot, determines the operating temperature of the QCA cell.

Cells and Logic

The first QCA cell was demonstrated in 1997 [36]. This device had a junction overlap area of approximately 50 × 50 nm, giving an operating temperature of 70 mK. As it was the first demonstration, the layout was very conservative and optimized for high yield, which resulted in a relatively large overlap and low operating temperature. In this first demonstration the goal was to use gate electrodes to move an electron between the top and bottom dots on the left side of a cell. The electron in the right half of the cell would move in the opposite direction to maintain the lowest energy configuration. To measure the polarization of the cell, single-electron transistors, which are the most sensitive electrometers demonstrated to date [37], are used to measure the potential of the dots in the right half of the cell. Measurements of the output of the two electrometers move in opposite directions, confirming that an electron in the right half moves in the opposite direction to the electron in the left half, confirming QCA operation. Full details of the experimental methods are given elsewhere [38–41].

The next step in the development of metal dot QCA was the demonstration of a logic gate [42]. The basic logic element in the QCA paradigm is the majority gate, where three inputs vote on the polarization of a QCA cell. For this experiment we again used metal dots defined by the Dolan bridge method. For the inputs of the majority gate we applied voltages to the input electrodes that mimicked the potentials of three input cells. The applied voltages were varied to step through the logic truth table. The polarization of the cell was measured by electrometers and the output of the cell confirmed proper operation of the gate.

These experiments showed the basic functionality of QCA cells. The next experiment [43] showed that a QCA line could switch without getting stuck in a metastable, partially switched, state. In this experiment three 2-dot cells were fabricated in a line, and an input applied to the left side of the line. Electrometers coupled to the output side of the line confirmed the proper switching.

Power Gain

These initial experiments used unclocked QCA cells, but clocking is an important element in QCA systems. Clocking of QCA cells is crucial to achieve perhaps the

most important quantity in a logic device: power gain. Without power gain the input signal would degrade in a line, due to the unavoidable energy dissipation at each stage, and fan-out would be impossible. Clocking in QCA requires a variable barrier to control the tunneling of electrons between dots. Since the tunnel barrier in metal dot QCA is a fixed aluminum oxide layer whose barrier height cannot be modulated, clocking dots are introduced into the QCA cell as intermediate dots. These dots are coupled to clock electrodes that control the potential of the central dots. A positive clock voltage pulls the electrons to the central dots to produce the null state. A negative voltage forces the electrons to leave in a direction that is determined by the cell's input. In the initial experiment, a differential input voltage is applied to the left side of the cell and electrometers measure the potential of the top and bottom dots of the right half of the cell. Measured output waveforms confirmed proper operation of the cell [44]. A clocked QCA cell can also be used as a latch, a short-term memory element, as demonstrated in our experiments [45, 46].

As shown by theory, the power gain of a QCA cell is not fixed. If the input is weak, the cell pulls power from the clock to restore the signal level. Since the signal energy is fixed for a given cell, the amount of power pulled from the clock will depend on the weakness of the input. An experimental demonstration of power gain involves a measurement of the charge on the dots of the QCA cell [43] as the inputs and clock are moved through one clock period. In this way the work done by the input on the cell can be calculated, along with the work done by the cell on the next cell. If the work done by the cell exceeds the work done on the cell, then the cell has demonstrated power gain. In our experiment an input with one-half the normal potential swing was applied to the input. The resulting experiment demonstrated a power gain of 2.07, in agreement with theory [47].

Shift Registers

Clocking in QCA enables not only power gain, but also the control of the flow of information in the computational system, needed for data pipelining. The basic element in a data flow structure is the shift register. A QCA shift register consists of a row of cells controlled by different clock phases. In our experiment we fabricated a shift register of two cells. Although this is a very short shift register, it can be used as a long register. For our experiment we fabricated the two clocked QCA cells with electrometers coupled to each cell so that we could measure the polarization of each cell independently. In the experiment a bit of information is latched into the first cell, and the input removed. The bit is then copied into the second cell, and erased in the first. Then the bit is copied back into the first cell and erased in the second. In this way the bit is shifted between cells, just as it would be in a long shift register. The experiment demonstrated 5 bit transfers, limited only by thermally induced errors [48–50].

Fan-Out

An important element in a general logic system is fan-out, where the output of one element is sent to the inputs of two or more elements. Since the energy of the output is split, power gain in the following logic elements is needed to restore the signal level. To demonstrate fan-out in QCA we fabricated a circuit with three cells. In the

experiment the input is latched into the central cell, which then acts as an input to the top and bottom cells. When a clock is applied to the top and bottom cells the bit is copied into both cells, and full signal strength is produced in these cells [51], confirming the operation of the circuit.

3.3 Molecular QCA

Molecules represent the smallest artificial structures that can be engineered by humans. To form switchable QCA molecules, at least two charge centers are required that can be reversibly occupied or unoccupied by an electron. The field of mixed-valence chemistry [52] concerns itself with molecules that have at least two charge centers connected by a bridging group through which tunneling can occur. Ongoing investigation concerns the questions of what makes a good dot and what makes a good bridge.

Several early theoretical investigations used model electronic π-systems as dots [53–61]. These molecules are often radical ions containing unpaired electrons and would be very reactive and likely unstable in real systems. Their use was to establish the fundamentals. Electrons in molecules can exhibit bistable switching and the perturbation due to a similar molecule at a reasonable distance (such that the dots form a square) is sufficient to switch the molecule. Energy levels are such that these effects survive room temperature operation. Kink energies are large enough that molecular QCA is robust against variations in position and orientation of molecules. Groups surrounding the charge centers can effectively insulate them from conducting substrates but do not screen the field. Applied electric fields which vary at a much larger length scale can effectively clock molecules (with three appropriately arranged charge centers).

Molecular synthesis by the Fehlner [62–66] and Lapinte [67–69] groups have succeeded in creating molecules that show the requisite bistability. These dots use Fe and/or Ru charge centers. Electronic measurements of the Fehlner molecules attached to a surface showed distinctive bistable behavior as the electron was switched by an applied electric field. This demonstrated both the bistable character of the molecules, and the potential for clocked control of the charge configuration by an applied (and inhomogeneous) field. The Lapinte molecules have been imaged with STM by the Kandel group [67–70] and show the desired charge localization on one end of a symmetric double-dot molecule. Triple-dot molecules, of the sort required for clocked control, have also been made and imaged. More recently ferrocene-base double dot molecules have been made by the Henderson group and imaged by the Kandel group.

There is much chemistry yet to be understood in designing QCA molecules. One issue is what makes the ideal dot. Transition metal atoms have the advantage of using d-orbitals that participate less in bonding and so may be more isolated. Carbon-base π-systems, on the other hand, can be chosen such that they involve anti-bonding orbitals and may spread the charge out more and therefore yield a lower reorganization energy. The reorganization energy is the energy associated with the relaxation of the surrounding atoms and may in some cases trap the charge and inhibit switching. Creating appropriate bridging groups involves choosing a system that is either long enough or opaque enough to be an effective barrier to through-bond tunneling. Conjugated systems may be too conducting.

Another approach that combines single-atom realization with lithographic control is the STM-base lithography of the Wolkow group [71]. They have created room-temperature QCA cells using a remarkable approach involving removing single electrons from dangling bonds on a silicon surface. As with molecules, the single-atom sizes easily yield room temperature operation, yet the placement and orientation of the cells can be controlled lithographically using the STM tip.

3.4 Other Implementations

Nanomagnetic QCA was first introduced by Cowburn's group [72] and developed extensively by the Porod group [73] and the Bokar group [74]. The mapping from QCA cells that represent an electric quadrupole to those that represent a magnetic quadrupole is straightforward. Nanomagnetic implementations are discussed elsewhere in this volume.

Some have proposed cell-cell coupling based on an electron exchange interaction [75], and indeed the earliest calculations showed a small splitting between the singlet and triplet spin states [76]. This approach has two serious drawbacks: the exchange splitting is quite small, and it is zero if there is not tunneling from cell to cell. If there is tunneling from cell to cell, the information is no longer localized and spin-wave solutions predominate.

It is interesting to consider the fundamental question of what sort of systems could implement QCA action. There are two basic features of QCA that must be satisfied:

1. A bit is to be represented completely by the local state of a cell composed of atoms.
2. The interaction between cells is through a field, rather than by transport.

The cell's binary information must therefore be represented by the positional or spin degrees of freedom of the electrons and nuclei in the cell. Nuclear positions could be used to encode the information—for molecules this would entail a conformational change, for larger cells we would call it mechanical. Lighter mass electrons have an advantage in that they can switch positions faster than nuclei. Spin states of either nuclei or electrons could switch quickly.

The field connecting cells must be electromagnetic because the other candidate fields are either too short range (the nuclear strong or weak forces) or too weak (the gravitation force). Direct spin-spin interaction energies are very small, so for magnetic coupling we need many nuclear or electronic spins acting collectively. Nanomagetic QCA thus must be sufficiently large that the coupling is adequate at room temperatures. The direct Coulomb interaction is quite strong and allows molecule-to-molecule coupling between multipole moments of the charge distribution. Since by assumption there is no transport from cell to cell, the charge of the cell cannot change and higher moments must be used to encode the information. QCA thus far has used dipole coupling and quadrupole coupling—the difference being what one chooses to define as a cell. It is also possible to use contact potentials along the cell surface to coupled mechanical (nuclear) degrees of freedom. The possibilities then can be seen to reduce to these categories:

1. Mechanical cells coupled by electrostatic contact potential forces. These would suffer from the slower response of atoms compared with electrons, but remain largely unexplored.
2. Magnetic cells (collective nuclear or electronic spins) coupled by magnetic fields. This is the basis of nanomagnetic QCA described elsewhere in this volume.
3. Electronic cells coupled by Coulomb multipole-multipole interactions. The charge distribution could be the result of either mobile atoms or mobile electrons, and could involve a few or several charges.

Few-electron QCA cells, as has been noted above, have an intrinsic bistability due to charge quantization. If there are many charges forming the charge multipole, the bistability must be provided by another mechanism. One example is a CMOS cell that is an analog to QCA and switches adiabatically [77].

4 Issues in QCA Development

4.1 The Role of Quantum Mechanics in QCA

To achieve robustness against fabrication variations, the QCA paradigm uses only a classical degree of freedom, the electric (or magnetic) quadrupole moment of the cell. It does not use quantum phase information nor interference effects. QCA involves bits not qubits. It is quantum mechanical precisely in that it relies on quantum tunneling for cell switching. This is crucial because if quantum mechanics were "turned off" ($\hbar = 0$) there would be no tunneling and a QCA cell could not switch. If the barriers to tunneling were removed so that classical switching were allowed, a QCA cell would oscillate and settle into a particular configuration randomly depending on the details of the trajectory and energy dissipation. Switching a classical double-well system is much more prone to error because the system can get caught in a metastable state if the timing is not perfect. Reliance on quantum tunneling stabilizes the bit information.

4.2 Power Gain

In molecular, metal dot (discussed above), or other implementations, power gain is crucial because there is always some dissipation of energy as information moves from stage to stage in a computation. This dissipation is the microscopic version of friction in mechanical devices. It can be minimized, and by moving gradually can be reduced to whatever level is desired, but cannot be completely eliminated. Therefore, unless there is a way to restore the signal energy, it will eventually be completely attenuated. In conventional devices the source of the energy is usually the constant-voltage power supply. In QCA the restoring energy is provided by the clock; it automatically supplies enough energy to restore the signal levels.

4.3 Metastability, Memory, and Coherence

For a physical system to act as a memory its state cannot be determined by only its boundary conditions. A Hamiltonian system in a unique ground state, for example,

cannot act as a memory. A device with even short-term memory must therefore be in a metastable state. It could be in a state representing a 1 or be in a state representing a 0. Which state the device is in is depends not just on its boundary conditions, but also on its past. If there is a large enough kinetic barrier between these two states they can often be justifiably treated as distinct energetically degenerate states, but they are actually metastable states very weakly coupled and with a very long Rabi oscillation period.

In clocked QCA wires (i.e., shift registers), information is represented by bit packets, a few cells in the line that are polarized in the 1 or 0 state [78]. Since they could also be in the opposite state energetically, it is true that if the time evolution was completely unitary, the bit could quantum mechanically oscillate from one state to the other. There is a considerable kinetic barrier to doing that, however, just as in the case of a CMOS bit. Moreover, in real systems entanglement with the environment stabilizes the bit packet by loss of quantum phase in the system [79]. Decoherence is precisely this sort of entanglement with the environment and, though it is detrimental for quantum computing, it actually stabilizes QCA bits. Further exploration of the roles of environmentally-induced decoherence and energy dissipation are part of the broader question of the transition between the quantum and classical worlds.

4.4 Wire Crossings

QCA is naturally an in-plane technology; it does not require going out of the plane. How can one therefore cross wires, that is move one bit independently across the path of another? Several proposals have been made that accomplish this. (1) The original wire-crossing proposal was to use the symmetry of cells and the second-near-neighbor coupling (suitably strengthened by duplication) to allow one cell line to communicate across the path of another. The limitation here is the amount of control in placement and orientation required. (2) A permuter is a logical function which simply switches inputs A and B to output B and A. This can be done with logic, though it does take several cells to implement [80]. (3) With expanded clocking timing one can have one bit packet cross a wire intersection horizontally at one time, and vertically at a later time. The cost is in the added complexity of the clocking circuitry. (4) A bridge crossover, similar to the CMOS via structure can be constructed that takes QCA cells out of the plane to cross. (5) In many instances the crossing is for distribution of signals to different parts of a logic array. Tougaw and Khatun have designed a general matrix distribution scheme, again using augmented clocking patterns [81].

4.5 Computational Architecture

It is clear that QCA requires rethinking circuit and computer architecture on the basis of this new device paradigm. Nevertheless because QCA still supports Boolean logic function, it is natural that the first designs are taken over from usual logic circuitry. Much design work is underway, supported crucially by the design tool QCADesigner produced by the Walus group [82]. The goals of this important effort engaging many research groups are both to capitalize on the functional density that QCA cells allow

(molecular cells are ~1 nm square), and to exploit the clocking paradigm wherein all communication is via shift registers. Kogge and Niemier have been early pioneers of this [83–88], suggesting a universal clocking floorplan which supports general data-flow. Tougaw et al. [89] and Niemier et al. [90] have explored programmable QCA logic arrays, a promising area. Exploring QCA architectures and circuit ideas is a very active area, as witnessed by other contributions to this volume.

4.6 Lithograph and Self-assembly

QCA circuits need to have a designed layout that reflects the circuit function; entirely regular arrays do not have interesting behavior. As a consequence the information contained in the circuit layout must be imposed and this is usually done, as with all extant semiconductor circuits, through lithography. It is possible, however, to have self-assembly take care of constructing the dots and forming the dots into cells, and perhaps even forming the cells into lines or other functional groups. For molecular QCA, this is particularly promising because bottom-up self-assembly of molecules into supra-molecular structures is a common, albeit demanding, strategy. Building in some level of self-assembly from below, where the cell size is about 1 nm, and imposing circuit structure from above using lithographic techniques, which can reach below 10 nm, is an appealing match of technologies.

Another approach that has received some attention is to use DNA or PNA [91] self-assembled structures as "molecular circuit-boards." DNA structures with surprising amount of inhomogeneous patterning have been synthesized using Seeman tiles [92], or the more recent DNA origami techniques [93]. The long-range concept would be to engineering attachment sites in the DNA scaffold which would covalently bond appropriate QCA molecules or supramolecular assemblies [94–97]. The geometric information that defines the circuit layout would in this way be expressed through the sequencing of base-pairs that self-assemble into the scaffold. Many issues remain, of course, including the requirements of geometric matching to the DNA repeat distance and the polyanionic nature of DNA, which could interfere with QCA operation. PNA scaffolds are neutral and could potentially solve this problem.

4.7 Energy Dissipation

QCA has two fundamental motivators: ultra-small devices in large functional-density arrays, and low power dissipation. Power dissipation has been a major driver in every stage of the evolution of microelectronics. Adiabatic switching between instantaneous ground states allows the absolute minimum dissipation of energy to the environment. As Landauer [22] and Bennett [98] showed, there is no fundamental lower limit to the amount of energy that needs to be dissipated as heat in order to compute a bit of information. If information is erased, however, a minimum about of energy equal to $k_BT \log(2)$ must be dissipated. The combination of these two ideas is known as Landauer's Principle (LP) and is connected to the Maxwell Demon [99, 100]. Though there is a substantial consensus on the correctness of LP, it has come under criticism

from both industrial researchers [101, 102] and philosophers of science [103, 104]. The low power dissipation in QCA is an example of LP in action [105]. We have recently demonstrated experimentally that a binary switch can be operated with dissipation of 0.01 k_BT, in agreement with LP [106–108].

5 Future Prospects

QCA research activity continues on several fronts. Molecular QCA will require improved understanding of the chemistry of mixed-valence molecules. This includes exploring linker and dot moieties, the role of ligand relaxation in charge transfer, surface attachment and molecular-scale patterning. Significant progress in nanomagnetic implementations is reported by several other contributors in this collection. Metal-dot QCA deserves more exploration, even though it requires cryogenic operation. New fabrication methods may raise the operating temperature considerably. Exploration of circuits and computational architectures is crucial for fully exploiting the potential of locally interconnected nanodevices.

References

1. Wharam, D., Thornton, T., Newbury, R., Pepper, M., Ahmed, H., Frost, J., Hasko, D., Peacock, D., Ritchie, D., Jones, G.: One-dimensional transport and the quantisation of the ballistic resistance. J. Phys. C: Solid State Phys. **21**, L209 (1988)
2. Van Wees, B., Van Houten, H., Beenakker, C., Williamson, J.G., Kouwenhoven, L., Van der Marel, D., Foxon, C.: Quantized conductance of point contacts in a two-dimensional electron gas. Phys. Rev. Lett. **60**, 848 (1988)
3. Kastner, M.: The single electron transistor and artificial atoms. Ann. Phys. (Leipzig) **9**, 885–894 (2000)
4. Meurer, B., Heitmann, D., Ploog, K.: Single-electron charging of quantum-dot atoms. Phys. Rev. Lett. **68**, 1371 (1992)
5. Goodnick, S.M., Bird, J.: Quantum-effect and single-electron devices. IEEE Trans. Nanotechnology **2**, 368–385 (2003)
6. Landauer, R.: Nanostructure physics: fashion or depth. In: Reed, M.A., Kirk, W.P. (eds.) Nanostructure Physics and Fabrication, pp. 17–30. Academic Press, New York (1989)
7. Fulton, T.A., Dolan, G.H.: Observation of single-electron charging effects in small tunnel junctions. Phys. Rev. Lett. **59**, 109–112 (1987)
8. Averin, D., Likharev, K.: Coulomb blockade of single-electron tunneling, and coherent oscillations in small tunnel junctions. J. Low Temp. Phys. **62**, 345–373 (1986)
9. Lent, C.S.: A simple model of Coulomb effects in semiconductor nanostructures. In: Kirk, W.P., Reed, M.A. (eds.) Nanostructures and Mesoscopic Systems, p. 183. Academic Press, San Diego (1992)
10. Toffoli, T., Margolus, N.: Cellular Automata Machines: A New Environment for Modelling. MIT Press, Cambridge (1987)
11. Lent, C.S., Tougaw, P.D., Porod, W.: A bistable quantum cell for cellular automata. In: International Workshop Computational Electronics, pp. 163–166 (1992)

12. Lent, C.S., Tougaw, P.D.: Lines of interacting quantum-dot cells: a binary wire. J. Appl. Phys. **74**, 6227–6233 (1993)
13. Lent, C.S., Tougaw, P.D., Porod, W., Bernstein, G.H.: Quantum cellular automata. Nanotechnology **4**, 49 (1993)
14. Tougaw, P.D., Lent, C.S.: Logical devices implemented using quantum cellular automata. J. Appl. Phys. **75**, 1818–1825 (1994)
15. Lent, C.S., Tougaw, P.D.: Bistable saturation due to single electron charging in rings of tunnel junctions. J. Appl. Phys. **75**, 4077–4080 (1994)
16. Tougaw, P.D., Lent, C.S.: Dynamic behavior of quantum cellular automata. J. Appl. Phys. **80**, 4722–4736 (1996)
17. Landauer, R.: Is quantum mechanics useful? Philos. Trans. R. Soc. Lond. Ser. A: Phys. Eng. Sci. **353**, 367–376 (1995)
18. Lent, C.S., Tougaw, P.D., Porod, W.: In: Proceedings of the Workshop on the Physics and Computation, PhysComp'94, pp. 5–13. IEEE (1994)
19. Lent, C.S., Tougaw, P.D.: A device architecture for computing with quantum dots. Proc. IEEE **85**, 541–557 (1997)
20. Korotkov, A.N., Likharev, K.K.: Single-electron-parametron-based logic devices. J. Appl. Phys. **84**, 6114–6126 (1998)
21. Tóth, G., Lent, C.S.: Quasiadiabatic switching for metal-island quantum-dot cellular automata. J. Appl. Phys. **85**, 2977–2984 (1999)
22. Keyes, R.W., Landauer, R.: Minimal energy dissipation in logic. IBM J. Res. Dev. **14**, 152–157 (1970)
23. Gardelis, S., Smith, C., Cooper, J., Ritchie, D., Linfield, E., Jin, Y.: Evidence for transfer of polarization in a quantum dot cellular automata cell consisting of semiconductor quantum dots. Phys. Rev. B **67**, 033302 (2003)
24. Perez-Martinez, F., Farrer, I., Anderson, D., Jones, G., Ritchie, D., Chorley, S., Smith, C.: Demonstration of a quantum cellular automata cell in a GaAs/AlGaAs heterostructure. Appl. Phys. Lett. **91**, 032102–032103 (2007)
25. Smith, C., Gardelis, S., Rushforth, A., Crook, R., Cooper, J., Ritchie, D., Linfield, E., Jin, Y., Pepper, M.: Realization of quantum-dot cellular automata using semiconductor quantum dots. Superlattices Microstruct. **34**, 195–203 (2003)
26. Macucci, M., Gattobigio, M., Bonci, L., Iannaccone, G., Prins, F., Single, C., Wetekam, G., Kern, D.: A QCA cell in silicon-on-insulator technology: theory and experiment. Superlattices Microstruct. **34**, 205–211 (2003)
27. Single, C., Augke, R., Prins, F., Wharam, D., Kern, D.: Towards quantum cellular automata operation in silicon: transport properties of silicon multiple dot structures. Superlattices Microstruct. **28**, 429–434 (2000)
28. Single, C., Augke, R., Prins, F., Wharam, D., Kern, D.: Single-electron charging in doped silicon double dots. Semicond. Sci. Technol. **14**, 1165 (1999)
29. Single, C., Prins, F., Kern, D.: Simultaneous operation of two adjacent double dots in silicon. Appl. Phys. Lett. **78**, 1421–1423 (2001)
30. Mitic, M., Cassidy, M., Petersson, K., Starrett, R., Gauja, E., Brenner, R., Clark, R., Dzurak, A., Yang, C., Jamieson, D.: Demonstration of a silicon-based quantum cellular automata cell. Appl. Phys. Lett. **89**, 013503-013503-013503 (2006)
31. Dzurak, A.S., Simmons, M.Y., Hamilton, A.R., Clark, R.G., Brenner, R., Buehler, T.M., Curson, N.J., Gauja, E., McKinnon, R.P., Macks, L.D.: Construction of a silicon-based solid state quantum computer. Quantum Inf. Comput. **1**, 82–95 (2001)
32. Davies, J.H., Nixon, J.A.: Fluctuations in submicrometer semiconducting devices caused by the random positions of dopants. Phys. Rev. B **39**, 3423 (1989)

33. Lafarge, P., Pothier, H., Williams, E.R., Esteve, D., Urbina, C., Devoret, M.H.: Direct observation of macroscopic charge quantization. Z. Phys. B **85**, 327–332 (1991)
34. Pothier, H., Lafarge, P., Orfila, P.F., Urbina, C., Esteve, D., Devoret, M.H.: Single electron pump fabricated with ultrasmall normal tunnel junctions. Phys. B **169**, 573 (1991)
35. Zimmerman, N.M., Huber, W.H., Simonds, B., Hourdakis, E., Fujiwara, A., Ono, Y., Takahashi, Y., Inokawa, H., Furlan, M., Keller, M.W.: Why the long-term charge offset drift in Si single-electron tunneling transistors is much smaller (better) than in metal-based ones: two-level fluctuator stability. J. Appl. Phys. **104**, 033710 (2008)
36. Orlov, A.O., Amlani, I., Bernstein, G.H., Lent, C.S., Snider, G.L.: Realization of a functional cell for quantum-dot cellular automata. Science **277**, 928–930 (1997)
37. Aassime, A., Gunnarsson, D., Bladh, K., Delsing, P., Schoelkopf, R.: Radio-frequency single-electron transistor: toward the shot-noise limit. Appl. Phys. Lett. **79**, 4031–4033 (2001)
38. Amlani, I., Orlov, A.O., Snider, G.L., Bernstein, G.H.: Differential charge detection for quantum-dot cellular automata. J. Vac. Sci. Technol. B **15**, 2832–2835 (1997)
39. Amlani, I., Orlov, A.O., Snider, G.L., Lent, C.S., Bernstein, G.H.: Demonstration of a six-dot quantum cellular automata system. Appl. Phys. Lett. **72**, 2179–2181 (1998)
40. Bernstein, G.H., Bazan, G., Chen, M., Lent, C.S., Merz, J.L., Orlov, A.O., Porod, W., Snider, G.L., Tougaw, P.D.: Practical issues in the realization of quantum-dot cellular automata. Superlattices Microstruct. **20**, 447–459 (1996)
41. Snider, G.L., Orlov, A.O., Amlani, I., Zuo, X., Bernstein, G.H., Lent, C.S., Merz, J.L., Porod, W.: Quantum-dot cellular automata: review and recent experiments (invited). J. Appl. Phys. **85**, 4283–4285 (1999)
42. Amlani, I., Orlov, A.O., Toth, G., Bernstein, G.H., Lent, C.S., Snider, G.L.: Digital logic gate using quantum-dot cellular automata. Science **284**, 289–291 (1999)
43. Toth, G., Orlov, A.O., Amlani, I., Lent, C.S., Bernstein, G.H., Snider, G.L.: ARTICLES-semiconductors II: surfaces, interfaces, microstructures, and related topics-Conductance suppression due to correlated electron transport in coupled double quantum dots. Phys. Rev.-Section B-Condens. Matter **60**, 16906–16912 (1999)
44. Orlov, A.O., Amlani, I., Kummamuru, R.K., Ramasubramaniam, R., Toth, G., Lent, C.S., Bernstein, G.H., Snider, G.L.: Experimental demonstration of clocked single-electron switching in quantum-dot cellular automata. Appl. Phys. Lett. **77**, 295–297 (2000)
45. Kummamuru, R.K., Liu, M., Orlov, A.O., Lent, C.S., Bernstein, G.H., Snider, G.L.: Temperature dependence of the locked mode in a single-electron latch. Microelectron. J. **36**, 304–307 (2005)
46. Orlov, A.O., Kummamuru, R.K., Ramasubramaniam, R., Toth, G., Lent, C.S., Bernstein, G.H., Snider, G.L.: Experimental demonstration of a latch in clocked quantum-dot cellular automata. Appl. Phys. Lett. **78**, 1625–2627 (2001)
47. Kummamuru, R.K., Timler, J., Toth, G., Lent, C.S., Ramasubramaniam, R., Orlov, A.O., Bernstein, G.H., Snider, G.L.: Power gain in a quantum-dot cellular automata latch. Appl. Phys. Lett. **81**, 1332–1334 (2002)
48. Kummamuru, R.K., Orlov, A.O., Ramasubramaniam, R., Lent, C.S., Bernstein, G.H., Snider, G.L.: Operation of a quantum-dot cellular automata (QCA) shift register and analysis of errors. IEEE Trans. Electron Devices **50**, 1906–1913 (2003)
49. Orlov, A.O., Kummamuru, R., Lent, C.S., Bernstein, G.H., Snider, G.L.: Clocked quantum-dot cellular automata shift register. Surf. Sci. **532–535**, 1193–1198 (2003)
50. Orlov, A.O., Kummamuru, R., Ramasubramaniam, R., Lent, C.S., Bernstein, G.H., Snider, G.L.: A two-stage shift register for clocked quantum-dot cellular automata. J. Nanosci. Nanotechnol. **2**, 351–355 (2002)

51. Yadavalli, K.K., Orlov, A.O., Timler, J.P., Lent, C.S., Snider, G.L.: Fanout gate in quantum-dot cellular automata. Nanotechnology **18**, 1–4 (2007)
52. Demadis, K.D., Hartshorn, C.M., Meyer, T.J.: The localized-to-delocalized transition in mixed-valence chemistry. Chem. Rev. **101**, 2655–2686 (2001)
53. Blair, E., Lent, C.S.: In: Third IEEE Conference on Nanotechnology, 2003. IEEE-NANO 2003, vol. 1, pp. 402–405. IEEE (2003)
54. Blair, E.P., Lent, C.S.: In: International Conference on Simulation of Semiconductor Processes and Devices, 2003. SISPAD 2003, pp. 14–18. IEEE (2003)
55. Hennessy, K., Lent, C.S.: Clocking of molecular quantum-dot cellular automata. J. Vacuum Sci. Technol. B: Microelectron. Nanometer Struct. **19**, 1752–1755 (2001)
56. Isaksen, B., Lent, C.S.: In: Third IEEE Conference on Nanotechnology, 2003. IEEE-NANO 2003, vol. 1, pp. 5–8. IEEE (2003)
57. Lent, C.S.: In: APS Meeting Abstracts, vol. 1, pp. 14002 (2000)
58. Lent, C.S., Isaksen, B.: Clocked molecular quantum-dot cellular automata. IEEE Trans. Electron Devices **50**, 1890–1896 (2003)
59. Lent, C.S., Isaksen, B., Lieberman, M.: Molecular quantum-dot cellular automata. J. Am. Chem. Soc. **125**, 1056–1063 (2003)
60. Lieberman, M., Chellamma, S., Varughese, B., Wang, Y., Lent, C.S., Bernstein, G.H., Snider, G.L., Peiris, F.C.: Quantum-dot cellular automata at a molecular scale. Ann. N. Y. Acad. Sci. **960**, 225–239 (2002)
61. Lieberman, M., Chellamma, S., Wang, Y., Hang, Q., Bernstein, G., Lent, C.S.: In: Abstracts of Papers of the American Chemical Society. (Amer Chemical Soc 1155 16th ST, NW, Washington, DC 20036 USA, 2002), vol. 224, p. U474 (2002)
62. Joyce, R.A., Qi, H., Fehlner, T.P., Lent, C.S., Orlov, A.O., Snider, G.L.: In: Nanotechnology Materials and Devices Conference, 2009. NMDC'09, pp. 46–49. IEEE (2009)
63. Qi, H., Gupta, A., Fehlner, T.P., Snider, G.L., Lent, C.S.: In: Abstracts of Papers of the American Chemical Society. (Amer Chemical Soc 1155 16TH ST, NW, WASHINGTON, DC 20036 USA, 2006), vol. 232 (2006)
64. Qi, H., Gupta, A., Noll, B.C., Snider, G.L., Lu, Y., Lent, C.S., Fehlner, T.P.: Dependence of field switched ordered arrays of dinuclear mixed-valence complexes on the distance between the redox centers and the size of the counterions. J. Am. Chem. Soc. **127**, 15218–15227 (2005)
65. Qi, H., Li, Z., Snider, G.L., Lent, C.S., Fehlner, T.: Field driven electron switching in a silicon surface bound array of vertically oriented two-dot molecular quantum cellular automata (2003)
66. Qi, H., Sharma, S., Li, Z., Snider, G.L., Orlov, A.O., Lent, C.S., Fehlner, T.P.: Molecular quantum cellular automata cells. Electric field driven switching of a silicon surface bound array of vertically oriented two-dot molecular quantum cellular automata. J. Am. Chem. Soc. **125**, 15250–15259 (2003)
67. Lu, Y., Quardokus, R., Lent, C.S., Justaud, F., Lapinte, C., Kandel, S.A.: Charge localization in isolated mixed-valence complexes: an STM and theoretical study. J. Am. Chem. Soc. **132**, 13519–13524 (2010)
68. Quardokus, R.C., Lu, Y., Wasio, N.A., Lent, C.S., Justaud, F., Lapinte, C., Kandel, S.A.: Through-bond versus through-space coupling in mixed-valence molecules: observation of electron localization at the single-molecule scale. J. Am. Chem. Soc. **134**, 1710–1714 (2012)

69. Wasio, N.A., Quardokus, R.C., Forrest, R.P., Corcelli, S.A., Lu, Y., Lent, C.S., Justaud, F., Lapinte, C., Kandel, S.A.: STM imaging of three-metal-center molecules: comparison of experiment and theory for two mixed-valence oxidation states. J. Phys. Chem. C **116**, 25486–25492 (2012)
70. Quardokus, R.C., Wasio, N.A., Forrest, R.P., Lent, C.S., Corcelli, S.A., Christie, J.A., Henderson, K.W., Kandel, S.A.: Adsorption of diferrocenylacetylene on Au (111) studied by scanning tunneling microscopy. Phys. Chem. Chem. Phys. **15**, 6973–6981 (2013)
71. Haider, M.B., Pitters, J.L., DiLabio, G.A., Livadaru, L., Mutus, J.Y., Wolkow, R.A.: Controlled coupling and occupation of silicon atomic quantum dots at room temperature. Phys. Rev. Lett. **102**, 046805 (2009)
72. Cowburn, R., Welland, M.: Room temperature magnetic quantum cellular automata. Science **287**, 1466–1468 (2000)
73. Imre, A., Csaba, G., Ji, L., Orlov, A., Bernstein, G., Porod, W.: Majority logic gate for magnetic quantum-dot cellular automata. Science **311**, 205–208 (2006)
74. Lambson, B., Carlton, D., Bokor, J.: Exploring the thermodynamic limits of computation in integrated systems: magnetic memory, nanomagnetic logic, and the Landauer limit. Phys. Rev. Lett. **107**, 010604 (2011)
75. Bandyopadhyay, S., Das, B., Miller, A.: Supercomputing with spin-polarized single electrons in a quantum coupled architecture. Nanotechnology **5**, 113 (1994)
76. Tougaw, P.D., Lent, C.S., Porod, W.: Bistable saturation in coupled quantum-dot cells. J. Appl. Phys. **74**, 3558–3566 (1993)
77. Hänninen, I., Lu, H., Lent, C.S., Snider, G.L.: Energy recovery and logical reversibility in adiabatic CMOS multiplier. In: Dueck, G.W., Miller, D. (eds.) RC 2013. LNCS, vol. 7948, pp. 25–35. Springer, Heidelberg (2013)
78. Blair, E.P., Liu, M., Lent, C.S.: Signal energy in quantum-dot cellular automata bit packets. J. Comput. Theor. Nanosci. **8**, 972–982 (2011)
79. Blair, E.P., Lent, C.S.: Environmental decoherence stabilizes quantum-dot cellular automata. J. Appl. Phys. **113**, 124302-124302-124316 (2013)
80. Fijany, A., Toomarian, N., Spotnizt, M.: Implementing permutation matrices by use of quantum dots. NASA Technical Briefs **25** (2001)
81. Tougaw, D., Khatun, M.: A scalable signal distribution network for quantum-dot cellular automata. IEEE Trans. Nanotech. **12**, 215–224 (2013)
82. Walus, K., Dysart, T.J., Jullien, G.A., Budiman, R.A.: QCADesigner: a rapid design and simulation tool for quantum-dot cellular automata. IEEE Trans. Nanotechnol. **3**, 26–31 (2004)
83. Dysart, T.J., Kogge, P.M., Lent, C.S., Liu, M.: An analysis of missing cell defects in quantum-dot cellular automata. In: IEEE International Workshop on Design and Test of Defect-Tolerant Nanoscale Architectures (NANOARCH) (2005)
84. Frost, S.E., Rodrigues, A.F., Giefer, C.A., Kogge, P.M.: In: IEEE Computer Society Annual Symposium on VLSI 2003. Proceedings, pp. 19–25. IEEE (2003)
85. Frost, S.E., Rodrigues, A.F., Janiszewski, A.W., Rausch, R.T., Kogge, P.M.: In: First Workshop on Non-Silicon, Computing, vol. 2 (2002)
86. Niemier, M.T., Kogge, P.M.: In: The 6th IEEE International Conference on Electronics, Circuits and Systems, 1999. Proceedings of ICECS'99, vol. 3, pp. 1211–1215. IEEE (1999)
87. Niemier, M.T., Kogge, P.M.: Exploring and exploiting wire-level pipelining in emerging technologies. ACM SIGARCH Comput. Archit. News **29**, 166–177 (2001)
88. Niemier, M.T., Kogge, P.M.: In: Proceedings of the Third Petaflops Workshop, with Frontiers of Massively Parallel Processing (1999)

89. Tougaw, D., Johnson, E.W., Egley, D.: Programmable logic implemented using quantum-dot cellular automata. IEEE Trans. Nanotechnol. **11**, 739–745 (2012)
90. Niemier, M.T., Rodrigues, A.F., Kogge, P.M.: In: 1st Workshop on Non-silicon Computation, pp. 38–45 (2002)
91. Lukeman, P.S., Mittal, A.C., Seeman, N.C.: Two dimensional PNA/DNA arrays: estimating the helicity of unusual nucleic acid polymers. Chem. Commun. **2004**, 1694–1695 (2004)
92. Winfree, E., Liu, F., Wenzler, L.A., Seeman, N.C.: Design and self-assembly of two-dimensional DNA crystals. Nature **394**, 539–544 (1998)
93. Rothemund, P.W.: Folding DNA to create nanoscale shapes and patterns. Nature **440**, 297–302 (2006)
94. Gu, H., Chao, J., Xiao, S.-J., Seeman, N.C.: Dynamic patterning programmed by DNA tiles captured on a DNA origami substrate. Nat. Nanotechnol. **4**, 245–248 (2009)
95. Kim, K.N., Sarveswaran, K., Mark, L., Lieberman, M.: DNA origami as self-assembling circuit boards. In: Calude, C.S., Hagiya, M., Morita, K., Rozenberg, G., Timmis, J. (eds.) Unconventional Computation. LNCS, vol. 6079, pp. 56–68. Springer, Heidelberg (2010)
96. Sarveswaran, K., Huber, P., Lieberman, M., Russo, C., Lent, C.S.: In: Third IEEE Conference on Nanotechnology, 2003. IEEE-NANO 2003, vol. 1, pp. 417–420. IEEE (2003)
97. Sarveswaran, K., Russo, C., Robinson, A., Huber, P., Lent, C.S., Lieberman, M.: In: Abstracts of Papers of the American Chemical Society. (Amer Chemical Soc 1155 16TH ST, NW, Washington, DC 20036 USA, 2002), vol. 224, p. U421 (2002)
98. Bennett, C.H.: Logical reversibility of computation. IBM J. Res. Dev. **17**, 525–532 (1973)
99. Leff, H., Rex, A.F.: Maxwell's Demon 2 Entropy, Classical and Quantum Information, Computing, vol. 2. CRC Press, Boca Raton (2010)
100. Timler, J., Lent, C.S.: Maxwell's demon and quantum-dot cellular automata. J. Appl. Phys. **94**, 1050–1060 (2003)
101. Cavin, R.K., Zhirnov, V.V., Hutchby, J.A., Bourianoff, G.I.: Energy barriers, demons, and minimum energy operation of electronic devices. Fluctuation Noise Lett. **5**, C29–C38 (2005)
102. Zhirnov, V.V., Cavin III, R.K., Hutchby, J.A., Bourianoff, G.I.: Limits to binary logic switch scaling - a Gedanken model. Proc. IEEE **91**, 1934–1939 (2003)
103. Earman, J., Norton, J.D.: Exorcist XIV: the wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond. Stud. History Phil. Sci. Part B: Stud. History Phil. Mod. Phys. **30**, 1–40 (1999)
104. Norton, J.D.: Waiting for Landauer. Stud. History Phil. Sci. Part B: Stud. History Phil. Mod. Phys. **42**, 184–198 (2011)
105. Lent, C.S., Liu, M., Lu, Y.: Bennett clocking of quantum-dot cellular automata and the limits to binary logic scaling. Nanotechnology **17**, 4240 (2006)
106. Orlov, A.O., Lent, C.S., Thorpe, C.C., Boechler, G.P., Snider, G.L.: Experimental test of Landauer's principle at the sub-kBT level. Jpn. J. Appl. Phys. **51**, 06FE10 (2012)
107. Snider, G.L., Blair, E.P., Thorpe, C.C., Appleton, B.T., Boechler, G.P., Orlov, A.O., Lent, C.S.: In: 12th IEEE Conference on Nanotechnology (IEEE-NANO), 2012, pp. 1–6. IEEE (2012)
108. Boechler, G.P., Whitney, J.M., Lent, C.S., Orlov, A.O., Snider, G.L.: Fundamental limits of energy dissipation in charge-based computing. Appl. Phys. Lett. **97**, 103502–103503 (2010)

Nanomagnet Logic (NML)

Wolfgang Porod[1(✉)], Gary H. Bernstein[1], György Csaba[1], Sharon X. Hu[2], Joseph Nahas[2], Michael T. Niemier[2], and Alexei Orlov[1]

[1] Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, USA
porod@nd.edu
[2] Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA

Abstract. We describe the background and evolution of our work on magnetic implementations of Quantum-Dot Cellular Automata (QCA), first called Magnetic QCA (MQCA), and now known as Nanomagnet Logic (NML).

Keywords: Nanomagnet logic · Quantum-Dot Cellular Automata · Field-coupled computing · Cellular Automata

1 Introduction

We all are familiar with the fact that magnetic phenomena are widely used for data storage, whereas electronic phenomena are used for information processing. This is based on the fact that *ferromagnetism* is nonvolatile (i.e. magnetization state can be preserved even without power), and that *electrons* and the flow of charge can be effectively controlled to perform logic. However, charge-based logic devices are volatile, which means that a power supply is needed to maintain the logic state and to stop information (electrons) from leaking away.

Presently, there are multiple research efforts underway to harness magnetic phenomena for logic in addition to storage. Motivation for this work is two-fold. First, the amount of static power dissipation in CMOS chips now rivals the levels of dynamic power dissipation. In other words, even if a chip does not perform any computation, stand-by power (i.e., needed to maintain volatile logic state, etc.) is similar to the power dissipated when performing useful work. Anecdotally, in 2004, Bernard Myerson – then chief technologist for IBM's system and technology group – likened the situation to "a car with a 10-gallon gas tank losing 5 gallons while parked with its motor turned off" [2]. The nonvolatile nature of magnetic phenomena means that there is no stand-by power dissipation. Second, the intrinsic switching energy of a magnetic device can be orders of magnitude lower than a charge-based CMOS transistor. While some drive circuitry overhead must be accounted for in magnetic systems, this suggests that magnetic logic could help to minimize dynamic power dissipation as well. Thus, while the multi-decade Moore's Law-based size scaling trends may continue, associated performance scaling trends are threatened by energy-related concerns that magnetic devices could help to alleviate.

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 21–32, 2014.
DOI: 10.1007/978-3-662-43722-3_2, © Springer-Verlag Berlin Heidelberg 2014

One of the driving forces behind this research is the semiconductor industry itself in the form of the Nanoelectronics Research Initiative (NRI) of the Semiconductor Research Corporation (SRC). The NRI was formed in response to the impending "red brick wall" in the industry's road map, which is primarily the result of the inability to manage dissipation associated with computation with field-effect transistors. In an effort to find alternative, *lower-power* device technologies, the NRI is searching for switches based on state variables other than charge. One possibility is the electron's spin, and associated magnetic phenomena. There now are several research efforts underway to explore switches where the logic state is represented by the magnetization of a nanomagnet. These various approaches have been summarized and reviewed in [3], and our work on nanomagnet logic is one of these efforts.

2 Single-Domain Magnets for NML

Nanomagnet Logic (NML) is based on patterned arrays of elongated nanomagnets that are sufficiently small to contain only a single magnetic domain. The magnetization state of a device – i.e. whether it is magnetized along one direction or another, commonly referred to as "up or down" – can be used to represent binary information in the same way that magnetic islands are used to store information in magnetoresistive random access memory (MRAM). Elongated single-domain magnets are essentially tiny bar magnets with poles on each end, that generate strong stray fields that can be used to couple to other nearby magnets. While such magnetic interactions between neighboring nanomagnets are undesirable for data-storage applications, we have demonstrated that these interactions can be exploited to perform logic operations.

It should be emphasized here that such single-domain behavior is rather special and specific to magnets with certain sizes and shapes. For our work with patterned ferromagnetic thin-film permalloy, these sizes are on the order of hundreds down to tens of nanometers (nm). If the magnet is too large, its magnetization state breaks up into multiple internal domains, and the poles at the end – along with their strong fringing fields – disappear. If the magnet is too small, its magnetization state can be switched by random thermal fluctuations, and it no longer has a stable magnetization state; this is the so-called superparamagnetic limit. As a fascinating side comment, Nature has learned to exploit the stray fields associated with single domain magnets for navigation in the Earth's magnetic field. Specialized, so-called magnetotactic bacteria grow perfectly single-domain nanomagnets, that are specific to a particular animal species [4]. By the way, much of our work on NML uses nanomagnets with sizes and shapes between those characteristic for the pigeon and the tuna.

Figure 1 shows a magnetic force microscope (MFM) image of an array of nanoscale magnets with varying sizes and aspect ratios. The coloring represents magnetic contrast, and dark and light spots indicate magnetic poles. It can be seen that these magnets display single-domain behavior if they are sufficiently small and narrow (left side of image), and these poles generate magnetic flux lines that can interact with external magnetic fields, or couple to neighboring magnets. Otherwise, their magnetization state breaks up into several internal domains (right side of image), and there is

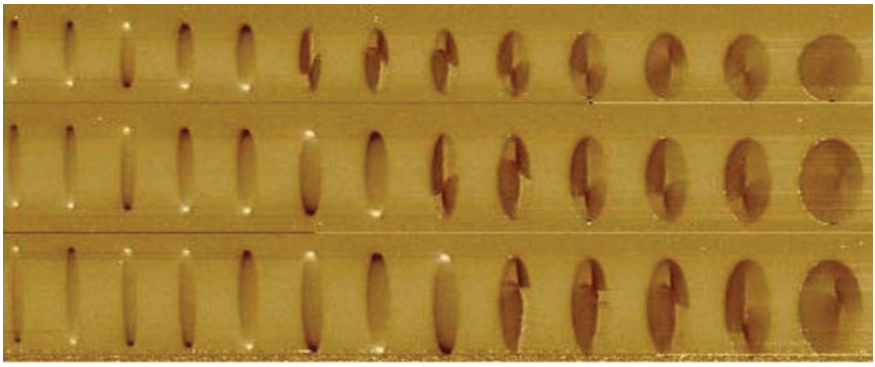


Fig. 1. Nano-scale magnets with varying sizes and aspect ratios. Single-domain behavior is observed if the magnets are sufficiently small and narrow. (Source: A. Imre, Ph.D. dissertation, University of Notre Dame, 2005 [5].)

flux closure inside, without strong coupling to the exterior. Such single-domain magnets, with typical size scales of tens to hundreds of nanometers, form the physical basis for NML.

3 From Quantum-Dot Cellular Automata to Nanomagnet Logic

Our current work on nanomagnet logic grew out of our previous work on Quantum-Dot Cellular Automata (QCA), which was an attempt to base computation on physically-interacting cellular arrays of quantum dots occupied by a few electrons [6]. Instead of wires, neighboring devices interact through direct Coulomb interactions between electrons on neighboring quantum dots. We have shown that such physical interactions in appropriately structured arrays of quantum dots can also be used to realize logic gates. Electronic implementations of QCA proved difficult due to technological limitations of quantum-dot fabrication (such as size variations) and electronic stray charges. For a review of electronic QCA, see Refs. [7, 8].

Nanomagnet logic can be viewed as a magnetic implementation of QCA. (In earlier publications, we used the term magnetic QCA (MQCA), but we now prefer to use NML in order to avoid confusion with quantum dots.) Early theoretical work on magnetic QCA is given in the Ph.D. Dissertation of György Csaba [9]. These simulations demonstrated the feasibility of using field-coupled single-domain nanomagnets for realizing basic logic functionality [10–12]. In subsequent experimental work stimulated by these simulations, and which constituted the Ph.D. Dissertation of Alexandra Imre [5], we first demonstrated magnetic wires formed by chains of near-by magnetic islands. Since the individual dots have elongated shapes, there are two basic types of wire arrangements. In one type, the magnets are lined up side-by-side and, as one's intuition would tell from two bar magnets next to each other with their long sides, the individual magnets prefer to be magnetized in the opposite direction; we call

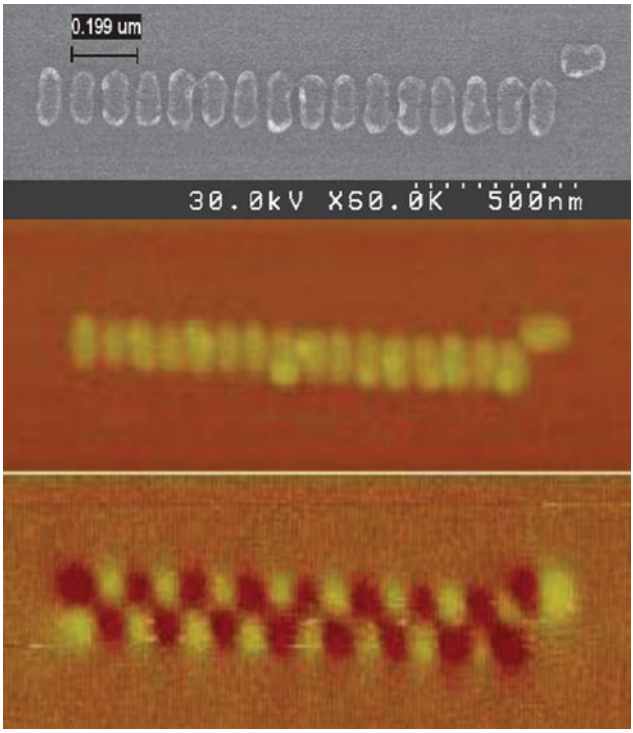


Fig. 2. Demonstration of a magnetically-ordered line of dots. Top panel: SEM, middle panel: AFM, bottom panel: MFM.

this antiferromagnetic coupling. The other type of wire consists of the individual magnets lined up in the same direction, and as one would expect, the individual magnets also prefer to be magnetized in the same direction; we call this ferromagnetic coupling.

Figure 2 shows an example of an antiferromagnetically-coupled wire consisting of a chain of 16 dots. The top portion of the image shows an SEM (scanning electron microscope) image of magnetic islands fabricated from a 30-nm thin film of permalloy using electron beam lithography and standard lift-off techniques. The bottom portion of the figure shows the structural AFM (atomic force microscope) image, and the MFM (magnetic force microscope) image showing the magnetic contrast. The dot on the top right (aligned in the horizontal direction) serves as an input that determines if the wire is aligned up-down-up-down or down-up-down-up.

In these experiments, a magnetic field is required to aid the switching of the array of nanomagnets [12]. When one dot is switched, the fringing fields are not sufficiently strong to switch a neighboring magnet. However, the fringing fields can be used to bias the switching event when an additional switching field is applied. Due to the elongated shape (magnetic shape anisotropy), the dots have very stable magnetization states along the long (magnetic easy) axis. They can be magnetized along the short (magnetic hard) axis by the application of a sufficiently strong magnetic field in that

direction. However, when this field is removed, the dots will "snap" back into the preferred "up" or "down" easy-axis direction, and the fringing fields from the neighbors can bias which way they switch. This switching field acts as a magnetic clock.

It turns out that the "native" logic element for NML is a three-input majority-logic gate, just like for the original electronic QCA. As shown schematically below, this gate consists of a cross-shaped arrangement of five dots, where three of the arms (labeled "A," "B," and "C") represent the inputs, the center dot (labeled "M") calculates the majority vote of these inputs, and the fourth arm (labeled "out") represents the output. This arrangement can also be viewed as the intersection between an antiferromagnetic and a ferromagnetic wire segment. Note that the majority vote is "calculated" through magnetic interactions in this physics-driven NML computing scheme (Fig. 3).

It is interesting to note that such a three-input majority gate can be reduced to either a binary AND or OR gate by viewing one of the inputs as a set-input, which selects the functionality of the gate. For example, if we view "C" as the set-input, and "A" and "B" as the data inputs, then a "0" on "C" means that both "A" and "B" have to be "1" in order to have a majority vote of "1" at the output. In other words, setting "C" to "0" reduces the three-input majority gate to an AND gate for the data inputs "A" and "B." Conversely, setting "C" to "1" results in an OR gate since only either "A" or "B" have to be "1" in order to have a majority vote of "1." This programmability offers interesting possibilities from a computer science perspective since the functionality of this gate can be determined by the current state of the computation.

We have experimentally demonstrated functioning majority-logic gates working properly at room temperature [13]. The figure below [from the Science paper] shows the eight possible input combinations for the three-input majority-logic gate. Note that here the different input combinations were realized by different arrangements of the horizontal input dots. However, the shape-dependent switching behavior of such nanoscale magnets [14] can be exploited to individually address specific inputs, thus providing programmability. We have since fabricated gates with input devices of varying aspect ratios, which has allowed a *single gate structure* to be successfully tested with all eight possible input combinations [15] (Fig. 4).

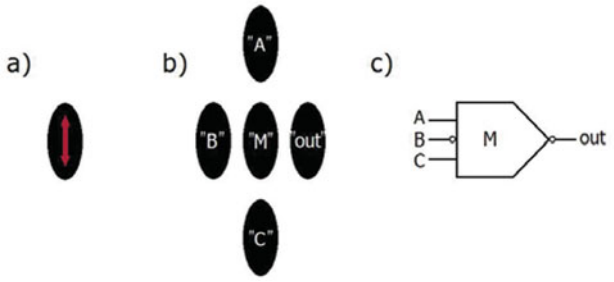


Fig. 3. Schematic of a magnetic three-input majority-logic gate, which consists of a cross-shaped arrangement of five dots. Panel (a) shows the basic dot, (b) the basic logic gate, and (c) the logic-gate symbol.

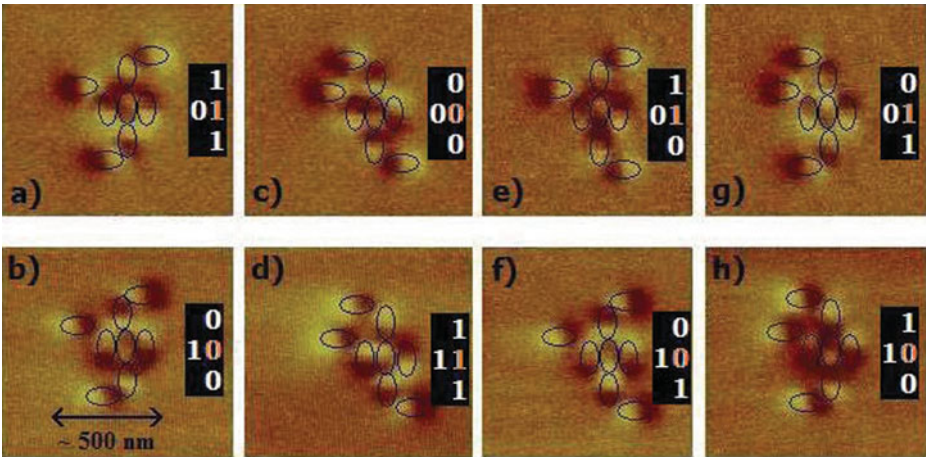


Fig. 4. Experimental demonstration of an NML three-input majority-logic gate. (Source: Imre et al., Ref. [13].)

Thus far [16–18], our work has provided a proof-of-concept demonstration of NML – i.e. the feasibility of performing digital logic with physically-coupled nanomagnets. However, so far we have used MFM to read the state of the dots, and *externally generated* magnetic fields to switch the dots. Of course, this is not practical for real applications. Below, we describe on-going work to develop electronic input and output (I/O) and mechanisms for generating local magnetic fields for "on-chip" clocking.

Before proceeding, we want to emphasize that our experimental proof-of-concept demonstrations satisfy five "tenets" that are considered essential for a digital system: (1) NML devices have non-linear response characteristics due to the magnetic hysteresis loop. (2) NML can deliver a functionally complete logic set enabled by the *3-input* majority gate and the NOT operation naturally achieved by the antiferromagnetic dot-to-dot coupling. (3) Signal amplification/gain greater than 1 has been experimentally demonstrated by showing the feasibility of 1:3 fanout [19], where the energy for the gain is provided by an external clocking field discussed further below. (4) The output of one device can drive another as the fringing fields from individual magnets can bias a neighbor. (5) Unwanted feedback is preventable through clocking.

4 Nanomagnet Logic: Towards System Integration

Recent and ongoing work addresses electronic means for both NML I/O and clocking. Our approach leverages existing MRAM technologies for READ/WRITE operations. After all, setting an input for NML, i.e. setting the state of an input magnet, is similar to writing the state of an MRAM bit. Similarly, reading the state of an NML output dot is just like reading an MRAM bit. These similarities to MRAM suggest that an NML circuit is analogous to a patterned ensemble of the free layers in an MRAM stack.

In this way, NML can leverage much existing technological know-how, and also benefit from future development in MRAM technologies [20].

Under the umbrella of the DARPA Non-Volatile Logic (NVL) program, we worked on approaches to on-chip clocking. As mentioned above, externally supplied switching energy is needed to re-evaluate a magnet ensemble with new inputs. To date, most NML circuits have been "clocked" by an external source. However, it is essential that clock functionality be moved "on-chip." Thus far, the most commonly employed clock is a magnetic field applied along the hard axis of an NML ensemble, which places the magnets into a metastable state such that they are sensitive to the fringing fields from their neighbors. Such magnetic fields can be generated on-chip by current-carrying wires for local control of NML circuits. In recent work, we have fabricated copper wires clad with ferromagnetic material on the sides and bottom (like field-MRAM word and bit lines), and we have demonstrated that NML magnets, interconnect, and logic gates can be switched (i.e. re-evaluated) in this way [21, 22].

Also under the umbrella of the DARPA Non-Volatile Logic (NVL) program, we worked on approaches for integrated electronic I/O. Electronic output can be achieved (similar to MRAM) by a magnetoresistance measurement, where the NML output dot is the free layer in a magnetic tunnel junction (MTJ) stack. Similarly, electronic input can be achieved using the spin-torque transfer (STT) effect, where the NML input dot is the free layer in an STT stack [23, 24].

As is well known from field-MRAM, there is an energy overhead associated with generating local magnetic fields using current-carrying wires. Early on, simulations showed that the overhead associated with such clocking is a major component of the total energy requirement, and that the dissipation associated with the switching of the magnets is rather small [25]. For NML, the clock energy could be amortized over 100,000s of devices as a single clock line could control many parallel ensembles [26]. Clock lines could be placed in series and in multiple planes to minimize driver overhead. Moreover, at cryogenic temperatures, clock lines could be made from superconducting niobium, and I^2R losses could drop to zero. In principle, this opens the door to extremely low energy information processing hardware/memory that could be integrated with RSFQ and SQL logic.

Also inspired by field-MRAM, another approach to lowering the energy overhead associated with clocking is to engineer the dielectric medium between the dots, which influences the coupling strength and thus the switching energy. Specifically, one can enhance the permeability of a dielectric by the controlled inclusion of superparamagnetic particles that increase the dielectric permeability, and thus lower the current required to achieve a certain switching field [27]. Following this approach, we have successfully fabricated such enhanced permeability dielectrics and demonstrated the lowering of switching fields and associated power dissipation [28–30].

Another possible approach to clocking is to exploit the strong local fields associated with a domain wall. We have shown that the motion of domain walls can be controlled [31, 32], and that their local fringing fields can assist in the switching of nearby magnets [33]. This is an interesting approach to NML clocking that deserves further investigation.

Multiferroics, magnetostriction, and spin-torque transfer have also been proposed as potential clocking mechanisms for NML. Multiferroic materials (e.g. BFO) could

allow for electric field control of magnetism, which would be highly attractive for NML. Included in this volume is a contribution from the group of Sayeef Salahuddin at UC Berkeley that addresses this interesting possibility [34].

Another important issue for NML is whether or not the magnets that form a circuit ensemble can be switched reliably – or whether or not devices placed into a meta-stable state by a clock are adversely affected by thermal noise (which could induce premature switching). The group of Jeff Bokor at UC Berkeley has shown that magnets with an extra biaxial anisotropy exhibit superior switching characteristics [35]. Essentially, such an "engineered-in" magnetic anisotropy helps to stabilize the magnets in the "vulnerable" metastable state against random fluctuations. We have shown that shape engineering, i.e. exploiting the influence of geometry on magnetic properties, can be used to not only enhance the reliability of switching, but also to design logic gates with reduced foot print [36, 37].

At Notre Dame, all of our work to date has been based on patterned thin-film permalloy dots, which have in-plane magnetization. An attractive alternative is to use structures with out-of-plane magnetization, such as Co/Pt multi-layer films, where the magnetic properties are due to the Co-Pt interfaces. In collaboration with Doris Schmitt-Landsiedel and her group at the Technical University of Munich (TUM), we are exploring the utility of this material system for NML. It has been shown that such Co/Pt structures can be patterned with a focused ion beam (FIB) instrument, where the ion beam destroys the interfaces, and thus the magnetization at these locations. In this fashion, a film can be patterned into islands, and sufficiently small islands also exhibit single-domain behavior. The TUM group has demonstrated magnetic coupling between neighboring islands [38], and they have shown magnetic ordering in arrays of coupled islands. Moreover, they have realized directional signal propagation in lines, and basic NML logic gates [39], as well as domain-wall assisted switching [40].

All our fabrication work so far has been based on using electron-beam lithography (EBL) to define the NML devices and structures. EBL is a flexible and useful tool for research, but not suitable for large-scale manufacturing. To this end, we collaborate with Paolo Lugli and his group at the TUM to explore the use of nanoimprint lithography and nanotransfer of permalloy structures for the fabrication of large-scale NML arrays [41].

NML represents a technology quite different from CMOS, with its own "pros" and "cons." Undoubtedly, this new technology will likely necessitate new circuit and architecture approaches [42]. Along these lines, we have worked to identify specific application spaces for NML. Our immediate focus is on low energy hardware accelerators for general-purpose multi-core chips, and application spaces that demand information processing hardware that can function with an extremely low energy budget. As an example, we anticipate that NML-based hardware might be used to implement a systolic architecture that can improve the performance of compute-bound applications, provide very high throughput at modest memory bandwidth, and eliminate global signal broadcasts. (Systolic solutions exist for many problems including filtering, polynomial evaluation, discrete Fourier transforms, matrix arithmetic and other non-numeric applications.) Moreover, as devices are non-volatile, information can be stored directly and indefinitely throughout a circuit (e.g. at a gate input)

without the need for explicit storage hardware (and the associated area and static/dynamic power dissipation associated with it).

Architectural-level design techniques such as these should allow us to minimize the "cons" of NML (nearest neighbor dataflow and higher latency devices when compared to CMOS FETs) and exploit the "pros" (inherently pipelined logic with no overhead). As a representative example [43], our projections suggest that hardware for finding specific patterns in incoming data streams could be $\sim$60-75X more energy efficient (at iso-performance) than CMOS hardware equivalents. Moreover, these projections include clock energy overheads.

5 Summary and Discussion

In this chapter, we have presented an overview of our work over the years on nanomagnet logic, which can be viewed as a magnetic implementation of the original QCA field-coupled computing idea. We discussed NML basics, as well as approaches and issues related to the realization of integrated systems. This review was Notre-Dame-centric by design, to provide a somewhat historical perspective on the work of our group.

Finally, we would like to mention a couple of other related research efforts that also use nanomagnets to represent logic state, but that employ different mechanisms to couple and switch these magnets. One such effort, the Spin-Wave Bus proposed by a group at UCLA [44], is based on spin waves propagating in a layer underneath the magnets. Since spin waves (plasmons) decay, this scheme requires amplifying elements to restore the signals. Another scheme, the All-Spin Logic proposed by a group at Purdue [45], is based on nanomagnet coupling by spin diffusion in a magnetic layer underneath the magnets. This scheme requires wires to be connected to the magnetic dots in order to inject spin-polarized electrons that then diffuse and provide the coupling mechanism. These approaches are interesting, and further research is warranted. However, in our opinion, it is hard to see how coupling between dots using either spin waves or spin diffusion can be more efficient or lower power than coupling by direct magnetic fringing fields.

We end with a historical note. It was recognized in the very early days of digital computer design that magnetic phenomena are attractive for several reasons [46]: They possess an inherent high reliability; They require in most applications no power other than the power to switch their state; They are potentially able to perform all required operations, i.e., logic, storage and amplification. In fact, some of the very early computers used ferrite cores not only for memory, but also for logic. Ferrite cores were coupled by wires strung in specific ways between them so as to achieve logic functionality. For example, the Elliott 803 computer used germanium transistors and ferrite core logic elements. Of course, this kind of magnetic logic technology based on stringing wires between bulky magnetic cores was not competitive against emerging semiconductor technology. However, with the advent of modern fabrication technology, which allows the fabrication of arrays of nanometer-size single-domain magnets, the old quest for magnetic logic might become a reality.

References

1. Calamia, J.: Magnetic logic attracts money. IEEE Spectrum (2011)
2. Markoff, J.: TECHNOLOGY; Intel's Big Shift After Hitting Technical Wall, New York Times, 17 May 2004
3. Bernstein, K., Cavin III, R.K., Porod, W., Seabaugh, A., Welser, J.: Device and architecture outlook for beyond-CMOS switches. Proc. IEEE **98**(12), 2169–2184 (2010)
4. Blakemore, R.: Magnetotactic bacteria. Science **190**(4212), 377–379 (1975)
5. Imre, A.: Ph.D. Dissertation, University of Notre Dame (2005)
6. Lent, C.S., Douglas Tougaw, P., Porod, W., Bernstein, G.H.: Quantum cellular automata. Nanotechnology **4**, 49–57 (1993)
7. Porod, W.: Quantum-dot cellular automata devices and architectures. Int. J. High Speed Electron. Syst. **9**(1), 37–63 (1998)
8. Snider, G.L., Orlov, A.O., Amlani, I., Zuo, X., Bernstein, G.H., Lent, C.S., Merz, J.L., Porod, W.: Quantum-dot cellular automata: review and recent experiments (invited). J. Appl. Phys. **85**(8), 4283–4285 (1999)
9. Csaba, G.: Ph.D. Dissertation, University of Notre Dame (2002)
10. Csaba, G., Porod, W.: Simulation of field-coupled computing architectures based on magnetic dot arrays. J. Comput. Electron. **1**, 87–91 (2002)
11. Csaba, G., Imre, A., Bernstein, G.H., Porod, W., Metlushko, V.: Nanocompting by field-coupled nanomagnets. IEEE Trans. Nanotechnol. **1**(4), 209–213 (2002)
12. Csaba, G., Porod, W., Csurgay, A.I.: A computing architecture composed of field-coupled single-domain nanomagnets clocked by magnetic fields. Int. J. Circ. Theory Appl. **31**, 67–82 (2003)
13. Imre, A., Csaba, G., Ji, L., Orlov, A., Bernstein, G.H., Porod, W.: Majority logic gate for magnetic quantum-dot cellular automata. Science **311**, 205–208 (2006)
14. Imre, A., Csaba, G., Orlov, A., Bernstein, G.H., Porod, W., Metlushko, V.: Investigation of shape-dependent switching of coupled nanomagnets. Superlattices Microstruct. **34**, 513–518 (2003)
15. Varga, E., Niemier, M.T., Bernstein, G.H., Porod, W., Hu, X.S.: Non-volatile and reprogrammable MQCA-based majority gates. Presented at the Device Research Conference (DRC), Penn State University, June 2009
16. Bernstein, G.H., Imre, A., Metlushko, V., Orlov, A., Zhou, L., Ji, L., Csaba, G., Porod, W.: Magnetic QCA systems. Microelectron. J. **36**, 619–624 (2005)
17. Imre, A., Csaba, G., Bernstein, G.H., Porod, W.: Magnetic quantum-dot cellular automata (MQCA). In: Macucci, M. (ed.) Quantum Cellular Automata: Theory, Experimentation and Prospects, pp. 269–276. Imperial College Press, London (2006)
18. Orlov, A., Imre, A., Ji, L., Csaba, G., Porod, W., Bernstein, G.H.: Magnetic quantum-dot cellular automata: recent developments and prospects. J. Nanoelec. Optoelec. **3**(1), 55–68 (2008)
19. Varga, E., Orlov, A., Niemier, M.T., Hu, X.S., Bernstein, G.H., Porod, W.: Experimental demonstration of fanout for nanomagnet logic. IEEE Trans. Nanotechnol. **9**(6), 668–670 (2010)
20. Bedair, S.M., Zavada, J.M., El-Masry, N.: Spintronic memories to revolutionize data storage. IEEE Spectrum (2010)
21. Alam, M.T., Siddiq, M.J., Bernstein, G.H., Niemier, M.T., Porod, W., Hu, X.S.: On-chip clocking for nanomagnet logic devices. IEEE Trans. Nanotechnol. **9**(3), 348–351 (2010)

22. Alam, M.T., Kurtz, S., Siddiq, M.J., Niemier, M.T., Bernstein, G.H., Hu, X.S., Porod, W.: On-chip clocking of nanomagnet logic lines and gates. IEEE Trans. Nanotechnol. **11**(2), 273–286 (2012)
23. Liu, S., Hu, X.S., Nahas, J.J., Niemier, M., Bernstein, G., Porod, W.: Magnetic-electrical interface for nanomagnet logic. IEEE Trans. Nanotechnol. **10**(4), 757–763 (2011)
24. Liu, S., Hu, X.S., Niemier, M.T., Nahas, J.J., Csaba, G., Bernstein, G., Porod, W.: Exploring the Design of the Magnetic-Electrical Interface for Nanomagnet Logic. IEEE Trans. Nanotechnol. **12**(2), 203–214 (2013)
25. Csaba, G., Lugli, P., Csurgay, A., Porod, W.: Simulation of power gain and dissipation in field-coupled nanomagnets. J. Comput. Electron. **4**(1/2), 105–110 (2005)
26. Niemier, M.T., Alam, M.T., Hu, X.S., Bernstein, G.H., Porod, W., Putney, M., DeAngelis, J.: Clocking structures and power analysis for nanomagnet-based logic devices. In: Proceedings of International Symposium on Low Power Electronics and Design, pp. 26–31 (2007)
27. Pietambaram, S., et al.: Appl. Phys. Lett. **90**, 143510 (2007)
28. Li, P., Csaba, G., Sankar, V.K., Ju, X., Hu, X.S., Niemier, M.T., Porod, W., Bernstein, G.H.: Switching behavior of lithographically fabricated nanomagnets for logic applications. J. Appl. Phys. **111**(7), 07B911-1–07B911-3 (2012)
29. Li, P., Csaba, G., Sankar, V.K., Ju, X., Varga, E., Lugli, P., Hu, X.S., Niemier, M., Porod, W., Bernstein, G.H.: Direct measurement of magnetic coupling between nanomagnets for nanomagnetic logic applications. IEEE Trans. Magn. **48**(11), 4402–4405 (2012)
30. Li, P., Sankar, V.K., Csaba, G., Hu, X.S., Niemier, M., Porod, W., Bernstein, G.H.: Magnetic properties of enhanced permeability dielectrics for nanomagnetic logic circuits. IEEE Trans. Mag. **48**(11), 3292–3295 (2012)
31. Imre, A., Csaba, G., Metlushko, V., Bernstein, G.H., Porod, W.: Controlled domain wall motion in micron-scale permalloy square rings. Physica E **19**, 240–245 (2003)
32. Varga, E., Csaba, G., Imre, A., Porod, W.: Simulation of magnetization reversal and domain-wall trapping in submicron permalloy wires with different wire geometries. IEEE Trans. Nanotechnol. **11**(4), 682–686 (2012)
33. Varga, E., Csaba, G., Bernstein, G.H., Porod, W.: Domain-wall assisted switching of single-domain nanomagnets. IEEE Trans. Magn. **48**(11), 3563–3566 (2012)
34. Salahuddin, S., group contribution, this volume
35. Lambson, B., Gu, Z., Monroe, M., Dhuey, S., Scholl, A., Bokor, J.: Concave nanomagnets: investigation of anisotropy properties and applications to nanomagnetic logic. Appl. Phys. A-Mater. Sci. Process. **111**(2), 413–421 (2013)
36. Kurtz, S., Varga, E., Niemier, M., Porod, W., Bernstein, G.H., Hu, X.S.: Non-majority magnetic logic gates: a review of experiments and future prospects for 'shape-based' logic. J. Phys. Condens. Matter **23**(5), 053202 (2011)
37. Niemier, M.T., Varga, E., Bernstein, G.H., Porod, W., Alam, M.T., Dingler, A., Orlov, A., Hu, X.S.: Shape-engineering for controlled switching with nanomagnet logic. IEEE Trans. Nanotechnol. **11**(2), 220–230 (2012)
38. Becherer, M., Csaba, G., Porod, W., Emling, R., Lugli, P., Schmitt-Landsiedel, D.: Magnetic ordering of focused-ion-beam structured cobalt-platinum dots for field-coupled computing. IEEE Trans. Nanotechnol. **7**(3), 316–320 (2008)
39. Breitkreutz, et al.: IEEE Trans. Mag. **48**(11) (2012)
40. Csaba, G., Kiermaier, J., Becherer, M., Breitkreutz, S., Ju, X., Lugli, P., Schmitt-Landsiedel, D., Porod, W.: Clocking magnetic field-coupled devices by domain walls. J. Appl. Phys. **111**(7), 07E337-1–07E337-3 (2012)
41. AtyabImtaar et al., to be presented at IEEE-NANO 2013

42. Niemier, M.T., Bernstein, G.H., Csaba, G., Dingler, A., Hu, X.S., Kurtz, S., Liu, S., Nahas, J., Porod, W., Siddiq, M., Varga, E.: Nanomagnet logic: progress toward system-level integration. J. Phys. Condens. Matter **23**(49), 493202 (2011)
43. Ju, X., Niemier, M., Becherer, M., Porod, W., Lugli, P., Csaba, G.: Systolic pattern matching hardware with out-of-plane nanomagnet logic devices. IEEE Trans. Nanotechnol. **12**(3), 399–407 (2013)
44. Khitun, A., Nikonov, D.E., Wang, K.L.: Magnetoelectric spin wave amplifier for spin wave logic circuits. J. Appl. Phys. **106**(12), 123909 (2009)
45. Behin-Aein, B., Datta, D., Salahuddin, S., Datta, S.: Proposal for an all-spin logic device with built-in memory. Nat. Nanotechnol. **5**, 266–270 (2010)
46. Gschwind, H.W.: Design of Digital Computers © 1967 by Springer

Silicon Atomic Quantum Dots Enable Beyond-CMOS Electronics

Robert A. Wolkow[1,2,3(✉)], Lucian Livadaru[3], Jason Pitters[2], Marco Taucer[3], Paul Piva[3], Mark Salomons[2], Martin Cloutier[2], and Bruno V.C. Martins[1]

[1] Department of Physics, University of Alberta, 11322-89 Avenue, Edmonton, AB T6G 2G7, Canada
rwolkow@ualberta.ca
[2] National Institute for Nanotechnology, National Research Council of Canada, 11421 Saskatchewan Drive, Edmonton, AB T6G 2M9, Canada
[3] Quantum Silicon Inc., 11421 Saskatchewan Drive, Edmonton, AB T6G 2M9, Canada

Abstract. We review our recent efforts in building atom-scale quantum-dot cellular automata circuits on a silicon surface. Our building block consists of silicon dangling bond on a H-Si(001) surface, which has been shown to act as a quantum dot. First the fabrication, experimental imaging, and charging character of the dangling bond are discussed. We then show how precise assemblies of such dots can be created to form artificial molecules. Such complex structures can be used as systems with custom optical properties, circuit elements for quantum-dot cellular automata, and quantum computing. Considerations on macro-to-atom connections are discussed.

Keywords: Silicon dangling bonds · Quantum-dot cellular automata

1 Preliminaries

There are two broad problems facing any prospective nano-scale electronic device building block. It must have an attractive property such as to switch, store or conduct information, but also, there must be an established architecture in which the new entity can be deployed and wherein it will function in concert with other elements. Nanoscale electronic device research has in few instances so far led to functional blocks that are ready for insertion into existing device designs. In this work we discuss a range of atom-based device concepts which, while requiring further development before commercial products can emerge, have the great advantage that an overall architecture is well established that calls for exactly the type of building block we have developed.

The atomic silicon quantum dot (ASiQD) described here fits within ultra low power schemes for beyond CMOS electronics based upon quantum dots that have been refined over the past 2 decades. The well known quantum dot cellular automata (QCA) scheme due to Lent and co-workers [1,2] achieves classical

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 33–58, 2014.
DOI: 10.1007/978-3-662-43722-3_3, © Springer-Verlag Berlin Heidelberg 2014

binary logic functions without the use of conventional current-based technology. Within this scheme, the binary states "1" and "0" are encoded in the position of electric charge. Variants exist but most commonly the basic cell consist of a square (or rectangular) quantum dot ensemble occupied by 2 electrons. Electrons freely tunnel among the quantum dots in a cell, while electron tunneling between cells does not occur. Within a cell, two classically equivalent states exist, each with electrons placed on the diagonal of the cell. Multiple cells couple and naturally mimic the electron configuration of nearest-neighbour cells. In general, cell-cell interactions must be described quantum mechanically but to a good approximation they are described simply by electrostatic interactions. A line of coupled cells serves as a binary wire. When a terminal cell is forced by a nearby electrode to be in one of its two polarized states, adjacent cells copy that configuration to transfer that input state to the other terminus. This transfer can happen spontaneously or can be zonally regimented by a clock signal that controls inter-dot barriers, or some other parameter. The last key feature of QCA is that three binary lines acting as computation inputs and one line acting as output can converge on a node cell to create a majority gate. If two of the three input lines are of one binary state, the fourth side of the node cell will output the majority state. Variants of such an arrangement allow for the realization of a full logical basis. To date, all manner of digital circuits have been designed, from memories to multipliers to even a microprocessor.

While complex working circuits have not yet been realized, all the rudimentary circuit elements have been already experimentally demonstrated [3,4]. Furthermore, the input state of a QCA circuit has been externally controlled and the output has been successfully read-out by a coupled single electron transistor [5,6]. Until the present work, all available quantum dots, typically consisting of thousands of atoms, had narrowly spaced energy levels requiring ultra-low temperature to exhibit desired electronic properties. Moreover, approximately as many wires as quantum dots were required to adjust electron filling, a scenario that would greatly limit the complexity of circuitry that could be explored.

A prospect for highly complex and room temperature operational QCA circuitry suddenly emerged with the discovery of atomic silicon quantum dots. Figure 1 shows a schematic 4-dot QCA cell on the left occupied with two electrons (indicated by blackened circles). On the right is an STM image of a real atom-scale cell made of 4 ASiQDs, the cell being less than 2 nm on a side. The darker of the two dots are predominantly electron occupied.

The ultimate small size of the ASiQDs leads to ultimate wide spacing of energy levels indeed sufficiently widely spaced to allow room temperature device operation. The ASiQDs can be prepared in a native 1− charge state (charge is expressed in elementary charge units henceforth). Close placement of dots causes Coulombic repulsion and even removal of an electron to the silicon substrate conduction band. By fabricating dots at an appropriate spacing, a desired level of electron occupation can be predetermined, eliminating the need for many wires. As all atomic dots are identical, and their placement occurs in exact registry with the regular atomic structure of the underlying crystalline lattice,

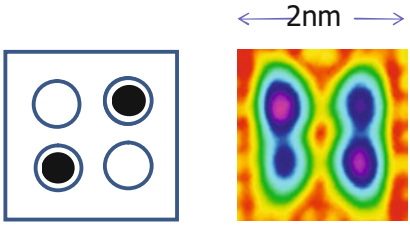


Fig. 1. Left: schematic representation of a square QCA cell with 2 electrons (blackened circles) positioned on the diagonal configuration. Right: STM map of an actual ASiQD structure with 4 dots in a square pattern as an embodiment of the QCA cell on the left. Electron population is predominant on the same diagonal as indicated on the schematic on the left.

structures with uniquely homogeneous and reproducible characteristics can be in principle fabricated. A further advantage lies in the fact that these dots are entirely made of and upon silicon, enabling compatibility with silicon CMOS circuitry. This allows the merging of established and new technologies, greatly easing the path to deployment.

Challenges in the precise positioning of single silicon atom dots previously limited creation of more than a 4-atom ensemble. New developments have enabled patterns with hundreds of atoms to be fabricated with error rates close to those required for functioning computation circuit elements. A path to further improvements appears to be in hand.

Information storage, transfer and computation without use of conventional electrical current, with several orders lower power consumption than CMOS appear within reach. Prospects for extremely small size and weight appear good, too, as are those for extreme speed. Existing true 2-dimensional circuit layouts indicate a great reduction in the need for multilayer interconnects. The all-silicon aspect of this approach leads to a natural CMOS compatibility and therefore an early entry point via a hybrid CMOS-ASiQD technology. Room temperature as opposed to cryo operation is very attractive. The materials stability of the system up to >200 °C is comparable to conventional electronics. Furthermore, the possibility of deployment in an analog mode broadens the appeal and power of the approach. A natural ability to merge with Si-based sensor circuitry is desirable too. As discussed below, potential applications in quantum information are also very appealing.

2 Preparing and Visualizing Silicon Surface Dangling Bonds

A silicon dangling bond, DB, exists at a silicon atom that is under-coordinated, that is where a silicon atom has only 3, rather the regular 4 bonding partners. In this discussion we will focus on the DB on the hydrogen terminated (100) face of

a silicon crystal, abbreviated H-Si(100). Atomically flat, ordered H-termination is ordinarily achieved by cracking Hydrogen gas, H_2, into H atoms by collision with a hot tungsten filament and allowing those H atoms to react with a clean silicon surface in a vacuum chamber. If the H-termination process is incomplete, or if an H atom is removed by some chemical or physical means, a DB is created. H atoms removed by the local action of a scanning tunneling microscopy (STM) are the focus here. Broadly speaking, the scanning motion of the tip can be halted to direct an intense electrical current in the vicinity of a single Si-H surface bond [7–10]. At approximately a 2 V bias between tip and sample it is understood that multiple vibrational excitations lead to dissociation of the Si-H bond. At near 5 V bias it is thought the Si-H bond can be excited to a dissociative state, as in a photochemical bond breaking event. Other not well understood factors are at play, such as a catalytic effect, intimately depending on particular tip apex structure and composition that might ease the Si-H bond apart as a substantial H atom-tip bond forms while the Si-H bond lengthens and weakens. The fate of removed H atoms is unclear though there is substantial evidence, in the form of H atom donation to the surface that some atoms reside on the STM tip [11,12]. Many details related to exact position of the tip and precise metering of the energetic bond breaking process so as to create just the change desired and not other surface alterations will be touched upon in the section on Quantum Silicon Incorporated and the commercial drive to fabricate atom scale silicon devices.

Figure 2a shows a model of a H-Si(100) surface. Silicon atoms are yellow. Hydrogen atoms are white. Note the surface silicon atoms are combined with H atoms in a 1 to 1 ratio. Note also that the surface silicon atoms deviate from the bulk structure not only in that they have H atom partners, but also in that each surface silicon atom is paired-up into a dimer unit. The dimers exist in rows.

Figure 2b shows a constant current STM image of a H-Si(100) surface. The dimer units are 3.84 Å separated along a dimer row. The rows are separated by twice that distance, 7.68 Å. The overlaid grid of black bars marks the position of the silicon surface dimer bonds. To reiterate, there is an H atom positioned at both ends of each dimer unit.

Figure 3 indicates the localized creation of a dangling bond upon action directed by a scanned probe tip. Figure 3b shows an STM image of several DBs so created [13].

3 The Nature of Silicon Dangling Bonds

Figure 4 shows two silicon surfaces imaged under the same conditions [14]. Both surfaces have a scattering of DBs. The left image is of a moderately n-type doped sample. It has been shown that DBs on such a surface are on average neutral. The DBs in that case are visible as white protrusions. The right hand image is of a relatively highly n-type doped sample. In that case each DB has a dark "halo" surrounding it. These DBs are negative. This results as the high concentration of electrons in the conduction band naturally "fall into" relatively low-lying DB surface state to make it fully, that is 2 electron, occupied. This localization of a

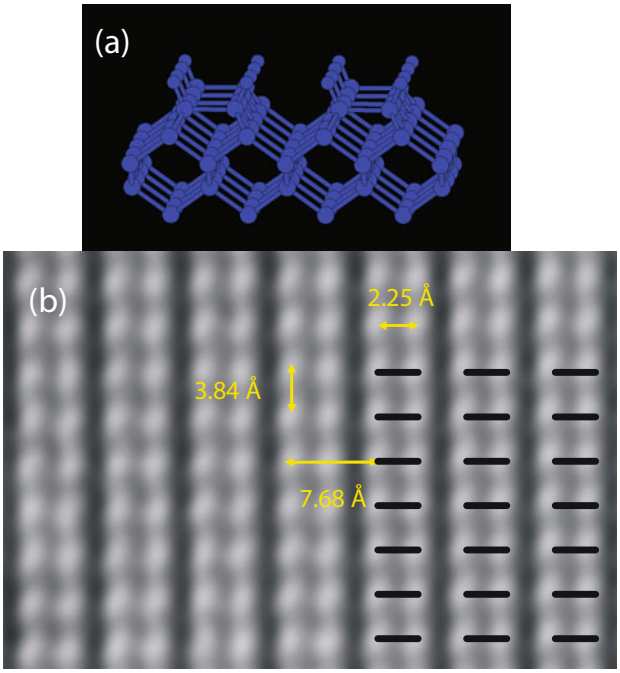


Fig. 2. (a) Model of a H-Si(100) surface with silicon atoms in yellow and hydrogen atoms in white. (b) Constant-current STM image of a H-Si(100) surface. Dimer rows are visible in the vertical direction and the atom separation along and across dimer rows are marked. Some dimer bonds are also marked by an overlaid grid of black bars (Color figure online).

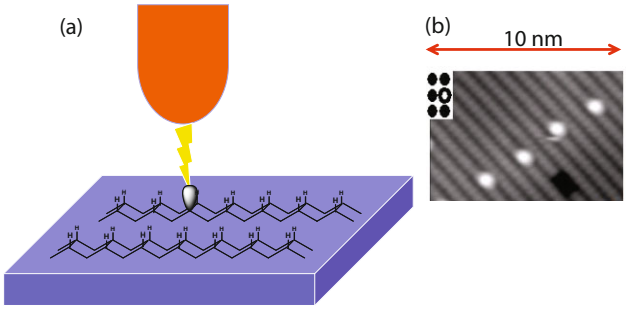


Fig. 3. (a) Schematic of the fabrication of a dangling bond on the H-Si(100) surface by a scanned probe tip at a chosen location. (b) STM image of several DBs created in a line.

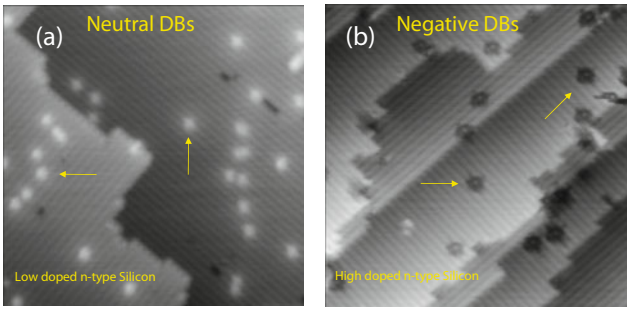


Fig. 4. Two silicon surfaces imaged under the same conditions. (a) A moderately n-type doped sample where DBs are on average neutral. (b) A highly doped sample where DBs are on average negatively charged.

negative charge at a DB causes destabilization of electron energy levels referred to as upward band bending. To a first approximation it is the inaccessibility of empty states for the STM tip to tunnel into that causes the highly local darkening of the STM image, that is, the halo. A much fuller description of the competing process involved in the imaging process have been described [15].

It is evident that the DB is effectively a dopant with a deep acceptor level. In accord with that character, a DB acts to compensate bulk n-type doping, causing the bands to shift up with respect to the Fermi level in the direction of a p-doped material. Most recently it has been shown that the neutral, single electron occupied DB can donate its charge to become positive thereby acting as a deep n-type dopant [16]. Summarizing, single electron occupation corresponds to neutral state. Two electron occupation corresponds to 1− charge. The absence of electrons in the DB leaves it in a 1+ charge state. The combination of dopant type, concentration, DB concentration on the surface, local electric field, and finally current directed through a DB, all contribute to determining its instantaneous charge state [15].

A clean silicon surface, where every site has a dangling bond, is very reactive toward water, oxygen and unsaturated hydrocarbons like ethylene and benzene. While H atoms immediately react with a clean silicon surface, H_2 does not [17]. It is a remarkable fact that single DBs interact only weakly with most molecules, resulting in no attachment at room temperature. Most often, a second immediately adjacent DB is required in order for a molecule to become firmly bonded to the surface. Two DBs typically act together to form two strong bonds to an incoming molecule. This has the practical consequence that a protective layer can be formulated and applied to encapsulate and stabilizes DBs against environmental degradation.

A special class of molecules, typified by styrene, C_8H_8 attach to silicon via a self-directed, chain reaction growth mechanism [18]. As shown in Fig. 5a, a terminal C reacts with a DB, thereby creating an unpaired electron at the adjacent C on the molecule. That species follows one of two paths. It either desorbs, or the

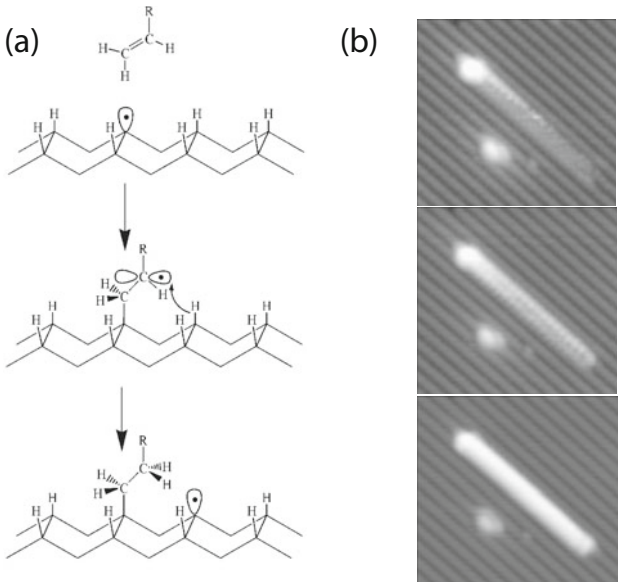


Fig. 5. (a) Various processes in a chain reaction resulting in the growth of lines a special class of molecules (e.g. styrene) on a H-Si(100) surface. R here denotes various possible radicals. (b) A line of about 20 styrene molecules grown in this fashion and imaged at different STM setpoints. The charge on the terminal DB at the top of the line depends on the STM setpoint and therefore causes (or not) a Stark shift in the molecular line.

radical C abstracts an H atom from an adjacent surface site to create a stably attached molecule, and a regenerated DB positioned one lattice step removed from the original DB position. The process repeats and repeats to create a multi-molecular line that gains a degree of order from the crystalline substrate.

Figure 5b shows an approximately 20 molecule long line of styrene grown in this way. The bright feature at the end of the line is a DB. It has been shown that under conditions where the terminal DB is negatively charged that charge acts to gate (Stark shift) the molecular energy levels causing conduction through the molecule where ordinarily it would not occur [19]. In other words the ensemble forms single-electron gated, one-molecule field effect transistor. The point of this discussion is to show there is precedent for microscopic observation of DBs at different charge-states, and to point out that a structure like the molecule transistor arrangement could be a useful detector of DB charge state [20].

4 Dangling Bonds Are Atomic Silicon Quantum Dots

The atoms in a silicon crystal enter into bonding and anti-bonding relationships with neighbouring and distant silicon atoms to form bands that span the crystal. In doing so the atoms give up their zero dimensional electronic character. Si

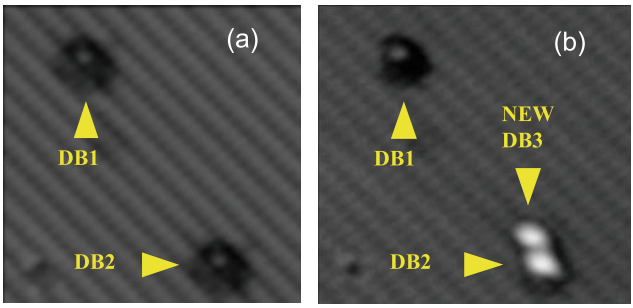


Fig. 6. (a) STM micrograph (10×10 nm, 2 V, 0.2 nA) of a H-Si(100) surface with two DBs. The distinct dark halo indicates each DB is negatively charged. (b) An additional DB is created at a site near DB2 causing both the new DB3 and DB2 to appear very differently.

atoms sharing in three ordinary Si-Si bonds and containing one dangling bond have a special mixed character. Like 4-coordinate silicon atoms, such atoms are very strongly bonded to the lattice and have an intimate role in the dispersive bands that delocalize electrons. At the same time, 3-coordinate atoms have one localized state, approximately of sp^3 character. This state is localized because it is in the middle of the band gap and mixes poorly with the valence and conduction band continua. The DB-containing atom is odder still in that the DB is partly directed toward the vacuum where it has a relatively limited spatial extent but is also partially contained within the silicon crystal where, because of dielectric immersion, is somewhat larger in spatial reach.

It was stated above that a DB state is like a deep dopant. Whereas a typical dopant has an ionization or affinity energy of several tens of meV, the DB has corresponding energies an order of magnitude larger. Consistent with that difference, the spatial extend of the DB state within the solid reaches several bond lengths, much less than the size of a common dopant atom [21].

The zero dimensional character of the DB, combined with the capacity to exhibit several (specifically 3) charge states leads us to think of the DB as a quantum dot. This may at first seem a bit odd as a quantum dot is often described as an artificial atom whereas we have a genuine atom, actually one part of an atom, forming our dot. But if a quantum dot is most fundamentally a vessel for containing and configuring electrons then, as subsequent examples will show, the ASiQD naturally and ably fits the definition, especially as the ease and precision of fabrication allows complex interactive ensembles of identical quantum dots to be made.

5 Fabricating and Controlling a Quantum Dot Cellular Automata Cell

Figure 6a shows two DBs. The distinct dark halo indicates the DBs are negatively charged. Figure 4b shows that when an additional DB is created by a tip directed H removal at a nearby site, both the new DB and the nearby pre-existing DB appear very differently, while the somewhat removed DB is unaltered. After extensive study it became clear that such a closely placed pair of DBs experiences a great Coulombic repulsive interaction, destabilizing the bound electrons and enabling one electron to leave the ensemble [12,14]. The reduced net charge simultaneously stabilizes the remaining bound electron and creates an unoccupied energy level on one of the atoms. Because the barrier separating the DBs is low, of order several 100 meV, and is also very narrow, of order 2 nm, tunneling to the vacant state is very facile. Such a pair of DBs may be referred to as tunnel coupled. Our WKB and ab initio calculations agree that the tunneling rate for the 3.84 Å separated DBs corresponds to an extremely short tunneling period of order 10 fs [22,23]. Conventional relatively large and necessarily widely spaced dots would have a tunnel rate many orders of magnitude lower. Figure 7 shows the energy landscape schematically [14]. Each DB is represented by a potential well. The well is within the silicon bandgap. In Fig. 7a the separation between DBs is sufficiently large for the Coulombic interaction to be diminished by distance and by screening by conduction band electrons. In Fig. 7b the high energy repulsive relationship existing between two negatively charged DBs is represented. Figure 7b also shows the relaxed situation resulting after removal of one electron to the conduction band. In that final scenario one vacant electron state is shown. That state and the low and narrow barrier enables tunneling between the DBs.

The pairing result demonstrates a "self-biasing" effect. That is, by using fabrication geometry and repulsion to adjust electron filling, the need for capacitively coupled filling electrodes is removed [14]. Figure 8a shows several pairs of DBs of different separations and therefore different average net occupations. It can be readily seen that closer spaced DBs more fully reject one electron, leading to less local charge induced band bending and therefore to a lighter appearance in the STM image. The increasingly widely spaced pairs look increasingly dark as the net charge approaches 2 electrons. A statistical mechanical model of the paired DBs reproduces the effect as shown in Fig. 8b. The graph stresses that occupation is a time averaged quantity and that pairs in the cross-over region will at any instant be either 1− or 2− charged [14].

Figure 9 shows a 4 dot ensemble or artificial molecule. The 4 dot cell was fabricated to result in an average net filling of 2 extra electrons. The graphs in Fig. 9b show the result of a statistical mechanical description of average occupation versus distance of separation in such a square cell at different temperatures [12].

One way to localize and thereby visualize the occupying electrons is to make an irregular shaped cell as is shown in Fig. 10 [14]. Figure 10a shows three dots, two of which look darker indicating greater negative charge localization Upon

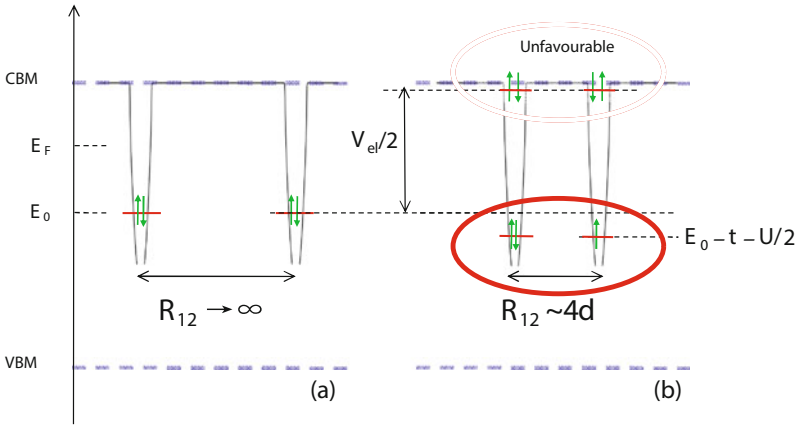


Fig. 7. The schematic energy landscape of a DB pair with each DB represented by a potential well with the ground state in the band gap. In (a), the separation between DBs is very large for the Coulombic interaction to be negligible and each DB is negatively changed (doubly occupied). In (b), DBs are much closer together (d is the dimer-dimer spacing) and a great Coulombic repulsion is associated with the doubly occupied configuration on both DBs. The diagram also shows the relaxed situation resulting after removal of one electron to the conduction band thus enabling tunneling of the remaining excess electron between the DBs.

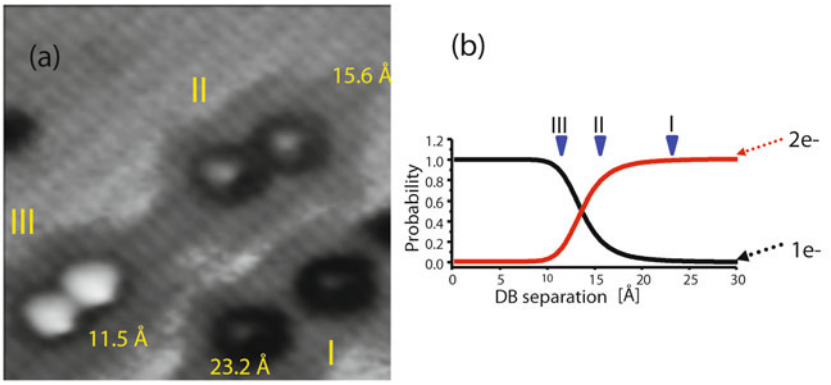


Fig. 8. (a) Several tunnel-coupled pairs of DBs fabricated at different separations (specified in each case) on the H-Si(100) surface. (b) Average occupation probability of a DB pair with 1 and 2 excess electrons as a function of DB separation. The three cases labeled in (a) are marked here with blue arrows (Color figure online).

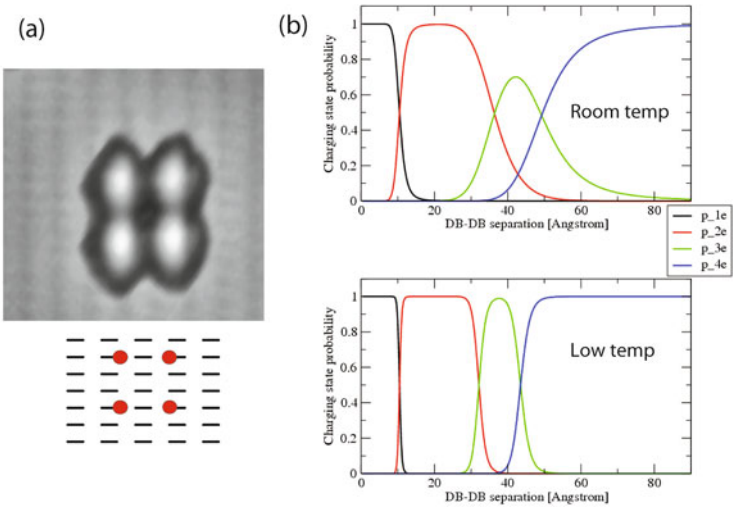


Fig. 9. (a) A fabricated ensemble (cell) of 4 tunnel-coupled DBs, or artificial molecule, calibrated to result in an average net filling of 2 extra electrons. A corresponding dimer lattice diagram is shown below. (b) The result of a statistical mechanical description of average occupation versus distance of separation in such a square cell at different temperatures (300 K top graph, 100 K bottom graph). The occupation probabilities with 1, 2, 3, 4 extra electrons are plotted for each case.

adding a fourth dot the previously darker sites become relatively light in appearance. This is due to the electrons attaining a lower energy configuration along a newly available longer diagonal. In a symmetric square or rectangular cell the freely tunneling electrons equally occupy the degenerate diagonal configurations. On the slow time scale of the STM measurement no instantaneous asymmetry can be seen.

In order to embody the QCA architecture it must be possible to break that symmetry electrostatically and thereby to polarize electrons within a cell. This capacity is illustrated first by referral to a 2 dot cell. Figure 11 shows the sequential building of a 2 dot cell occupied by one extra electron and the polarization of that cell by one perturbing charge [14]. Figure 11a shows a small area, 3 nm across, of H-terminated silicon at room temperature. Figure 11b shows the creation of one ASiQD, while Fig. 11c shows the creation of a second ASiQD and the concomitant reduction in charge and darkness as seen by the STM. Upon charge removal, rapid tunnel exchange ensues. The coupled entity resulting may be described as an artificial homonuclear diatomic molecule. Like in an ordinary molecule, the Born-Oppenheimer approximation is valid. In other words, the electron resides so very briefly on one atom that nuclear relaxation does not have time to occur. On the electronic time scale, the nuclei are frozen. Finally in Fig. 11d another charged DB is created. Using the knowledge displayed in Fig. 6, the last DB is placed near enough to the molecule to affect it electrostatically,

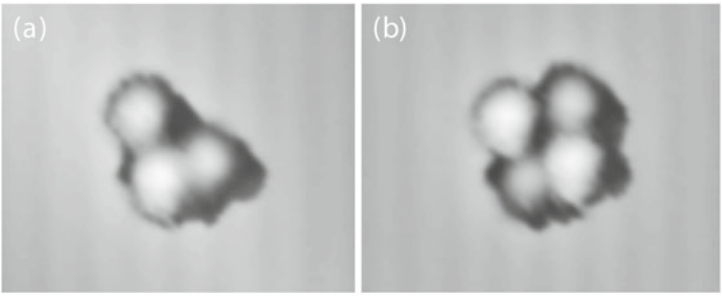


Fig. 10. (a) Three dots, two of which look darker indicating greater negative charge localization. (b) Upon adding a fourth dot in the upper right region the previously sites which previously looked dark become relatively light in appearance. This is due to the electrons attaining a lower energy configuration along the newly available longer diagonal.

but not so close as to be tunnel coupled and a direct participant in the molecule. It is clear that the perturbation creates an analog of a heteronuclear diatomic molecule. The bond is polarized, it has ionic character. The perturbation controllably positions the electron. This is a key result. It shows that the type of coupling required between QCA cells, and between an electrode and a QCA cell is possible by this new approach. And all of this is possible at room temperature with an all silicon system.

Figure 12 shows the result of placing two perturbing electrons along one diagonal to place a 4 dot, 2 electron cell into one polarized binary state [14]. It is stressed again that this result was obtained at room temperature.

One could wonder about ways to increase the chemical stability of such complex QCA structures against environmental damage. As already discussed, for certain species such as styrene, there are known stable attachment mechanisms that require only a single DB. Such reactions can lead to molecular line growth on the surface. However the class of molecules that undergo such a process is rare. A passivating layer can be formed of molecules containing functional groups known not to react with a single DB. A yet broader class of molecules chemisorbs when two closely spaced DBs are provided. However, in order to cooperatively react with an incoming molecule, two DBs must be immediately adjacent, that is, co-located on the same underlying silicon dimer unit. Such DBs are separated by approximately 2.3 Å. After reaction initiates at one of the two DBs, the second DB is just within reach by a C atom centered radical, allowing a second Si-C bond to form and resulting in a stable species and DB annihilation. When a second DB is not within reach, the single bonded species very quickly releases its grip because the interaction strength is only of order 0.1 eV. The DB in that circumstance is left unchanged. There is evidence of reactivity involving paired DBs separated by 3.84 Å. The particular molecular functions able to form such a bond are readily avoided. In any case, none of the QCA patterns to be

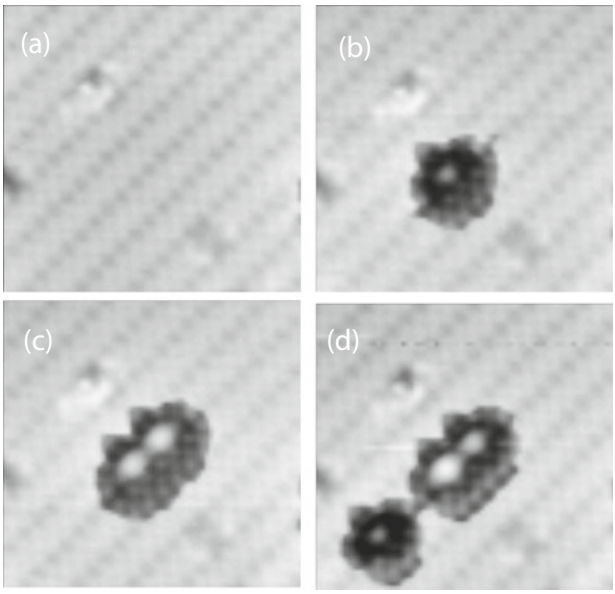


Fig. 11. The sequential building of a 2-dot cell occupied by one extra electron and the polarization of that cell by one perturbing charge. (a) A room temperature STM image approximately 3 nm across of H-terminated silicon surface. (b) The creation of one ASiQD. (c) The creation of a second ASiQD and the concomitant reduction in charge and darkness as seen by the STM. This tunnel coupled entity may be described as an artificial homonuclear diatomic molecule. (d) Another charged DB is created. This last dot is placed near enough to the molecule to affect it electrostatically, but not so close as to be tunnel coupled. It is seen that the perturbation creates an analog of a heteronuclear diatomic molecule as the bond is polarized.

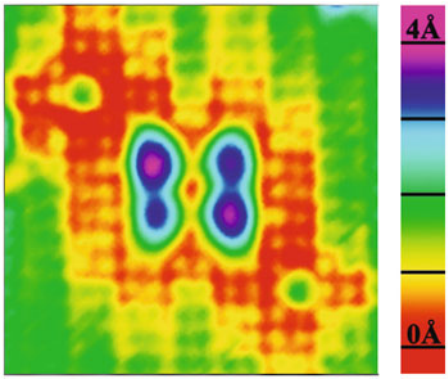


Fig. 12. Two perturbing electrons are positioned along one diagonal to place a 4-dot, 2-electron cell into one polarized binary state at room temperature.

fabricated will include either of those reactive DB pairings. All of the DB cell structures will involve DBs spaced by approximately 8 Å, or more.

6 The ASiQD as a New Element: Building Molecules

Our primary focus is on ultra low power, ultra fast field controlled computing as embodied by the QCA architecture or a variant thereof. We are well aware though of multiple other possibilities created by the ASiQDs. It is like having a new element to build with.

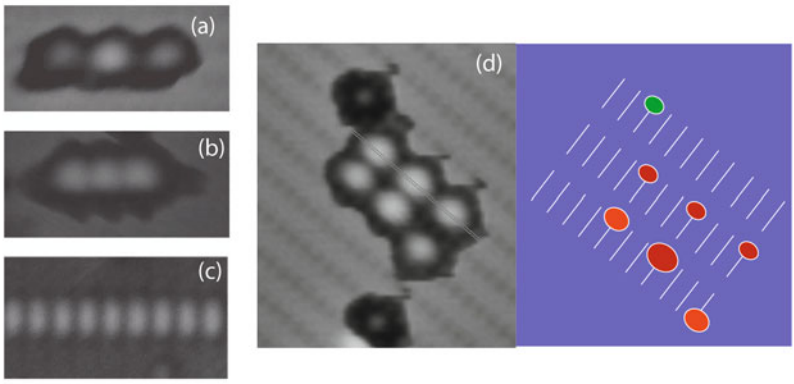


Fig. 13. A collection of diverse assemblies or molecules composed of atomic silicon quantum dots. The three-atom structures in (a) and (b) have slightly differently spacing. Different electronic structure results. The widely spaced grouping has a distinct, less negative (brighter appearing) central atom while within the tightly spaced assembly all atoms appear the same. The wider spaced molecule attains a larger negative charge compared to the tightly spaced molecule. The latter has one electron. The widely spaced molecule has two electrons and those repel one another, leaving the central atom approximately neutral on average, and brighter in appearance. (c) Is an image of a linear chain of ASiQDs. Such a structure will delocalize charge and allow biasing wires to be fabricated where needed. The complex arrangement labeled (d) uses both tunnel coupled atoms and perturbing atoms on the periphery to locally change the character of individual atoms and thereby the properties of whole ensemble. The perturbing atoms approximate the varied effects of different functional groups as used to tune organic molecular properties.

Figure 13 shows a collection of diverse assemblies or molecules [12]. The three-atom structures in (a) and (b) have slightly differently spacing. It is clear that different electronic structure results. The widely spaced grouping has distinct, less negative (brighter appearing) central atom while within the tightly spaced assembly all atoms appear the same. An in depth discussion is outside of our present scope. We will only note that the key difference between these structures

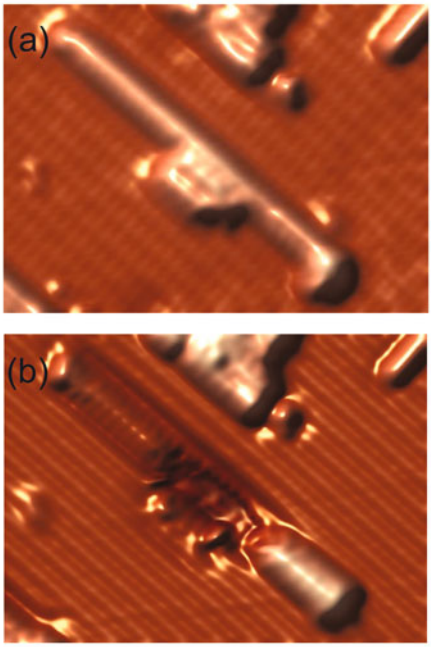


Fig. 14. STM images of single-triple CF_3-styrene/OCH_3-styrene heterostructure. (a) $V_s = 2\,V$. (b) $V_s = -2\,V$. Single OCH_3-styrene and CF_3-styrene lines image above H-Si surface, while triple CF_3-styrene chains image below H-Si surface.

is occupation. The wider spaced molecule tolerates a larger negative charge compared to the tightly spaced molecule. The latter has one electron. The widely spaced molecule has two electrons and those repel one another, leaving the central atom approximately neutral on average, and brighter in appearance. The complex arrangement labeled (d) uses both tunnel coupled atoms and perturbing atoms on the periphery to locally change the character of individual atoms and thereby the properties of whole ensemble.

Figure 13c shows a linear chain of ASiQDs. Such a structure will delocalize charge and allow biasing wires to be fabricated where needed. Such wires will bridge between relatively large lithographically created structures and the atom scale, allowing intimate input. It is a compelling feature of the ASiQD approach to circuit fabrication that high density ensembles enable passive components such as wires to be made while somewhat more widely spaced structures allow for the creation of the active elements in a circuit.

Molecules of designed optical properties can be made. This has been touched upon in a recent charge qubit characterization discussion [24]. More complex control is however possible. Not only the absorption energy, but the mode of adsorption can be pre-defined, whether electric dipolar, magnetic dipolar, or complex multipolar. The particular polarization dependence of absorption can

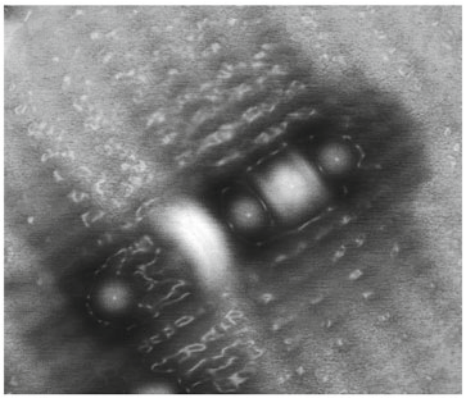


Fig. 15. A six atom molecule provides an example of the rich electronic structure within reach when molecules are designed and built in place where and as needed. The atoms are as close spaced as the lattice allows except between the 2nd and 3rd atoms counting from the left. Despite the uniformity of the constituents, the particular configuration results in a richly varied electronic property across the ensemble. Because the electronic structural maxima are not simply conformal to the positions of the constituent atoms it is not obvious without foreknowledge how many atoms are involved.

be narrowly constrained too. Precise arrays of absorbers or emitters can be fabricated to achieve collective enhancements. Regular molecules and nano-clusters can also be used alone or in combination with ASiQD molecules to achieve designed function.

In addition to the perturbing effect of a nearby charge centre, attached molecules can be placed nearby to source a substantial and highly local multipolar field having the effect of making nearby ASiQDs more or less electron rich, rather like the approach of adding functional groups in chemistry. The pronounced effect of polarizing methoxy groups was experimentally and theoretically described in [25,26] and is shown in Fig. 14.

Finally we show as Fig. 15 an example of the kind of rich electronic structure within reach. In this case 6 ASiQDs were prepared in a line. The atoms are as close spaced as the lattice allows except between the 2nd and 3rd atoms counting from the left. Because the electronic structural maxima are not simply conformal to the positions of the constituent atoms it is not obvious without foreknowledge how many atoms are involved. A catalogue of such molecules is being prepared in concert with modeling to extend known structure-property relationships.

For over two decades molecular electronics research has sought to transfer molecules from a bottle to a surface to achieve desirable circuitry. In practice, studied configurations between molecules and electrodes are uncertain and variable, as are the properties of the ensemble. While progress toward self-assembling molecular structures has been made, for the most part the linking forces directing assembly of molecules have been focused on geometric arrangements without consideration or control of electrical connectivity. In some instances when single

molecules have been rendered to conduct electrical current, molecular decomposition has promptly occurred. We are considering initiating what could be considered as "phase 2" of molecular electronics. By building artificial molecules where and as needed, and by precisely building contacts also, it is possible to ensure interactions and properties are exactly as designed. Working only with the limited palette described here, it is evident that artificial molecules of diverse properties can be designed and made in place. Networks of high complexity and function can be formed. At the cost of greater fabrication complexity, for example through the addition of atoms other than H and Si, a far broader range of properties will be available. It is anticipated that efficient substrate coupling will diminish decomposition pathways and that artificial molecules as envisioned here will not fail as organics will as a result of undesired excitation and heating.

7 Instrumentation and Custom Lithography to Make Prototype QCA Circuitry

State of the art instrumentation is required to make advances in this area. Years of ordinary STM investigations were hampered by at least two problems. One is non-ideal scanning and fabrication control something we will refer again to in the next section. The other aspect is a lack of a bridge between the atom sale and the macro scale.

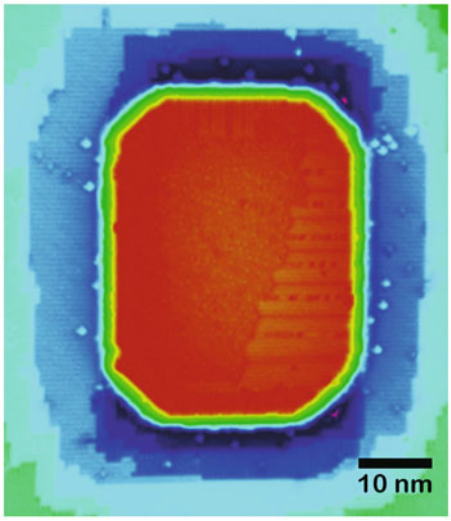


Fig. 16. A crystalline Titanium Silicide island on a silicon substrate. Crucially, the silicide-silicon boundary is atomically abrupt. The silicon nearby is well ordered as required. Ultra fine extensions from such contacts with linear wire structures composed of closely spaced ASiQDs will enable precise electrostatic addressing of atomic structures.

Our first approach to making sufficiently fine lithographic features to controllably interact with atom scale structures began nearly 20 years ago. Titanium silicide contacts were prepared using a normal optical lithography and lift-off approach [27,28]. When examined at the atomic scale, lithographic features prepared in this way are unacceptably rough and crudely defined for our purposes. The transition from pure silicide to pure silicon is not abrupt, but spans 10s of nm, and is of unknown composition and of uncontrolled electronic character. However, it is possible to grow crystalline Titanium Silicide features on a silicon substrate with atomically-precise boundaries by simply evaporating titanium and annealing in ultra high vacuum. Figure 16 shows a highly crystalline Titanium Silicide island on a silicon substrate. The boundary is atomically abrupt. The silicon nearby is well ordered as required. No spatial patterning is imposed. Refinements of this technique will enable the type and quality of contacts required. Ultra fine extensions from such contacts with linear wire structures composed of closely spaced ASiQDs will enable precise electrostatic addressing of atomic structures.

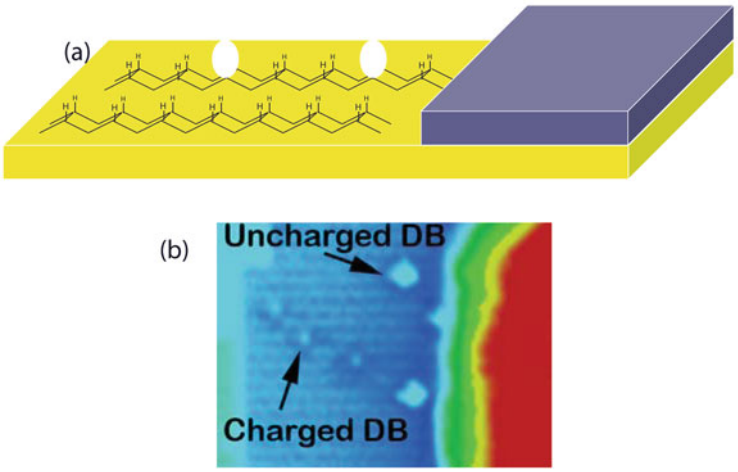


Fig. 17. (a) A schematic diagram of a silicide contact and nearby DBs. The spatially varying potential built-in at the silicide-silicon interface causes DBs nearby to take position dependent charge states. (b) An experimental verification of the scheme.

The potential to controllably alter the charge state of individual DBs has been demonstrated [29]. Figure 17 shows a schematic diagram of a silicide contact and nearby DBs. Because of the spatially varying potential built-in at the silicide-silicon interface (and because the H-terminated surface is not pinned) the DBs nearby take position dependent charge states. Figure 17b shows an experimental verification of the scheme. Going forward, combined lithographic approaches will allow for suitably small and high quality lithographic features. Multiple

contacts will be connected and active while a device is in the STM fabrication and inspection tool allowing prototyping methods and device testing to advance substantially over what has been available to date.

In parallel with the development of nano-lithographic methods, a multi-probe STM has been developed to allow nano-scale electrical characterization that has until now been out of reach. The instrument shown in Fig. 18 has three independently scannable tips, watched over by a scanning electron microscope. Each tip can be quickly redeployed as a scanned probe for imaging or touched down as a current source or as a voltage probe. Initial applications have enabled the first detection of a potential step on a crystalline silicon surface [30], the absolute measurement of the Si(111)-7 × 7 surface conductivity [31], the absolute value of the Si(111)-7 × 7 step resistivity [31] and the conductivity of a DB wire on the H-Si(100) surface [32]. Various measurements related to and enabling of atomic silicon surface electronic circuitry will follow.

8 Quantum Silicon Incorporated and First Proto-Circuits

A spin-off company has been created to exploit the recently discovered DB/ASiQD properties reviewed here. The goal is to undertake focused development of all the many components needed to demonstrate working atom scale field controlled computing elements. In so doing, fundamental components and methods will be developed that also enable analog and quantum circuitry.

It has been 2 decades since Lyding, Tucker and associates first made DB wires [33]. Though many subsequent studies explored other aspects of that system limited progress has been made in fabricating atomically precise patterns. Making precise DB patterns is difficult with standard equipment. The first ASiQD paper several years ago made the key step forward [14]. Despite years of work by many researchers the evidence of coupling among DBs and their quantum dot character was a surprising and disruptive event. It changed entirely what was known to be possible with DBs.

There were numerous other difficulties slowing the recognition of DBs as circuit building blocks. Though dangling bond type defects were identified by Bardeen in the 1940s only in the last several years has a robust theory describing the properties of the DBs and importantly how those are manifest in STM images been completed [12,15,34–36]. While clear indications that the DB can exhibit a positive charged state were recorded by us one decade ago those results went unreported for want of fuller theoretical or experimental verification. Additional experimental proof has appeared just in the past year and been reported together with the original observations [16].

While the above advances have established the value of DBs as circuit elements it has remained clear that wished-for ensembles of DBs would never be studied or deployed until advances in fabrication were made. The 2006 ASiQD paper reported on rare successful fabrication efforts. It was then impractical to make even a 2 cell QCA structure [13].

Efforts to better understand and control STM scanning have led to some improvements [37,38] but fell short of control necessary to fabricate atom scale

QCA circuitry. Recently, our reevaluation of non-idealities inherent to the scanned probe fabrication process and in the character of the scanned probe tip itself have led to a large improvement. Yields of a tiny fraction of 1 % have jumped to 80 %. The various refinements will not be discussed here but the results can be seen in Fig. 19. A 16 atom, 4 QCA cell is shown with one atom out of place. Further study has uncovered the dominant reasons for this remnant fabrication error and it is anticipated that as new fabrication tools come on line in the year ahead that improved yields will result. As a test of larger pattern fabrication capacity, and despite the errors still produced, two well studied QCA circuits from the literature [39] were made. We see good overall pattern fidelity and a clear demonstration that we have broken free of the 4 atom limit of a few years ago and may soon be able to make 100 atom structures with excellent fidelity. The circuits in Fig. 19 required about 1 min to fabricate and were automatically made by computer upon input of pattern required.

Substrate defect elimination, automatic fabrication error correction, as-needed on the fly rerouting and circuit redundancy strategies are all parts of the plan to achieve functional devices within the next several years. Numerous

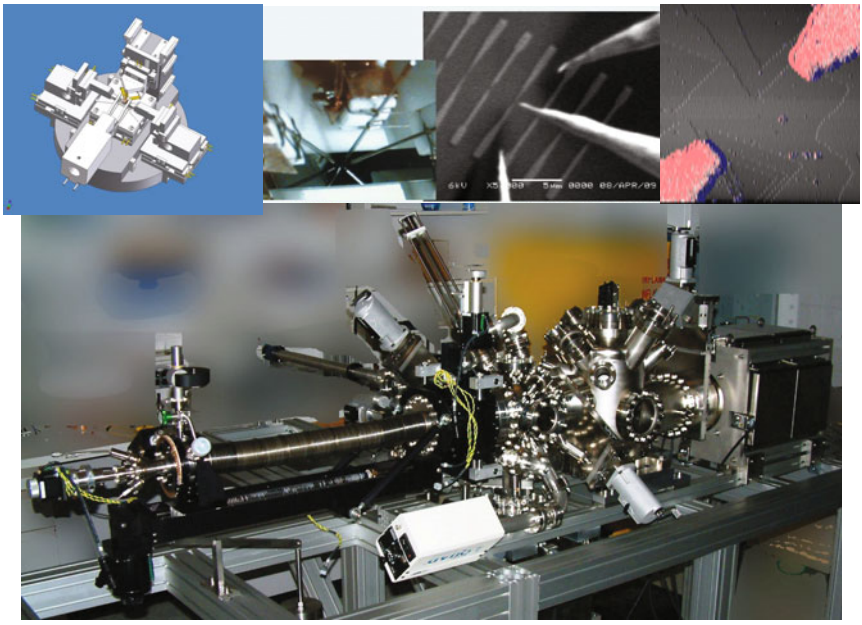


Fig. 18. A multi-probe STM has been developed to allow nano-scale electrical characterization that had previously been beyond reach. The instrument has three independently scannable tips, watched over by a scanning electron microscope. Each tip can be deployed as a scanned probe for imaging or touched down as a current source or as a voltage probe. Nanoscale potentiometry and transport measurements have been made with the instrument.

contact, input/output, encapsulation and packaging issues among others which we have charted courses for will not be discussed in this document. Likewise approaches to eventual very rapid parallel fabrication processes are being developed but are beyond the present scope of discussion.

9 ASiQDs for Quantum Computing

In the QCA mode of operation and in analog electronic strategies the resulting device is not quantum in that it does not depend upon quantum coherence or superposition. However, it is anticipated that ASiQD-based qubits for eventual quantum computing applications can be made - both charge-based and electron spin-based qubits are possible.

The key attribute of the ASiQD-based charge qubit is that the rate of tunneling is very large [23,24]. It is estimated that the tunnel rate can be as much as 106 times larger than the rate of decohering events. Practically this means that there could be sufficient time to undertake a coherent operation before a disruption occurs. Though this is an attractive situation, it is also challenging as it means that precisely phase-controlled operations on the qubit must be done with fraction tunnel period time control which is not possible with any conventional electronics. Some form of optical control is suggested therefore. A new strategy for characterization of the charge qubit has been presented [24].

The spin-type qubit would minimally use one ASiQD per qubit. Unlike in all of the above discussion where it was presumed the isolated DB would have a 1− charge, the DB would be prepared in a neutral one-electron state. Application of a magnetic field differentiates the up and down spin states to create a suitable Zeeman-split two-level system. This is similar to the P atom dopant approach. There, the P atom is studied at very low temperature where it does not ionize and so is not in fact a dopant. It retains its one extra electron. The resulting paramagnetic centre has desirable electron spin properties as proven by ESR measurements of very large numbers of P atoms.

An architecture for spin-based quantum computing using P dopants was first put forward by Kane [40]. In that scheme, the nuclear spins of P atoms and the spins of bound electrons serve as qubits, which interact via hyperfine- and exchange-interactions. In pursuit of that goal, great effort has been devoted over the last decade to place single P dopant atoms into a silicon lattice either by a chemical process that achieves nm position control but not perfect atom scale control, or by ion implantation which is highly uncontrolled and not scalable but which has led to the most impressive results so far [41–43]. The fabrication challenges facing P dopant approaches to quantum computation are significant, and it is known that quantum computation cannot work even in principle until these are overcome [44]. A later proposal by Loss and DiVincenzo [45] makes possible an all-electron-spin approach to quantum computation, which could do away with the need for the nuclear spins of P donors. Paramagnetic ASiQDs are a suitable platform for this architecture or some variant of it, and immediately

offer the advantage of atomically-precise fabrication, in addition to the important features of electronic control and silicon compatibility.

Presented in this way, the DB appears to be a far more attractive electron-spin qubit than the implanted P atom in silicon. That is our belief. But this writing is the first to our knowledge to point out the many advantages. The main disadvantage of the DB route to spin-based quantum computing is that passivation or encapsulation is required, but that seems a surmountable problem. The advantages to using DBs are many. Unlike P atom insertion through a multi-step process, DBs can be made instantly. While P atoms cannot be placed exactly and reproducibly with respect to other P atoms, any number of ASiQDs can be perfectly juxtaposed, just as designed. Some of the latest strategies for achieving robust qubits by combining multiple physical qubits into one logical qubit [46] are greatly aided by this precise multi qubit fabrication facility.

The DB has another advantage related to its large ionization energies. First, the paramagnet species easily exists at room temperature. While low temperature will still be required to avoid lattice-induced decoherence effects, the deeply

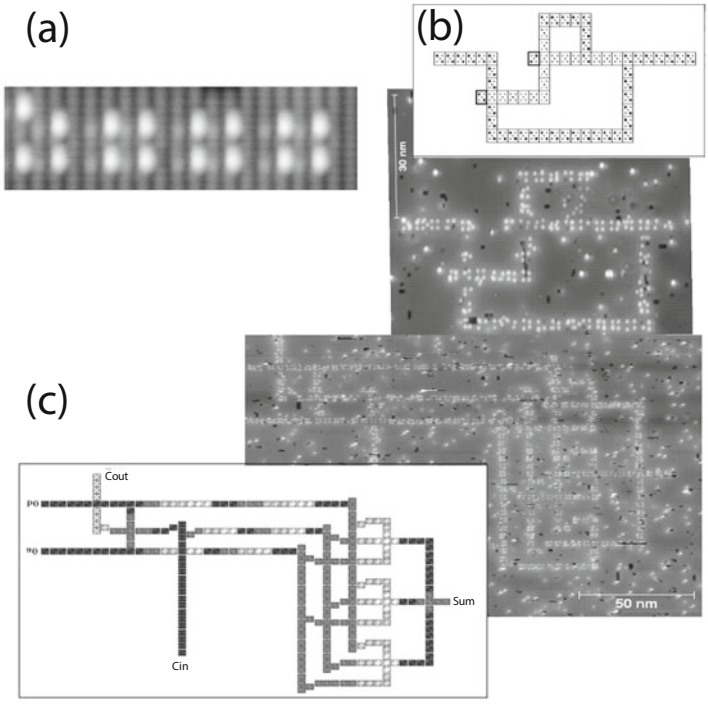


Fig. 19. (a) A 16 atom, 4 QCA cell is shown with one atom out of place, demonstrating the near perfection, but occasional errors resulting at this time. (b), (c) Larger pattern fabrication capacity demonstration. The circuits required about 1 min to fabricate and were automatically made by computer upon input of pattern required.

held electron in the DB will be advantageous over the relatively weakly held P electron. An important advantage emerges trivially because of the relative electronic sizes of the DB and P atoms. Being small and well confined, the DB interacts with fewer surrounding nuclei than the P atom which encompasses an order of magnitude larger volume. The DB as a result will experience fewer nuclear spin-electron spin decoherence effects.

10 Conclusions

In this paper, we outlined our recent and current efforts in building atom-scale quantum-dot cellular automata circuits on a silicon surface. As a building block we use the silicon dangling bond on a H-Si(001) surface, shown to act as a quantum dot. The fabrication, experimental STM imaging, and charging characteristics of the dangling bond and their assemblies are discussed. We then show how precise assemblies of such dots can be created to form artificial molecules. Such complex structures can be used as systems with custom optical properties, circuit elements for quantum-dot cellular automata, and quantum computing. Considerations on macro-to-atom connections are discussed.

References

1. Lent, C.S., Tougaw, P.D., Porod, W., Bernstein, B.H.: Quantum cellular automata. Nanotechnology **4**, 49–57 (1993)
2. Lent, C.S., Tougaw, P.D.: A device architecture for computing with quantum dots. Proc. IEEE **85**(4), 541–557 (1997)
3. Orlov, A.O., Kummamuru, R., Timler, J., Lent, C.S., Snider, G.L., Bernstein, G.H.: Experimental Studies of Quantum Dot Cellular Automata Devices. Research Signpost, Trivandrum (2004)
4. Yadavalli, K.K., Orlov, A.O., Timler, J.P., Lent, C.S., Snider, G.L.: Fanout gate in quantum-dot cellular automata. Nanotechnology **18**(37), 375401 (2007)
5. Tang, Y., Orlov, A.O., Snider, G.L., Fay, P.J.: Towards real-time testing of clocked quantum dot cellular automata. In: Nanotechnology Materials and Devices Conference, 2009. NMDC'09. IEEE, pp. 76–79. IEEE (2009)
6. Joyce, R.A., Qi, H., Fehlner, T.P., Lent, C.S., Orlov, A.O., Snider, G.L.: A system to demonstrate the bistability in molecules for application in a molecular QCA cell. In: Nanotechnology Materials and Devices Conference, 2009. NMDC'09. IEEE, pp. 46–49. IEEE (2009)
7. Foley, E.T., Kam, A.F., Lyding, J.W., Avouris, P.: Cryogenic UHV-STM study of hydrogen and deuterium desorption from Si(100). Phys. Rev. Lett. **80**(6), 1336 (1998)
8. Shen, T.-C., Wang, C., Abeln, G.C., Tucker, J.R., Lyding, J.W., Avouris, P., Walkup, R.E.: Atomic-scale desorption through electronic and vibrational excitation mechanisms. Science **268**, 1590–1592 (1995)
9. Stokbro, K., Thirstrup, C., Sakurai, M., Quaade, U., Hu, B.Y.-K., Perez-Murano, F., Grey, F.: STM-induced hydrogen desorption via a hole resonance. Phys. Rev. Lett. **80**(12), 2618 (1998)

10. Soukiassian, L., Mayne, A., Carbone, M., Dujardin, G.: Atomic-scale desorption of H atoms from the Si(100)-2 × 1: H surface: inelastic electron interactions. Phys. Rev. B **68**(3), 035303 (2003)
11. Kuramochi, H., Uchida, H., Kuwahara, Y., Watanabe, K., Aono, M.: Site-independent adsorption of hydrogen atoms deposited from a scanning tunneling microscope tip onto a Si(111)-7 × 7 surface. Jpn. J. Appl. Phys. **36**(10A), L1343–L1346 (1997)
12. Pitters, J.L., Livadaru, L., Haider, M.B., Wolkow, R.A.: Tunnel coupled dangling bond structures on hydrogen terminated silicon surfaces. J. Chem. Phys. **134**(6), 064712 (2011)
13. Tong, X., Wolkow, R.A.: Electron-induced H atom desorption patterns created with a scanning tunneling microscope: implications for controlled atomic-scale patterning on H:Si(100). Surf. Sci. **600**(16), L199–L203 (2006)
14. Haider, M., Pitters, J.L., DiLabio, G., Livadaru, L., Mutus, J., Wolkow, R.: Controlled coupling and occupation of silicon atomic quantum dots at room temperature. Phys. Rev. Lett. **102**(4), 046805 (2009)
15. Livadaru, L., Pitters, J.L., Taucer, M., Wolkow, R.A.: Theory of nonequilibrium single-electron dynamics in STM imaging of dangling bonds on a hydrogenated silicon surface. Phys. Rev. B **84**(20), 205416 (2011)
16. Taucer, M., Livadaru, L., Piva, P.G., Achal, R., Labidi, H., Pitters, J.L., Wolkow, R.A.: Single electron charging dynamics of atomic silicon quantum dots on the H-Si(100) surface. arXiv:1305.3597
17. Oura, K., Lifshits, V.G., Saranin, A.A., Zotov, A.V., Katayama, M.: Hydrogen interaction with clean and modified silicon surfaces. Surf. Sci. Rep. **35**, 1–69 (1999)
18. Lopinski, G.P., Wayner, D.D.M., Wolkow, R.A.: Self-directed growth of molecular nanostructures on silicon. Nature **406**, 48–51 (2000)
19. Piva, P.G., DiLabio, G.A., Pitters, J.L., Zikovsky, J., Rezeq, M., Dogel, S., Hofer, W.A., Wolkow, R.A.: Field regulation of single-molecule conductivity by a charged surface atom. Nature **435**(7042), 658–661 (2005)
20. Piva, P.G., DiLabio, G.A., Pitters, J.L., Wolkow, R.A.: Electrostatically regulated atomic scale electroconductivity device
21. Shockley, W., et al.: Electrons and Holes in Semiconductors, vol. 1. van Nostrand, New York (1963)
22. Baseer Haider, M., Pitters, J.L., DiLabio, G.A., Livadaru, L., Mutus, J.Y., Wolkow, R.A.: Controlled coupling and occupation of silicon atomic quantum dots. arXiv:0807.0609
23. Livadaru, L., Xue, P., Shaterzadeh-Yazdi, Z., DiLabio, G.A., Mutus, J., Pitters, J.L., Sanders, B.C., Wolkow, R.A.: Dangling-bond charge qubit on a silicon surface. New J. Phys. **12**(8), 083018 (2010)
24. Shaterzadeh-Yazdi, Z., Livadaru, L., Taucer, M., Mutus, J., Pitters, J.L., Wolkow, R.A., Sanders, B.C.: On measuring coherence in coupled dangling-bond dynamics (2013). arXiv:1305.6359
25. Piva, P.G., Wolkow, R.A., Kirczenow, G.: Nonlocal conductance modulation by molecules: scanning tunneling microscopy of substituted styrene heterostructures on H-terminated Si(100). Phys. Rev. Lett. **101**(10), 106801 (2008)
26. Kirczenow, G., Piva, P.G., Wolkow, R.A.: Modulation of electrical conduction through individual molecules on silicon by the electrostatic fields of nearby polar molecules: theory and experiment. Phys. Rev. B **80**(3), 035309 (2009)
27. Tong, X., Wolkow, R.A.: Scanning tunneling microscopy characterization of low-profile crystalline $TiSi_2$ microelectrodes on a Si(111) surface. Appl. Phys. Lett. **86**(20), 203101 (2005)

28. Palermo, V., Buchanan, M., Bezinger, A., Wolkow, R.A.: Lateral diffusion of titanium disilicide as a route to contacting hybrid Si/organic nanostructures. Appl. Phys. Lett. **81**(19), 3636 (2002)
29. Pitters, J.L., Dogel, I.A., Wolkow, R.A.: Charge control of surface dangling bonds using nanoscale Schottky contacts. ACS Nano **5**(3), 1984–1989 (2011)
30. Zikovsky, J., Salomons, M.H., Dogel, S.A., Wolkow, R.A.: Silicon surface conductance investigated using a multiple-probe scanning tunneling microscope. In: Joachim, C. (ed.) Atomic Scale Interconnection Machines: Proceedings of the 1st AtMol European Workshop Singapore 28th–29th June 2011. Advances in Atom and Single Molecule Machines, vol. 1, pp. 167–180. Springer, Berlin (2012). ISBN 978-3-642-28171-6
31. Martins, B.V.C., Smeu, M., Guo, H., Wolkow, R.A.: Conductivity of Si(111)-7 × 7: the role of a single atomic step. arXiv:1310.1313
32. Martins, B.V.C., Wolkow, R.A.: (2013, in preparation)
33. Lyding, J.W., Shen, T.C., Hubacek, J.R., Tucker, J.R., Abeln, G.C.: Nanoscale patterning and oxidation of H-passivated Si(100)-2 × 1 surfaces with an ultrahigh vacuum scanning tunneling microscope. Appl. Phys. Lett. **64**(15), 2010–2012 (1994)
34. Raza, H.: Theoretical study of isolated dangling bonds, dangling bond wires, and dangling bond clusters on a H:Si(001)-(2 × 1) surface. Phys. Rev. B **76**(4), 045308 (2007)
35. Ye, W., Min, K., Martin, P.P., Rockett, A.A., Aluru, N.R., Lyding, J.W.: Scanning tunneling spectroscopy and density functional calculation of silicon dangling bonds on the Si(100)-2 × 1: H surface. Surf. Sci. **609**, 147–151 (2013)
36. Schofield, S.R., Studer, P., Hirjibehedin, C.F., Curson, N.J., Aeppli, G., Bowler, D.R.: Quantum engineering at the silicon surface using dangling bonds. Nat. Commun. **4**, 1649 (2013)
37. Jahromi, S.A.Z., Salomons, M., Sun, Q., Wolkow, R.A.: Prediction of the resonant frequency of piezoelectric tube scanners through three-dimensional finite element modeling of a tube assembly. Rev. Sci. Instrum. **79**(7), 3 (2008)
38. Randall, J.N., Von Ehr, J.R., Ballard, J.B., Owen, J.H.G., Fuchs, E.: Automated Scanning Tunneling Microscope image analysis of Si(100): H 2 × 1 surfaces. Microelectron. Eng. **98**, 214–217 (2012)
39. Hänninen, I., Takala, J.: Arithmetic design on quantum-dot cellular automata nanotechnology. In: Bereković, M., Dimopoulos, N., Wong, S. (eds.) SAMOS 2008. LNCS, vol. 5114, pp. 43–52. Springer, Heidelberg (2008)
40. Kane, B.E.: A silicon-based nuclear spin quantum computer. Nature **393**(6681), 133–137 (1998)
41. Fuechsle, M., Miwa, J.A., Mahapatra, S., Ryu, H., Lee, S., Warschkow, O., Hollenberg, L.C., Klimeck, G., Simmons, M.Y.: A single-atom transistor. Nat. Nanotechnol. **7**(4), 242–246 (2012)
42. Fuechsle, M., Miwa, J.A., Mahapatra, S., Ryu, H., Lee, S., Warschkow, O., Hollenberg, L.C.L., Klimeck, G., Simmons, M.Y.: Spectroscopy of a deterministic single-donor device in silicon. Proc. SPIE **8400**, 840006 (2012)
43. Tan, K.Y., Chan, K.W., Mottonen, M., Morello, A., Yang, C., van Donkelaar, J., Alves, A., Pirkkalainen, J.M., Jamieson, D.N., Clark, R.G., Dzurak, A.S.: Transport spectroscopy of single phosphorus donors in a silicon nanoscale transistor. Nano Lett. **10**(1), 11–15 (2010)
44. Kane, B.E.: Can we build a large-scale quantum computer using semiconductor materials? MRS Bull. **30**, 105–110 (2005)

45. Loss, D., DiVincenzo, D.P.: Quantum computation with quantum dots. Phys. Rev. A **57**(1), 120 (1998)
46. Fowler, A.G., Mariantoni, M., Martinis, J.M., Cleland, A.N.: Surface codes: towards practical large-scale quantum computation. Phys. Rev. A **86**(3), 032324 (2012)

Circuits and Architectures

A Clocking Strategy for Scalable and Fault-Tolerant QDCA Signal Distribution in Combinational and Sequential Devices

Douglas Tougaw(✉)

Valparaiso University, Valparaiso, IN, USA
Doug.Tougaw@valpo.edu

Abstract. A signal distribution network (SDN) for Quantum-dot Cellular Automata (QDCA) devices is described. This network allows the distribution of a set of inputs to an arbitrary number of outputs in any desired order, overcoming the challenges associated with wire crossings that have faced QDCA systems in the past. The proposed signal distribution network requires only four distinct clock signals, regardless of the number of inputs or outputs, and those clock signals each repeat a very simple pattern. This network is highly scalable, completing the distribution of N inputs to an arbitrary number of distributed signals in 4N – 2 clock cycles. The operation of this device is demonstrated by applying it to a two-input/one-output XOR gate and a three-input/two-output full adder. A modified SDN customized for use with sequential devices is also shown.

Keywords: Quantum-dot Cellular Automata · Field-Coupled Nanocomputing · Signal distribution network · Wire crossing

1 Introduction to Quantum-dot Cellular Automata

Quantum-dot Cellular Automata (QDCA) is an emerging computing architecture based on Field-Coupled Nanocomputing (FCN) that has demonstrated the ability to use quantum mechanical interactions to implement both combinational and sequential logic devices [1–9]. Devices implemented using QDCA cells show the potential to be faster and smaller than conventional microelectronic devices. Perhaps even more importantly, they are predicted to operate at a tiny fraction of the power required by current devices.

QDCA devices are designed by carefully selecting the placement of cells and the timing with which their tunneling barriers are raised and lowered. As shown in Fig. 1, a QDCA cell is typically composed of four quantum dots located at the four corners of a square. Tunneling is allowed between adjacent dots, and the cell occupancy is controlled by a back plane voltage so that a total of two electrons occupy each four-site cell. The electrons in each cell Coulombically repel each other, so the cell will typically exhibit one of two states, either pointing diagonally left (a binary 0) or diagonally right (a binary 1). Electrons in nearby cells will also interact with each other Coulombically, causing adjacent cells along the same linear axis to align in the

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 61–72, 2014.
DOI: 10.1007/978-3-662-43722-3_4, © Springer-Verlag Berlin Heidelberg 2014

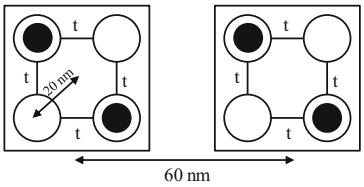


Fig. 1. The QDCA geometry studied in this work. Each cell is composed of four quantum dots, with two electrons sharing each four-site cell. Sites are each 20 nm from the center of the cell, and adjacent cells are 60 nm apart. Tunneling is allowed between adjacent sites within the cell, and electron repulsion leads to a bistable alignment into one of two states, which are used to encode a binary 0 and a binary 1.

same direction. In the system studied in this work, the corners of the cells are 20 nm from the center, and the centers of adjacent cells are 60 nm apart.

By carefully arranging the location of QDCA cells, it is possible to create a majority logic gate, which is further capable of functioning as either an AND or an OR gate [5]. Combining these gates with the QDCA cell configuration of an inverter, any combinational logical device can be constructed from QDCA cells [6]. A very active area of QDCA research involves the design and verification of complex devices composed of the fundamental QDCA gates [10–15]. Invariably, these more complex devices require a number of binary wires to transmit signals between the logic gates, and these wires often need to cross each other. Due to the coplanar nature of the QDCA architecture, these wire crossings present a special challenge. The purpose of this work is to demonstrate a comprehensive solution to this wire crossing problem.

In addition to carefully considering the placement of each cell in a QDCA device, the designer must also determine the timing of clock signals applied to the device, along with which cells will be attached to those clock signals. These clock signals control the tunneling barriers within each cell, and they are switched quasi-adiabatically. The tunneling barriers of each cell are raised and lowered smoothly to control how and when each cell responds to its neighbors. The tunneling barriers are typically modulated through four stages: the locking state, in which the tunneling barriers are raised; the locked state, in which tunneling is entirely restricted by the presence of high barriers; the relaxing state, during which tunneling barriers are lowered; and the relaxed state, in which tunneling barriers are held very low, allowing nearly free electron movement within a cell.

This natural sequence of four clock signals is shown in Fig. 2, which illustrates the flow of data through four adjacent clocking regions by the careful timing of the clock signals. The arrows demonstrate the flow of information from one region to the next.

2 The Signal Distribution Network

The signal distribution network (SDN) is a possible solution to the wire crossing problem. This device has many significant benefits over the wire-crossing solutions that have been previously presented. Most importantly, it relies entirely on nearest

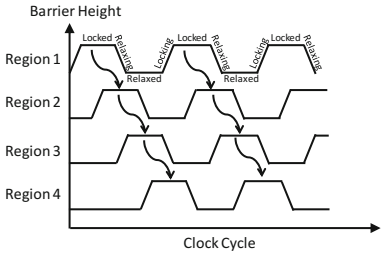


Fig. 2. Data propagation through four adjacent quasi-adiabatically switched regions. The clock signals are phase-shifted to ensure that as each region enters the "locking" phase, the cells in the previous region are in the "locked" phase. This causes the data to be shifted from one region to the next, as indicated by the arrows.

neighbor interactions, which will increase the excitation energy of the device and in turn will improve its thermal behavior and its tolerance for fabrication imperfections. Furthermore, the SDN can be scaled to handle combinational devices with any number of inputs and any number of outputs, typically with fewer cells and using less surface area than other candidate solutions. Additionally, it requires only four clock signals, each of which exhibits a very regular pattern of transitions regardless of the function being implemented. Finally, we will see that the SDN is guaranteed to handle the signal distribution functionality for an N-input device in no more than 4N – 2 clock cycles, meaning that the entire combinational device will require approximately 4N clock cycles, depending on how the clock signals are used in the binary logic portion of the device [16].

The block diagram shown in Fig. 3 illustrates the operation of the SDN. The system separates the wire crossings from the logical operations, allowing each of the two blocks to be optimized. Once the SDN has done its work, the distributed signals are applied to the combinational logic gates, and there is no need for any wires to cross within that region. Separating the signal distribution from the combinational logic gates allows the optimization of the SDN and the combinational logic gates separately, leading to a more efficient and much more scalable solution.

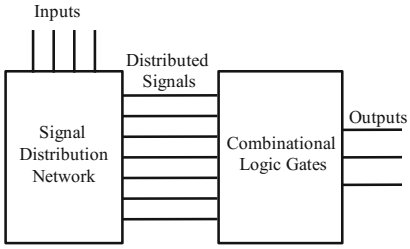


Fig. 3. Separation of the function into signal distribution and combinational logic. This separation allows for the use of an optimized, highly regular signal distribution network, followed by a similarly optimized combinational logic block without the need for interspersed wire crossings. The number of inputs, distributed signals, and outputs are entirely arbitrary in this figure.

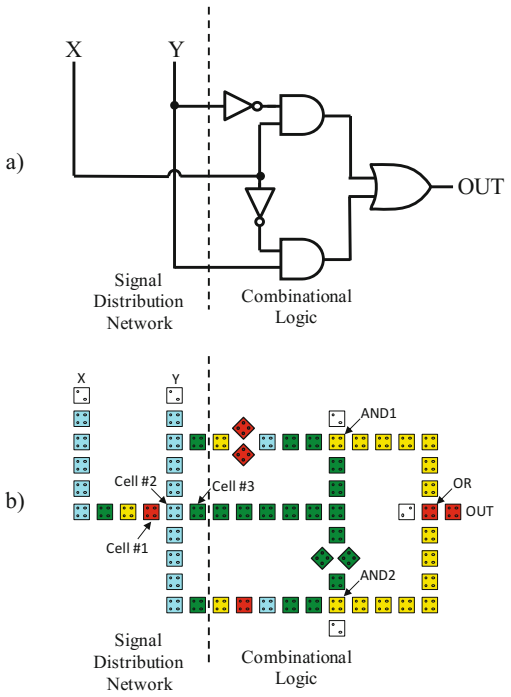


Fig. 4. An implementation of an XOR device using a minimal signal distribution network. (a) A schematic representation of the function being implemented, which is logically identical to that shown in Fig. 4. (b) The layout of QDCA cells used to implement the function. Cell #2 is driven by input Y in the first clock cycle. Input X is conveyed along the horizontal line for four cycles until it reaches Cell #1. Cell #1 is then used to drive Cell #2 (and the vertical line it is a part of) in the fifth cycle. This is made possible by removing input Y during the second cycle. This implementation uses only four different clock signals and eight clock cycles to complete the calculation.

As an example of this type of structure, consider the circuit shown in Fig. 4, which separates the signal distribution function from the logic gates as described above.

The resulting benefits of this change are quite significant. The SDN-based device shown in Fig. 4 will require only four clock signals, rather than the seven that were required for an earlier design of this circuit that did not use the SDN [16]. Furthermore, as can be seen in Fig. 5, the clocking pattern of these four clock signals is entirely regular, which significantly simplifies the clock generation.

The operation of this new type of wire crossing merits a detailed description. The two inputs, labeled X and Y in Fig. 4(b), each drive a long vertical binary wire in the first clock cycle. The rightmost wire (initially carrying input Y) is already at the right edge of the SDN, so any required copies of that signal can be simply be connected from that vertical line at the required location. However, any necessary copies of the leftmost vertical binary wire (initially carrying input X) will need to cross over the other vertical binary wire before they can be used in the combinational logic gates.

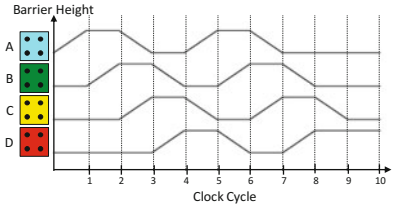


Fig. 5. The clock signals applied to the SDN-based XOR gate shown in Fig. 4. These signals are highly regular, repeating the same pattern. The color-coded cells to the left of the y-axis labels correspond to the coloring in Fig. 4(b).

This will be handled by the clock signals shown in Fig. 5 and by carefully assigning individual cells to the corresponding clock signals.

During clock cycle #2, the X input signal propagates one cell to the right (into the cell colored green in the color version of Fig. 4(b)). In clock cycle #3, it moves one more cell to the right into the yellow cell, and in clock cycle #4, it moves one more cell to the right into the red cell, which is also labeled cell #1 in Fig. 4(b). Clock cycle #5 is the key to this device's operation. Since the inputs X and Y were only applied during the first clock cycle and were then removed, the right-hand vertical wire is now available to be driven by cell #1. Through the interaction between cell #1 and cell #2, the entire right-hand vertical wire is driven to contain the value of input X during clock cycle #5. Note that this signal was originally contained in the left-hand vertical wire during clock cycle #1. In clock cycle #6, the signal is available along the entire length of the right-hand vertical wire, so copies can be made at any needed location, such as at cell #3.

At this point, the signals have been fully distributed, and no additional wire crossings will be needed. However, input Y, which was originally on the right-hand vertical wire, will arrive at its destination five clock cycles before input X unless we take corrective action. This problem is solved by carefully assigning the cells on each horizontal wire to clocking regions of varying widths. The wires containing signal Y take very small steps, with each clock signal assigned to just one or two cells. By contrast, signal X moves very quickly from cell #3 to the AND gates because those entire lines all belong to the same clock signal. In this way, the signals corresponding to X and Y each arrive at the AND gates during clock cycle #6, the AND calculations are completed during clock cycle #7, and the OR calculations yield the final output during clock cycle #8.

The contents of each of the labeled cells in Fig. 4(b) for each clock cycle can be seen in Fig. 6, and Fig. 7 shows a series of snapshots of the device at key phases of its operation.

Although this device requires eight clock cycles to complete the XOR function, only six of these are used by the SDN. The remaining two clock cycles are used to implement the (relatively simple) combinational logic. The device was simulated for all four possible combinations of the two inputs, and each was found to yield the correct result.

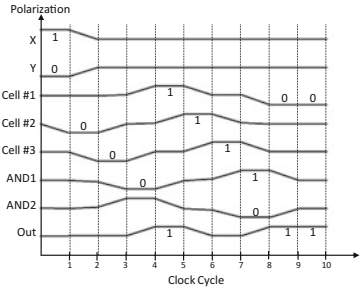


Fig. 6. The states of labeled cells throughout the operation of the device in Fig. 4(b) when X = 1 and Y = 0. Inputs X and Y are only applied during the first clock cycle, and their contents are propagated through the device at different speeds so that they arrive at AND1 and AND2 at the same time. The output correctly shows that $1 \oplus 0 = 1$.

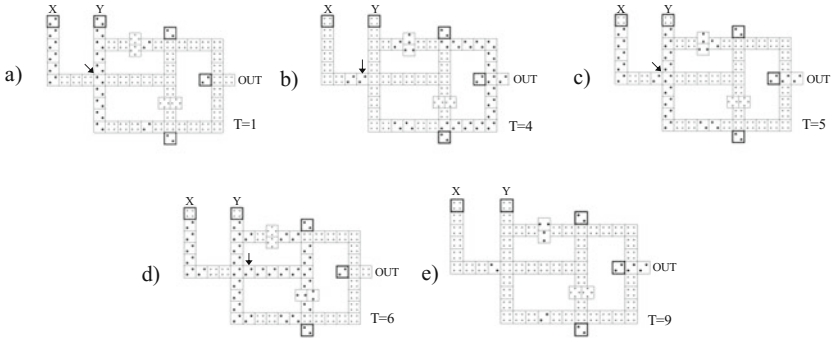


Fig. 7. Snapshots of the XOR gate with SDN at key points during its operation. (a) At T = 1, the left-hand vertical wire contains X = 1, while the right-hand vertical wire contains Y = 0. (b) At T = 4, input X has propagated to the cell indicated with the arrow. (c) At T = 5, this cell drives the right vertical wire by its interaction with the indicated cell. Note that input Y is unpolarized at this time. (d) At T = 6, the value of input X has propagated through the vertical wire and is quickly transmitted to both AND gates, where input Y just has arrived. (e) At T = 9, the device delivers the correct output.

3 Expanding the Signal Distribution Network

Although the previous example focused on a two-input device, the benefits of the SDN become much more apparent when this is expanded to a three-input device. Figure 8 shows a schematic representation of a full adder, which was originally introduced in Ref. [5].

Figure 8 once again demonstrates a separation of the signal distribution function from the combinational logic gates, and Fig. 9 illustrates how the two-input SDN can be expanded to three inputs in order to implement the circuit shown in Fig. 8.

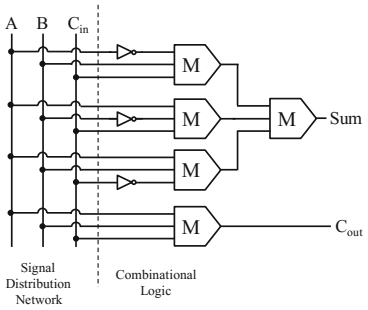


Fig. 8. A schematic of the circuit used to perform one-bit full addition using QDCA cells. This circuit takes full advantage of the majority logic functionality of the QDCA cell to minimize the number of gates needed.

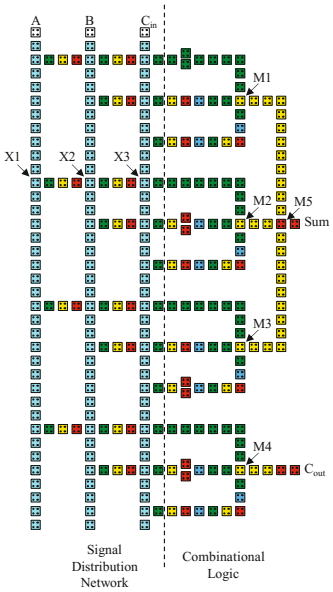


Fig. 9. A QDCA implementation of the schematic circuit shown in Fig. 8. The three vertical wires carry the three inputs, and these three inputs must all arrive at the majority logic gates (M1–M4) at the same time. This is done by adjusting the clock phases of the horizontal lines.

Figure 10 shows that the three-input SDN with associated combinational logic gates can still be implemented using only four clock signals, and that those clock signals still exhibit the highly regular pattern of behavior first seen in Fig. 5.

Figure 11 shows the data traces for labeled cells in Fig. 9 for the case where $A = 1$, $B = 0$, and $C_{in} = 1$. This figure illustrates that the three inputs are once again applied only during the first clock cycle, and that these signals flow from left to right along the horizontal wires (such as the one including cells X1, X2, and X3).

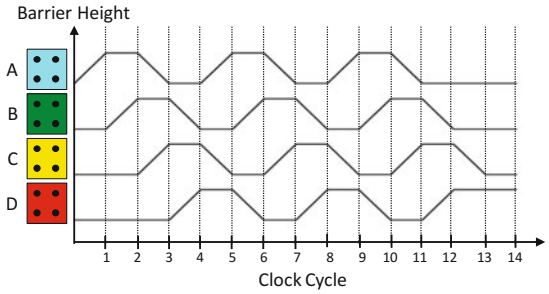


Fig. 10. The four clock signals used to drive the device shown in Fig. 9. Again, note the highly regular nature of these clock signals, which repeat the pattern shown in Fig. 5 three times each.

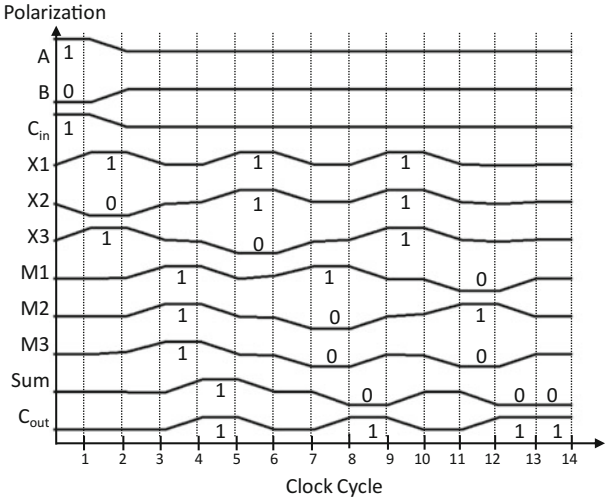


Fig. 11. The states of labeled cells throughout the operation of the device in Fig. 9 when A = 1, B = 0, and C_{in} = 1. These inputs are only applied during the first clock cycle, and their contents are propagated through the device at different speeds so that they arrive at M1–M4 at the same time. The correct output, shown at the bottom right corner of the figure, first appears during the 12^{th} clock cycle. These results are the outcome of simulating the device in Fig. 9 using the clock signals in Fig. 10 for 14 clock cycles.

Each time one of the input signals reaches another vertical wire, it is allowed to drive the entire length of that wire, since the input signal that originally drove it in clock cycle #1 has been removed. Whereas in the two-input XOR gate, only one vertical wire had to be traversed, now the leftmost signal must cross two vertical wires. This is not a problem, but it does require four more clock cycles for the signal to progress through the three-input SDN. In this way, the SDN for this device delivers 12 distributed copies of its three inputs during clock cycle #10 (as compared to clock cycle #6 for the two-input device). The combinational logic gates then require two

clock cycles to complete their work, and the correct output first appears during clock cycle #12. More generally, the time to complete signal distribution will be 4N – 2 clock cycles, where N is the number of inputs. This relationship is valid for all values of $N \geq 2$, including the devices shown in Fig. 4 ($N = 2$) and Fig. 9 ($N = 3$). It has also been shown to hold for a device with $N = 4$ [16].

The device shown in Fig. 9 was simulated for all eight possible combinations of the three inputs, and it yielded the correct outputs for each of the eight cases.

4 Sequential Signal Distribution Network

The SDN structure shown in the previous section is optimized for use with combinational logic devices. The signal distribution requirements for a sequential device have particular characteristics that require modifications to this design in order to optimize the complete system.

It is possible, as shown in Fig. 12, to consider a sequential device to be composed of two combinational devices (a next state decoder and an output decoder), along with a number of D flip-flops to store the current state of the system. In this case, each of the two combinational devices could be implemented using an SDN followed by the necessary combinational logic gates. The problem comes in the region between the next state decoder and the output decoder, which requires a large number of wire crossings in order to route the four bits of the current state back to the input of the next state decoder.

These wire crossings can be eliminated by using a modified SDN to route the current state signals back to the input of the next state decoder, as shown in Fig. 13. Not only does this modified SDN eliminate the need for wire crossings between the two decoders, it also fills the role of the SDN for the next state decoder. As shown in Fig. 13, the outputs of the next state decoder (the current state of the device) all leave the next state decoder in the same clock phase, and these horizontal lines (labeled S0–S3) are available to be applied to the output decoder, if one is required. In addition, they are also routed vertically through a series of one-cell regions that propagate

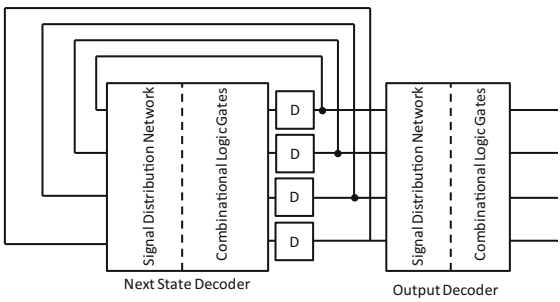


Fig. 12. Schematic representation of a sequential logic device implemented using two combinational devices (a next state decoder and an output decoder) as well as four (in this case) D flip-flops to store the current state of the device. Note the large number of wire crossings between the next state decoder and the output decoder.

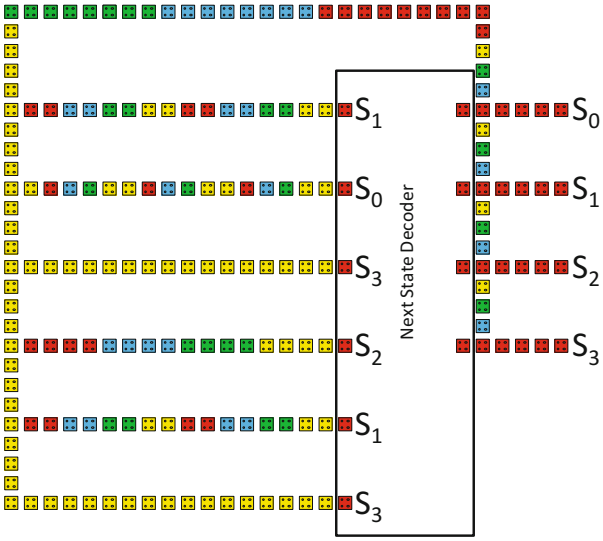


Fig. 13. The Sequential Signal Distribution Network (SSDN). This device routes and distributes the bits of the current state (available at the output of the next state decoder) and makes them available (20 clock cycles later) at the inputs of the next state decoder. The output decoder, which is not required in all cases, is not shown in this figure.

the signal vertically through unpolarized horizontal wires (just like the SDN in the previous section, but moving in upward instead of rightward). These signals then complete the loop by moving leftward along the top of the next state decoder, and then copies of the signals are distributed as needed to the next state decoder by a number of horizontal lines (six, in this case). The particular signal that is applied to each of the inputs of the next state decoder is determined by the pattern of clock phases that appears on each horizontal line. S_0, which has the shortest distance to travel, will appear on the far-left vertical line after seven clock cycles, and so it needs to move more slowly along the horizontal line than, say, S_3, which appears on the far-left vertical line after 19 clock cycles. This Sequential Signal Distribution Network (SSDN) delivers all four bits of the current state to all necessary inputs of the next state decoder 20 clock cycles after they appear at the output of the next state decoder. By reassigning clock regions along the topmost wire down to the S_0 output, this number could be reduced to as few as 14 clock cycles or, more generally, $3N + 2$ clock cycles, where N is the number of bits in the current state.

5 Conclusions and Summary

This work has introduced two versions of a signal distribution network for QDCA systems, one for combinational devices and the other for sequential devices. The combinational SDN block can accept an arbitrary number of inputs and yield an

equally arbitrary number of distributed signals, which can then be forwarded to a combinational logic block without the need for additional wire crossings.

Regardless of the number of inputs, the combinational SDN requires only four highly regular clock signals, and the signal distribution function for a device with N inputs is always completed in 4N – 2 clock cycles.

The wire crossings in the combinational SDN rely only on the strongest interactions available in QDCA systems—the interactions between near-neighbor cells along the same line. For this reason, the excitation energy of the system should not be degraded by the use of the SDN, which means that systems using it will exhibit thermal behavior and tolerance of fabrication imperfections very similar to that of other robust QDCA wires and gates.

The sequential SDN fills the same role, distributing multiple copies of an arbitrary number of bits, but it does so in an optimized manner for sequential devices. It implements the required wire crossings that must occur at the output of the next state decoder, while also distributing those bits to the necessary locations at the input of the next state decoder. It can perform this task in as little as 3N + 2 clock cycles, where N is the number of bits in the current state.

Acknowledgment. This work was supported by the Leitha and Willard Richardson Professorship of Engineering, which is provided by the Valparaiso University College of Engineering.

References

1. Lent, C.S., Tougaw, P.D., Porod, W.: Bistable saturation in coupled quantum dots for quantum cellular automata. Appl. Phys. Lett. **62**(7), 714–716 (1993)
2. Lent, C.S., Tougaw, P.D.: Lines of interacting quantum-dot cells: a binary wire. J. Appl. Phys. **74**(10), 6227–6233 (1993)
3. Lent, C.S., Tougaw, P.D., Porod, W.: Bistable saturation in coupled quantum-dot cells. J. Appl. Phys. **74**(5), 3558–3566 (1993)
4. Lent, C.S., Tougaw, P.D., Porod, W., Bernstein, G.H.: Quantum cellular automata. Nanotechnology **4**(1), 49–57 (1993)
5. Tougaw, P.D., Lent, C.S.: Logical devices implemented using quantum cellular automata. J. Appl. Phys. **75**(3), 1818–1825 (1994)
6. Lent, C.S., Tougaw, P.D.: A device architecture for computing with quantum dots. Proc. IEEE **85**(4), 541–557 (1997)
7. Tougaw, P.D., Lent, C.S.: Dynamic behavior of quantum cellular automata. J. Appl. Phys. **80**(8), 4722–4736 (1996)
8. Niemier, M.T.: The effects of a new technology on the design, organization, and architectures of computing systems. Ph.D. Dissertation, University of Notre Dame (2003)
9. Orlov, A.O., Amlani, I., Kummamuru, R., Rajagopal, R., Toth, G., Lent, C.S., Bernstein, G.H., Snider, G.L.: Experimental demonstration of clocked single-electron switching in quantum-dot cellular automata. Appl. Phys. Lett. **77**, 295–297 (2000)
10. Pasky, J.R., Henry, L., Tougaw, P.D.: Regular arrays of quantum-dot cellular automata macrocells. J. Appl. Phys. **87**(12), 8604–8609 (2000)
11. Gin, A., Williams, S., Meng, H., Tougaw, P.D.: Hierarchical design of quantum cellular automata. J. Appl. Phys. **85**(7), 3713–3720 (1999)

12. Janulis, J., Tougaw, D., Henderson, S., Johnson, E.: Serial bit stream analysis using quantum-dot cellular automata. IEEE Trans. Nanotechnol. **3**(1), 158–164 (2004)
13. Tougaw, D., Johnson, E.W., Egley, D.: Programmable logic implemented using quantum-dot cellular automata. IEEE Trans. Nanotechnol. **11**(4), 739–745 (2012)
14. Wood, J.D., Tougaw, D.: Matrix multiplication using quantum-dot cellular automata to implement conventional microelectronics. IEEE Trans. Nanotechnol. **10**(5), 1036–1042 (2011)
15. Graunke, C.R., Wheeler, D.I., Tougaw, D., Will, J.D.: Implementation of a crossbar network using quantum-dot cellular automata. IEEE Trans. Nanotechnol. **4**(4), 435–440 (2005)
16. Tougaw, D., Khatun, M.: A scalable signal distribution network for quantum-dot cellular automata. IEEE Trans. Nanotechnol. **12**, 215–224 (2013)

Electric Clock for NanoMagnet Logic Circuits

Marco Vacca[1], Mariagrazia Graziano[1(✉)], Alessandro Chiolerio[2,3], Andrea Lamberti[2,3], Marco Laurenti[2,3], Davide Balma[4], Emanuele Enrico[5], Federica Celegato[5], Paola Tiberto[5], Luca Boarino[5], and Maurizio Zamboni[1]

[1] Dipartimento di elettronica e telecomunicazioni, Politecnico di Torino, Turin, Italy
mariagrazia.graziano@polito.it
[2] Istituto Italiano di Tecnologia (IIT), Center for Space Human Robotics, Turin, Italy
[3] Chilab - Materials and Microsystems Laboratory, Chivasso, Italy
[4] Ceramics Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[5] Istituto Nazionale per la Ricerca Metrologica (INRIM), Turin, Italy

Abstract. Among Field-Coupled technologies, NanoMagnet Logic (NML) is one of the most promising. Low dynamic power consumption, total absence of static power, remarkable heat and radiations resistance, in association with the possibility of combining memory and logic in the same device, make this technology the ideal candidate for low power, portable applications. However, the necessity of using an external magnetic field to locally control the circuit represents, currently, the weakest point of this technology. The high power losses in the clock generation system adopted up to now wipes out the most important advantages of this technology.

In this chapter we discuss a clock system based on a piezoelectric actuator that allows electrical control of NanoMagnet Logic circuits. The low power consumption coupled with the fact that electric fields are easier to generate at the nanoscale level makes this clock system a strong candidate as the final and effective clocking mechanism for this technology. Another remarkable advantage of the proposed solution resides in its compatibility with currently available technology.

1 Introduction on Clocking For NML Logic

Recent years have shown a rapidly rising interest in field-coupled technologies, like Quantum dot Cellular Automata (QCA) [1]. In these devices information is not represented using voltage or current values, but rather with identical cells that can assume different charge configurations [2]. Information propagates therefore through the electrostatic interaction among neighbor cells [3]. Different means can be used to map the QCA theoretical principle on practical application. This can be obtained choosing appropriate materials and structures to implement the basic cells needed in this technology, leading to different types of QCA, like metallic QCA [3], semiconductor QCA [4] or molecular QCA [5–7]. Metallic QCA were the first practical implementation of the QCA principle, but

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 73–110, 2014.
DOI: 10.1007/978-3-662-43722-3_5, © Springer-Verlag Berlin Heidelberg 2014

they work only at temperatures near absolute zero. The same thing applies also to semiconductor QCA, which use a complex semiconductor structure as the basic cell. Molecular QCA instead use the charge stored in complex molecules to represent logic values, they can work at room temperature and at very high speed. Another possibility is instead to use rectangularly shaped magnets as the basic cell, as shown in Fig. 1. If magnet sizes are reduced to the nanoscale level, around 50–100 nm, they are forced into the single domain condition and their hysteresis cycle changes as shown in Fig. 1. The consequence is that, at the equilibrium, only two states are possible and therefore they can be used to represent logic values '0' and '1'. This particular QCA implementation is then called NanoMagnet Logic (NML) [8–10]. The characteristic of NML logic is that devices work at room temperature and they can be fabricated with currently available technological process. As a consequence very advanced experimental activity is going on with it [11–13]. Since NML circuits are based on magnets, they have some specific advantages over other QCA implementations. They have no static power consumption and potentially a very low dynamic power absorption. They have an high heat and radiation resistance, making them ideal for hard environment applications. Finally, due to their magnetic nature, they combine logic and memory in the same device, opening up completely new possibilities in the development of logic circuits. Their only disadvantage is that they work at relatively low speed, around 50–100 MHz [14].

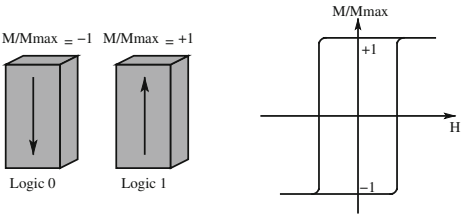


Fig. 1. Reducing the size of a rectangular shaped magnet at nanoscale level (50–100 nm) force it in the single domain condition. Their hysteresis cycle changes and at the equilibrium only two states are possible, thus representing the logic values '0' and '1'.

In NML technology, as it happens in general for QCA, circuits are built placing cells on the same plane. Information propagation and logic computation are caused by magnetic interaction between neighbor cells. For example in Fig. 2(A) a NML wire is shown. Magnets are aligned horizontally and they align antiferromagnetically to reach the minimum energy state. Since every magnet is in the opposite state of its neighbors, an inverter can be simply built making a wire with an odd number of magnets (Fig. 2(B)). The basic logic gate available on this technology is the majority voter, shown in Fig. 2(C). It is a three inputs gate where the value of the central magnet is equal to the majority of the inputs. While this gate coupled with the inverter allows to design any kind of logic

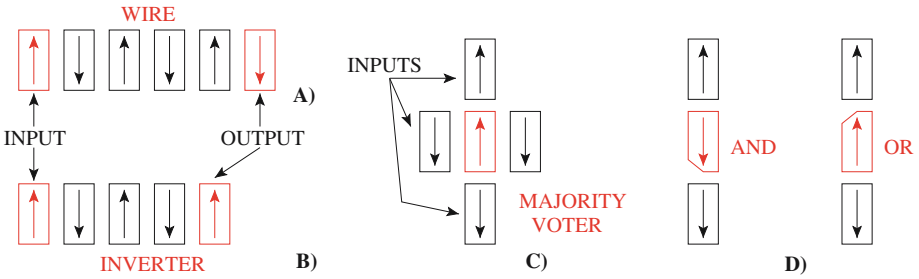


Fig. 2. NML basic logic gates. (A) Wire. (B) Inverter. It is made simply using a wire with an odd number of magnets. (C) Majority voter. The value of the central magnets is equal to the majority of the inputs. (D) AND/OR gates built changing the shape of the central magnet.

function, its asymmetry may cause errors in the signal propagation as shown in [15,16]. However, it is possible to expand the logic set available implementing two inputs AND/OR gates, that are shown in Fig. 2(D). Changing the shape of the central magnet gives it a preferential state, so that only when both magnets are up or down they are able to influence the central magnet [17].

Unfortunately the magnetic field generated by a magnet is not strong enough to switch its neighbors. As a consequence an external mean, like an external magnetic field, is necessary to help magnets switching [18–20]. Considering for example the simple QCA wire of Fig. 3(A), if only the input magnet is switched from '1' to '0' (Fig. 3) the second magnet will remain in the same state without switching. However an external magnetic field can be applied to force the other magnets in the RESET state (Fig. 3(B)). The RESET state is an unstable state so, when the magnetic field is removed, the first magnets will switch according to the input element, while the other magnets will remain temporary in the RESET state (Fig. 3(C)). Magnets will therefore switch one by one (Fig. 3 from (D) to (E)) with a domino-like effect, therefore propagating the new input value.

As demonstrated in [21,22] only a limited number of magnets can be cascaded without errors generation during the realignment phase. Consequently, in order to build complex circuits a multiphase clock system must be used. Circuits are divided in small areas, called clock zones, each of them made by a limited number of magnets. At each clock zones a different clock signal is applied. For example in Fig. 4 a 3 phase clock system is depicted. Each clock zone is subjected to one of the three clock signals showed in Fig. 4(A). The clock waveform is always the same, but signals have a phase difference of 120 °. The circuit state evolution is showed in Fig. 4(B). When magnets belonging to a clock zone are in the SWITCH state, that means the magnetic field is slowly removed, magnets on the clock zone on their left are in the HOLD phase (no magnetic field applied) and act like an input. Magnets in the clock zone on their right are in the RESET state, so they have no influence on the signals propagation. Thanks to this mechanism signals propagate correctly in a specific direction.

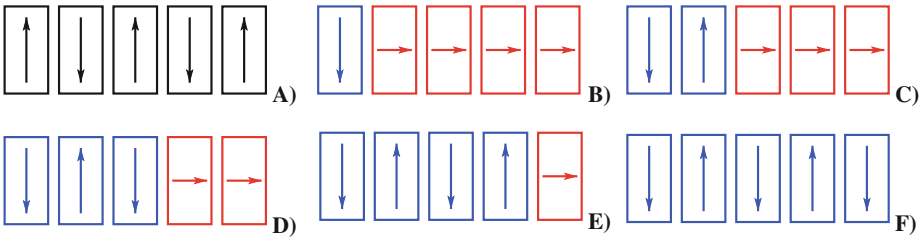


Fig. 3. Clocking mechanism for NML technology. (A) Starting from an initial configuration, (B) the value of the input element change and the other magnets are forced in the RESET state by an external mean, like a magnetic field. (C)–(F) When the magnetic field is removed magnets realign themselves following the input element, propagating therefore the information through the circuit.

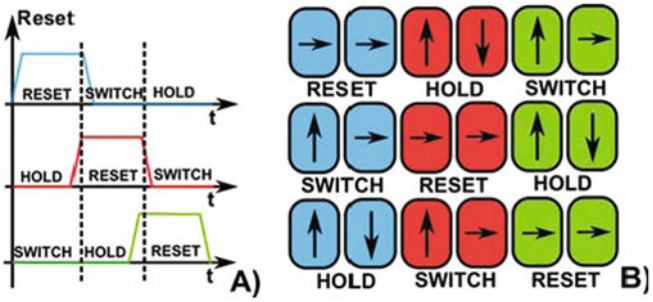


Fig. 4. 3-phase clock system. (A) Clock signal waveforms. 3 clock signals, with a phase difference of 120 °, are used to assure a correct signals propagation. (B) Detailed signal propagation through a simple NML wire. When magnets of a clock zone are switching (SWITCH) magnets on their left are in the HOLD state and act as an input, while magnets on their right are in the RESET state so they have no influence.

1.1 Comparison of Main Clock Solutions

Clocking is the most important feature of NML circuits, giving them a characteristic pipelined behavior [23–25] that represents a considerable source of problems. The most important of these problems is power consumption. While the intrinsic power consumption of NML circuits is extremely low [18] the losses in the clock generation system can be quite high, wiping out one of the major advantages of NML technology.

Different clock systems have been proposed. The first of this mechanisms is shown in Fig. 5(A). A magnetic field directed along the shorter magnets side is generated by a current flowing through a wire placed under the magnets plane. The wire, made of copper, is buried in a ferrite yoke to obtain a better distribution of the magnetic flux lines. This system works and it has already been theoretically demonstrated [26], however its power consumption is very high. In [11] a current of 545 mA on a wire 1 µm wide was necessary to successfully reset magnets. A magnetic field is also used in another NML implementation,

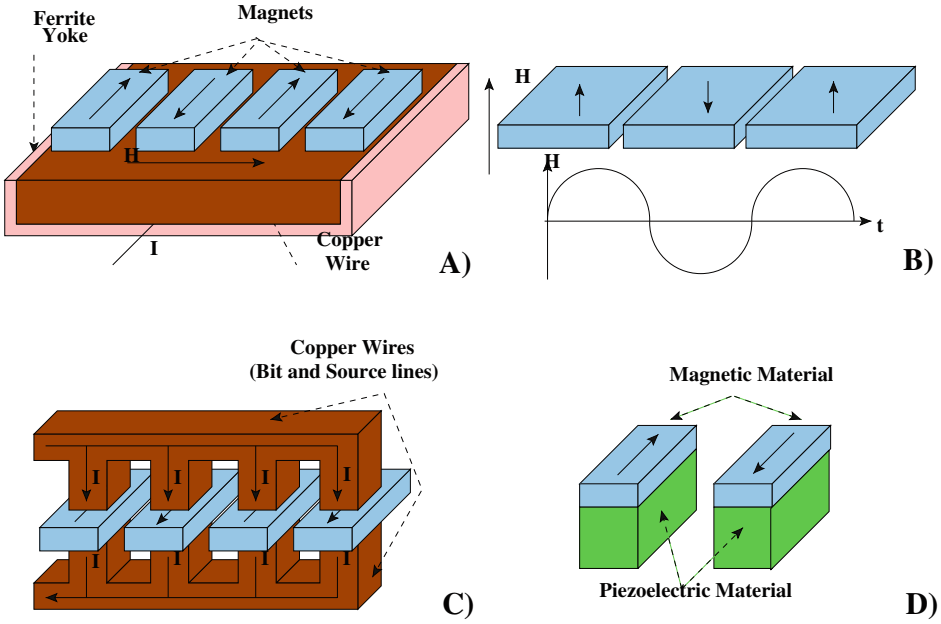


Fig. 5. NML clock systems. (A) Magnetic field clock. A magnetic field is generated by a current flowing through a wire placed under the magnets plane. (B) Out-of-plane NML clock. In this particular implementation of NML, the magnetization of each dots points out-of-plane. In this case the clock is an oscillating magnetic field applied perpendicularly to the plane. No clock zones are present because clock field is applied uniformly to the entire circuit. (C) STT-current clock. Spin-Torque coupling due to a current flowing through the magnets, which in this case are MTJ junctions. (D) Strain based clock. Magnets are forced in the RESET state through a mechanical deformation.

called Out-of-Plane NML [27], where magnets are made by Cobalt-Platinum structures and the magnetization lies perpendicularly to the plane. In this case the magnetic field is perpendicular to the plane (Fig. 5(B)). A different approach uses instead STT-coupling with a current flowing through the magnets as clock mechanism (Fig. 5(C)). The basic element in this case is a Magneto-Tunnel Junction (MTJ) and wires are used to contact them and to drive the current [28,29]. This solution as the remarkable advantage that is based on the Magnetic RAM memories structure, a technology that is already commercially available, moreover the value of clock current required is lower than in the magnetic field approach. Unfortunately all current based clock systems have a relatively high power consumption, mainly due to the Joule losses in the clock wires. A different solution was proposed in [30,31] and it is shown in Fig. 5(D). Magnets are multilayer structures made by a piezoelectric layer and a magnetic layer, and they can be forced in the RESET state applying a proper electric field to them. The use of an electric field allows an important reduction of the losses in the clock generation system obtaining therefore true low power operations.

2 Electric Control of Magnetic Circuits

It is clear from the analysis presented in Sect. 1.1 that to obtain a true low power system NML circuits must be controlled with an electric field, therefore with a voltage, instead of using currents or magnetic fields. To control the magnetization of a material through a voltage the easiest way is to exploit the magnetoelastic effect. The magnetoelastic effect is a physical property of magnetic materials that bonds the magnetization with a mechanical stress. A mechanical stress of proper intensity and direction can change the magnetization of a magnetic material. The solution presented in [30] and reported in Fig. 5(D) is based on the magnetoelastic effect. When an electric field is applied to the dots the piezoelectric material is strained inducing therefore a mechanical stress on the magnetic layer. However, some problems arise with that solution. The first problem is related to the piezoelectric layer. We have a good experience about the fabrication of piezoelectric films, both for MEMS applications [32] and not [33], and we can say that it is not easy to obtain piezoelectric films of just 40 nm, as proposed in [30]. Moreover the surface of piezoelectric films is normally quite rough. The second problem arises with magnets sizes. In-plane NML circuits rely on the shape anisotropy, which is a properties of magnetic materials. Shape anisotropy depends on magnets shape, particularly one magnets side must be bigger than the other to obtain a stable enough magnetization. Moreover shape anisotropy also influences the energy necessary to force magnets in the reset state. The lower the magnets aspect ratio (the ratio between the longer and the shorter magnet side) is, the lower is the reset energy but also the stability of the magnets when they are in one of the stable states. In [30] magnets are $99 \times 101\,\text{nm}^2$, but with the best currently available fabrication processes it is quite difficult to obtain a resolution of just 1 nm. Moreover, with those sizes the tolerance toward process variations will be very low. The consequence is that, with a so low magnets aspect ratio, any kind of fabrication error or defect would drastically change the magnetic anisotropy. As a result magnets would be unstable, if the aspect ratio is reduced, or it would be impossible to force them in the reset state, if the aspect ratio increases. Finally, in [30] no feasible system to generate and distribute the electric field is proposed. Furthermore, while the power consumption estimated is quite low, it does not take into account the losses in the clock distribution network.

The clock system here described is also based on the magnetoelastic effect, however it is built around our experience with piezoelectric materials. The aim is to provide a clock solution where both the layout of NML circuits and of the clock distribution network are considered. We want to choose a solution that (i) takes into account the limitations of currently available fabrication processes, (ii) that allows to obtain a true low power circuit, considering also the losses in the distribution network. The idea is to take a piezoelectric substrate and to deposit magnets on top of it, as shown in Fig. 6. When an electric field is applied to the piezoelectric layer, it is strained and therefore a mechanical stress is induced on the magnets, forcing them in the RESET state. To generate the electric field it is sufficient to place two electrodes near the piezoelectric substrate

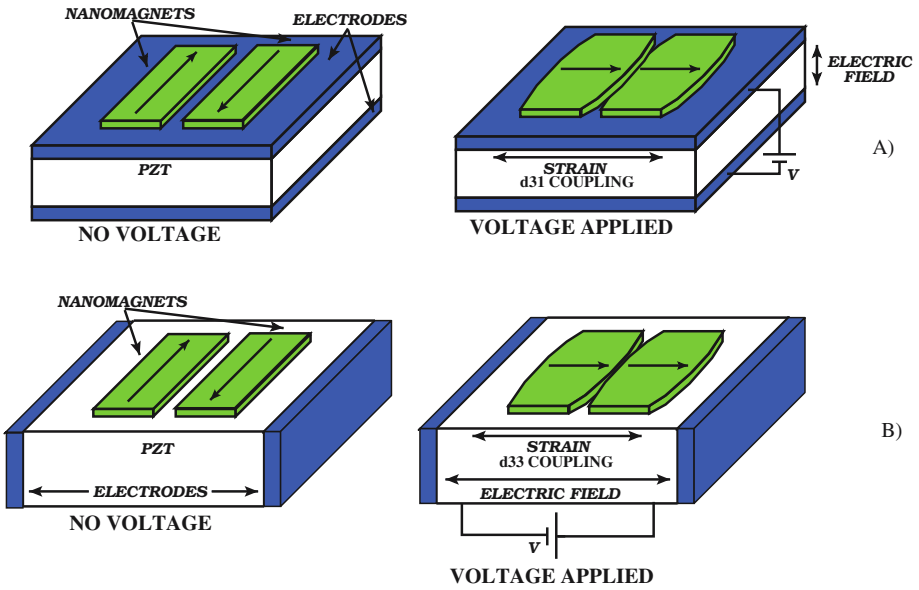


Fig. 6. Magnetoelastic clock system for NML logic. (A) Electrodes are placed on top and on the bottom of the piezoelectric layer. The electric field and the piezoelectric substrate are coupled through the d_{31} coefficient. Moreover electrodes interfere with the mechanical coupling and the magnets fabrication process. (B) Electrodes are placed on the side of the piezoelectric layer. The electric field and the piezoelectric substrate are coupled through the d_{33} coefficient. Electrodes do not interfere with the mechanical coupling and magnets fabrication.

and then to apply a voltage across them. They can be placed on top and on the bottom of the substrate (Fig. 6(A)) or at both sides of the substrate (Fig. 6(B)). The first solution was partially experimentally demonstrated in [34], successfully showing the magnetization control with an applied voltage. However, two problems arise when it must be applied to NML circuits. The generated electric field is perpendicular to the plane, while the strain of the piezoelectric layer is directed along the plane. The consequence is that the electric field and the strain are coupled through the d_{31} coefficient of the piezoelectric material. d coefficients are physical constants of piezoelectric materials that bond the mechanical stress and the applied voltage, as can be seen from Eqs. 1 and 2, that describe the mathematical theory behind the piezoelectricity.

$$\{S\} = \left[s^E\right] \cdot \{T\} + [d] \cdot \{E\} \tag{1}$$

$$\{D\} = \left[d^t\right] \cdot \{T\} + \left[\epsilon^E\right] \cdot \{E\} \tag{2}$$

Equations 1 and 2 are normally called *coupled equations* in the *strain-charge* form. They provide the link between the deformation and the electric field. S is

the strain matrix and D is the electric charge matrix, s is the compliance (the inverse of the stiffness), T is the stress matrix, d is the matrix for the direct piezoelectric effect while d^t is the matrix for the converse piezoelectric effect, finally E is the electric field and ϵ is the permittivity. The matrix of d coefficients therefore links the various components of the applied electric field to the resulting strain. The two most important d coefficients for this kind of applications are the d_{31} and the d_{33}. When the applied electric field is perpendicular to the strain the d_{31} coefficient is used, while, when the electric field and the strain lie in the same plane and in the same direction, the d_{33} coefficient is instead used. d_{31} coefficient is normally smaller than the d_{33} coefficient.

The second problem is related to the electrodes placement. One of the electrodes is placed on top of the piezoelectric layer, between the piezoelectric material and the magnets, reducing the mechanical coupling between them. Moreover, considering the fabrication of a complex NML circuit, these electrodes must be contacted with wires to distribute the clock voltage, but this is difficult considering that electrodes are covered by magnets. In the second solution (Figure 6(B)) electrodes are placed at both sides of the piezoelectric layer, so the electric field and the strain lie along the same plane and are coupled through the d_{33} coefficient, which is higher than the d_{31} as can be seen from Table 1. The coupling between piezoelectric layer and magnets is stronger and electrodes can be contacted freely because there is no physical interference caused by magnets. We therefore choose this second structure to build our magnetoelastic clock system for NML technology.

Table 1. Comparison between piezoelectric thin films. d_{31} and d_{33} are the two main piezoelectric coefficients. Ef_{MAX} is dielectric strength and ϵ_r is the dielectric constant.

	d_{31} (pm/V)	d_{33} (pm/V)	Ef_{MAX} (MV/m)	ϵ_r
PZT	$-30 \div -80$	50÷150	>50	300÷1300
PVDF	23	−33	5	12
ZnO	0.26	5.9	25÷40	10.9
BT	−33	82	2	1250÷10000

Table 1 shows a comparison between some of the piezoelectric material most commonly used. d_{31} and d_{33} coefficients bond the mechanical strain with the applied voltage, and as can be clearly seen, d_{33} is normally higher than d_{31}. Ef_{MAX} is the maximum electric field that the material can tolerate while ϵ_r is the relative dielectric constant. PZT (Lead-Zirconate-Titanate) is one of the most commonly employed material and, as can be seen from Table 1, it has also the best performance. The other materials shown in Table 1 are PVDF, which is a polymer, Zinc Oxide (ZnO) and Barium-Titanate (BT), but they show worst performance than PZT, so in this work we use PZT as our reference choice for the piezoelectric substrate.

2.1 Magnetic Material Choice

While there is little room for different choices of piezoelectric material, the selection of the magnetic material to be used is more complex. The reason behind this fact lies in the limited amount of mechanical stress that can be applied to the magnets. In current-based clock systems there is virtually no limit to the amount of current that can be supplied to the circuit. There is indeed a limit to the value of current that can be used but it is very high, as shown in [11] where a current of 545 mA is used to generate the magnetic field. In this clock system, instead, the maximum stress that can be applied to the magnets is quite limited. The reason is simple: If the mechanical stress is to high there is a mechanical damage on the structure. To evaluate the maximum stress that can be used, two contributions must be considered, the contribution of the piezoelectric substrate and the contribution of the magnets.

Since the piezoelectric material is an insulator, there is a maximum value of electric field that can be applied without breaking it. As a consequence the maximum strain that it is possible to have is given by Eq. 3.

$$\xi_{MAX\_RIG} = Ef_{MAX} \cdot d \tag{3}$$

Ef_{MAX} is the maximum electric field tolerated by the piezoelectric substrate, while $d = d_{33}$ is the longitudinal piezoelectric coefficient that relates the strain induced with the applied voltage. As can be seen from Table 1, PZT is one of the material that can tolerate the highest values of electric field, bigger than 50 MV/m. However, structural limitations of the piezoelectric materials must be considered, because, also if the applied electric field is lower than the maximum tolerated, the generated strain can be bigger than the maximum strain due to material structural limitations (ξ_{MAX_STRUCT}). As a consequence, the maximum strain of the piezoelectric material (ξ_{MAX})) is the minimum between these two contributions, as shown in Eq. 4.

$$\xi_{MAX} = min(\xi_{MAX\_RIG}, \xi_{MAX\_STRUCT}) \tag{4}$$

PZT has a very high tolerance to structural deformations, because ξ_{MAX_STRUCT} is equal to $500 \cdot 10^{-6}$ [35], so the lower bound is due to the maximum electric field that can be applied. Once the maximum strain tolerated by the piezoelectric material is evaluated, it can be converted to a stress as shown in Eq. 5:

$$\sigma_{MAX\_PIEZO} = Y_{Magnet} \cdot \xi_{MAX} \tag{5}$$

Multiplying the maximum value of strain (ξ_{MAX}) for the Young modulus (Y_{Magnet}) of the magnetic material gives the maximum stress that can be applied to the magnets (σ_{MAX_PIEZO}). This assumption is valid only if the thickness of the magnetic material is much smaller than the thickness of the piezoelectric substrate, because only if this condition is satisfied the mechanical stress generated by the substrate is completely transferred to the magnets. Table 2 shows the comparison between some magnetic materials, where the Young modulus is in the range of 80–209 GPa.

Table 2. Magnetic materials comparison. Ms is the saturation magnetization, $\lambda_1 00$ and $\lambda_1 11$ are the magnetostrictive coefficients, while Y is the Young modulus and σ_{MAX_STRUCT} is the fracture stress.

	Ms (10^6 A/m)	λ_S ($\cdot 10^{-6}$)	Y (GPa)	σ_{MAX_STRUCT} (MPa)
Cobalt	1.45	50	209	225
Iron	1.71	14	211	540
Nickel	0.49	33	214	100
Terfenol	0.8	600	80	28

To complete the analysis the value of maximum stress previously evaluated (σ_{MAX_PIEZO}) must be compared with the maximum stress that can be tolerated by the magnetic material (σ_{MAX_STRUCT}). The maximum stress that can be applied to the circuit is therefore the minimum between these two values of stress as shown in Eq. 6.

$$\sigma_{MAX} = min(\sigma_{MAX\_STRUCT}, \sigma_{MAX\_PIEZO}) \tag{6}$$

Between these two values, the lower bound is normally given by the maximum structural stress that can be tolerated by the magnetic material (σ_{MAX_STRUCT}). As can be seen from Table 2 this value is in the range of 28–540 MPa.

Once that the maximum stress is evaluated, the next step is to understand the minimum stress that must be applied to the magnets to force them in the RESET state. The minimum stress can be evaluated considering two important properties of magnetic materials: Shape and stress anisotropy. The value of shape anisotropy is related to the sizes and aspect ratio of magnets and can be evaluated as shown in Eq. 7.

$$E_A = \frac{1}{2}\mu_0 N_d M_s^2 V \tag{7}$$

where μ_0 is the magnetic permeability, N_d is the demagnetization factor, M_s is the magnetization saturation and V is the volume. Stress anisotropy is instead related to the applied mechanical stress and can be evaluated as shown in Eq. 8.

$$E_A = \frac{3}{2}\lambda_s \sigma V \tag{8}$$

where λ_s is the magnetostrictive coefficient of the magnetic material, σ is the applied stress and V is the volume. The minimum stress is therefore the value of stress that generates a stress anisotropy at least equal to the shape anisotropy and can be calculated equaling Eqs. 7 and 8, as shown in Eq. 9.

$$\sigma_{MIN} = \frac{\mu_0 N_d M_s^2}{3\lambda_s} \tag{9}$$

Once the minimum and maximum stresses that can be applied are know, it it possible to analyze different magnetic materials and to understand which of them is suited for this kind of application. The analysis is performed using Octave,

evaluating how the minimum stress changes with magnet sizes. Since in NML circuits magnets are either $50 \times 100\,\text{nm}^2$ or $60 \times 90\,\text{nm}^2$, the shorter magnet side is chosen equal to 50 nm, and the longest magnet side is changed from 50 nm to 100 nm, changing therefore magnets aspect ratio (from 1 to 2). Magnets thickness is chosen equal to 10 nm, to obtain the maximum mechanical coupling between the substrate and the magnets. Minimum and maximum stresses are therefore evaluated for these values of magnet sizes. Figure 7 shows the results of the analysis for two common magnetic materials, Iron and Cobalt.

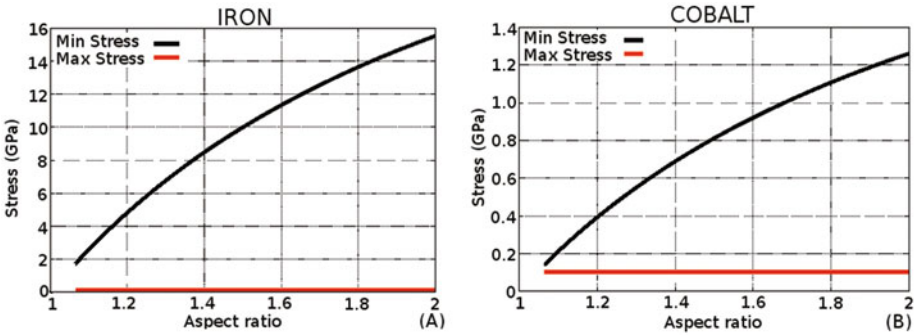


Fig. 7. Minimum and maximum stress evaluation changing magnets sizes for (A) Iron and (B) Cobalt. The maximum stress is always bigger than the minimum required stress, so these two materials cannot be used for this particular application.

To assure the magnets stability in presence of thermal noise the value of shape anisotropy must be at least equal to $30K_bT$, as shown in Eq. 10, generating therefore a lower bound on the value of aspect ratio:

$$\frac{1}{2}\mu_0 N_d M_s^2 V = 30K_bT \tag{10}$$

The minimum value of aspect ratio that assures thermal stability is equal to 1.06, that means magnets must be at least $50 \times 53\,\text{nm}^2$. However from Fig. 7 it is possible to see that either Iron (Fig. 7(A)) or Cobalt (Fig. 7(B)) cannot be used for this application. The value of minimum stress is always bigger than the maximum applicable stress.

Things are different if high magnetostrictive materials are considered. Figure 8 shows the analysis results applied to nickel and terfenol. Terfenol is an alloy of Terbium, Disprosium and Iron. As can be seen Nickel shows a sizes range where the minimum stress is effectively lower than the maximum applicable stress. The maximum aspect ratio is 1.28, that is equal to magnets of $50 \times 64\,\text{nm}^2$. Terfenol shows better performance, since the maximum aspect ratio is 1.57 that means magnets of $50 \times 78\,\text{nm}^2$. It appears that Terfenol is the best choice for this application.

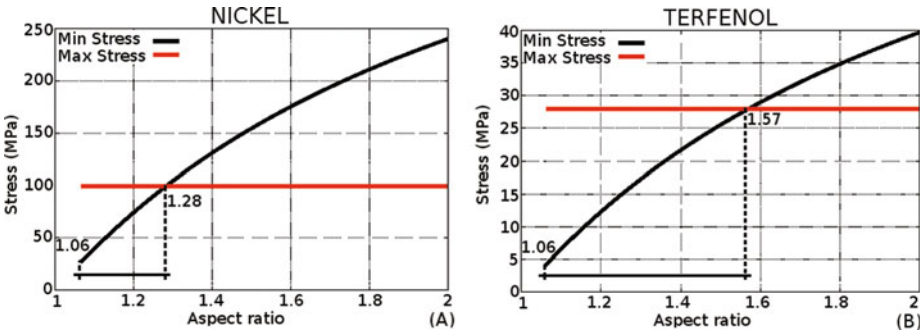


Fig. 8. Minimum and maximum stress evaluation changing magnets sizes for (A) Nickel and (B) Terfenol. As can be seen there is a range where the minimum stress is effectively lower than the maximum stress, but the range for terfenol is bigger.

2.2 Process Variations

To better understand which of the two proposed materials, Nickel and Terfenol, is the best candidate for this application, the impact of process variations must be evaluated as well. The same analysis done before was repeated considering a random variation on the shorter magnets side, previously considered fixed at 50 nm. Figure 9 shows the results. With a process variation of ±5 % nickel still shows a good working range (Fig. 9(A)) but with a variation of ±10 % there is just one value of aspect ratio where magnets can be successfully reset (Fig. 9(B)). Clearly with bigger variations, like a ±20 % variation, there is no working range. Better results are obtained for terfenol because with a variation of ±5 % (Fig. 9(D)) and ±10 % (Fig. 9(E)) it still shows a good working range. Also considering a variation of ±20 % (Fig. 9(F)) there is still a value of aspect ratio in which magnets can be successfully reset.

The consequences are blatant, Terfenol can tolerate higher process variations than Nickel, making it the best candidate for this application, and it is therefore used as a reference for the results presented in this chapter. The chosen magnet sizes are $50 \times 65 \times 10\,\text{nm}^3$, with an height of 65 nm, correspondent to an aspect ratio of 1.3, which is exactly in the middle of terfenol working range. These sizes give therefore the best tolerance to process variations.

2.3 Building Circuits

To design a generic NML circuit the clock zones layout must comply with two important requisites:

- It should allow signals propagation in every direction.
- It should take into account the physical constraints of the clock generation network, i.e. the fabrication and physical placement of the wires used to route currents or voltages required to implement the multiphase clock system used.

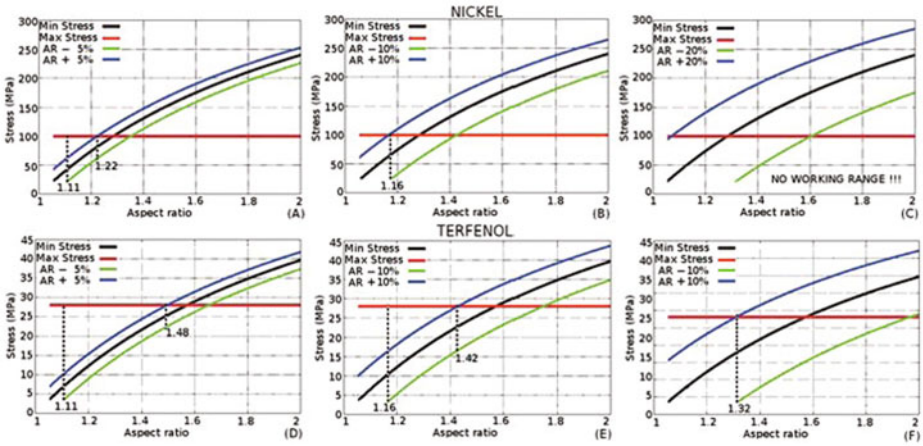


Fig. 9. Process variation evaluation. The analysis is extended considering process variations on the shorter magnets side, for nickel of ±5 % (A), ±10 % (B), ±20 % (C), then for terfenol of ±5 % (D), ±10 % (E) and ±20 % (F).

To comply to these rules the NML circuit design chosen for this particular clock system is based on NAND/NOR gates, since they are universal gates that can implement any kind of logic function. Figure 10 shows the layout of NAND gate. A NOR gate can be simply obtained changing the shape of the central magnet [17].

Every clock zone is made by a mechanical isolated island of piezoelectric material. Magnets are deposited directly on top of the piezoelectric layer to maximize the mechanical coupling. A NAND gate can be built coupling a AND gate [17] with an inverter, which is simply built by an horizontal wire with an odd number of magnets. Helper blocks [36] are used to help the signal propagation in presence of vertical NML wires. Every gate has two inputs and can have up to two outputs as outlined in Fig. 10.

Theoretically, the electrodes should be placed at both sides of the piezoelectric island, as shown in Fig. 6(B). However, this makes the fabrication process more complex. A first possible solution is to use electrodes placed at both sides of the piezoelectric island but buried under it (Fig. 11(A)). The fabrication of electrodes under the piezoelectric layer, however, is complex. The processes involved in the fabrication of piezoelectric materials works at very high temperature and they can therefore damage the electrodes, since the PZT is created on top of them. Moreover, if Copper is used to fabricate the electrodes, an interface layer is required to stick together electrodes and PZT. Figure 11(B) shows a simulation of the electric field distribution, obtained through Comsol Multiphysics [37], of the structure with buried electrodes. The structure is 350 nm width, the PZT is 100 nm thick and electrodes are $50 \times 50\,\text{nm}^2$. As can be seen the electric field is quite uniform between the two electrodes (3–4 MV/m with an applied voltage of 1 V) but it is lower near the area correspondent to the electrodes and

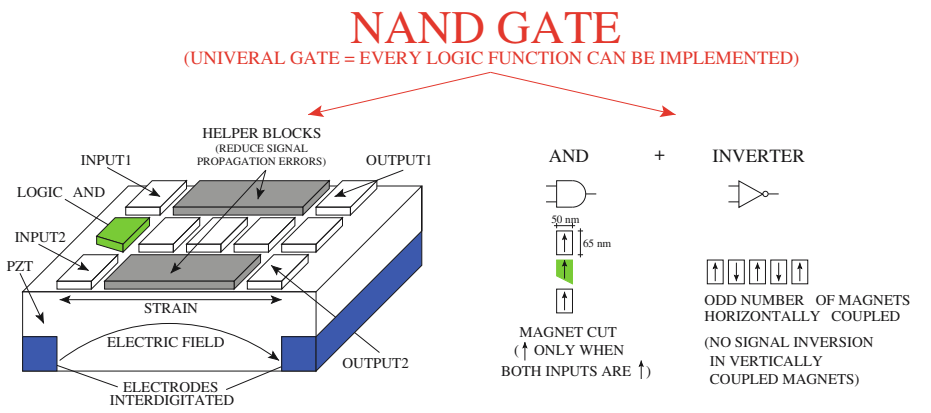


Fig. 10. NAND gate layout. A NAND can be obtained coupling a AND gate [17] with an inverter, which is simply made by an odd number of magnets horizontally aligned.

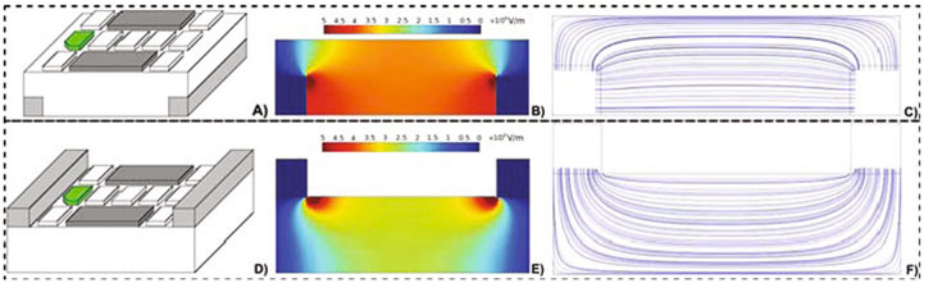


Fig. 11. Electrodes placement. Cell is 350 nm width and the piezoelectric layer is 100 nm thick, while the electrodes section is $50 \times 50\,\text{nm}^2$ (A) Electrodes are buried under the piezoelectric layer. (B) The electric field distribution is quite uniform between the electrodes (3–4 MV/m with an applied voltage of 1 V) but is lower near the area correspondent to the electrodes. The strain is proportional to the electric field value. (C) The electric field lines are quite uniform. (D) Electrodes placed on top of the piezoelectric island. (E) The distribution is still uniform but the electric field this time is higher near the two electrodes. (F) Also with this structure the electric field lines have a good uniformity.

it it 0 directly above them. The strain of the material depends on the electric field value, so it will be higher in the area between the electrodes and lower above them. Thanks to the mechanical continuity of the island, there will be a strain also where the electric field is 0, because, intuitively, the strain of the central part of the island will induce a strain also near the border of the island. Electrodes are not exactly placed at the island sides, however, as can be seen from Fig. 11(C), the electric field lines distribution is quite good between the two electrodes, with lines almost parallel to the surface.

A second solution is to deposit the electrodes on top of the piezoelectric island, as shown in Fig. 11(D). This solution is technology-friendly, because the PZT is fabricated before the electrodes, so they are not damaged by the high temperature processes involved in the PZT fabrication. Moreover, no interface layer is needed between the electrodes and the substrate. Considering also the global circuit layout, electrodes placed on top of the substrate are easier to contact. Interconnection wires used to connect the electrodes can be fabricated using additional layers, in the same ways as CMOS interconnections. This solution is therefore compatible with CMOS fabrication processes. Magnets are deposited on the islands after the deposition of the electrodes. This solution shows also a better electric field distribution (Fig. 11(E)) than the buried electrodes one. The electric field is still uniform between the electrodes, moreover near the borders its intensity increases. Figure 11(F) shows the electric field lines, similarly to the previous case they are quite uniform, especially near the surface which is the most important part. Since the use of electrodes on top of the island shows a better electric field distribution and is compatible with CMOS fabrication processes, this is the solution adopted hereinafter for this particular clock system.

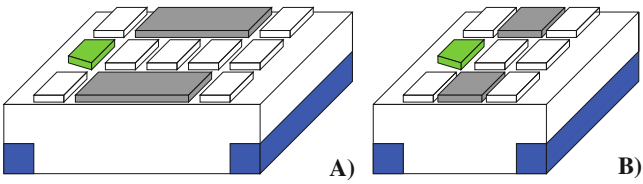


Fig. 12. Two possible sizes for the gates. (A) Gates can be 5 magnets width, simplifying the fabrication processes because the structure is bigger. Performance are however worse because there are more magnets in the critical path. (B) The minimum gate width is equal to 3 magnets to grant a proper signal propagation. Performance are better but higher resolution lithography processes are required.

The electrodes placement is the reason behind the choice of using AND/OR gates as basic logic gates, instead of using majority voters. Majority voters are three input gates (see Fig. 2(C)) where two inputs come from up and down directions and the central one comes from left direction. The use of a majority voter with this clock system is therefore not possible, because at the left of the gate there is one of the electrodes. However AND/OR gates has only two inputs coming from up and down directions, so there is no interference with the electrodes. Moreover they are smaller than a majority voter reducing the circuit area.

The minimum height of the gate is equal to 3 magnets, whilst the width can be equal to 5 (Fig. 12(A)) or 3 magnets (Fig. 12(B)). With 5 magnets the whole structure is bigger, therefore it is simpler to fabricate. Performance are worse because there are more magnets on the critical path and therefore the clock frequency is lower. Furthermore the structure is bigger, so the clock losses

are bigger. See Sect. 2.4 for further details on performance. With a width of just 3 magnets performance are improved but fabrication processes with higher resolution are required. Smaller widths cannot be reached because the required resolution cannot be achieved with current technological processes. Bigger values of width are also not possible because bigger structures have far too many magnets in the critical path, generating therefore errors in the signal propagation.

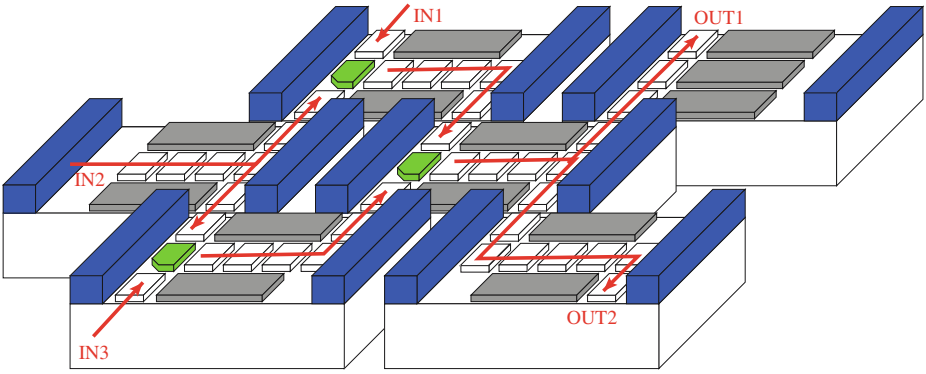


Fig. 13. Complex circuit layout example using the magnetoelastic clock.

Figure 13 shows a possible layout of a NML circuit based on this clock solution. Every piezoelectric island is equal to a clock zone. Particularly, the layout is made by interleaved lines of clock zones to correctly match the outputs coming from a clock zone with inputs of neighbor clock zones. Due to the presence of electrodes, signals propagation follows the north and south border of every clock zone. A clock zone then can be occupied by a logic gate or simply by a wire. Finally, a proper multiphase clock scheme can be applied to the layout to obtained a correct signals propagation.

The layout here proposed is based on mechanical isolated islands. To obtain islands of PZT starting from a uniform substrate, it can be patterned through lithography, selectively removing part of the piezoelectric material in a specific area [38,39], as shown in Fig. 14. PZT can be completely removed to obtain perfect isolation, however also removing just a small part of it provides a very good mechanical separation, so that the strain of an island does not produce a strain in neighbor cells. This is an important requisite in every clock system: When magnets of a specific zones are clocked there should be no crosstalk with their neighbor clock zones. Clearly the higher the resolution of the lithography process used is, the smallest the area that can be removed is. Theoretically, a gap of few nanometers (providing to have a process with high enough resolution) is sufficient, since the strain of every island is smaller than 1 nm.

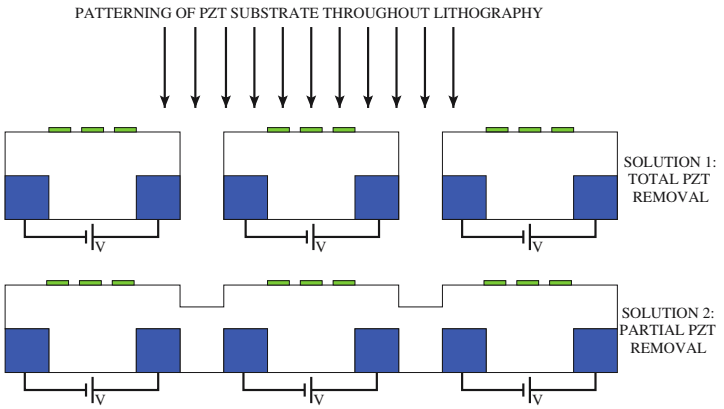


Fig. 14. Creation of mechanical isolated islands starting from a uniform piezoelectric layer. The substrate can be patterned with lithography to selectively remove part of it. The PZT can be removed partially or totally, in both cases a good mechanical isolation is obtained.

2.4 Performance: Timing and Power Estimation

Recently a dedicated simulator for NML technology, ToPoliNano, was developed [40,41]. While this tool represents a powerful instrument to automatically generate [42] and simulate NML circuits [43], it does not support low level physical simulations. To evaluate timing performance of NML circuit it is instead possible to use low level physical simulators like OOMMF [44], NMAG [45] or Magpar [46]. These finite element simulators are based on the Landau-Lifshitz-Gilbert (LLG) equation which describes the dynamic behavior of a generic magnetic structure. Among these simulators Magpar allows also the evaluation of an applied stress on the circuit dynamic. As a consequence, to understand the timing behavior of the proposed clock system, a NML wire was simulated with Magpar. Magnet sizes and the value of applied stress are taken from the theoretical analysis described in Sect. 2.1.

Figure 15 shows the simulation results obtained. The reset time, the times required for a magnet to rotate its magnetization vector from one stable state to the RESET state, is around 1 ns (Fig. 15(A)). The switching time, the time required for a magnet to go from the RESET state to one of the stable states, is slightly smaller but also in this case it is near 1 ns. This numbers can be used to estimate the circuits clock frequency. It can be said that in the best case the clock period must be at least equal to (Eq. 11):

$$T_{ck} = T_{RESET} + N * T_{SWITCH} \tag{11}$$

T_{RESET} is the reset time, T_{SWITCH} is the switching time and N represents the number of magnets in the critical path. The critical path is the total number of magnets in a clock zone between an input to the output. In case of a NAND gate with a width of 3 magnets, N is equal to 5, therefore the correspondent clock

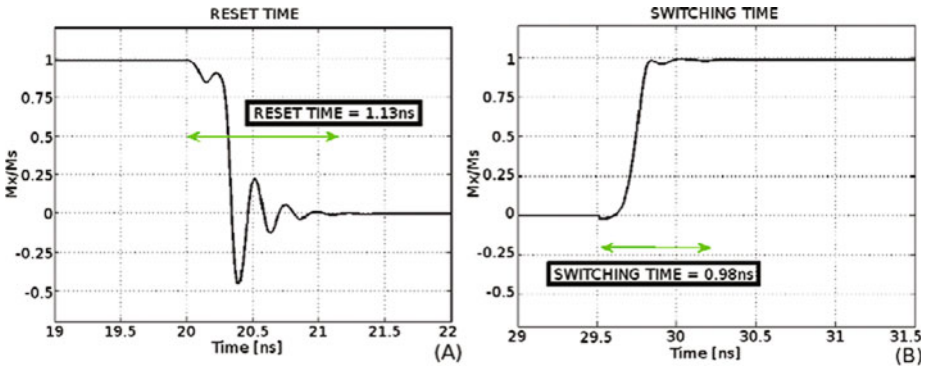


Fig. 15. Magpar simulation of a NML wire. (A) Time required to successfully reset a magnet. (B) Switching time, from RESET state to one of the two stable states.

frequency is 160 MHz. In case of a NAND gate with a width of 5 magnets, N is equal to 7, so the clock frequency is 125 MHz. Nonetheless, the evaluation of the real clock frequency is more complex as different factors must be considered. For example, the total switching time is not directly the sum of the switching time of every magnet, because one magnet starts to switch before its neighbor has completed the switching. The clock period is therefore lower and the frequency higher. As a consequence the upper bound for the clock frequency can be estimated in 200 MHz for gates with a width of 3 magnets and 150 MHz for gates with a width of 5 magnets. Anyway this is just an upper bound as the real clock frequency must be evaluated according to the multiphase clock scheme selected, and it is consequently partially independent from the clock technology itself. The clock frequency changes accordingly to the clock scheme selected, for example with a 3 phase clock [47,48] the frequency will be lower than a 2 phase clock [17]. For this reasons no further details are given on the clock frequency value because out of the scope of this chapter.

Albeit timing is a crucial point, the most important parameter that must be evaluated is the power consumption. Two contributions must be considered: The intrinsic energy consumption required to force magnets in the reset state, and the looses in the clock generation network. The intrinsic energy consumption is equivalent to the height of the energy barrier between stable and reset states. If an adiabatic switching is used, that means a very low rise time for the clock signal, this value of energy consumption is reduced to $30K_bT$. In case, on the contrary, an abrupt switching is used, that means a rise time for the clock signal as short as possible, this value of energy consumption is equal to the entire energy barrier. With magnetic field-based NML circuits, the value of energy barrier can be very high (thousands of times K_bT [16]) so there is a big difference between adiabatic and abrupt switching. In this case with magnets of $50 \times 65 \times 10\,\text{nm}^3$ the energy barrier is just about $180K_bT$, so an abrupt switching is used, because it allows to maximize performance.

This can be done because the intrinsic energy consumption is very small compared to the second source of losses, i.e. the energy loss in the clock generation network. The biggest contribution to this term of energy consumption is the energy required to charge the capacitor. Since piezoelectric materials are insulators, every NAND/NOR gate is equivalent to a capacitor (an insulator embraced by two electrodes). The capacitance can be estimated as shown in Eq. 12.

$$C = \frac{\epsilon_0 \cdot \epsilon_r \cdot t_{PZT} \cdot h_{NAND}}{w_{NAND}} \tag{12}$$

where ϵ_0 is the absolute dielectric constant, ϵ_r is the relative dielectric constant of the piezoelectric material, t_{PZT} is the thickness of the piezoelectric substrate while h_{NAND} and w_{NAND} are the NAND/NOR gate height and width. The voltage that must be applied to every gate can be calculated as in Eq. 13.

$$V = \frac{w_{NAND} \cdot \sigma}{Y \cdot d_{33}} \tag{13}$$

where σ is the applied stress, Y is the Young modulus of the magnetic material and d_{33} is the coefficient that couples the strain and the applied voltage in the piezoelectric substrate. With this gate sizes and materials, the required voltage is normally in the range of 0.7–1.3 V. The energy required to charge the capacitance is then (Eq. 14):

$$E_{clock} = \frac{1}{2} \cdot C \cdot V^2 \tag{14}$$

The bigger it the capacitance, the bigger is the energy consumption. The capacitance depends on the physical size of the NAND/NOR gate, like the height and the width of the island, and in particular it is directly proportional to the piezoelectric layer thickness. Nevertheless, while physical sizes do not depend on the particular piezoelectric material used, this is not true for the relative dielectric constant (ϵ_r). The capacitance depends directly on the value of ϵ_r, so the bigger it is the bigger the energy consumption is. As can be seen from Table 1 the PZT ϵ_r is quite high, so if this material is used power consumption is expected to be higher. This is probably the only real flaw of PZT used for this application. Finally, the value of voltage to be applied depends both on the width of the island and on physical properties of magnetic and piezoelectric materials used.

Since the geometrical characteristics of every island are determined by logic constraints (i.e. the number of magnets in the critical path and the structure of AND/OR gates) and materials are chosen as explained in Sect. 2, the only remaining free parameter is the thickness of the piezoelectric layer. Figure 16 indicates the total energy required to switch a NAND/NOR gate, showing the variation with the PZT thickness. Figure 16(a) highlights the difference between the energy consumption and the clock losses in a 3×3 NAND/NOR gate with magnets made of Terfenol. The intrinsic energy consumption is constant with different values of PZT thickness, because it depends only on the magnets characteristics. The clock losses increase instead linearly with the PZT thickness and most important, they are much bigger then the intrinsic energy loss (11000 K_bT

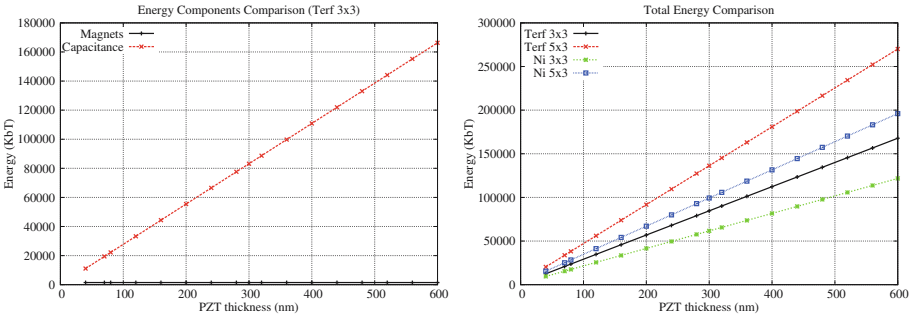


Fig. 16. (a) Comparison between energy consumption components for a 3×3 NAND/NOR with magnet of Terfenol. Energy required to reset the magnets is constant and much lower than energy lost to charge the capacitor. (b) Comparison between NAND/NOR with different sizes and different materials. Nickel has an overall lower energy consumption due to a higher Young modulus.

against 1400 Kb_T with a PZT thickness of 40 nm). To be fair also the losses due to Joule effect should be considered, however, since piezoelectric materials are insulators, the value of current is so small that the resulting losses are 7–8 orders of magnitude smaller than the energy required to charge the capacitance, that is the only really important source of energy loss. Moreover, the distance between electrodes is quite big so no relevant tunnel current can be expected.

Figure 16(b) shows instead a comparison between the total energy consumption with gates of different sizes, 3×3 and 3×5 (that means an height of 3 magnets and a width of 5 magnets), and considering both terfenol and nickel as magnetic materials. As can be clearly observed gates with a width of 5 magnets have a bigger energy consumption than gates with a width of 3 magnets. This is expected because with a bigger width the capacitance is smaller while the voltage is bigger. From Eq. 14 it is evident that voltage has a bigger impact compared to the capacitance, so reducing the capacitance and increasing the voltage overall increases the energy consumption. What is partially unexpected is that, changing the magnetic material employed changes the total energy consumption. As can be seen from Fig. 16(b), using Nickel leads to a lower energy consumption. This can be explained looking at Eq. 13. Nickel magnets require a bigger stress, nonetheless their Young modulus is also bigger, so overall the required voltage and the energy consumption are lower. While Nickel shows a reduced power consumption, Terfenol remains the privileged choice for its better tolerance to process variations.

PZT thickness is an important parameter that must be chosen carefully considering both energy consumption requirements and technology limitations. With magnets 10 nm thick the minimum thickness for the PZT substrate is 40 nm. The PZT must be much thicker than the magnets to be sure that the mechanical stress is completely transferred from the substrate to the magnets.

Table 3. NML total power comparison with different clock systems.

		Energy (fJ)
I	Magnetic field	62
II	STT-current	11
III	Multiferroic	0.004
IV	**Magnetoelastic without compensation**	**0.052**
V	**Magnetoelastic with compensation**	**0.006**
VI	CMOS LOP 21 nm	0.110
VII	Adiabatic CMOS LOP 21 nm	0.040

Increasing the thickness of the piezoelectric layer generally improves the material piezoelectric properties and simplifies the fabrication process at the cost of improved energy consumption.

Finally a comparison between different NML clock solutions and CMOS technology is presented in Table 3, comparing the total energy consumption of a NAND gate. In case of magnetic field clock (I in Table 3) an adiabatic switching is considered, therefore the intrinsic energy consumption of each magnet is assumed equal to $30K_bT$. Clock wires are made of copper and have a section of $400 \times 400\,\text{nm}^2$, with a length of 200 nm. The current value used is extrapolated from the experimental data shown in [11] and is assumed equal to 2 mA. With these values the total energy consumption is 62 fJ. For the STT-current clock (II in Table 3) the energy consumption is assumed equal to 1.6 fJ for each magnet [28], for a total of 11 fJ. In case of a pure multiferroic logic (III in Table 3), data shown in [30] assume a power consumption of 0.004 fJ for a NAND gate. With the clock solution here proposed the energy consumption in case of a 3×3 terfenol NAND gate is equal to 0.052 fJ (IV in Table 3). This value is orders of magnitude better than current-based clock systems. A comparison with a pure multiferroic approach is not possible because the data reported in [30] do not take into account the energy loss in the clock generation network. A comparison among the different clock solutions applied to a complex circuit is shown in [49], where a NML circuit for biosequences analysis [50,51] is presented. The advantage of this clock solution in terms of power consumption is plain.

An important fact must be highlighted. From a circuital point of view every mechanically isolated island is equivalent to a capacitor. As a consequence it is possible to set up an appropriate RLC resonant circuit to recover the energy used to charge the capacitance, reducing the NAND gate energy consumption from 0.052 fJ to 0.006 fJ (V in Table 3). Due to the presence of a parasitic resistance it is not possible to completely recover the energy, nonetheless it is still possible to obtain a good energy consumption reduction. This is particularly important if higher thickness values must be used for the PZT substrate to improve the piezoelectric properties. With higher thickness values the energy consumption is bigger, but it can be compensated by recovering it with setting up a RLC resonator. Considering instead a CMOS NAND gate (VI in Table 3) data extrapolated from ITRS roadmap indicates an energy consumption of 0.110 fJ for the

21 nm technology node. If adiabatic techniques to recover the dynamic component of the energy consumption are used in conjunction with CMOS, only 0.040 fJ are required to switch a NAND gate (VII in Table 3), mainly due to the presence of leakage currents. In both cases our solution shows a remarkable advantage over state of the art CMOS transistors and other NML clock systems in terms of energy consumption.

3 Experimental Fabrication

We are currently experimentally validate the proposed clock solution. Our aim is the demonstration of the working principle but also the characterization of the structure in terms of power consumption. To enrich the explanation we also present here a brief discussion on piezoelectric and magnetic materials required by this clock system.

3.1 Piezoelectric Materials

Due to the fundamental role played by piezoelectric material in this clock system, here an in-depth analysis of their state-of-the-art is presented. Moreover an insight on the fabrication techniques commonly employed and the problematics related to the dimensions scaling is given.

With the advent of the micro and nanotechnologies the development of MEMS based on piezoelectric materials has gained growing and growing importance in the world of scientific research and a lot of attention is currently devoted to the study of such materials. The possibility of combining piezoelectric sensors and actuators with silicon-based CMOS technology, and to integrate them into smart structures, allowing the implementation of a new class of devices which integrate sensors and operate autonomously, converting mechanical energy into dielectric displacement [52,53].

Among all functional ceramics, lead zirconate titanate (Pb(Zr1-xTix)O3, PZT) attracted considerable attention thanks to the presence of excellent ferroelectric and piezoelectric properties, becoming one of the most studied materials of the last decade. It consists of a solid solution of lead titanate and lead zirconate, which can be present in different relative quantities. Piezoelectricity and ferroelectricity of PZT are strictly related to the presence of a perovskite crystal structure. The symmetry of the perovskite phase depends on the stoichiometry of PZT, which influences also its electrical, dielectric, piezoelectric and ferroelectric properties. A titanium-rich phase is responsible for the presence of the tetragonal cell while a zirconium-rich composition gives rise to the presence of a rhombohedral phase. It has been shown that there exists a particular value of the ratio Zr/Ti for which the piezoelectric and ferroelectric properties are maximized [53–56].

This condition is known as the morphotrophic phase boundary (MPB). When Zr/Ti is around 53/47 the MPB is present and the dielectric/piezoelectric

response of the material and its ferroelectricity are enhanced [57,58]. Different works show that the piezoelectric constant d33, as well as permittivity and ferroelectricity, can reach their maximum values near the MPB condition [59]. In particular, it is reported that the (100)-oriented rhombohedral phase is expected to have excellent piezoelectric properties [58]. The transition between a centro-symmetrical paraelectric cubic phase and a non-centro-symmetrical ferroelectric tetragonal one is also related to temperature, occurring at a particular value, the so called Curie temperature (TC). For PZT TC is around 300 °C [60]. Piezoceramic bulk materials for the realization of MEMS/NEMS based technology must be used in the form of thin films. Generally, sputtering and sol-gel are the most diffused techniques for the deposition of PZT thin films. Independently of the deposition technique, the selection of the substrate material is quite deleterious for both electrical and piezoelectric properties of PZT. Generally polycrystalline PZT films can be simply grown using non-epitaxial Si substrates, in order to get the monolithic integration of PZT thin films in the Si-based technology. It is also reported that the epitaxial growth on substrates like MgO or SrTiO3 has to be preferred to enhance the piezoelectric response of the material [59].

However, large residual stress, either compressive or tensile, in the order of some MPa can arise, affecting the physical properties of the material itself [59,61,62]. Sufficiently thick and crack-free PZT films have been fabricated and their applications in MEMS devices have been fully demonstrated [63,64]. Ferroelectric PZT thin films are also widely used in the study and fabrication of nonvolatile ferroelectric random access memories (FeRAM) while the strong piezoelectric response of PZT, in accordance with the presence of high electromechanical coupling coefficients, is exploited in the realization of sensors and actuators, especially for acoustic applications, where large piezoelectric coefficients and electromechanical coupling factors are preferred. Especially in the field of ultrasonic imaging, where transducers operating at frequencies comprised between 20 and 50 MHz are required, PZT films have been recently used and transducers operating at 20 MHz have been demonstrated [65,66]. Furthermore, good electrical properties, i.e., a high dielectric constant and a strong dependence of the dielectric constant on the DC bias field, in accordance with ferroelectricity, makes PZT a good candidate for the realization of microwave devices applications, like voltage-controlled oscillators, varactors, delay lines and phase shifters [57].

Even though PZT is the most performing piezoceramic, thanks to its high piezoelectric constant, recently lots of efforts have been devoted to the study and development of innovative, non toxic, lead-free piezoelectric materials. Among them zinc oxide (ZnO), barium titanate (BT) and aluminum nitride (AlN) are by far the most studied. Zinc oxide is a lead-free, pyroelectric, piezoelectric, wide band-gap semiconductor (3.37 eV) with a large exciton binding energy (60 meV) at room temperature and a wurtzite hexagonal structure. It can be synthesized both in the form of dense thin films and different kinds of nanostructures (nanowires, nanobelts, nanorods, nanocombs) by physical vapor deposition techniques like sputtering, pulsed laser deposition, atomic layer deposition,

chemical vapor deposition, hydrothermal growth and sol-gel techniques. It has been demonstrated that ZnO can be successfully used for the realization of photodetectors, light-emitting diodes, and chemical sensors while the presence of piezoelectricity [67] makes ZnO suitable for the realization of surface acoustic wave resonators, nanogenerators and touch-sensitive piezoelectric sensors. However, as semiconductor ZnO has some advantages and a high electrical resistance, high breakdown voltages and consequently suffers from high electrical losses. Some of these problems, together with its poor piezoelectric properties, can be solved by properly doping ZnO. Aluminum nitride (AlN) can be used in place of ZnO, thanks to its high electrical resistance, high breakdown voltage and low dielectric loss [68–70]. Despite its smaller piezoelectric constant, [67] a lower dielectric permittivity respect to the one of PZT makes AlN competitive in the realization of MEMS devices. Similar to PZT, barium titanate belongs to the class of perovskite-based materials. The presence of a non centro-symmetrical tetragonal crystal structure at room temperature allows it to feature piezoelectricity. BT is a high dielectric material with a piezoelectric coefficient higher than those of ZnO and AlN [71]. As in the previous cases, also BT can be synthesized in different ways, both from physical and chemical vapor deposition methods, and by hydrothermal synthesis. The main problem of BT is the low TC, which is around 120 °C but can be increased by doping the material. A lot of work has to be done in order to improve ZnO, AlN and BT electromechanical properties to become comparable with the PZT ones. Despite all the recent developments in the synthesis of lead-free, non toxic functional materials, PZT still remains the best choice for the realization of sensors and actuators for the micro and nanotechnologies.

Fabrication of Bulk and Thin Film Electro-Ceramics. Ferroelectricity and piezoelectricity in ceramic materials were discovered in the early 1940s, after the observation of high dielectric constants in ceramic barium titanate capacitors [72]. From then, intensive studies have been done, especially about two compositional systems: barium titanate (BaTiO3 or BTO) and lead zirconate titanate (Pb[ZrxTi1-x]O3 where 0×1 or PZT) leading to many applications in medical ultrasonic, high deformation actuators, integrated circuits [73–76]. Some applications utilize chemical coprecipitation [77] or hydrothermal [78] techniques; however ferroelectric ceramics are traditionally obtained from powders formulated starting from individual oxides.

In the standard method for the fabrication of piezoelectric bulk ceramics, the key steps are calcination and sintering: during these steps there is a consistent redistribution of atoms in order to minimize the free energy of the system: new phases are created, the number of surfaces decreases and the grain size increases [79]. The fabrication process starts from the preparation of a specific composition in powder from the chosen precursors; then, the desired shape is created, before densification and finishing. The strong research and development activity in fuel cells, sensors, ferroelectric memories and MEMS technologies has stimulated efforts to fabricate electroceramic films of sub-micron thickness range.

The deposition techniques that allow growing thin films of ferroelectric materials can be divided into two main groups: physical vapor deposition (PVD); chemical processes involving chemical vapor deposition (CVD) and chemical solution deposition (CSD). PVD consists simply in removing atoms from a ceramic target and condensing the vapor onto a substrate. Sputtering technique is used for the deposition of ferroelectric thin films (oxides) because it is a relatively simple process, compatible with standard semiconductor processes; it has quite high deposition rates and good thickness uniformity [80]. Along with sputtering, very diffused is pulsed laser deposition (PLD), where short and intense laser pulses evaporate and ablate the surface of the target material. Metallorganic chemical vapor deposition (MOCVD) decomposes the gaseous precursors on a hot substrate to form a film of desired composition. MOCVD is one of the more promising methods for FeRAM applications (Ferroelectric Random Access Memories) because of its high deposition rate, good step coverage and conformal deposition on three-dimensional surfaces, uniformity of film composition and thickness over a large area, and compatibility with the standard semiconductor [80]. Still not so well established is molecular beam epitaxy (MBE). The chemical solution deposition (CSD) or sol-gel technique has been very much studied and employed in the growth of ferroelectric thin films: low cost, uniformity, excellent thickness and compositional control. Conventional method includes, first of all, the preparation of the desired solution, using metal oxide precursors, solvent and chemical additives to stabilize and control the solution. Then, after coating the substrate with the solution, a sequence of thermal steps is necessary to obtain the desired phase. Finally, another chemical method to deposit ferroelectric films on various substrates is the hydrothermal technique, where chemical reactions in aqueous medium are used. Thin films are usually employed in a composite structure [81], where the elastic properties are dominated by the other part of the device: this can be, for example, a silicon cantilever or membrane [32]. The piezoelectric material and the substrate are clamped along one surface and in the in-plane directions the strains are the same. By contrast, in the perpendicular dimension, the film is free to move. As a consequence of this anisotropy all the deformation modes involve more than one piezoelectric coefficient. This is the main difference between bulk and thin film materials. Besides bulk and thin film materials, we would like to point out that a new class of piezoelectrics is being studied in the last years, the so called hybrids, consisting in a polymeric matrix with ceramic micro/nanoparticles embedded in it, allowing to realize flexible sensing skins for smart applications [82].

Scaling Dimensions of Ferroelectrics and PZT. Ferroelectricity originates thanks to the parallel alignment of permanent dipole moments by means of chemical (short-range) and physical (long-range) forces. This cooperative phenomenon arises below a certain phase transition temperature, called the Curie temperature, as a result of the competition between these temperature-dependent interactions. As a consequence a size effect is inevitably involved [83]: a physical dimension reduction of ferroelectric structures leads to a decrease in transition

temperature and spontaneous polarization along with the increase in the coercive field. Changes in properties occur when the film thickness approaches the electrode depletion width or the grain size approaches the ferroelectric domain size [84].

The extent to which ferroelectric materials can be implemented in future generations of microelectronic structures is strictly dependent on the influence of size on the material properties. Several studies have investigated the finite size effects in ferroelectric materials focusing on a critical lateral size or film thickness below which ferroelectricity disappears [85]. The critical thickness of a ferroelectric thin film with perpendicular polarization comes close to about 1 nm [86]. At this value, surface reconstructions need to be considered, because they may ultimately prevent the emergence of ferroelectricity [87,88]. Regarding the minimum lateral dimension of a ferroelectric domain, it is important to underline that for atomic-scale ferroelectric islands the electric field is generally screened by free external charges that obstruct the inter-domains dipolar interactions. Recent advances [89], reveals that this is not an intrinsic behavior of ferroelectric materials, but an effect of the mechanical and electrical boundary conditions resulting from the synthesis methods used.

It seems clear that in many of the existing studies, size effects are controlled by fabrication process rather than intrinsic limits on the stability of the ferroelectric phase. Film thickness, grain size, and lateral size of the ferroelectric material are parameters that can be tuned within a certain range by an appropriate choice of processing condition [84]. Other quantities like the depletion region amplitude at surfaces, electrode-ferroelectric material interfaces, and grain boundaries can be manipulated operating on the chemical synthesis or annealing conditions. As a result, over time, the minimum thickness at which size effects are observed is continuously dropping. The tremendous number of contributions describing the size effects in ferroelectric thin films introduce high scatter in the reported thickness dependence of the high field properties. In particular scaling down PZT-film thickness while retaining good ferroelectricity has been a topic of great interest scientifically and this dimension has dramatically decreased during the past decade [90].

The achievement of size-independent coercive fields in PZT films down to 60 nm in thickness was reported by Bilodeau et al. [91]. The switchable polarization dropped somewhat for films below 100 nm in thickness. Below 60 nm the films were poor electrical insulators. Kim et al. [92] and Oikawa et al. [90] reported the preparation of 70 and 35 nm thick PZT with good ferroelectric properties. Using mean field theory, Li et al. [93] were able to calculate the critical thickness at which perovskite ferroelectric thin films showed a stable polar phase and applied this theory to PbTiO3 thin films. Song and No [94] implemented the same approach to PZT thin films, finding that the critical thickness was approximately 18 nm.

There are also reports that ferroelectricity is stable in films down to much smaller thicknesses. Tybell et al. [95] showed that ferroelectricity can be observed in PZT films as thin as 4 nm [96] and this statement was confirmed through

measurements of the piezoelectric properties with an atomic force microscope. Ferroelectricity in nanometer-thick layers [97] has also found evidence for copolymer ferroelectrics. Regarding the lateral dimension, interesting progress has been done in the achievement of nanometric ferroelectric islands obtained, for example, by ion milling or through self-assembled nanostructures [98]. Nevertheless these dimensions are still relatively large compared with the theoretical values discussed above. A promising route could be the study of 1D nanostructure: ferroelectric BaTiO3 nanorods with diameters as small as 5 nm have been achieved [99].

3.2 High Magnetostriction Materials

Magnetostrictive materials are broadly defined as materials that undergo a change in shape due to change in the magnetization state of the material. In principle, all ferromagnetic materials exhibit a change in shape resulting from magnetization change. In most common materials, nickel, iron, and cobalt, the change in length is on the order of 10 parts per million. This type of magnetostriction has been termed Joule magnetostriction after James P. Joules discovery in the 1850s. The relatively small change in shape of these materials limited their use in engineering.

Following the advances in permanent magnet research, novel magnetostrictive compounds containing iron, terbium and dysprosium have become available in commercial quantities. Rare earth- iron giant magnetostrictive material discovered by Clarke [100] feature magnetostrain that are two order of magnitude higher than nickel ((in the range of 500–1000 ppm), as shown in Table 4. However, to exploit such elastic strains a linear behavior vs applied field is required. Application for highly magnetostrictive materials are miniature magnetostrictive actuators and motors [101], magnetostrictive linear actuators [102], sonar, valve actuation and active vibration control.

More recently an emerging class of magnetostrictive materials iron-gallium alloys (Galfenol) [103] has been discovered that exhibit moderate magnetostriction (350×10^6) at low magnetic fields (8 kA/m) along with low magnetic

Table 4. Comparison between the magnetostriction coefficient of common mgnetic and high magnetostrictive materials.

Material	λ_S (ppm)
Fe	14
Ni	33
Co	50
$Ni_{80}Fe_{20}$	27
$DyFe_2$	650
$TbFe_2$	2630
$Tb_{0.6}Dy_{0.7}Fe_{1.9}$	2400

hysteresis and high tensile strength. The magnetostrictive effect can be used to design sensors for force measurement and energy harvesting devices [104].

It is well known that in low-dimensional systems, magnetic properties are strongly influenced by the reduced symmetry at substrate/film interface. Changes in the magnetization direction as a function of temperature, film thickness, or composition in ultrathin ferromagnetic multilayers consisting of few monolayers has attracted much interest. In this context, magnetoelastic properties of polycrystalline films can be strongly influenced by film thickness, and that in ultrathin films, surface anisotropy and surface magnetostriction terms may dominate over the volume terms. In particular, magnetostriction constant is seen to increase with increasing thickness [105]. Growth conditions may affect the magnetic properties of polycrystalline films due to such factors as film stress, in addition to film thickness. The stress may vary during the film growth depending on the growth model [106]. In our case, films are grown by dc sputtering that is usually known to induce quenched-in stress related to thermally activated process.

3.3 Building a Demonstrator

As our lithography process does not reach at the moment the necessary resolution, we are starting with a more simple structure that will allow us to demonstrate this clock mechanism as a proof of concept. The structure of the demonstrator is shown in Fig. 17(A). Two interdigitated electrodes are fabricated on top of a PZT layer. A NML wire, i.e. a simple chain of magnets, is used as test circuit. Magnets are located in the area between two electrodes arms. The aim of this structure is to demonstrate that, when a voltage is applied to contact pads, magnets are forced in the RESET state. Removing the voltage, magnets should align antiferromagnetically, demonstrating therefore the correctness of this clock solution.

The fabrication process of the demonstrator is shown in Fig. 17 from (B) to (E). First of all the piezoelectric layer is fabricated (Fig. 17(B)) as previously described elsewhere [32]. Briefly, the PZT solution (from E.C. Dept., Jozef Stefan Institute Ljubljana, Slovenia) is spin coated over a thin lead titanate (PT) layer. The PT layer was previously spin coated on the metal-patterned substrate, dehydrated at 200 °C and then annealed at 500 °C. This layer acts as seed layer for the crystallization of the PZT. Four PZT layers (about 70 nm each) were deposited as follows: spin coating, dehydration at 200 °C, pyrolization at 350 °C and finally, the multilayer was annealed at 600 °C. This annealing condition preserved the (110) preferential orientation of PZT. This process can be repeated until the desired thickness is obtained. As a second step the electrodes structure is created on top of the PZT substrate (Fig. 17(C)). A photoresist is deposited on the PZT layer and is successively patterned through direct writing lithography, partially removing the PZT. Metal is then deposited on top of the photoresist and finally the photoresist is removed obtaining the final electrodes structure. A magnetic film is deposited on the existing structure (Fig. 17(D)). Electron Beam Lithography (EBL) or Focused Ion Beam Lithography (FIB) is used to

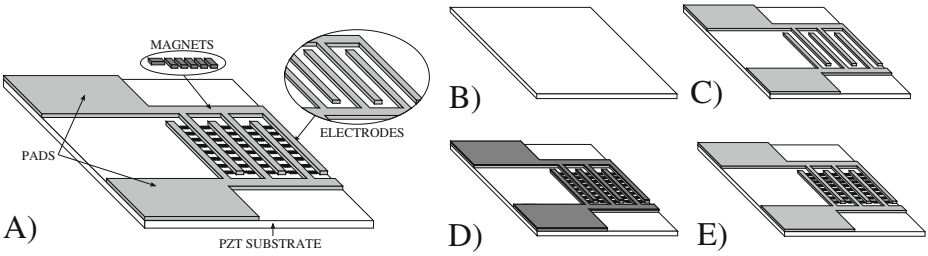


Fig. 17. (A) Structure of the proposed circuit demonstrator. Two interdigitated electrodes are fabricated on a PZT layer. Magnets are located in the area between two electrodes arms. Contact pads are used to apply the voltage to the structure. (B) Fabrication Process. Fabrication of the PZT through sol-gel techniques. (C) Electrodes fabrication through direct laser writing lithography. (D) Deposition of magnetic material through sputtering and patterning of magnets through EBL or FIB lithography. (E) Optional removal of magnetic material from the electrodes structure.

pattern the magnetic material to obtain magnets and to remove the magnetic material in excess to avoid unwanted contacts between electrodes. The magnetic material is normally a conductor so, also if it covers the electrodes, there will be an electric contact. However, magnetic materials are normally bad electrical conductors, so to improve the structure an optional step (Fig. 17(E)) can be used in order to remove the magnetic materials directly above the electrodes. Clearly the area where magnets are located must be properly masked in the process.

Due to technical limitations we have not obtained the complete structure yet, however in Fig. 18 we can illustrate some preliminary results. Figure 18(A) shows a scanning electron microscope (SEM) image of the electrodes structure. It is possible to observe the contact pads and the electrodes. Arms of the two electrodes are alternatively interleaved, in this way there will be a strain between each couple of arms. The maximum resolution that we can currently obtain with our lithography process is 2 μm, so, as it is possible to note from Fig 18(B), electrodes sizes are in the micrometer range. These sizes are much bigger than the desired ones, however, they are enough if the purpose is the demonstration of the effective magnets reset and switching. Figure 18(C) shows instead a SEM image of a typical PZT substrate [33]. It is possible to observe its typical grain structure, with grain sizes in the range of a hundred nanometers. The average roughness is around 3 nm. Figure 18(D) shows instead a film of Iron-Terbium. Iron-Terbium is a material similar to Terfenol with high magnetostriction that we are studying since our studies in [107]. Its surface is pretty rough for the moment and the thickness is quite high: 500 nm. Our efforts are now concentrated on the creation of thinner films with a lower roughness and its deposition on the PZT layer. In the meantime we are using Nickel as magnetic material for our experiments. Normally, an interface layer is used between the PZT and the magnets [34]. This approach reduces the mechanical coupling so we are trying to avoid this choice, depositing the magnetic layer directly on top of the PZT.

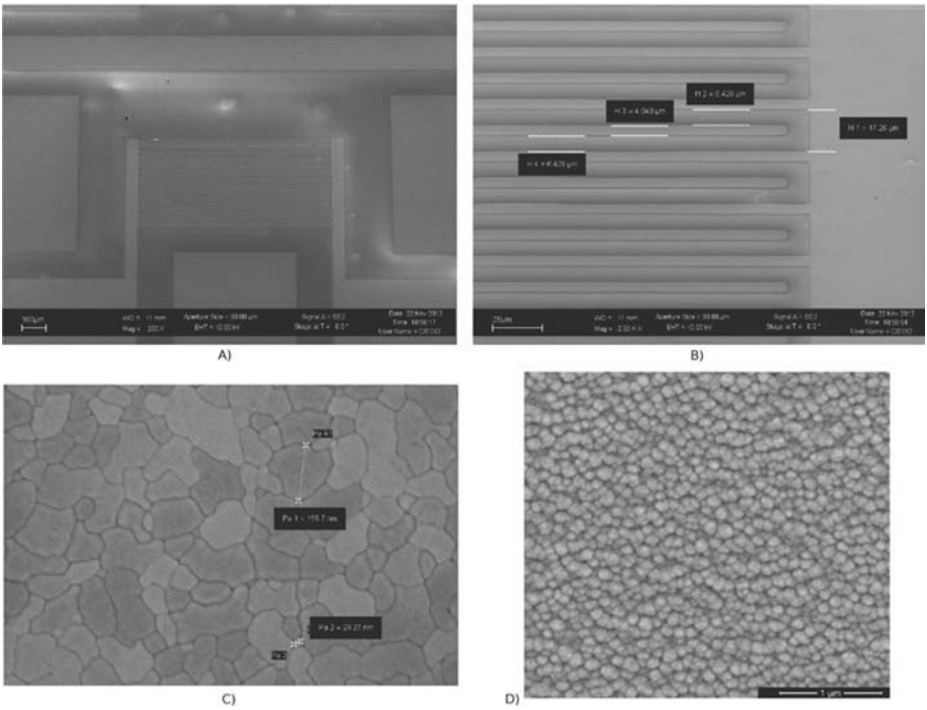


Fig. 18. First experimental results. (A) Interdigitated electrodes structure. Arms are interleaved so that between each couple of them there is a strain in the PZT layer. (B) Detail of the electrodes structure. Sizes are in the range of micrometers because the resolution limit of our lithography process is 2um. (C) SEM image of a PZT layer. Its grain structure is clearly visible. (D) SEM image of a FeTb film. The surface is pretty rough.

Magnets Fabrication. To fabricate magnets we firstly experimented Electron Beam Lithography (EBL). However we were not able to obtain magnets with the required resolution with this technique, due to the small dimensions to achieve and to the vicinity of the desired structures. In the case of small and close packed features, the proximity effect represents one of the hardest limits to overcome. The reason lies probably in this reason, in the relative low acceleration voltage of our FEI Dualbeam Quanta 3D (30 kV) and the high insulating nature of the PZT substrate, which generates many interferences with the electron beam. We therefore switched to Focus Ion Beam (FIB) lithography, obtaining far better results, some of them are shown in 19. Figure 19(A) shows a simple couple of magnets. The smallest obtainable gap among magnets is around 30–50 nm, but the minimum magnet size is around $65 \times 135\,\text{nm}^2$, as shown in Fig. 19(A). Maintaining the magnets sizes between 100 nm and 200 nm the FIB lithography is not critical and at the same time is sufficient for the construction of the demonstrator. Both our theoretical analysis and the results presented in

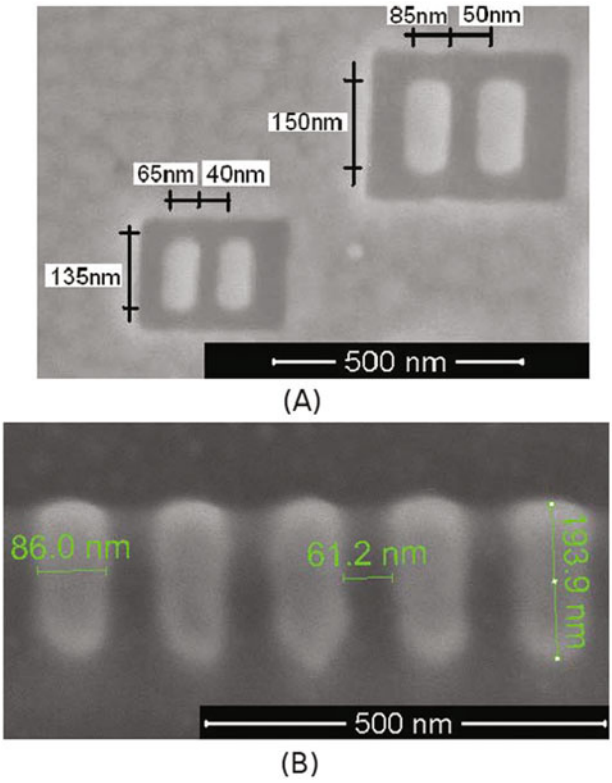


Fig. 19. Magnets fabrication results. (A) Simple couple of magnets with an high aspect ratio. (B) NML wire.

[34] indicate in fact that this clock system can work also with sizes of few hundred of nanometers, but with poorer performance because magnets will be no more single domain entities.

Figure 19(B) shows an example of NML wire, the smallest side is around 86 nm. Unfortunately when we tried to measure this structures with a Magnetic Force Microscope (MFM) we got no magnetic response. We suspect that this is due to multiple reasons, mainly the instrument resolution which is not enough and the influence of the ion beam annealing. The ion bombardment of FIB lithography generates a great amount of heat in an area of about 30 nm (considering an acceleration voltage of 30 kV) around its impact point. The generated heat could be above the Curie temperature so degrading the properties of the magnetic layer, leading to a weak response during the MFM measurements. We are trying to solve this problems in two ways, reducing the FIB current and consequently the damage of the magnetic layer and trying to introduce an intermediate layer between the PZT and the magnetic films. The introduction of an intermediate layer should reduce the interference with the electron beam improving therefore the resolution of the EBL process.

4 Conclusions

The clock solution here proposed represents an important breakthrough for NanoMagnet Logic. It achieves two important results:

- It allows to develop a true low power system, considering also the losses in the clock generation network.
- It can be fabricated with current technological processes because it is designed around the constraints arisen by the fabrication processes themselves.

However, some problems require further investigations. From the technological point of view issues related to the patterning of magnetic materials (for the magnets fabrication) on a piezoelectric substrate must be solved. That means to understand and to solve problems of the currently employed lithographic techniques for magnetic layer patterning. Secondly, it is important to understand how the physical properties of piezoelectric materials scale with dimensions. Probably a suitable compromise between performance and piezoelectric layer thickness must be found.

Moreover, some important questions arise, and they must be investigated both from a theoretical and experimental point of view. Are there any other sources of energy loss that are not considered here (like unwanted leakage currents)? Are there problems of crosstalk between neighbor clock zones? It is know that piezoelectric devices can work at frequency equal to tens of MHz, but what is the real maximum frequency that can be achieved? Since this system is based on mechanical deformations, do the vibrations caused by the environment influence it? Which is the expected reliability of this system?

We are trying to understand these problematics and to solve the related problems. Our goal is to destroy one of the most important barrier for the successful development of this technology, the clock power consumption, in order to finally get the first full-magnetic devices.

Acknowledgements. We thank Nanofacility Piemonte and Compagnia di San Paolo for the support. We would also like to thank χLab laboratory (Materials and Processes for Micro & Nano Technologies - Chivasso) for the technical support in PZT deposition.

References

1. Lent, C., Tougaw, P., Porod, W., Bernstein, G.: Quantum cellular automata. Nanotechnology **4**, 49–57 (1993)
2. Tougaw, P., Lent, C., Porod, W.: Bistable saturation in coupled quantum-dot cells. J. Appl. Phys. **74**, 3558–3566 (1993)
3. Kummamuru, R., Orlov, A., Ramasubramaniam, R., Lent, C., Bernstein, G., Snider, G.: Operation of a Quantum-dot Cellular Automata (QCA) shift register and analysis of errors. IEEE Trans. Electron Devices **50**, 1906 (2003)
4. Smith, C.G., Gardelis, S., Rushforth, A.W., Crook, R., Cooper, J., Ritchie, D.A., Linfield, E.H., Jin, Y., Pepper, M.: Realization of quantum-dot cellular automata using semiconductor quantum dots. Superlattices Microstruct. **34**, 195–203 (2003)

5. Pulimeno, A., Graziano, M., Piccinini, G.: Molecule interaction for QCA computation. IEEE International Conference on Nanotechnology, pp. 1–5 (2012)
6. Pulimeno, A., Graziano, M., Demarchi, D., Piccinini, G.: Towards a molecular QCA wire: simulation of write-in and read-out systems. Solid State Electron. **77**, 101–107 (2012)
7. Pulimeno, A., Graziano, M., Saginario, A., Cauda, V., Demarchi, D., Piccinini, G.: Bis-ferrocene molecular QCA wire: ab-initio simulations of fabrication driven fault tolerance. IEEE Trans. Nanotechnol. **12**(4), 498–507 (2013)
8. Porod, W.: Magnetic Logic Devices Based on Field-Coupled Nanomagnets. Nano & Giga (2007)
9. Imre, A.: Experimental study of nanomagnets for quantum-dot cellular automata(MQCA)logic applications. Ph.D. thesis, University of Notre Dame, Notre Dame, Indiana, December 2005
10. Graziano, M., Vacca, M., Zamboni, M.: Magnetic QCA design: modeling, simulation and circuits. Cellular Automata - Innovative Modelling for Science and Engineering, InTech, (2011) http://www.intechopen.com/articles/show/title/magnetic-qca-design-modeling-s
11. Niemier, M., et al.: Nanomagnet logic: progress toward system-level integration. J. Phys. Condens. Matter **23**, 34 (2011)
12. Karunaratne, D., Bhanja, S.: Study of single layer and multilayer nano-magnetic logic architectures. J. Appl. Phys. **111**, 07A928–07A928-3 (2012)
13. Varga, E., Csaba, G., Bernstein, G., Porod, W.: Implementation of a nanomagnetic full adder circuit. In: 2011 11th IEEE International Conference on Nanotechnology, August 2011
14. Rizos, N., Omar, M., Lugli, P., Csaba, G., Becherer, M., Schmitt-Landsiedel, D.: Clocking schemes for field coupled devices from magnetic multilayers. In: International Workshop on Computational Electronics, Beijin, China, IEEE. pp. 1–4 (2009)
15. Vacca, M., Vighetti, D., Mascarino, M., Amaru, L., Graziano, M., Zamboni, M.: Magnetic QCA majority voter feasibility analysis. In: 2011 7th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME), pp. 229–232 (2011)
16. Vacca, M., Graziano, M., Zamboni, M.: Majority voter full characterization for nanomagnet logic circuits. IEEE Trans. Nanotechnol. **11**(5), 940–947 (2012)
17. Niemier, M., Varga, E., Bernstein, G., Porod, W., Alam, M., Dingler, A., Orlov, A., Hu, X.: Shape engineering for controlled switching with nanomagnet logic. IEEE Trans. Nanotechnol. **11**(2), 220–230 (2012)
18. Niemier, M., Hu, X., Alam, M., Bernstein, G., Porod, W., Putney, M., DeAngelis, J.: Clocking structures and power analysis for nanomagnet-based logic devices. In: International Symposium on Low Power Electronics and Design, Portland-Oregon, USA, IEEE, pp. 26–31 (2007)
19. Graziano, M., Chiolerio, A., Zamboni, M.: A Technology Aware Magnetic QCA NCL-HDL Architecture. Genova, Italy, IEEE, pp. 763–766 (2009)
20. Graziano, M., Vacca, M., Chiolerio, A., Zamboni, M.: A NCL-HDL snake-clock based magnetic QCA architecture. IEEE Trans. Nanotechnol. **10**(5), 1141–1149 (2011)
21. Csaba, G., Porod, W.: Behavior of nanomagnet logic in the presence of thermal noise. In: International Workshop on Computational Electronics, Pisa, Italy, IEEE, pp. 1–4 (2010)
22. Ercan, I., Anderson, N.: Heat Dissipation Bounds for Nanocomputing: Theory and Application to QCA (2011)

23. Graziano, M., Vacca, M., Blua, D., Zamboni, M.: Asynchrony in quantum-dot cellular automata nanocomputation: elixir or poison? IEEE Des. Test Comput. **28**(5), 72–83 (2011)
24. Vacca, M., Graziano, M., Zamboni, M.: Asynchronous solutions for nano-magnetic logic circuits. ACM J. Emerg. Tech. Comp. Syst. **7**(4), 15 (2011)
25. Vacca, M., Graziano, M., Zamboni, M.: Nanomagnetic logic microprocessor: hierarchical power model. IEEE Trans. VLSI Syst. **21**(8), 1410–1420 (2012)
26. Alam, M., Siddiq, M., Bernstein, G., Niemier, M., Porod, W., Hu, X.: On-chip clocking for nanomagnet logic devices. IEEE Trans. Nanotechnol. **9**(3), 348–351 (2010)
27. Ju, X., Niemier, M., Becherer, M., Porod, W., Lugli, P., Csaba, G.: Systolic pattern matching hardware with out-of-plane nanomagnet logic devices. IEEE Trans. Nanotechnol. **12**(3), 399–407 (2013)
28. Das, J., Alam, S., Bhanja, S.: low power magnetic quantum cellular automata realization using magnetic multi-layer structures. J. Emerg. Sel. Topics Circuits Syst. **1**(3), 267–276 (2011)
29. Das, J., Alam, S., Bhanja, S.: Ultra-low power hybrid cmos-magnetic logic architecture. IEEE Trans. Circ. Syst. **59**(9), 2008–2016 (2012)
30. Fashami, M.S., Atulasimha, J., Bandyopadhyay, S.: Magnetization dynamics, throughput and energy dissipation in a universal multiferroic nanomagnetic logic gate with fan-in and fan-out. Nanotechnology **23**(10), 105201 (2012)
31. Atulasimha, J., Bandyopadhyay, S.: Hybrid spintronic/straintronics: A super energy efficient computing scheme based on interacting multiferroic nanomagnets. In: 2012 12th IEEE International Conference on Nanotechnology, August 2012
32. Balma, D., Lamberti, A., Marasso, S., Perrone, D., Quaglio, M., Canavese, G., Bianco, S., Cocuzza, M.: Piezoelectrically actuated MEMS microswitches for high current applications. Microelectron. Eng. **88**(8), 2208–2210 (2011)
33. Chiolerio, A., Quaglio, M., Lamberti, A., Celegato, F., Balma, D., Allia, P.: Magnetoelastic coupling in multilayered ferroelectric/ferromagnetic thin films: a quantitative evaluation. Appl. Surf. Sci. **258**, 8072–8077 (2012)
34. Chung, T., Keller, S., Carman, G.: Electric-field-induced reversible magnetic single-domain evolution in a magnetoelectric thin film. Appl. Phys. Lett. **94**, 132501 (2009)
35. Roy, K., Bandyopadhyay, S., Atulasimha, J.: Switching dynamics of a magnetostrictive single- domain nanomagnet subjected to stress. Phys. Rev. B **83**, 1–15 (2011)
36. Carlton, D., Emley, N., Tuchfeld, E., Bokor, J.: Simulation studies of nanomagnet-based logic architecture. Nanoletters **8**(12), 4173–4178 (2008)
37. Comsol Multiphysics http://www.comsol.com/
38. Guillon, S., Saya, D., Mazenq, L., Nicu, L., Soyer, C., Costecalde, J., Remiens, D.: Lead-zirconate titanate (PZT) nanoscale patterning by ultraviolet-based lithography lift-off technique for nano-electromechanical systems applications. In: 2011 International Symposium on Piezoresponse Force Microscopy and Nanoscale Phenomena in Polar Materials, July 2011
39. Huang, C., Chen, Y., Liang, Y., Wu, T., Chen, H., Chao, W.: Fabrication of nanoscale PtOx /PZT/PtOx capacitors by e-beam lithography and plasma etching with photoresist mask. Electrochem. Solid-State Lett. **9**(3), C51–C53 (2006)
40. Frache, S., Chiabrando, D., Graziano, M., Riente, F., Turvani, G., Zamboni, M.: ToPoliNano: nanoarchitectures design made real. in: IEEE International Symposium on Nanoscale Architectures (NANOARCH), pp. 160–167 (2012)

41. Vacca, M., Frache, S., Graziano, M., Zamboni, M.: ToPoliNano: a synthesis and simulation tool for NML circuits. In: IEEE International Conference on Nanotechnology, pp. 1–6, August 2012
42. Vacca, M., Frache, S., Graziano, M., Crescenzo, L., Cairo, F., Zamboni, M.: Automatic Place&Route of Nano-magnetic Logic circuits. pp. 58–63, July 2013
43. Frache, S., Chiabrando, D., Graziano, M., Graziano, M., Boarino, L., Zamboni, M.: Enabling design and simulation of massive parallel nanoarchitectures. J. Parallel Distrib. Comput. **74**, 2530–2541 (2014)
44. Donahue, M., Porter, D.: OOMMF User's Guide, Version 1.0. Technical Report Interagency Report NISTIR 6376, National Institute of Standards and Technology, Gaithersburg, September 1999
45. Fischbacher, T., Franchin, M., Bordignon, G., Fangohr, H.: A systematic approach to multiphysics extensions of finite-element-based micromagnetic simulations: nmag. IEEE Trans. Magn. **43**(6), 2896 (2007)
46. Scholz, W., Fidler, J., Schrefl, T., Suess, D., Dittrich, R., Forster, H., Tsiantos, V.: Scalable parallel micromagnetic solvers for magnetic nanostructures. Comp. Mat. Sci. **28**, 366–383 (2003)
47. Awais, M., Vacca, M., Graziano, M., Masera, G.: FFT Implementation using QCA. In: 2012 19th IEEE International Conference on Electronics, Circuits and Systems (ICECS), pp. 741–744 (2012)
48. Awais, M., Vacca, M., Graziano, M., Masera, G.: Quantum dot cellular automata check node implementation for LDPC decoders. IEEE Trans. Nanotechnol. **12**(3), 368–377 (2013)
49. Wang, J., Vacca, M., Graziano, M., Zamboni, M.: Biosequences analysis on nanomagnet logic. In: International Conference on IC Design and Technology, pp. 131–134, May 2013
50. Urgese, G., Graziano, M., Vacca, M., Awais, M., Frache, S., Zamboni, M.: Protein alignment HW/SW optimizations. In: The IEEE International Conference on Electronics, Circuits, and Systems (ICECS), pp. 145–148 (2012)
51. Graziano, M., Frache, S., Zamboni, M.: A hardware viewpoint on biosequence analysis: what's next? ACM J. Emerg. Technol. Comput. Syst. **9**(4), 29 (2013)
52. Chidambaram, N., Mazzalai, A., Muralt, P.: Comparison of lead zirconate titanate (PZT) thin films for MEMS energy harvester with interdigitated and parallel plate electrodes. IEEE Trans. Ultrason. Ferr. **59**(8), 1624–1631 (2012)
53. Ortner, K., Koeßler, D., Jung, T., Jacobsen, H., Quenzer, H.: Influence of bias voltage on the structure of lead zirconate titanate piezoelectric films prepared by gas flow sputtering. Plasma Process. Polym. **4**(1), S134–S138 (2004)
54. Park, C.S., Lee, J.W., Park, G.T., Kim, H.E.: Microstructural evolution and piezoelectric properties of thick Pb(Zr, Ti)O3 films deposited by the multi-sputtering method: Part II. Piezoelectric properties. J. Mater. Res. **22**(5), 1373–1377 (2007)
55. Du, X.H., Zheng, J., Belegundu, U., Uchino, K.: Crystal orientation dependence of piezoelectric properties of lead zirconate titanate near the morphotropic phase boundary. Appl. Phys. Lett. **72**(19), 2421–2423 (1998)
56. Kakimoto, K.I., Kakemoto, H., Fujita, S., Masuda, Y.: Control of crystal orientation and piezoelectric response of lead zirconate titanate thin films near the morphotropic phase boundary. J. Am. Ceram. Soc. **85**(4), 1019–1021 (2002)
57. Wu, J., Xiao, D., Wang, Y., Zhu, J., Zhu, J., Xie, R.: Crystal orientation dependence of piezoelectric properties of lead zirconate titanate near the morphotropic phase boundary. J. Am. Ceram. Soc. **91**(11), 3786–3788 (2008)

58. Park, C.S., Kim, S.W., Park, G.T., Choi, J.J., Kim, H.E.: Orientation control of lead zirconate titanate film by combination of sol-gel and sputtering deposition. J. Mater. Res. **20**(1), 243–246 (2005)
59. Kanno, I., Kotera, H., Wasa, K., Matsunaga, T., Kamada, T., Takayama, R.: Crystallographic characterization of epitaxial Pb(Zr, Ti)O3 films with different Zr/Ti ratio grown by radio-frequency-magnetron sputtering. J. Appl. Phys. **93**(7), 4091–4096 (2003)
60. Zhang, S., Xia, R., Lebrun, L., Anderson, D., Shrout, T.: Piezoelectric materials for high power, high temperature applications. Mater. Lett. **59**, 3471–3475 (2005)
61. Lee, J., Bosman, G., Green, K., Ladwing, D.: Model and analysis of gate leakage current in ultrathin nitrided oxide MOSFETs. IEEE Trans. Electron Devices **49**(7), 1232–1241 (2002)
62. Lee, J.W., Park, G.T., Park, C.S., Kim, H.E.: Enhanced ferroelectric properties of Pb(Zr, Ti)O3 films by inducing permanent compressive stress. Appl. Phys. Lett. **88**(7), 072908 (2006)
63. Park, C.S., Lee, J.W., Park, G.T., Kim, H.E.: Microstructural evolution and piezoelectric properties of thick Pb(Zr, Ti)O3 films deposited by multi-sputtering method: Part I. Microstructural evolution. J. Mater. Res. **22**(5), 1367–1372 (2007)
64. Xu, F., Trolier-McKinstry, S., Ren, W., Xu, B., Xie, Z.L., Hemker, K.: Domain wall motion and its contribution to the dielectric and piezoelectric properties of lead zirconate titanate films. J. Appl. Phys. **89**(2), 1336–1348 (2001)
65. Abellard, A., Kuscer, D., Holc, J., Levassort, F., Noshchenko, O., Lethiecq, M., Kosec, M.: Fabrication and modeling of piezoelectric transducers for high-frequency medical imaging. In: Proceeding of the 47th International Conference on Microelectronics, Devices and Materials (2011)
66. Hsu, H.S., Benjauthrit, V., Wei, Q., Huang, Y., Zhou, Q., Shung, K.: Silver doped 0.9PMN-PT-0.1PZT composite films for very high frequency ultrasonic transducer applications. Appl. Phys. A **111**(2), 459–463 (2013)
67. Fu, Y., Luo, J., Du, X., Flewitt, A., Li, Y., Markx, G., Walton, A., Milne, W.: Recent developments on ZnO films for acoustic wave based bio-sensing and microfluidic applications: a review. Sens. Actuat. B-Chem. **143**(2), 606–619 (2010)
68. Cleland, A., Pophristic, M., Ferguson, I.: Single-crystal aluminum nitride nanomechanical resonators. Appl. Phys. Lett. **79**(13), 2070–2072 (2001)
69. Dubois, M., Muralt, P.: Properties of aluminum nitride thin films for piezoelectric transducers and microwave filter applications. Appl. Phys. Lett. **74**(20), 3032–3034 (1999)
70. Piazza, G., Stephanou, P., Pisano, A.: Piezoelectric aluminum nitride vibrating contour-mode MEMS resonators. J. Microelectromech. Syst. **15**(6), 1406–1418 (2001)
71. Panda, P.: Review: environmental friendly lead-free piezoelectric materials. J. Mater. Sci. **44**, 5049–5062 (2009)
72. von Hippel, A., Breckenridge, R., Chesley, F., Tisza, L.: High dielectric constant ceramics. Ind. Eng. Chem. **38**(11), 1097–1109 (1946)
73. Muralt, P.: Ferroelectric thin films for micro-sensors and actuators: a review. J. Micromech. Microeng. **10**, 136–146 (2000)
74. Turner, R., Fuierer, P., Newnham, R., Shrout, T.: Materials for high temperature acoustic and vibration sensors: a review. Appl. Acoust. **41**, 299–324 (1994)
75. de Araujo, C., Taylor, G.: Integrated ferroelectrics. Ferroelectr. Taylor & Francis **41**, 299–324 (1991)
76. Pritchard, J., Bowen, C., Lowrie, F.: Multilayer actuators: review. Br. Ceram. Trans. **100**(6), 265–273 (2001)

77. Mazdiyasni, K.: Fine particle perovskite processing. Am. Ceram. Soc. Bull. **63**(4), 591–94 (1984)
78. Dawson, W.: Hydrothermal synthesis of advanced ceramic powders. Am. Ceram. Soc. Bull. **67**(10), 1673–38 (1988)
79. Moulson, A., Herbert, J.: Electroceramics: Materials, Properties, Applications, 2nd edn. Wiley, New York (2003)
80. Izyumskaya, N., Alivov, Y., Cho, S., Morkoç, H., Lee, H., Kang, Y.: Processing, structure, properties, and applications of PZT thin films. Crit. Rev. Solid State Mater. Sci. **32**, 111–202 (2007)
81. Trolier-McKinstry, S., Muralt, P.: Thin film piezoelectrics for MEMS. J. Electroceram. **12**, 7–17 (2004)
82. Chiolerio, A., Lombardi, M., Guerriero, A., Canavese, G., Stassi, S., Gazia, R., Cauda, V., Manfredi, D., Chiodoni, A., Verna, A., Cocuzza, M., Montanaro, L., Pirri, C.: Effect of the fabrication method on the functional properties of BaTiO3 - PVDF nanocomposites. J. Mat. Sci. (2013). doi:10.1007/s10853-013-7500-9
83. Gruverman, A., Kholkin, A.: Nanoscale ferroelectrics: processing, characterization and future trends. Rep. Prog. Phys. **69**, 2443–2474 (2006)
84. Shaw, T., Trolier-McKinstry, S., McIntyre, P.: The properties of ferroelectric films at small dimensions. Annu. Rev. Mater. Sci. **30**, 263–298 (2000)
85. Ishikawa, K., Nomura, T., Okada, N., Takada, K.: Size effect on the phase transition in PbTiO3 fine particles. Jpn. J. Appl. Phys. **35**, 5196 (1996)
86. Scott, J.: Nano-phase ferroelectric arrays for Gbit devices. Ferroelectrics **260**(1), 305 (2001)
87. Munkholm, A., Streiffer, S., Murty, M.R., Eastman, J., Thompson, C., Auciello, O., Thompson, L., Moore, J., Stephenson, G.: Antiferrodistortive reconstruction of the PbTiO3(001) surface. Phys. Rev. Lett. **88**(1), 016101 (2001)
88. Streiffer, S., Eastman, J., Fong, D., Thompson, C., Munkholm, A., Murty, M.R., Auciello, O., Bai, G., Stephenson, G.: Observation of nanoscale 180 stripe domains in ferroelectric PbTiO3 thin films. Phys. Rev. Lett. **89**(6), 067601 (2002)
89. Ahn, C., Rabe, K., Triscone, J.M.: Ferroelectricity at the nanoscale: local polarization in oxide thin films and heterostructures. Science **303**, 488 (2004)
90. Oikawa, T., Morioka, H., Nagai, A., Funakubo, H., Saito, K.: Thickness scaling of polycrystalline Pb(Zr, Ti)O3 films down to 35nm prepared by metallorganic chemical vapor deposition having good ferroelectric properties. Appl. Phys. Lett. **85**(10), 1754 (2004)
91. Bilodeau, S., Johnston, S., Russell, M., Vestyck, D., Buskirk, P.V.: Voltage scaling of ferroelectric thin films deposited by CVD. Integr. Ferroelectr. **26**(1–4), 119–135 (1999)
92. Kim, S.H., Koo, C., Park, D.Y., Lee, D.S., Yeom, J.H., Ha, J., Lim, J., Hwang, C.: Scaling issues of Pb(Zr, Ti)O3 capacitor stack for high density FeRAM devices. J. Korean Phys. Soc. **42**, S1417–S1419 (2003)
93. Li, S., Eastman, J., Vetrone, J., Foster, C., Newnham, R., Eric Cross, L.: Dimension and size effects in ferroelectrics. Jpn. J. Appl. Phys. **36**, 5169 (1997)
94. Song, H., No, K.: Characterization of the property degradation of PZT thin films with thickness. J. Korean Phys. Soc. **58**(4), 809–816 (2011)
95. Tybell, T., Ahn, C., Triscone, J.M.: Ferroelectricity in thin perovskite films. Appl. Phys. Lett. **75**(6), 856–858 (1999)
96. Bune, A., Fridkin, V., Ducharme, S., Blinov, L., Palto, S., Sorokin, A., Yudin, S., Zlatkin, A.: Two-dimensional ferroelectric films. Nature **391**, 874–877 (1998)

97. Tiedke, S., Schmitz, T., Prume, K., Roelofs, A., Schneller, T., Kall, U., Waser, R., Ganpule, C., Nagarajan, V., Stanishevsky, A., Ramesh, R.: Direct hysteresis measurements of single nanosized ferroelectric capacitors contacted with an atomic force microscope. Appl. Phys. Lett. **79**(22), 3678 (2011)
98. Szafraniak, I., Harnagea, C., Scholz, R., Bhattacharyya, S., Hesse, D., Alexe, M.: Ferroelectric epitaxial nanocrystals obtained by a self-patterning method. Appl. Phys. Lett. **83**(11), 2211 (2003)
99. Yun, W., Urban, J., Gu, Q., Park, H.: Ferroelectric properties of individual barium titanate nanowires investigated by scanned probe microscopy. Nano Lett. **2**(5), 447 (2002)
100. Clark, E.: Ferromagnetic materials, vol. 1. North Holland Publishing Co, Amsterdam (1980) (Ed. E.P. Wohlfarth)
101. Wang, W.: A high precision micropositioner based on magnetostriction principle. Rev. Sci. inst. **63**, 249–254 (1992)
102. Fukuda, T.: GMA applications to micromobile robot as microactuator without power supply cables. IEEE Micro-electro-mechanical Systems Proc., pp. 210–215 (1991)
103. Clark, A., Wun-Fogle, M., Restorff, J., Lograsso, T.: Magnetostrictive properties of galfenol alloys under compressive stress. Mater. Trans. **43**(5), 881–886 (2002)
104. Weng, L., Wang, B., Dapino, M., Sun, Y., Wang, L., Cui, B.: Relationships between magnetization and dynamic stress for Galfenol rod alloy and its application in force sensor. J. Appl. Phys. **113**(7), 17A917 (2013)
105. Song, O., Bllentine, C.A., O'Handley, R.: Giant surface magnetostriction in polycrystalline Ni and NiFe films. Appl. Phys. Lett. **64**(19), 2593–2595 (1994)
106. Koch, R., Hu, D., Das, A.: Compressive stress in polycrystalline volmer-weber films. Phys. Rev. Lett. **94**(14), 146101 (2005)
107. Magni, A., Celegato, F., Coisson, M., Olivetti, E., Pasquale, M., Sasso, C.: Magnetization properties of FeTb thin films. IEEE Trans. Magn. **46**(2), 487–490 (2010)

Majority Logic Synthesis Based on Nauty Algorithm

Peng Wang[1], Mohammed Niamat[1], and Srinivasa Vemuru[2(✉)]

[1] Department of Electrical Engineering and Computer Science, University of Toledo, Toledo, OH 43606, USA
peng.wang@rockets.utoledo.edu,
mohammed.niamat@utoledo.edu

[2] Department of Electrical and Computer Engineering and Computer Science, Ohio Northern University, Ada, OH 45810, USA
s-vemuru@onu.edu

Abstract. Quantum-dot Cellular Automata (QCA) based on majority logic is a promising technology for implementation of future integrated circuits. However, the current majority logic synthesis approaches based on three-feasible networks often lead to inefficient QCA circuit implementations. In this work, four-feasible networks are used as the starting point; therefore each node in the network can accommodate one extra variable. Using the Nauty algorithm, 222 standard functions along with their majority gate mapping are identified for Boolean logic functions with up to four-variables. In addition, all redundancies in the synthesized results are eliminated to reduce the size of the QCA circuit implementations. The proposed method leads to an average reduction of 7.94 % in terms of levels of logic and 8.13 % in terms of gates for the 24 Microelectronics Center of North Carolina (MCNC) benchmarks as compared to other methods.

Keywords: Majority logic synthesis · Quantum-dot cellular automata (QCA) · Nauty algorithm

1 Introduction

The feature size of the Complementary Metal Oxide Semiconductor (CMOS) components has been continuously reducing for the past five decades. However, this trend encounters serious difficulties as the fundamental physical limits of CMOS are quickly approached [1]. Fortunately, many new nanoscale devices have been developed, such as Quantum-dot Cellular Automata (QCA) [2–5], Single Electron Tunneling (SET) [6, 7] and Tunneling Phase Logic (TPL) [8]. While a lot of effort has already been put into understanding and developing new nanoscale devices, research efforts on logic synthesis techniques are also needed to fully utilize the potential of these new technologies.

CMOS technology uses NAND, NOR, NOT, and other functions as the basic logic units, whereas these new technologies are based on either majority or minority logic. For example, QCA uses majority logic, SET uses both majority and minority

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 111–132, 2014.
DOI: 10.1007/978-3-662-43722-3_6, © Springer-Verlag Berlin Heidelberg 2014

logic, and TPL uses minority logic in the implementation of digital circuits. Therefore, the traditional logic synthesis methods that target CMOS technology are not optimal in implementation of digital circuits in post-CMOS technologies. Since minority logic is the complementary form of majority logic, QCA technology and its majority logic basis will be used to explain the proposed method.

A majority logic synthesis method proposed in [9, 10] uses the geometric interpretation of Boolean functions. This results in thirteen standard Boolean functions. However, these functions can only solve three-variable problems. Several approaches to deal with Boolean logic functions with any number of variables have been proposed in [11–15]. The main idea of these synthesis methods is to first decompose the Boolean logic network to three-feasible networks (each node in a three-feasible network contains three or fewer inputs), and then convert each node to a corresponding majority expression. However, these methods are not efficient enough to obtain optimal majority expressions as they handle three-feasible functions only.

In this work, four-feasible networks are used as the starting point as they typically lead to simpler majority networks [16]. A complete list of four-variable primitives is developed and the process of obtaining standard functions using Nauty algorithm is introduced. By providing the optimal majority gate mapping to these standard functions, any Boolean function can be converted into majority gates by mapping it to the standard logic functions. Finally, the redundancies present in the overall logic circuit implementation are removed from the implemented majority networks. In Sect. 2, the necessary background information is presented. Section 3 presents the process of obtaining a full set of standard functions using Nauty algorithm. Section 4 introduces the synthesis procedure to convert standard Boolean functions into optimized majority networks. Section 5 shows the experimental results followed by a conclusion in Sect. 6.

2 Background Material

2.1 Quantum Cells

A QCA cell contains four quantum dots that contain two free electrons. These two free electrons can move through the tunnels between the dots. Due to Coulombic repulsion, the electrons occupy diagonal sites. Thus, there are two stable polarization states which can be encoded as logic “0” and logic “1” as shown in Fig. 1(a) and (b) respectively.

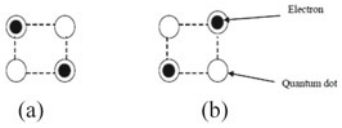


Fig. 1. (a) QCA cell logic “0” (b) QCA cell logic “1”

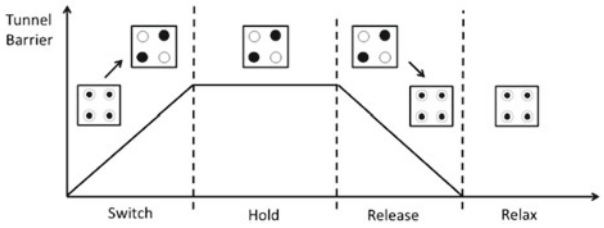


Fig. 2. Four phases of the QCA clock

2.2 QCA Clock

The QCA clock controls the barriers between the quantum dots in QCA cells. It works in a pipeline fashion [17], which is performed in four clock phases: Switch, Hold, Release and Relax, as shown in Fig. 2.

- Switch phase: During the switch phase, the tunnel barrier begins to rise slowly and the polarization of the cell is influenced by its neighbor cells (which are now in their Hold phase).
- Hold phase: The polarization is complete and the barrier reaches its peak. In the meantime, the cells affect their neighboring cells and pass information.
- Release phase: The barrier begins to lower and the cells lose their polarization. Signals carried by these cells should have been passed to their neighbors.
- Relax phase: The tunneling barrier is in the lowest state and the QCA cells will remain unpolarized.

All the four clock phases constitute one clock cycle, in which one bit of information is passed to its neighbor. As the QCA clock continues, the information propagates through the QCA circuit.

2.3 QCA Devices

A QCA wire is a sequence of QCA cells as shown in Fig. 3. The signal applied to the first cell propagates along the wire with the QCA clock.

A QCA majority voter can perform a three-input logic function following the rule that the output signal is always decided by the value of the majority of the inputs. The logic function for a majority voter, shown in Fig. 4, is given by Eq. (1). By forcing one of the three inputs of the majority gate to either logic “0” or “1”, a two-input AND or a two-input OR gate can be realized as shown in Eqs. (2) and (3), respectively.

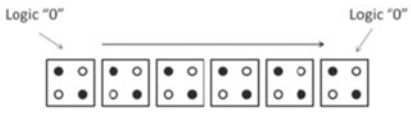


Fig. 3. A QCA wire

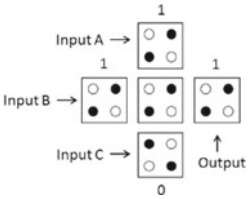


Fig. 4. A QCA majority voter

$$M(A, B, C) = AB + AC + BC \tag{1}$$

$$M(A, B, 0) = AB + B_*0 + A_*0 = AB \tag{2}$$

$$M(A, B, 1) = AB + B_*1 + A_*1 = AB + A + B = A + B \tag{3}$$

In a QCA inverter, shown in Fig. 5, the signal that comes in from the input cell separates into two parallel paths. Cells placed at the output at an orientation of 45° to these parallel paths are used to change the polarity of the incoming signal.

With implementations of AND, OR and NOT logic, any Boolean logic function can be constructed using majority voters and inverters. However, such implementations may not lead to optimal results in terms of the number of logic gates or the number of logic levels.

2.4 Graph and Subgraph

A graph G consists of a set of vertices represented by *V(G)*, a set of edges represented by *E(G)*, and an incidence function Ψ_G that associates each edge to a pair of (not necessarily distinct) vertices. For example, a graph $G = (V(G), E(G))$ can be described by Eq. (4).

$$\left.\begin{aligned} V(G) &= \{u, v, w, x, y, z\} \\ E(G) &= \{a, b, c, d, e, f, g\} \\ \Psi_G(a) &= \{wu\};\ \Psi_G(b) = \{wz\};\ \Psi_G(c) = \{uz\};\ \Psi_G(d) = \{uv\} \\ \Psi_G(e) &= \{vy\};\ \Psi_G(f) = \{wx\};\ \Psi_G(g) = \{zz\} \end{aligned}\right\} \tag{4}$$

By using points to indicate vertices and using lines to indicate the edges, a graphical representation for graph *G* is shown in Fig. 6. The shape of the graph is not unique since the positions of the points and the shapes of the lines usually have no significance.

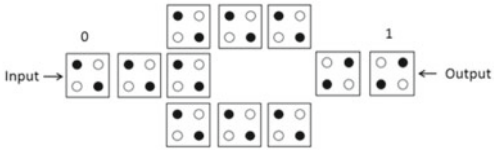


Fig. 5. Circuit diagram of an inverter

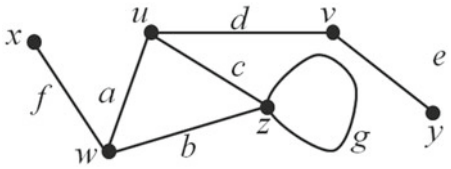


Fig. 6. A graphical representation for graph G

A graph F is considered a subgraph of a graph G if $V(F) \subseteq V(G)$, $E(F) \subseteq E(G)$ and Ψ_F is under restriction of Ψ_G. Such a graph F is contained by G and denoted as $F \subseteq G$. There are two ways of obtaining a subgraph from a graph G. One is to delete the vertices as well as the edges linked to them. The other one is to delete the edges as well as vertices attached to them. In this work, the former method is used. For example, by deleting the vertices x and z from $V(G)$ and updating $E(G)$ and Ψ_G correspondingly, $V(F)$, $E(F)$ and Ψ_F are obtained as shown in Eq. (5) and $E(F)$ a graphical representation is shown in Fig. 7. It can be seen from what has been discussed that a subgraph is a partition of a graph.

$$\left.\begin{array}{l} V(G) = \{u, v, w, y\};\ E(G) = \{a, d, e\} \\ \Psi_G(a) = \{wu\};\ \Psi_G(d) = \{uv\};\ \Psi_G(e) = \{vy\} \end{array}\right\} \tag{5}$$

2.5 Graph Isomorphism and Nauty Algorithm

Two different graphs that have essentially the same structure are isomorphic to each other. In other words, these two graphs are not distinguishable from one another in terms of the number of nodes and their edge connectivity. A more precise definition is: for two graphs G and H, if there are bijections θ: $V(G) \rightarrow V(H)$ and φ: $E(G) \rightarrow E(H)$ so that $\Psi_G(e) = uv$ if and only if $\Psi_H(\varphi(e)) = \theta(u)\ \theta(v)$, G and H are isomorphic [18]. Consider two graphs G and H as shown in Fig. 8(a) and (b). Though they look different, with the bijections θ and φ shown as Eq. (6), G and H are isomorphic to each other. According to the definitions, for example, in graph G, $\Psi_G(n8) = 1\ 5 \rightarrow \Psi_H(\varphi(n8)) = \theta(1)\ \theta(5) \rightarrow \Psi_H(m8) = a\ e$, which is exactly the case in graph H. All vertices and edges can be verified in this way to show that G and H are isomorphic.

However, testing of the graph isomorphism is a formidable task. Theoretically, for two graphs with n vertices, one needs to test all $n!$ bijections between G and H and

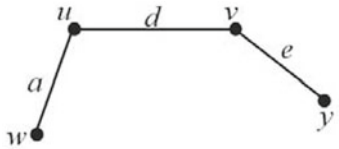


Fig. 7. A graphical representation for graph F (a subgraph of graph G)

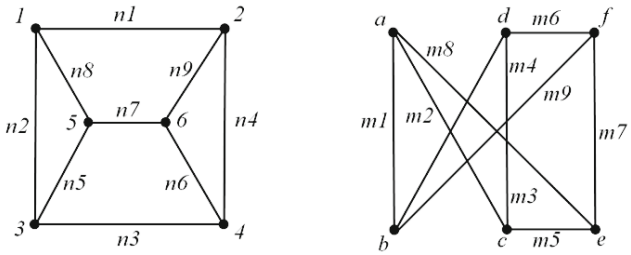


Fig. 8. (a) A graphical representation for graph G (b) A graphical representation for graph H

check if there exists a proper bijection that can make the graphs isomorphic. For a moderately small graph with $n = 100$, $n!$ is an unmanageably large number. A lot of research has been done regarding isomorphism testing and many algorithms have been proposed. Among them, Nauty is considered to be one of the fastest and the most comprehensive isomorphism testing programs.

$$\theta := \begin{pmatrix} 123456 \\ abcdef \end{pmatrix} \varphi := \begin{pmatrix} n1 & n2 & n3 & n4 & n5 & n6 \\ m1 & m2 & m3 & m4 & m5 & m6 \end{pmatrix} \tag{6}$$

Nauty, first proposed by McKay, uses a set of transformations to reduce the graphs, and test for the presence of isomorphism [19]. In this work, the details of the algorithm will not be discussed and the emphasis is put on the use of the software. The inputs to the software are two graphs and the output shows if they are isomorphic or not. Additionally, if they are isomorphic, Nauty will also give the bijection between them, which is a set of variable substitutions to make one graph to overlap over the other. For further information on graph theory and Nauty algorithm, the interested reader may refer to the material in Refs. [18, 19].

3 Identifying Standard Functions with Nauty Algorithm

3.1 Primitives

The Boolean logic functions that can be represented by only one majority gate are called primitives, which are the basic units to build all other majority expressions.

For four-variable functions, five types of primitives can be defined according to the number of variables used as inputs to a majority gate. The first type has no variable-inputs, which means this type includes two primitives, logic "1" and logic "0". The second type has a single variable-input like "a" or "(c')". Since four-variable Boolean functions are discussed here, there are 8 primitives that belong to this type.

The third type of primitive has two variable-inputs. Any combination of two inputs from "a", "b", "c", "d", "(a')", "(b')", "(c')" and "(d')" are allowed except these four pairs: "$(a\ a')$", "$(b\ b')$", "$(c\ c')$" and "$(d\ d')$". The number of possible

combinations should be multiplied by 2 because the third input can be either logic "1" or logic "0". The total number of primitives that belong to this type is calculated as shown in Eq. (7).

$$(C_2^8 - 4) * 2 = 48 \tag{7}$$

The last type of primitive has all of three variable-inputs. For this type, three of the four variables "a", "b", "c" and "d" will be picked up to fill a majority gate and each variable can appear as itself or its complement, thus giving a total number as calculated in Eq. (8). Therefore, the total number of primitives for four-variable functions is calculated as Eq. (9).

$$C_3^4 * 2^3 = 32 \tag{8}$$

$$2 + 8 + (C_2^8 - 4) * 2 + C_3^4 * 2^3 = 10 + 48 + 32 = 90 \tag{9}$$

The total number of primitives for three-variable problems can be obtained through the same method. Equation (10) shows the detailed calculation.

$$8 + (C_2^6 - 3) * 2 + 2^3 = 40 \tag{10}$$

In summary, there are ninety primitives for four-variable problems and forty for three-variable problems. The three-variable primitives are a subset of the four-variable primitives. Table 1 lists all four-variable primitives.

3.2 Majority Expression

The geometric interpretations of three-variable problems have been explored in [10]. However, four-variable functions are more complex because they cannot always be built within two levels. The basic principle will be introduced through an example described in Eq. (11):

$$f = ab'c'd' + abcd + ac'd \tag{11}$$

As a four-variable function can be mapped on to a four by four K-map as a collection of on-set and off-set squares, the K-map of the function f is shown in Fig. 9 with the filled squares as on-set and the white squares as off-set.

As shown in Fig. 10, the solution can be represented by a tree structure with each of its nodes containing at most three branches since a majority gate has only three inputs. The top node of the tree is the result which is formed by selecting the proper primitives as inputs to a majority gate. The rule for selecting primitives is stated as follows:

The first primitives should cover as many on-set squares as possible and as few off-set squares as possible. If several primitives meet this rule, any one of them can be selected. The second primitive must cover the on-set squares that are not covered by the first primitive and must not cover the off-set squares that are covered by the first primitive. For the second primitive, it is also an objective to cover as many on-set squares and as few off-set squares as possible. After the first two primitives are

Table 1. A list of primitives for four-variable functions

No.	Boolean logic	Majority logic	No.	Boolean logic	Majority logic
1	*0*	*0*	46	$d' + c$	$M(d',1,c)$
2	*1*	*1*	47	$a + b$	$M(a,1,b)$
3	a	a	48	$a + c$	$M(a,1,c)$
4	a'	a'	49	$a + d$	$M(a,1,d)$
5	b	b	50	$b + c$	$M(b,1,c)$
6	b'	b'	51	$b + d$	$M(b,1,d)$
7	c	c	52	$c + d$	$M(c,1,d)$
8	c'	c'	53	$a'b'$	$M(a,1,b)'$
9	d	d	54	$a'c'$	$M(a,1,c)'$
10	d'	d'	55	$a'd'$	$M(a,1,d)'$
11	ab	$M(a,0,b)$	56	$b'c'$	$M(b,1,c)'$
12	ac	$M(a,0,c)$	57	$b'd'$	$M(b,\ 1,d)'$
13	ad	$M(a,0,d)$	58	$c'd'$	$M(c,1,d)'$
14	bc	$M(b,0,c)$	59	$ab + ac + bc$	$M(a,b,c)$
15	bd	$M(b,0,d)$	60	$a'b' + a'c' + b'c'$	$M(a,b,c)'$
16	cd	$M(c,0,d)$	61	$a'b + a'c + bc$	$M(a',b,c)$
17	$a' + b'$	$M(a,0,b)'$	62	$ab' + ac' + b'c'$	$M(a,b',c')$
18	$a' + c'$	$M(a,0,c)'$	63	$ab' + ac + b'c'$	$M(a,b',c)$
19	$a' + d'$	$M(a,0,d)'$	64	$a'b + a'c' + bc'$	$M(a',b,c')$
20	$b' + c'$	$M(b,0,c)'$	65	$ab + ac' + bc'$	$M(a,b,c')$
21	$b' + d'$	$M(b,0,d)'$	66	$a'b' + a'c + b'c$	$M(a',b',c)$
22	$c' + d'$	$M(c,0,d)'$	67	$ab + ad + bd$	$M(a,b,d)$
23	ab'	$M(a,0,b')$	68	$a'b' + a'd' + b'd'$	$M(a,b,d)'$
24	ac'	$M(a,0,c')$	69	$a'b + a'd + bd$	$M(a',b,d)$
25	ad'	$M(a,0,d')$	70	$ab' + ad' + b'd'$	$M(a,b',d')$
26	bc'	$M(b,0,c')$	71	$ab' + ad + b'd$	$M(a,b',d)$
27	bd'	$M(b,0,d')$	72	$a'b + a'd' + bd'$	$M(a',b,d')$
28	$a'b$	$M(b,0,a')$	73	$ab + ad' + bd'$	$M(a,b,d')$
29	$a'c$	$M(c,0,a')$	74	$a'b' + a'd + b'd$	$M(a',b',d)$
30	$b'c$	$M(c,0,b')$	75	$ac + cd + ad$	$M(a,c,d)$
31	cd'	$M(c,0,d')$	76	$a'c' + c'd' + a'd'$	$M(a,c,d)'$
32	da'	$M(d,0,a')$	77	$a'c + cd + a'd$	$M(a',c,d)$
33	db'	$M(d,0,b')$	78	$ac' + c'd' + ad'$	$M(a,c',d')$
34	dc'	$M(d,0,c')$	79	$ac' + c'd + ad$	$M(a,c',d)$
35	$a' + b$	$M(a',1,b)$	80	$a'c + cd' + a'd'$	$M(a',c,d')$
36	$a' + c$	$M(a',1,c)$	81	$ac + cd' + ad'$	$M(a,c,d')$
37	$a' + d$	$M(a',1,d)$	82	$a'c' + c'd + a'd$	$M(a',c',d)$
38	$b' + a$	$M(b',1,a)$	83	$bc + cd + bd$	$M(b,c,d)$
39	$b' + c$	$M(b',1,c)$	84	$b'c' + c'd' + b'd'$	$M(b,c,d)'$
40	$b' + d$	$M(b',1,d)$	85	$b'c + cd + b'd$	$M(b',c,d)$
41	$c' + a$	$M(c',1,a)$	86	$bc' + c'd' + bd'$	$M(b,c',d')$
42	$c' + b$	$M(c',1,b)$	87	$bc' + c'd + bd$	$M(b,c',d)$
43	$c' + d$	$M(c',1,d)$	88	$b'c + cd' + b'd'$	$M(b',c,d')$
44	$d' + a$	$M(d',1,a)$	89	$bc + cd' + bd'$	$M(b,c,d')$
45	$d' + b$	$M(d',1,b)$	90	$b'c' + c'd + b'd$	$M(b',c',d)$

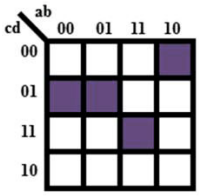


Fig. 9. K-map implementation for $f = ab'c'd' + abcd + ac'd$

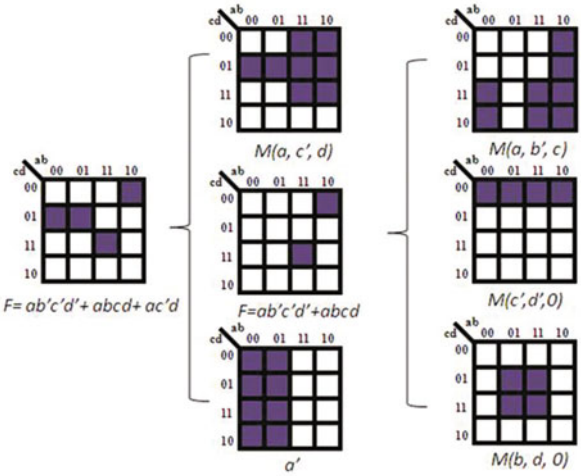


Fig. 10. Solution based on K-map

selected, there is often little choice for the third primitive. The third primitive must cover the on-set squares that are covered only once by the first two primitives and it cannot cover the off-set squares that have been covered by the first two squares. To summarize, any squares that are covered twice or more will be on-set in the resulting K-map. Any squares that are covered only once or not covered by any primitives will not be set in the resulting K-map.

Notice, not all the four-variable functions can be represented by only three majority voters. After the first two primitives have been selected, it is possible that no primitive can match the requirement to "cover the on-set squares which are covered only once by the first two primitives and don't cover the off-set squares which are covered once by the first two primitives". In this case, the required patterns can be expanded to another level, and another three primitives will be needed. The expansion of the level will lead to tree structure solutions for four variable logic functions.

For the example $f = ab'c'd' + abcd + ac'd$, the primitive *M(a, c', d)* (No. 79 in Table 1) is selected first because it covers more of the on-set squares and less of the off-set squares as compared to other primitives. Then "(a')" (No. 4 in Table 1) is selected as the second primitive. When trying to select the last primitive, it can be seen

that no single primitive has of the coverage of the needed pattern. Therefore, another level of majority gates is added and three more primitives are then selected to satisfy the required coverage of pattern. Following the same rule, all the qualified primitives can be found and the result is $f = M(M(a, c', d), a', M(M(a, b', c), M(c', d', 0), M(b, d, 0)))$. The resulting implementation is illustrated in Fig. 10. There may be more than one solution for any given problem. Most of four variable problems can be solved within three levels and all of them can be solved in four levels.

3.3 Four-Dimensional Cubes and Four-Variable K-maps

A four-variable Boolean function can be represented by a four dimensional (4-D) cube [20, 21], as shown in Fig. 11, with each literal in this function "a", "b", "c", and "d" being one of the four dimensions of a 4-D cube. The sixteen different minterms in the Boolean logic function represent sixteen vertices in a 4-D cube, which is similar to the sixteen cells in a four-variable K-map. For example, the minterm "$abc'd$" can be represented by vertex (1, 1, 0, 1) and also as an on-set square in K-map. For a four-variable function, the set of all the on-set vertices in a 4-D cube, which represent minterms, can be considered a specific spatial structure which can be rotated, flipped, or mirrored to represent many other four-variable functions. Thus, one structure could describe many different functions.

As an example, consider two logic functions $f = a'c' + a'b'd'$ and $f = b'd + acd$. After mapping these two logic functions to a 4-D cube and corresponding K-map, as shown in Fig. 12(a) and (b), it can be seen that the spatial structures formed by their

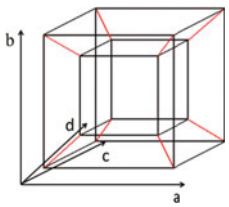
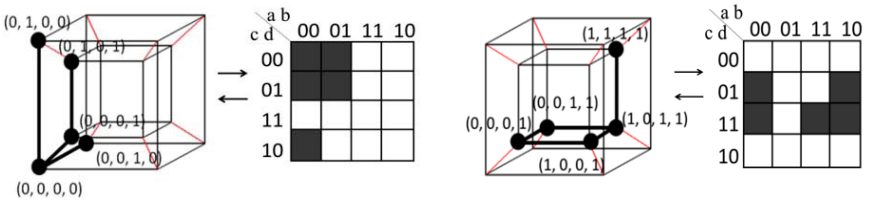


Fig. 11. A 4-D cube



(a) (b)

Fig. 12. (a) 4-D cube and K-map interpretation of $f = b'd + acd$ (b) 4-D cube and K-map interpretation of $f = a'c' + a'b'd'$

on-set vertices are exactly the same (one face and another vertex above it), and the only difference between them is they are in different positions.

If one can find all possible on-set structures in a 4-D cube/K-map and determine the most simplified majority expression for each structure, any four-variable problem could be targeted to these solutions to get the most simplified majority expression. These structures are referred to as standard functions.

3.4 Identifying Standard Functions

In this work, a 4-D cube with 16 vertices is considered as a graph. Its subgraphs are the 4-D structures that are constructed by all of the on-set vertices in Boolean logic functions. If two 4-D structures are isomorphic, they are essentially the same structures, but just located in different positions in a 4-D cube. One structure can be moved to overlap another structure. So if two 4-D structures of Boolean logic functions are isomorphic, these two functions belong to the same category, which means they have a common standard function. If all graph isomorphism categories can be identified, then a full set of standard functions are obtained.

The Nauty algorithm is used to identify if two subgraphs are isomorphic. Since a subgraph is a graph itself, the term "graph" will be used for convenience. Based on Nauty, a new method has been developed to find all standard functions, as shown in Fig. 13. All possible four-variable Boolean logic functions, a total of $2^{16} = 65536$, are

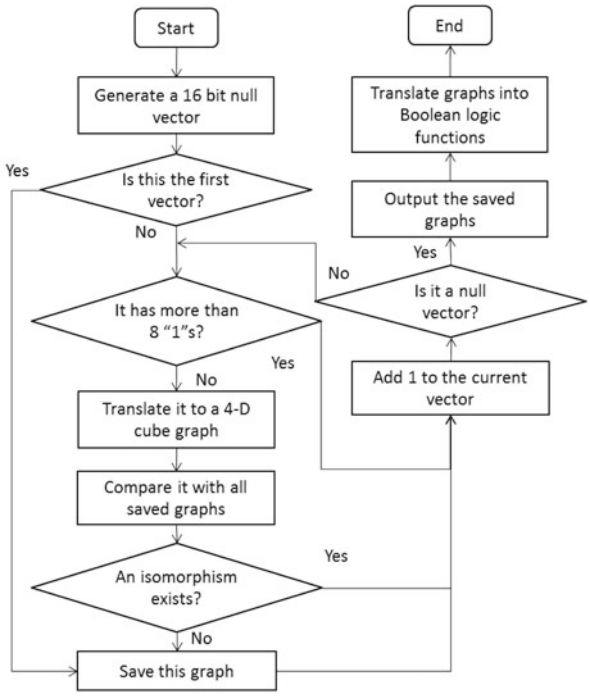


Fig. 13. Flowchart for finding standard functions

used as inputs. By comparing them and selecting the distinct ones, this method will generate an exhaustive list of standard functions.

At the beginning of the process, a Boolean logic function needs to be encoded into a graph. For the first step, a 16-bit vector is used to represent the on-set and off-set vertices for a function. If a particular minterm appears in a function, the corresponding bit in the vector is set to "1" otherwise it is set to "0". For example, the vector for $f = a'bd' + ac'd + abd + ab'cd'$ is $\{1,0,1,0,0,1,1,0,0,1,0,1,0,0,0,0\}$ as shown in Fig. 14. The rightmost bit in the vector is "0" because the minterm $a'b'c'd'$ does not appear in the function. Similarly, the 13th bit is "1" because the minterm $abc'd$ is included in f. Then each bit in the vector is mapped to a corresponding vertex in the 4-D cube as shown in Fig. 15(a). For example, the mapping of the vector in Fig. 14 on to a 4-D cube is shown in Fig. 15(b). In this figure, the vertices labeled as 4, 6, 9, 10, 13 and 15 are mapped as on-set vertices because the 4th, 6th, 9th, 10th, 13th and 15th bits in the vector are "1". The rest vertices are off-set because their corresponding bits in the vector are "0".

As shown in Fig. 13, a 16-bit null vector $\{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0\}$ is generated. It is translated to a graph and be directly saved since there is nothing to compare it with. Then "1" is added to the current vector, which makes the vector become $\{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1\}$. Now the new vector becomes the current vector and it is mapped to a graph and compared with all saved graphs. If an isomorphism exists in the saved graphs, this vector is discarded. Otherwise the graph of the vector is added to the saved graph. The method then adds "1" to the current vector and repeats the same graph isomorphism check procedure repeatedly. The method ends when all possible 16-bit vectors have been used as an input once,

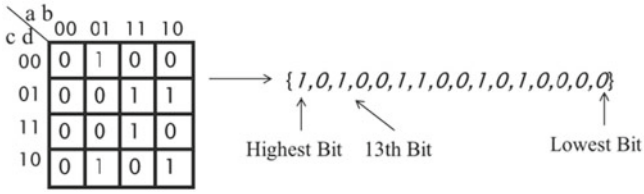


Fig. 14. 16-bit vector for $f = a'bd' + ac'd + abd + ab'cd'$

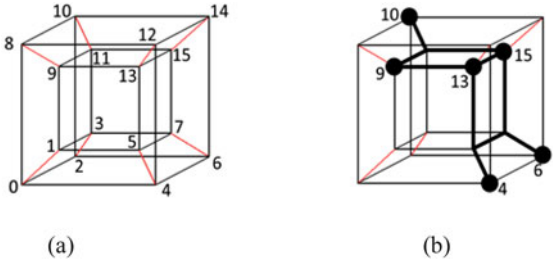


Fig. 15. (a) Rules for mapping a vector to a cube (b) 4-D structure for vector $\{1,0,1,0,0,1,1,0,0,1,0,1,0,0,0,0\}$

i.e., the vector adds "1" 65536 times and overflows back to the null vector. This resulted in 222 standard functions. The majority expressions for these standard functions are obtained using the method introduced in Sect. 3.2.

Notice, only standard functions with fewer than eight minterms need to be considered because functions with nine to sixteen minterms are just the complementary form of functions comprising of seven to zero minterms. So, this method filters out all vectors that have more than eight "1"s.

4 Proposed method

4.1 Synthesis Method Overview

An overview of proposed synthesis method is shown in Fig. 16. The input to the algorithm is an arbitrary network of Boolean functions and the output is an optimized majority expression network. The input network must be preprocessed and decomposed to a four-feasible network. This step is done with the SIS tool [22]. For each node in the resulting decomposed network, the standard functions will be applied to find its optimal majority expression. Then all redundancies are removed from the preliminary majority expression network to obtain the final solution.

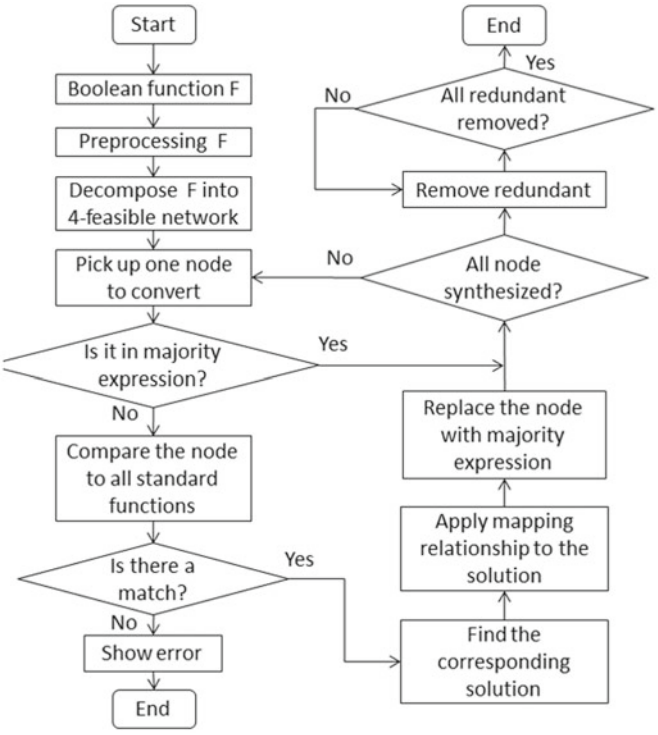


Fig. 16. An overview of proposed method

4.2 Preprocessing and Decomposing

Preprocessing will provide a good start for decomposition by algebraically factoring out the common terms and removing the redundant terms from the Boolean equations. Decomposing will split the larger-input nodes in the network into a set of nodes with lesser number of inputs; thus it makes them easy to convert into majority expressions. In particular, we are interested in obtaining all nodes with less than four inputs. These two steps are accomplished by using proper sequence of commands in SIS.

4.3 Application of Standard Functions

The resulting decomposed four-feasible network is then be mapped to the standard functions. This step is illustrated with an example node given by Eq. (12).

$$f = ad' + bd' + cd' + bcd \tag{12}$$

The function is transformed into a graph and used as an input to Nauty, which compares the function with the whole set of standard functions in order to determine which one is its isomorphic function. After this step, the standard function and its majority expression are found as shown in Eqs. (13) and (14), respectively.

$$f = a'bcd + ab'c'd + ab'cd' + abc'd' + abc'd + abcd' + abcd' + abcd \tag{13}$$

$$f = M(M(a, b, 0), M(a, c, d), a) \tag{14}$$

Nauty also gives the bijections between the given function and the standard function, which is $a \rightarrow d'$, $b \rightarrow c$, $c \rightarrow a$, $d \rightarrow b$ for this problem. By substituting the variables in Eq. (14), the majority expression for the Eq. (12) can be found as shown in Eq. (15).

$$f = M(M(d', c, 0), M(d', a, b), d') \tag{15}$$

The whole process is explained in Fig. 17. Applying this principle to all four feasible nodes obtained from SIS decomposition, a complete majority gate network solution is obtained for the Boolean function.

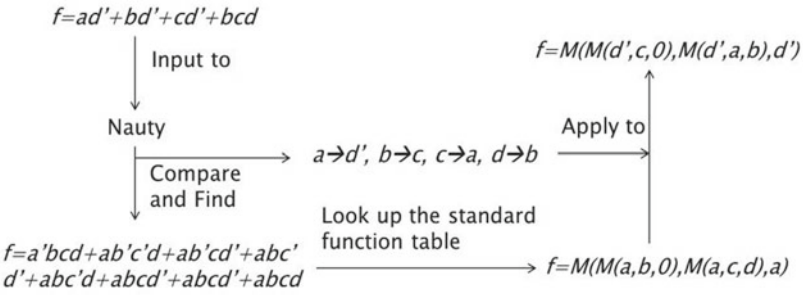


Fig. 17. Application of Nauty based standard functions

4.4 Redundancy Removal

Redundancies may exist in the preliminary majority expression network and the circuit can be further simplified by removing them. This process is explained through an example shown in Eq. (16). The following network has three inputs *a, b* and *c*, and one output *f* as shown in Eq. (16), where *[3]*, *[4]*, *[5]* and *[6]* are internal nodes.

$$\left.\begin{array}{l} f=M(b,[3]',[4]);\ [3]=M(1,[5]',[6]) \\ [4]=M(0,a,c');\ [5]=M(1,a',c);\ [6]=M(0,a,c') \end{array}\right\} \quad (16)$$

The repeated nodes have to be removed. By selecting each node and comparing its original form and complementary form with the rest of the nodes, it can be found that *[4]* and *[6]* are identical while *[4]* and *[5]* are complementary to each other. By substituting *[5]* and *[6]* with $[4]'$ and *[4]*, Eq. (17) is obtained.

$$\left.\begin{array}{l} f=M(b,[3]',[4]);\ [3]=M(1,[4]'',[4]) \\ [4]=M(0,a,c') \end{array}\right\} \quad (17)$$

Then the inverters have to be optimized. In this network, one of the inputs in node *[3]* has doubled inverters and can be eliminated out. The last step is to simplify the nodes with duplicated inputs. Two inputs to the majority function node *[3]* are *[4]*, therefore *[3]* will be identical to the logic value of *[4]*. By substituting $[3] = [4]$, it can be seen that *f* now has two same inputs and can be simplified. The optimized output is as shown in Eq. (18). Note that, the redundancy removal process should be carried out for multiple iterations until no more changes can be made to the network.

$$f = b \quad (18)$$

The simplified function after taking redundancies into consideration requires no majority gate as compared to 3 levels and 5 majority gates in the original network.

5 Results and Comparisons

The proposed method has been applied to MCNC benchmark circuits and compared to other synthesis approaches. The circuit "mux" in MCNC benchmark suite is used to explain the process in detail. The original circuit equations are shown in Eq. (19). It has 1 primary output, *v*, which is in curly brackets, and 21 primary inputs *a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u*. The rest of the variables are the internal nodes.

$$
\begin{aligned}
&\{v\} = j0\ u \\
&j0 = b0'\,e0'\,f0'\,g0' + b0'\,e0'\,q' + b0'\,f0'\,r' + b0'\,q'\,r' \\
&\qquad + e0'\,g0'\,r + e0'\,q'\,r + f0'\,g0'\,q + f0'\,q\,r' + g0'\,q\,r \\
&b0 = m'\,n'\,o'\,p' + m'\,n'\,s + m'\,o'\,t + m'\,s\,t + n'\,p'\,t' \\
&\qquad + n'\,s\,t' + o'\,p'\,s' + o'\,s'\,t + p'\,s'\,t' \\
&e0 = i'\,j'\,k'\,l' + i'\,j'\,s + i'\,k'\,t + i'\,s\,t + j'\,l'\,t' + j'\,s\,t' \\
&\qquad + k'\,l'\,s' + k'\,s'\,t + l'\,s'\,t' \\
&f0 = e'\,f'\,g'\,h' + e'\,f'\,s + e'\,g'\,t + e'\,s\,t + f'\,h'\,t' + f'\,s\,t' \\
&\qquad + g'\,h'\,s' + g'\,s'\,t + h'\,s'\,t' \\
&g0 = a'\,b'\,c'\,d' + a'\,b'\,s + a'\,c'\,t + a'\,s\,t + b'\,d'\,t' \\
&\qquad + b'\,s\,t' + c'\,d'\,s' + c'\,s'\,t + d'\,s'\,t'
\end{aligned}
\tag{19}
$$

The SIS tool is used to simplify the circuit and decompose it into a four-feasible network. Then the process introduced in Sect. 4.4 is applied to each node of the network to obtain its majority expression. After removing the redundancies, the final result is shown in Eq. (20). Logic "1" and "0" are represented as "*<1>*" and "*<0>*" in the majority expression network because some circuits use "1" or "0" to name their variables. For comparison purpose, the output from the synthesis method in [15] is shown in Eq. (21)

$$
\begin{array}{lll}
\{v\}=M(u,[2],[3]) & [2]=M(<0>,q,[4]) & [3]=M(<0>,q',[7]) \\
[4]=M(<1>,[5],[6]) & [5]=M(<1>,[14],[17]) & [6]=M(<1>,[20],[23]) \\
[7]=M(<1>,[8],[9]) & [8]=M(<1>,[26],[29]) & [9]=M(<1>,[32],[35]) \\
[10]=M(<0>,s,t) & [11]=M(<0>,s,t') & [12]=M(<0>,s',t) \\
[13]=M(<1>,s,t)' & [14]=M([13],[15],[16]) & [15]=M(<0>,d,r) \\
[16]=M(<0>,h,r') & [17]=M([12],[18],[19]) & [18]=M(<0>,r,c) \\
[19]=M(<0>,r',g) & [20]=M([11],[21],[22]) & [21]=M(<0>,r,b) \\
[22]=M(<0>,r',f) & [23]=M([10],[24],[25]) & [24]=M(<0>,r,a) \\
[25]=M(<0>,r',e) & [26]=M([13],[27],[28]) & [27]=M(<0>,r,l) \\
[28]=M(<0>,r',p) & [29]=M([12],[30],[31]) & [30]=M(<0>,r,k) \\
[31]=M(<0>,r',o) & [32]=M([11],[33],[34]) & [33]=M(<0>,r,j) \\
[34]=M(<0>,r',n) & [35]=M([10],[36],[37]) & [36]=M(<0>,r,i) \\
[37]=M(<0>,r',m) & &
\end{array}
\tag{20}
$$

```
{v}=M(<0>,u,[44])   [2]=M(<1>,[13],[3])   [3]=M(<0>,[6],[10])
[4]=M(<1>,[30],[5])   [5]=M(<0>,[6],[27])   [6]=M(<0>,s,t)
[7]=M(<0>,s,t')   [8]=M(<0>,s',t)   [9]=M(<1>,s,t)'
[10]=M(<0>,[11],[12])   [11]=M(<1>,a,r')   [12]=M(<1>,e,r)
[13]=M(<1>,[18],[14])   [14]=M(<0>,[7],[15])   [15]=M(<0>,[16],[17])
[16]=M(<1>,r',b)   [17]=M(<1>,r,f)   [18]=M(<1>,[23],[19])
[19]=M(<0>,[8],[20])   [20]=M(<0>,[21],[22])   [21]=M(<1>,r',c)
[22]=M(<1>,r,g)   [23]=M(<0>,[9],[24])   [24]=M(<0>,[25],[26])
[25]=M(<1>,r',d)   [26]=M(<1>,r,h)   [27]=M(<0>,[28],[29])
[28]=M(<1>,r',i)   [29]=M(<1>,r,m)   [30]=M(<1>,[35],[31])
[31]=M(<0>,[7],[32])   [32]=M(<0>,[33],[34])   [33]=M(<1>,r',j)
[34]=M(<1>,r,n)   [35]=M(<1>,[40],[36])   [36]=M(<0>,[8],[37])
[37]=M(<0>,[38],[39])   [38]=M(<1>,r',k)   [39]=M(<1>,r,o)
[40]=M(<0>,[9],[41])   [41]=M(<0>,[42],[43])   [42]=M(<1>,r',l)
[43]=M(<1>,r,p)   [44]=M(<0>,[45],[46])   [45]=M(<1>,q',[2])
[46]=M(<1>,q,[4])
```

(21)

Comparing Eqs. (20) and (21), it can be seen that the proposed method uses 6 levels and 37 majority gates while the method in [15] uses 9 levels and 46 gates. The number of levels is reduced by 33.3 % and the number of gates used is reduced by 19.6 %. The circuit diagrams for the results obtained by both methods are shown in Figs. 18 and 19.

More benchmark comparisons are provided in Table 2. It can be seen that on average, there is 8.13 % reduction in terms of majority gates and 7.94 % reduction in terms of logic levels by the proposed method as compared to the best existing method [15].

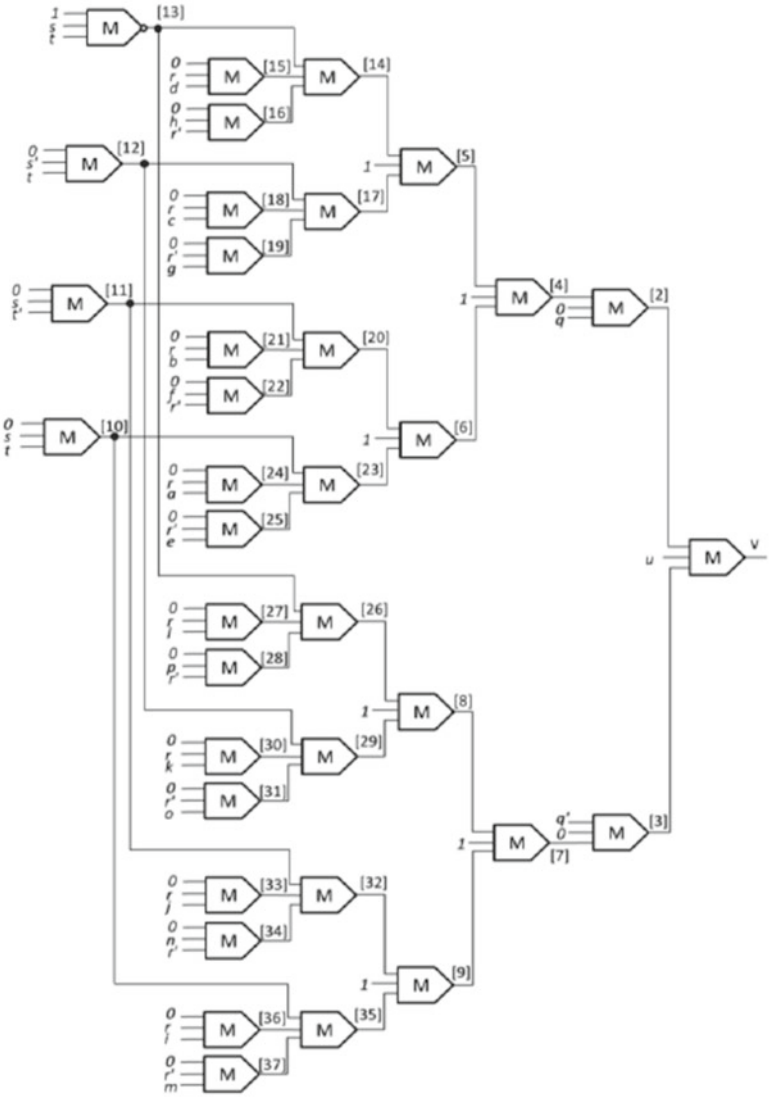


Fig. 18. Circuit diagram for "mux" obtained by the proposed method

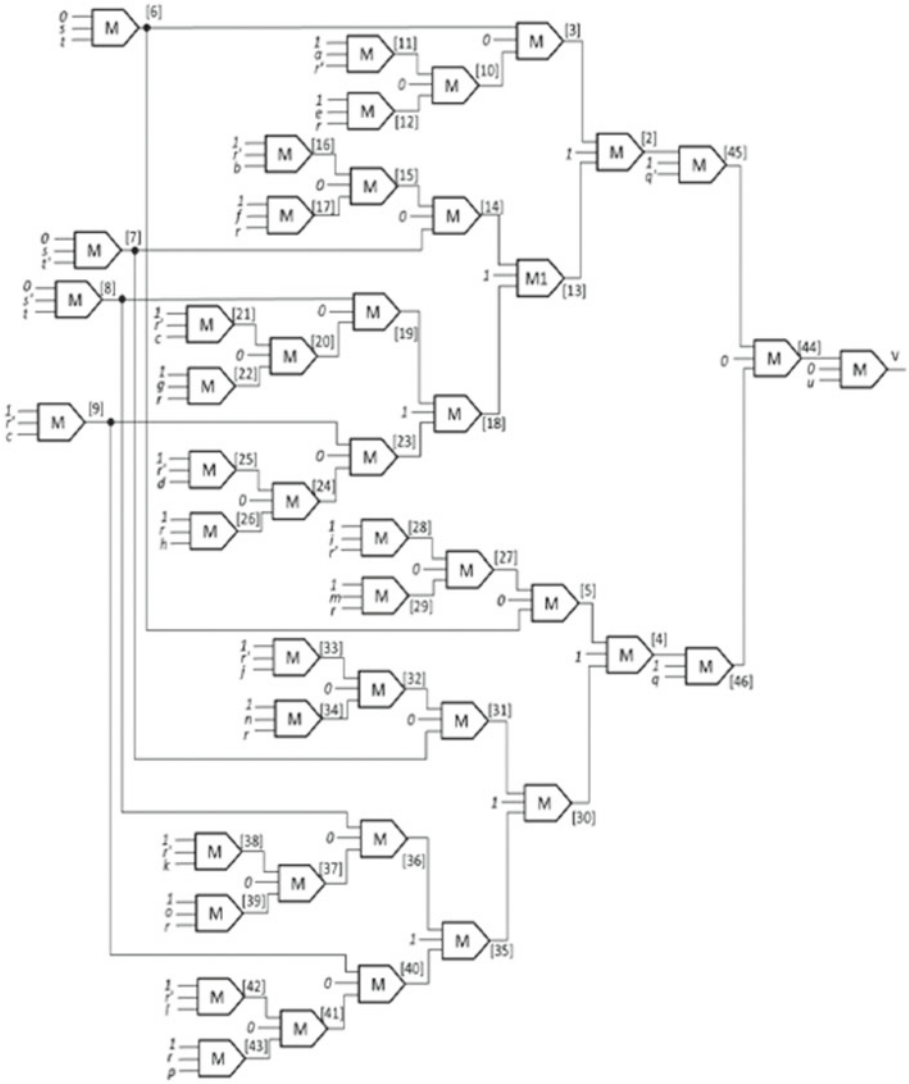


Fig. 19. Circuit diagram for "mux" obtained by the method in [15]

Table 2. Overall comparisons

	Existing method [15]		Proposed method		Reduction %	
	Level	Gate	Level	Gate	Level	Gate
b1	2	7	2	6	0.00	14.29
cm82a	3	7	3	6	0.00	14.29
majority	4	6	3	6	25.00	0.00
alu2	18	340	16	347	11.11	−2.06
x2	7	37	6	36	14.29	2.70
cm152a	6	21	6	17	0.00	19.05
cm85a	6	26	6	19	0.00	26.92
cm151a	7	23	7	20	0.00	13.04
cu	7	40	6	40	14.29	0.00
cm163a	7	38	7	32	0.00	15.79
cmb	4	28	4	26	0.00	7.14
pm1	6	35	6	32	0.00	8.57
vda	15	700	14	670	6.67	4.29
cm150a	9	46	6	37	33.33	19.57
mux	9	46	6	37	33.33	19.57
ttt2	11	145	10	144	9.09	0.69
i1	6	36	6	35	0.00	2.78
frg1	18	105	17	102	5.56	2.86
term1	11	106	10	89	9.09	16.04
k2	19	1301	19	1193	0.00	8.30
x1	11	264	11	253	0.00	4.17
example2	10	247	9	241	10.00	2.43
apex6	17	662	15	677	11.76	−2.27
frg2	14	582	13	600	7.14	−3.09
Average reductions					7.94	8.13

6 Conclusion

The approach to generate optimal majority expressions for all functions with four or fewer variables is presented in this paper. In this approach, Nauty algorithm is used to identify the standard functions and give the relationship between an arbitrary Boolean function and its standard function. In order to convert a Boolean logic network, SIS tool is used for preprocessing and decomposition. Using the standard functions, the decomposed network is converted into a preliminary majority network. The redundancy removal can then be applied to further simplify the result.

By summarizing these steps, an automated computer program is developed for majority logic synthesis. Since the minority logic can be easily obtained by applying De Morgan's Law to majority logic, this approach can also be used to build a minority network. This synthesis method is integrated into SIS synthesis tool for better usability.

Acknowledgements. This work was supported in part by National Science Foundation awards 0958298 and 0958355.

References

1. International technology roadmap for semiconductors (ITRS). http://www.itrs.net
2. Snider, G.L., Orlov, A.O., Amlani, I., Zuo, X., Bernstein, G.H., Lent, C.S., Merz, J.L., Porod, W.: Quantum-dot cellular automata: review and recent experiments. J. Appl. Phys. **85**, 4283–4285 (1999)
3. Lent, C.S., Tougaw, P.D.: A device architecture for computing with quantum dots. Proc. IEEE **85**, 541–557 (1997)
4. Walus, K., Jullien, G., and Dimitrov, V.S.: Computer arithmetic structures for quantum cellular automata. In: Proceedings of the Asilomar Conference on Signals, Systems, and Computers, vol. 2, pp. 1435–1439, IEEE Press (2003)
5. Tougaw, P.D., Lent, C.S.: Logical devices implemented using quantum cellular automate. J. Appl. Phys. **75**, 1818–1825 (1994)
6. Oya, T., Asai, T., Fukui, T., Amemiya, Y.: A majority-logic nanodevice using a balanced pair of single-electron boxes. J. Nanosci. Nanotechnol. **2**, 333–342 (2002)
7. Oya, T., Asai, T., Fukui, T., Amemiya, Y.: A majority-logic nanodevice using an irreversible single-electron boxes. IEEE Trans. Nanotechnol. **2**, 15–22 (2003)
8. Fahmy, H.A., Kiehl, and R.A.: Complete logic family using tunneling-phase-logic devices. In: Proceedings of the International Conference on Microelectronics, pp. 22–24 (1999)
9. Zhang, R., Walus, K., Wang, W., Jullien, G.A.: A method of majority logic reduction for quantum cellular automata. IEEE Trans. Nanotechnol. **3**, 443–450 (2004)
10. Walus, K., Schulhof, G., Jullien, G.A., Zhang, R., Wang, w.: Circuit design based on majority gates for applications with quantum-dot cellular automata. In: IEEE Asilomar Conference on Signals, Systems, and Computers, pp. 1354–1357. IEEE Press (2004)
11. Bonyadi, M., Azghadi, S., Rad, N., Navi, K., Afjei, E.: Logic optimization for majority gate-based nanoelectronic circuits based on genetic algorithm. In: Proceedings of the International Conference on Electrical Engineering, pp. 1–5 (2007)
12. Rai, S.: Majority gate based design for combinational quantum cellular automata (QCA) circuits. In: Proceedings of 40th Southeastern Symposium on System Theory. pp. 222–224 (2008)
13. Zhang, R., Gupta, P., Jha, N.K.: Majority and minority network synthesis with application to QCA-, SET-, and TPL-based nanotechnologies. IEEE Trans. Comput. Aided Des. Integr. Circ. Syst. **26**, 1233–1245 (2007)
14. Huo, Z., Zhang, Q., Haruehanroengra, S., Wang W.: Logic optimization for majority gate-based nanoelectronic circuits. In: Proceedings of the International Symposium on Circuits and Systems, pp. 1307–1310 (2006)
15. Kong, K., Shang, Y., Lu, R.: An optimized majority logic synthesis methodology for quantum-dot cellular automata. IEEE Trans. Nanotechnol. **9**, 170–183 (2010)
16. Wang, P., Niamat, M., Vemuru, S.: Minimal majority gate mapping of 4-variable functions for quantum cellular automata. IEEE Conference on Nanotechnology, pp. 1307–1312 (2011)
17. Blair, E.P., Lent, C.S.: Quantum-dot cellular automata: an architecture for molecular computing. In: International Conference on Simulation of Semiconductor and Devices, pp. 14–18 (2003)
18. Bondy, J.A., Murty, U.S.R.: Graph Theory. Springer, London (2008)
19. McKay, B.D.: Practical graph isomorphism. Congressus Numerantium **30**, 45–87 (1981)
20. Wolf, W.: Modern VLSI Design: System-on-Chip Design. Prentice-Hall, Upper Saddle River (2002)

21. Brayton, R.K., McMullen, C., Hachtel, G.D., Sangiovanni-Vincentelli, A.: Logic Minimization Algorithms for VLSI Synthesis. Kluwer, Norwell (1984)
22. Sentovich, E.M., Singh, K.J., Lavagno, L., Moon, C., Murgai, R., Saldanha, A., Savoj, R., Stephan, P.R., Brayton, R.K., Sangiovanni-Vincentelli, A.: SIS: a system for sequential circuit synthesis. Technical Report, University of California, Berkeley (1992)

Reversible Logic Based Design and Test of Field Coupled Nanocomputing Circuits

Himanshu Thapliyal(✉), Nagarajan Ranganathan, and Saurabh Kotiyal

Department of Computer Science and Engineering,
University of South Florida, Tampa, FL, USA
{hthapliy,ranganathan}@cse.usf.edu, skotiyal@mail.usf.edu

Abstract. Reversible computing is based on logic circuits that can generate unique output vector from each input vector, and vice versa, that is, there is a one-to-one mapping between the input and the output vectors. Reversible computing is the only solution for non-dissipative ultra low power green computing. Conservative reversible circuits are a specific type of reversible circuits, in which there would be an equal number of 1s in the outputs as there would be on the inputs, in addition to one-to-one mapping. This work illustrates the application of reversible logic towards testing of faults in traditional and reversible field coupled nanocircuits (Portions of this chapter are based on [2]. The enhancement is comprehensive treatment of: basics of reversible computing, motivation for reversible computing, background on conservative logic, basics of QCA computing, such as QCA logic devices and QCA clocking, related work etc. Several new reversible testable designs are introduced such as design of testable reversible T latch, design of testable asynchronous set/reset D latch and master-slave D flip-flop, design of testable reversible complex sequential circuits. QCA layouts of conservative logic gates are introduced with internal design details of QCA logic devices. Complete fault patterns information and analysis are provided for conservative logic gates. The synthesis of non-reversible testable design based on MX-cqca gate is extended to MX-cqca based implementation of standard functions. The significance of this work and broader prospective for future directions is also presented.). We propose the design of two vectors testable sequential circuits based on conservative logic gates. The proposed sequential circuits based on conservative logic gates outperform the sequential circuits implemented in classical gates in terms of testability. Any sequential circuit based on conservative logic gates can be tested for classical unidirectional stuck-at faults using only two test vectors. The two test vectors are all 1s, and all 0s. The designs of two vector testable latches, master-slave flip-flops, double edge triggered flip-flops, asynchronous set/reset D latch and D flip-flop are presented. The importance of the proposed work lies in the fact that it provides the design of reversible sequential circuits completely testable for any stuck-at fault by only two test vectors, thereby eliminating the need for any type of scan-path access to internal memory cells. The reversible designs of the double edge triggered flip-flop, ring counter and Johnson Counter are proposed for the first time in literature. We are showing the application of the proposed approach towards 100 % fault coverage for single

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 133–172, 2014.
DOI: 10.1007/978-3-662-43722-3_7, © Springer-Verlag Berlin Heidelberg 2014

missing/additional cell defect in the QCA layout of the Fredkin gate. We are also presenting a new conservative logic gate called Multiplexer Conservative QCA gate (MX-cqca) that is not reversible in nature but has similar properties as the Fredkin gate of working as 2:1 multiplexer. The proposed MX-cqca gate surpasses the Fredkin gate in terms of complexity (the number of majority voter), speed and area.

Keywords: Conservative logic · Reversible logic · Quantum dot cellular automata

1 Introduction

Reversible logic has applications in emerging technologies such as quantum computing, quantum dot cellular automata, optical computing, etc [37,38,44,47,63]. Reversible logic based circuits satisfy the property that there is one-to-one mapping between the input and the output vectors, that is for each input vector there is a unique output vector and vice-versa. Conservative logic is a logic family which exhibit the property that there are equal number of 1s in the outputs as in the inputs. Conservative logic may or may not be in reversible in nature. Conservative logic is called reversible conservative logic when there is a one-to-one mapping between the input and the output vectors along with the property that there are equal number of 1s in the outputs as in the inputs. Conservative logic circuits are not considered reversible, if one-to-one mapping between the input and the output vectors is not preserved.

Researchers have proved that if the computation is performed in a irreversible manner each bit of information lost will produce $kT \ln 2$ Joules of heat energy [31]. From thermodynamic point of view, in order to avoid this limit, Bennett showed that $kT \ln 2$ energy dissipation would not occur, if a computation is carried out in a reversible way [5]. Thus, from thermodynamic considerations, a firm lower limit on dissipation of $E_{diss} = kT \ln 2 \approx 18\,\text{meV}$ (in room-temperature environment) is a necessity for conventional (irreversible) logic, even if reliability issues could be ignored. Reversible logic can be useful to design non-dissipative circuits if the physical implementation of the logic is also physically reversible. CMOS cannot be considered as a practical implementation platform as CMOS is not physically reversible. In modern CMOS technology, voltage-coded logic signals have an energy of $E_{sig} = (1/2)CV^2$, and whenever the node voltage is changed, it leads to dissipation of this energy and is order of magnitude higher than the $kT \ln 2$ factor. In contrast, there are emerging nanotechnologies such as Quantum Dot Cellular automata (QCA) computing, Optical Computing, and Superconductor Flux Logic (SFL) family, etc., where the energy dissipated due to information destruction will be a significant factor of the overall heat dissipation of the system [4,13,19,30,55,56,58]. Thus, one of the primary motivations for adopting reversible logic lies in the fact that it can provide a logic design methodology for designing ultra-low power circuits beyond $kT \ln 2$ limit for those emerging nanotechnologies in which the energy dissipated due to information destruction

will be a significant factor of the overall heat dissipation. For example, in new Superconductor Flux Logic (SFL) family based on nSQUID gates, the energy dissipation in conventional logically irreversible architectures is close to few kT ln2 per logic operation. By employing reversible logic, the energy dissipation per nSQUID gate per bit measured, at 4 K temperature is below the thermodynamic threshold limit of kT ln2 [55]. Therefore, reversible logic is being investigated for its promising applications in power-efficient nanocomputing [17,32,35].

Further, quantum dot cellular automata (QCA) is one of the emerging field coupled nanotechnologies in which it is possible to implement reversible logic gates [19,37,38]. QCA makes it possible to achieve circuit densities and clock frequencies beyond the limits of existing CMOS technology [70,71]. The logic states of 1 and 0 are represented by the position of the electrons inside the QCA cell as illustrated in Fig. 2(b). Thus, when the bit is flipped from 1 to 0 there is no actual discharging of the capacitor as in conventional CMOS. Hence, QCA does not have to dissipate all its signal energy during transition. Further, propagation of the polarization from one cell to another is because of interaction of electrons in adjacent QCA cells. As there is no movement of electrons from one QCA cell to the other, hence there is no current flow. Thus, QCA has no dissipation in signal propagation. Therefore, QCA has significant advantage compared to CMOS technology in terms of power dissipation. Further in contrast to CMOS, the cells in QCA are connected to 4 clocking zones, each lagging behind by 90 ° in phase. QCA clocking helps in the successive transfer of information from one clock zone to the next [25,34]. Therefore, we have information flow from the input to the output in a pipelined fashion. Thus, QCA cells are inherently suitable for pipeline and systolic designs [15]. QCA computing can be implemented in semiconductor, molecular and magnetic platforms. Researchers are currently targeting magnetic and molecular QCA, several smaller circuits in magnetic and molecular QCA have been fabricated and tested [3,7,26,33,52,53]. Theoretical studies have also shown that molecular QCA can operate at room temperature at THz of speed [36].

Several works can be found in the literature for QCA design such as adders, multipliers, shifters, memories, FPGA, synthesis etc. [9,10,24,28,45,64,73,76]. Due to high error rates in nano-scale manufacturing, the major goal in QCA and other nanotechnologies is to have devices with reduced error rates [37]. In the manufacturing process for QCA, defects can occur in the synthesis and deposition phases. However, defects are more likely to occur during the deposition phase [62].

In this work, we propose the design of testable sequential circuits based on conservative logic gates. It has shown in [61] that the combinational circuits based on conservative logic gates outperform all the circuits implemented using classical gates in the area of testing. Further, any combinational circuit based on conservative logic gates can be tested for classical unidirectional stuck-at faults using only two test vectors. The two test vectors are all 1s, and all 0s [61]. This is because whenever there are unidirectional faults in combinational conservative logic circuits, the number of 1s in the inputs will differ from the number of 1s in

the outputs. Thus for unidirectional stuck-at faults in conservative logic circuits, counting the number of 1s in the inputs and the outputs would be the fault detection scheme. The feedback in sequential circuits makes them untestable by all 0s and all 1s test vectors. Moreover, in reversible logic fanout is not allowed. Hence, we propose a technique that will take care of the fanout at the output of the reversible latches and can also disrupt the feedback to make them suitable for testing by only two test vectors, all 0s and all 1s. In other words, circuits will have feedback while executing in the normal mode. However, in order to detect faults in the test mode, our proposed technique will disrupt feedback to make conservative reversible latches testable as combinational circuits. The proposed technique is extended towards the design of two vectors testable master-slave flip-flops, double edge triggered flip-flops, asynchronous set/reset D latch, D flip-flop and counters. Thus, our work is significant because we are providing the design of reversible sequential circuits completely testable for any unidirectional stuck-at faults by only two test vectors. The reversible design of the double edge triggered flip-flop, ring counter and Johnson counter are proposed for the first time in literature.

Field coupled quantum dot cellular automata (QCA) computing is based on majority voting, hence the designs based on conservative logic will be completely different from those based on conventional CMOS. Single missing/additional cell defects are prominent permanent defects in QCA circuits. We implemented the Fredkin gate in the QCA technology and observed that all 0s and all 1s test vectors cannot provide 100 % fault coverage for single missing/additional cell defect in the QCA layout of the Fredkin gate. Thus, to have the 100 % fault coverage for single missing/additional cell defect by all 0s and all 1s test vectors, we identified the QCA devices in the QCA layout of the Fredkin gate that can be replaced with fault tolerant components to provide the 100 % fault coverage. Further, while designing a QCA sequential circuit, the designer may sometimes prefer to sacrifice the reversibility to save the number of QCA cells while keeping the test strategy to be the same, that is the design can still be tested by two test vectors. Thus, we also propose a new conservative logic gate called Multiplexer Conservative QCA gate (MX-cqca) which is not reversible in nature but has similar properties as the Fredkin gate of working as 2:1 multiplexer. The proposed MX-cqca gate surpasses the Fredkin gate in terms of complexity(the number of majority voter), speed and area. MX-cqca can implement the multiplexer function with 1 less majority gate than the Fredkin gate; it also requires a smaller area and fewer QCA cells in QCA layout. The design and verification of the QCA layouts were performed using the QCADesigner and HDLQ tools.

The chapter is organized as follows: Sect. 2 presents the background on conservative logic, the basics of QCA computing, such as QCA logic devices and QCA clocking, related work etc.; Sect. 3 presents design of testable reversible latches; Sect. 4 describes design of testable reversible master-slave flip-flops; Sect. 5 presents design of testable reversible double edge triggered flip-flop; Sect. 6 shows the design of testable reversible complex sequential circuits; Sect. 7 discusses the application of the proposed two vectors, all 0s and all 1s, testing

approach to QCA computing; Sect. 8 presents the proposed multiplexer conservative QCA gate; Sect. 9 discusses design methodology for non-reversible testable design based on MX-cqca gate; and Sect. 10 provides some discussions and conclusions.

2 Background

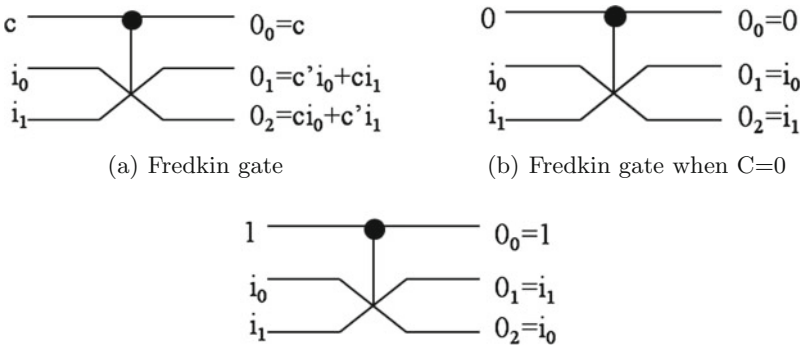
A conservative logic gate is a multiple-output logic element in which the number of 1s at the inputs is equal to that of the corresponding outputs. According to [18,61], a conservative logic circuit can be considered as a directed graph whose nodes are conservative logic gates, and the edges are wires of arbitrary lengths. The fanout at the output is not allowed in conservative logic circuits. A conservative logic network can be reversible in nature if the one-to-one mapping is maintained between the inputs and the outputs, while it will be irreversible in nature if one-to-one mapping is not preserved. Researchers in [27,60,61] have proved that: (i) in the event of unidirectional stuck-at-faults in a conservative logic network, either the number of 1s in its output set will differ from the number of 1s in its input set, or the output set is correct; (ii) in a conservative logic network the two vector test set, all 1s and all 0s, provide 100 % coverage for unidirectional stuck-at faults. Any stuck-at-1 fault in the conservative logic circuit can be detected by setting all inputs to 0s followed by subsequent checking of the outputs for the presence of any 1s. Any stuck-at-0 faults can be detected by setting all inputs to 1s followed by subsequent checking of outputs for the presence of any 0s. The comprehensive proofs can be referred in [27,60,61].

2.1 Conservative Reversible Fredkin Gate

The Fredkin gate is a popularly used reversible conservative logic gate, first proposed by Fredkin and Toffoli in [18]. The Fredkin gate shown in Fig. 1(a) can be described as a mapping (c, i_0, i_1) to ($o_0 = c$, $o_1 = c'i_0 + ci_1$,$o_2 = ci_0 + c'i_1$), where c, i_0, i_1 are the inputs and o_0, o_1 and o_2 are the outputs, respectively. Table 1 shows the truth table for the Fredkin gate which demonstrates that Fredkin gate is reversible and conservative in nature, that is, it has unique input and output mapping and also has the same number of 1s in the outputs as in the inputs. Fredkin gate is also called as controlled swap gate as it can swap two input bits i_0, i_1 when $c_0 = 1$. The controlled swap operation of the Fredkin gate is illustrated in Fig. 1(b), (c).

2.2 Basics of QCA Computing

In this work, the conservative logic gates are implemented in the QCA nanotechnology, thus we are also providing the introductory material on QCA computing. A QCA cell is a coupled dot system in which four dots are at the vertices of a square. The cell has two extra electrons that occupy the diagonals within the cell due to electrostatic repulsion. The cell polarization P measures the charge



(a) Fredkin gate

(b) Fredkin gate when C=0

(c) Fredkin gate when C=1

Fig. 1. Fredkin gate and its working mode

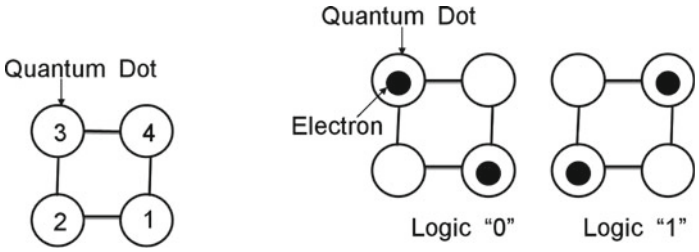
Table 1. Truth table for Fredkin gate

c	i_0	i_1	o_0	o_1	o_2
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	1	1	1	0
1	1	0	1	0	1
1	1	1	1	1	1

distribution along diagonal axes and is given by Eq. 1 (here Pi denotes the electronic charge at dot i). When electrons are in dots 1 and 3, P = −1 (Logic '0') and when electrons in dots 2 and 4, P = +1 (Logic '1') [70]. Figure 2(a) and (b) illustrate the 4 quantum dots in a QCA cell, and the implementation of logic '0' and logic '1' in a QCA cell, respectively.

$$P = \frac{(P2 + P4) - (P1 + P3)}{P1 + P2 + P3 + P4} \tag{1}$$

The basic QCA device is the majority voter (MV) or majority gate, which is represented as $Y = X_1X_2 + X_2X_3 + X_1X_3$, where Y is the majority of the inputs X_1, X_2 and X_3. Figure 3(a) shows the majority voter. The majority voter can be made to work as an AND gate or as an OR gate, by setting one of the inputs as '0' and '1', respectively (For example, if $X_3 = 0$ we will get $Y = X_1X_2$. Similarly if $X_3 = 1$, we will get $Y = X_1 + X_2$). Another important gate in QCA is the inverter, which is formed when a QCA cell, say cell-1 is placed 45 ° to another QCA cell, for example cell-0, cell-1 gets the inverse value of cell-0. There can be many ways of designing the QCA inverter, one of which is shown in Fig. 3(b). In QCA computing, signal transfer is made through wires that



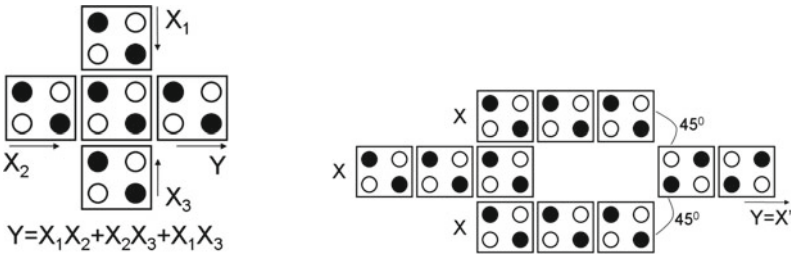
(a) QCA 4 Dots (b) QCA cell working as logic '1' and logic '0'

Fig. 2. QCA cell and logic operation

are of two types (i) Binary wire, (ii) Inverter chain. The electrons in adjacent QCA cells interact with each other resulting in propagation of the polarization from one cell to another. Thus, a QCA wire can be formed by arranging the QCA cells in a series in which all the neighboring cells will get the polarization of the driver cell (input). The binary wire is shown in Fig. 4(a). Two wires in QCA can cross without interaction. This is because QCA provides an inverter chain of QCA cells, in which the dots in each cell are rotated by 45° (This is not the same as in QCA inverter). Each cell in this arrangement has opposite polarization of their neighbors as they interact inversely. The inverter chain is shown in Fig. 4(b). In QCA, when a binary wire crosses the inverter chain, there is no interaction between the two; hence the signals in the inverter chain and binary wire can pass over each other. In QCA computing, the clock helps in the synchronization of circuits and provides the power required for functionality. QCA clocking consists of four phases: switch, hold, release and relax, as shown in Fig. 5 [25,34]. During the switch phase, the barriers are raised and the cells become polarized, depending on the state of its adjacent cell. The states of the cells are fixed during this stage. During the hold phase, the barriers are maintained at a high value. This helps the outputs to drive the inputs of the next stage, which is in the switching phase. In the release phase, the barriers are lowered and the cells are allowed to relax to an unpolarized state. During the relaxed phase, the cells remain in an unpolarized neutral state. The cells in QCA are connected to 4 clocking zones, each lagging behind by 90° in phase. QCA clocking helps in the successive transfer of information from one clock zone to the next. Therefore, we have information flow from the input to the output in a pipelined fashion [1].

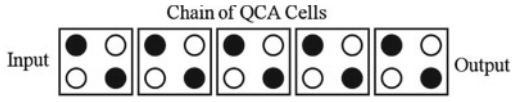
2.3 Related Work

The research on reversible logic is expanding towards both design and synthesis. In the synthesis of reversible logic circuits there has been several interesting attempts in the literature such as in [22,40,51,59,75]. The researchers have addressed the optimization of reversible logic circuits from the perspective of quantum cost and the number of garbage outputs. Recently, in [20,21]

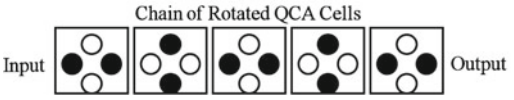


(a) Majority voter (b) Inverter

Fig. 3. QCA majority voter (MV) and inverter devices



(a) Binary wire



(b) Inverter chain

Fig. 4. QCA binary wire and inverter chain

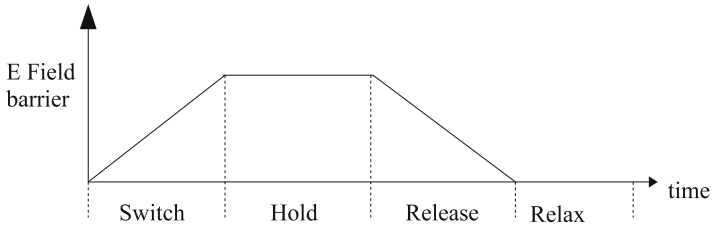


Fig. 5. QCA 4 phase clocking

interesting contributions are made towards deriving exact minimal elementary quantum gate realization of reversible combinational circuits. Any nanotechnology having applications of reversible logic such as nano-CMOS devices, NMR based quantum computing, or low power molecular QCA computing, all are susceptible to high device error rates. This attracted the attention of researchers towards testing of reversible logic circuits. In [48], it has been shown that for reversible logic circuits, the test set that detects all single stuck-at faults can also detect multiple stuck-at faults. In [50], four fault models for reversible circuits, viz., single missing gate fault, the repeated-gate fault, the multiple missing gate fault and the partial missing-gate fault are proposed based on ion-trap quantum

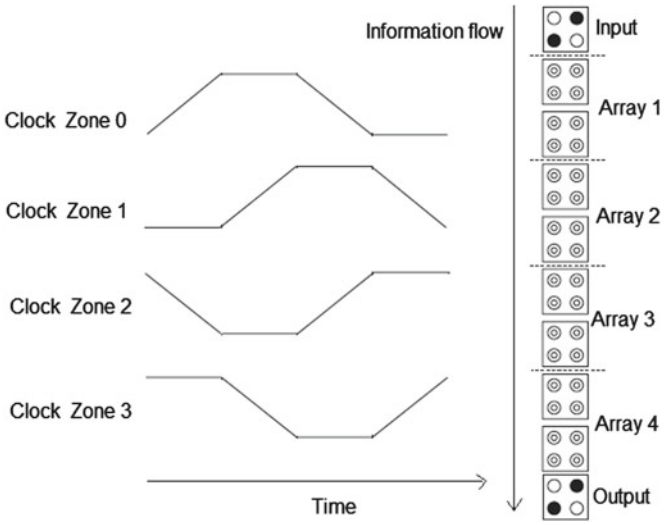
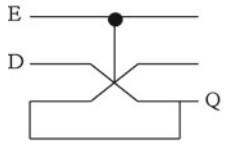


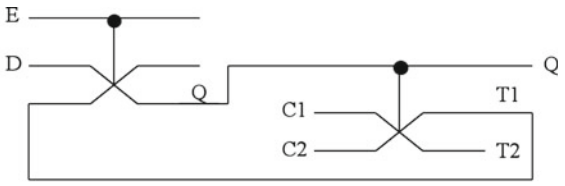
Fig. 6. Information flow in a QCA wire

computing at logical level. In [78], a new fault model called crosspoint fault model is proposed along with the ATPG method. In [54], a universal test set is proposed for detection of missing-gate faults in reversible circuits. In [8], a DFT methodology for detecting bridging faults in reversible logic circuits is proposed. Recently, the design of reversible finite field arithmetic circuits with error detection is also proposed [41]. An online testing methodology for reversible circuits using a combination of R1 gate along with R2 gate (a 4*4 Feynman Gate) is proposed in [72] while in [39] an automatic conversion of any given reversible circuit into an online testable circuit that can detect online the single-bit errors, including soft errors in the logic blocks is presented. The online testing methodology of reversible logic circuits is also addressed in [14]. In our recent work we addressed the concurrent testing of single missing faults in QCA circuits based on reversible logic [65,66]. With respect to the work on design of reversible sequential circuits, various interesting contributions are made in which the designs are optimized in terms of various parameters such as the number of reversible gates, garbage outputs, quantum cost, delay etc. [11,18,39,43,57,67,69].

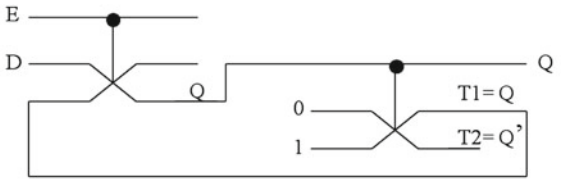
To the best of our knowledge the offline testing of faults in reversible sequential circuits is not addressed in the literature. In this work, we present the designs of reversible sequential circuits that can be tested by using only two test vectors, all 0s and all 1s for any unidirectional stuck-at-faults. Further, the approach of fault testing based on conservative logic is extended towards the design of non-reversible sequential circuits based on a new conservative logic gate called multiplexer conservative QCA gate (Mx-cqca).



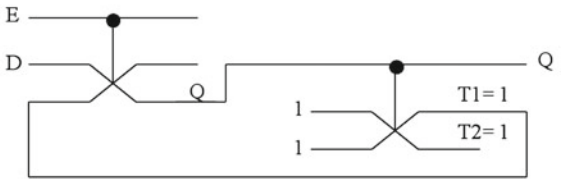
(a) Fredkin gate based D latch



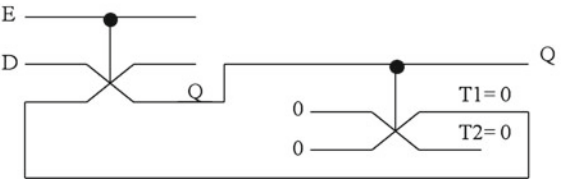
(b) Fredkin gate based D latch with control signals C1 and C2



(c) Fredkin gate based D Latch in normal mode: C1=0 and C2=1



(d) Fredkin gate based D latch in test mode for stuck-at-0 fault : C1=1 and C2=1



(e) Fredkin gate based D latch in test mode for stuck-at-1 fault: C1=0 and C2=0

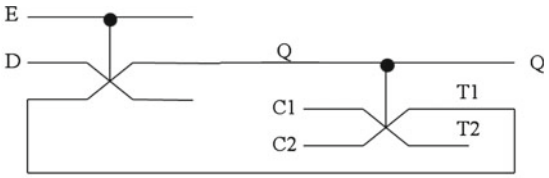
Fig. 7. Design of testable reversible D latch using conservative Fredkin gate

3 Design of Testable Reversible Latches

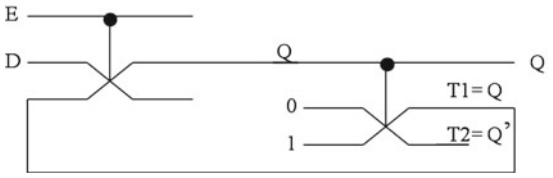
The characteristic equation of the D latch can be written as $Q^+ = D \cdot E + \bar{E} \cdot Q$. In the proposed work, E (Enable) refers to the clock and is used interchangeably in place of clock. When the enable signal (clock) is 1, the value of the input D is reflected at the output that is $Q^+ = D$. While, when E = 0 the latch maintains its previous state, that is $Q^+ = Q$. The reversible Fredkin gate has two of its outputs working as 2:1 MUxes, thus the characteristic equation of the D latch can be mapped to the Fredkin gate (F). Figure 7(a) shows the realization

of the reversible D latch using the Fredkin gate. But fan-out is not allowed in conservative reversible logic. Moreover, the design cannot be tested by two input vectors all 0s and all 1s because of feedback, as the output Q would latch 1 when the inputs are toggled from all 1s to all 0s and could be misinterpreted as stuck-at-1 fault.

In this work, we propose to cascade another Fredkin gate to output Q as shown in Fig. 7(b). The design has two control signals, C1 and C2. The design can work in two modes: (a) normal mode; (b) test mode. *Normal Mode*: The normal mode is shown in Fig. 7(c) in which we will have C1C2 = 01 and the design will work as a D latch without any fanout problem. *Test Mode (Disrupt the Feedback)*: In test mode , when C1C2 = 00 as shown in Fig. 7(d) the design will be testable with all 0s input vectors, as output T1 will become 0 resulting in the testable design with all 0s input vectors. Thus any stuck-at-1 fault can be detected. When C1C2 = 11 as shown in Fig. 7(e) the output T1 will become 1 and the design will become testable with all 1s input vectors for any stuck-at-0 fault. It can seen from the illustration that C1 and C2 will disrupt the feedback in test mode, and in normal mode will take care of the fan-out. Thus our proposed design works as a reversible D latch and can be tested with only two test vectors, all 0s and all 1s, for any stuck-at fault by utilizing the inherent property of conservative reversible logic.



(a) Fredkin gate based negative enable D latch with control signals C1 and C2



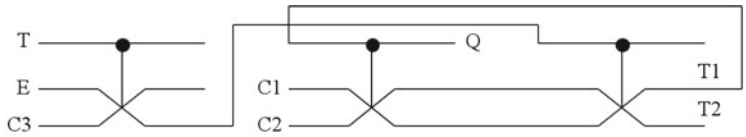
(b) Fredkin gate based negative enable D latch in normal mode

Fig. 8. Design of testable negative enable D latch using conservative Fredkin gate

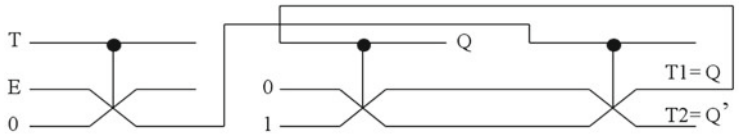
3.1 Design of Testable Negative Enable Reversible D Latch

A negative enable reversible D latch will pass the input D to the output Q when E = 0; otherwise maintains the same state. The characteristic equation of the negative enable D latch is $Q^+ = D \cdot \bar{E} + E \cdot Q$. This characteristic equation of

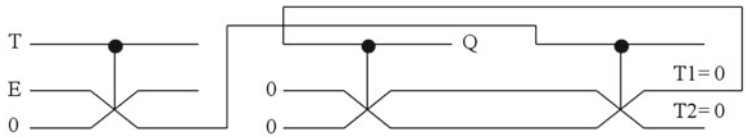
the negative enable reversible D latch can be mapped to the 2nd output of the Fredkin gate as shown in Fig. 8(a). The second Fredkin gate in the design take cares of the fanout. The second Fredkin gate in the design also makes the design testable by two test vectors all 0s and all 1s by breaking the feedback based on control signals C1 and C2, as illustrated above for positive enable reversible D latch. The working of the testable negative enable reversible D latch in normal mode is illustrated in Fig. 8(b). The negative enable D latch is helpful in the design of testable reversible master-slave flip-flops. This is because it can work as a slave latch in the testable reversible master-slave flip-flops in which no clock inversion is required. The details of which are discussed in the section describing reversible master-slave flip-flops.



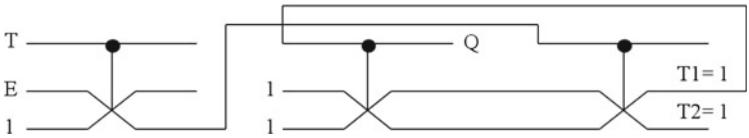
(a) Fredkin gate based T latch with control signals C1, C2 and C3, where C3 helps in realizing the AND function while C1 and C2 operates the test mode as well as the normal mode



(b) Fredkin gate based T latch in normal mode: C1=0 and C2=1



(c) Fredkin gate based T Latch in test mode for detecting any stuck-at-0 fault: C1=1 and C2=1



(d) Fredkin gate based T Latch in test mode for detecting any stuck-at-1 fault: C1=0 and C2=0

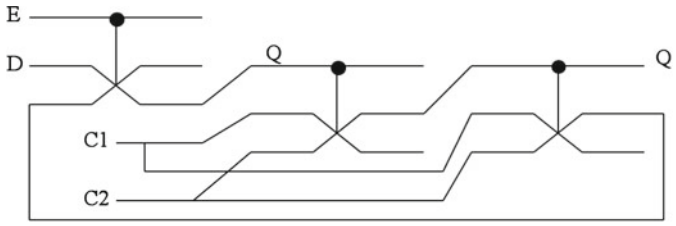
Fig. 9. Design of testable reversible T Latch using conservative Fredkin gate

3.2 Design of Testable Reversible T Latch

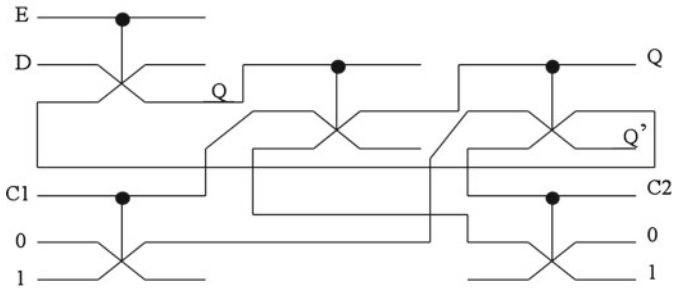
The characteristic equation of the T latch can be written as $Q^+ = (T \cdot Q) \cdot E + \bar{E} \cdot Q$. But the same result can also be obtained from $Q^+ = (T \cdot E) \oplus Q$. The T(toggle) latch is a complementing latch which complements its value when T = 1, that is when T = 1 and E = 1 we have $Q^+ = Q'$. When T = 0, the T latch maintains its state and we have no change in the output. Figure 9(a) shows the proposed design of reversible testable T latch with C1, C2, and C3 as control signals. The control signal C3 helps to realize the reversible AND function as we can generate $T \cdot E$ when C3 = 0, at one of the outputs of the Fredkin gate as illustrated in Fig. 9(b). C1 and C2 are the main control signals that help in breaking the feedback to make the design testable as well as in enabling the normal mode of operation. In normal mode, as illustrated in Fig. 9(b), the values of the control signals will be C1 = 0 and C2 = 1 thus helping in realizing the function $(T \cdot E) \oplus Q$. In test mode, when C1 = 0 and C2 = 0 as shown in Fig. 9(d) it will break the feedback and test the design with all 0s test vector for any stuck-at-1 fault, while when C1 = 1 and C2 = 1 as shown in Fig. 9(c) it will break the feedback and helps in testing the design with all 1s test vector for any stuck-at-0 fault. The other types of reversible testable latches based on conservative reversible logic such as the JK latch and the SR latch can be designed similarly, thus are not discussed in this work.

3.3 Design of Testable Asynchronous Set/Reset D Latch

The design of the asynchronously set/reset D latch is shown in Fig. 10(a). The design has 3 Fredkin gates. We can observe that the first Fredkin gate maps the D latch characteristic equation, while the second and the third Fredkin gates take care of the fan-out and also help in asynchronous set/reset of the output Q. The design has two control inputs C1 and C2. When C1 = 0 and C2 = 1, the design works in normal mode implementing the D latch characteristic equation. When C1 = 0 and C2 = 0, the second and third Fredkin gates will reset the output Q to 0. When C1 = 1 and C2 = 1, the design will be set to Q = 1. Thus, the control inputs help the design to work in various modes. But the design shown in Fig. 10(a) has fan-out of more than one in C1 and C2 inputs which is prohibited in reversible logic. Thus, a modified design of the D latch with asynchronous set/reset capability in which there is no fan-out is shown in Fig. 10(b). There is a special characteristic of the reversible D latch design shown in Fig. 10(b). The design shown in Fig. 10(b) has the control signals C1 and C2 which helps in disrupting the feedback. For example, the feedback is disrupted when C1C2 = 00; the feedback output Q resets to 0 which makes the reversible D latch testable with all 0s test vector for any stuck-at-1 fault. Similarly, when C1C2 = 11 the output Q sets to 1 and the design becomes testable with all 1s test vector for any stuck-at-0 fault. Thus, the proposed reversible D latch design with asynchronous set/reset significantly reduces the testing cost. Thus if we design asynchronous set/reset D latch only with Fredkin gates we can have the significant testing benefits.



(a) Design of testable Fredkin gate based asynchronous set/reset D Latch with fanout

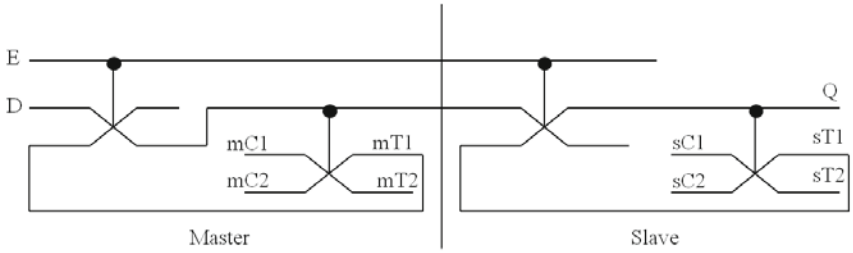


(b) Design of testable Fredkin gate based reversible asynchronous set/reset D latch without fanout. For C1C2=01, the asynchronous set/reset D latch operates in normal mode. For C1C2=00, the asynchronous set/reset D latch operates in test mode for detecting any stuck-at-1 faults. For C1C2=11, the asynchronous set/reset D latch operates in test mode for detecting any stuck-at-0 faults.

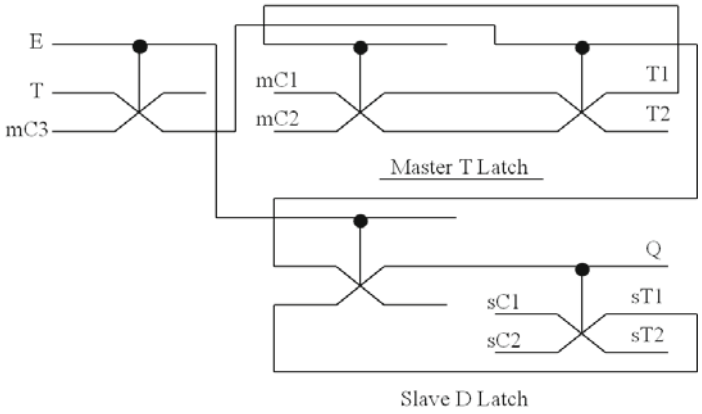
Fig. 10. Design of testable reversible asynchronous set/reset D latch

4 Design of Testable Master-Slave Flip-Flops

In the existing literature, the master-slave strategy of using one latch as a master and the other latch as a slave is used to design the reversible flip-flops [11,57, 68,69]. In this work, we have proposed the design of testable flip-flops using the master-slave strategy that can be detected for any stuck-at faults using only two test vectors all 0s and all 1s. Figure 11(a) shows the design of the master-slave D flip-flop in which we have used positive enable Fredkin gate based testable D latch shown in Fig. 7(b) as the master latch, while the slave latch is designed from the negative enable Fredkin gate based testable D latch shown earlier in Fig. 8(a). The testable reversible D flip-flops has four control signals mC1, mC2, sC1 and sC2. mC1 and mC2 control the modes for the master latch, while sC1 and sC2 control the modes for the slave latch. In the normal mode, when the design is working as a master-slave flip-flop the values of the control signals will be $mC1 = 0$ and $mC2 = 1$, $sC1 = 0$ and $sC2 = 1$ (similar to values of the control signals C1 and C2 earlier described for the testable D latches).



(a) Fredkin gate based testable reversible master-slave D flip-flop



(b) Fredkin gate based testable reversible master-slave T flip-flop

Fig. 11. Fredkin gate based testable reversible master-slave flip-flops

In the test mode:

1. to make the design testable with all 0s input vectors for any stuck-at-1 fault, the values of the control signals will be mC1 = 0 and mC2 = 0, sC1 = 0 and sC2 = 0. This will produce the outputs mT1 and sT1 as 0 which results in disrupting the feedback and the design becomes testable with all 0s input vectors for any stuck-at-1 fault.
2. to make the design testable with all 1s input vectors for any stuck-at-0 fault, the values of the control signals will be mC1 = 1 and mC2 = 1, sC1 = 1 and sC2 = 1. This will result in outputs mT1 and st1 to have the value of 1, disrupting the feedback, and resulting in the design testable with all 1s input vectors for any stuck-at-0 fault.

The other type of master-slave flip-flops such as the testable master-slave T flip-flop, testable master-slave JK flip-flop and testable master-slave SR flip-flop can be designed similarly in which master is designed using the positive enable corresponding latch, while the slave is designed using the negative enable Fredkin gate based D latch. For example, as illustrated in Fig. 11(b), in the design of

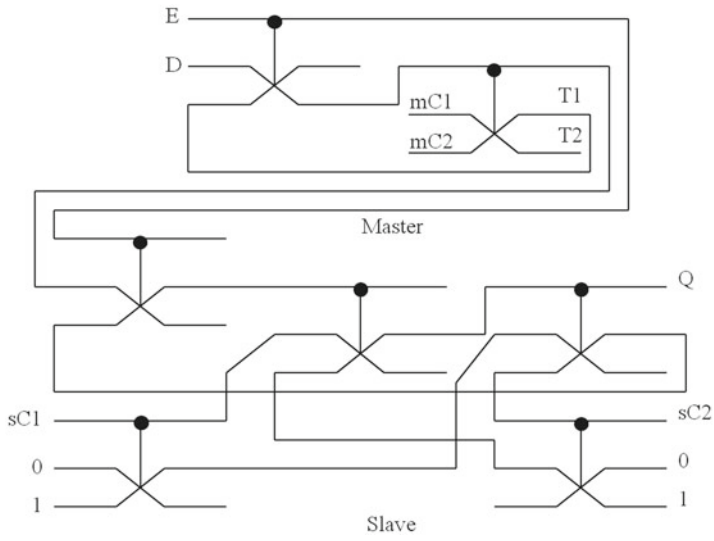


Fig. 12. Fredkin gate based testable reversible asynchronous set/reset master-slave D flip-flop

master-slave T flip-flop the master is designed using the positive enable T latch, while the slave is designed with the negative enable D latch.

The reversible design of the master-slave D flip-flop with asynchronous set/reset is shown in Fig. 12. The design contains positive enable testable D latch shown in Fig. 7(b) as the master latch and negative enable asynchronous set/reset D latch shown in Fig. 10(b) as the slave latch.

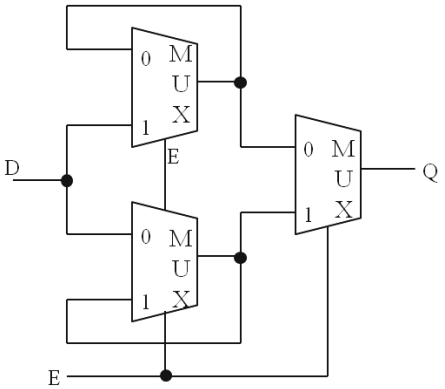
5 Design of Testable Reversible Double Edge Triggered(DET) Flip-Flops

The double edge triggered flip-flop is a computing circuit that sample and store the input data at both the edges, that is at both the rising and the falling edge of the clock. The master-slave strategy is the most popular way of designing the flip flop. In the proposed work E (Enable) refers to the clock and is used interchangeably in place of clock. In the negative edge triggered master-slave flip-flop when $E = 1$ (the clock is high), the master latch passes the input data while the slave latch maintains the previous state. When $E = 0$ (the clock is low), the master latch is in the storage state while the slave latch passes the output of the master latch to its output. Thus, the flip-flop does not sample the data at both the clock levels and waits for the next rising edge of the clock to latch the data at the master latch.

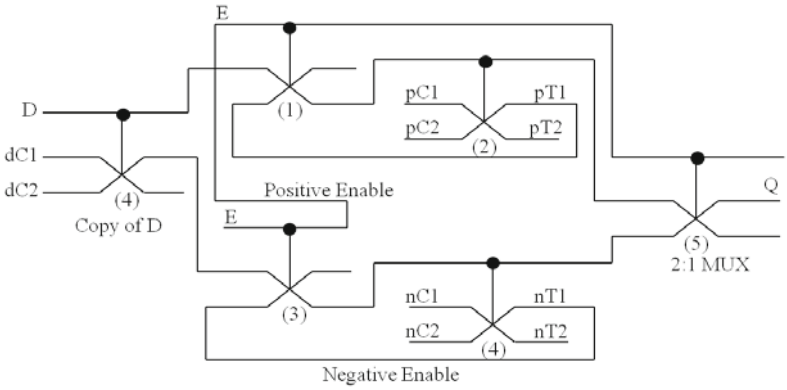
In order to overcome the above problem, researchers have introduced the concept of double edge triggered (DET) flip-flops which sample the data at both the edges. Thus DET flip-flops can receive and sample two data values in a clock

period thus frequency of the clock can be reduced to half of the master-slave flip flop while maintaining the same data rate. The half frequency operations make the DET flip flops very beneficial for low power computing as frequency is proportional to power consumption in a circuit. The DET flip-flop is designed by connecting the two latches, viz., the positive enable and the negative enable in parallel rather than in series. The conventional design of the DET flip-flop is illustrated in Fig. 13(a) [49]. The 2:1 MUX at the output transfer the output from one of these latches which is in the storage state (holding its previous state). The equivalent testable reversible design of the DET flip flop is proposed in this work and is shown in Fig. 13(b).

In the proposed design of testable reversible DET flip-flop, the positive enable testable reversible D latch and the negative enable testable reversible D latch are arranged in parallel. The Fredkin gates labeled as 1 and 2 form the positive

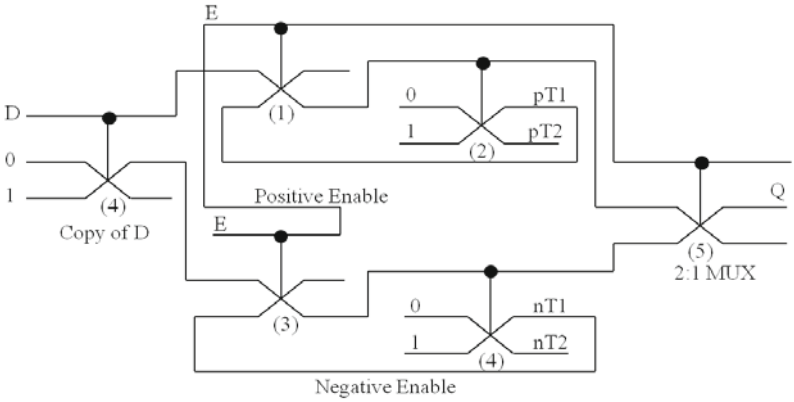


(a) Conventional DET flip-flop

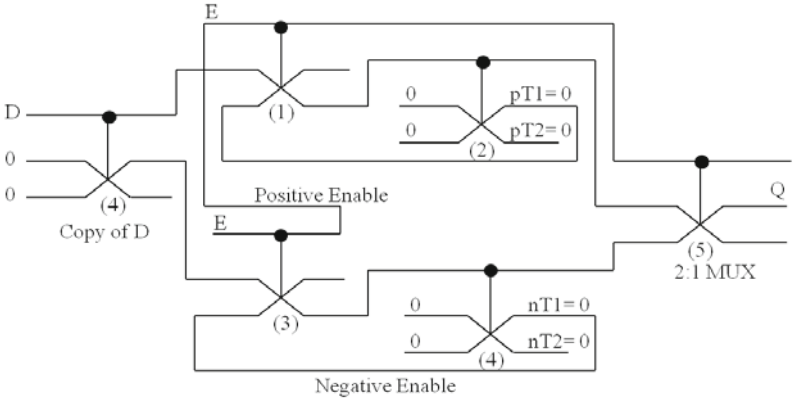


(b) Fredkin gate based DET flip-flop

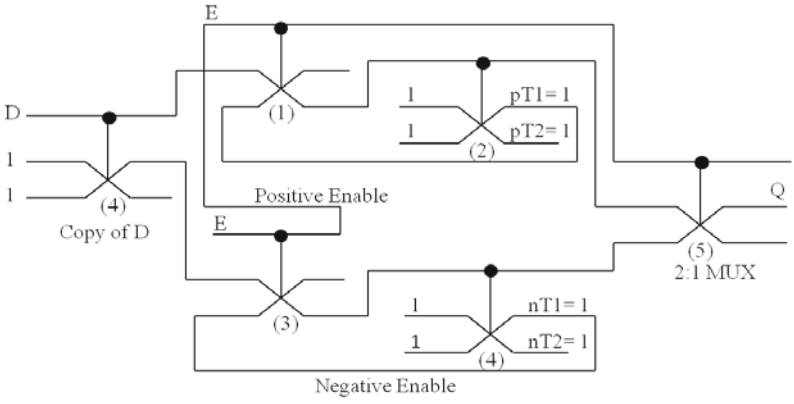
Fig. 13. Fredkin gate based double edge triggered (DET) flip-flop



(a) Normal mode



(b) Test mode for stuck-at-1 fault



(c) Test mode for stuck-at-0 fault

Fig. 14. Working of Fredkin gate based double edge triggered flip-flop

enable testable D latch while the Fredkin gates labeled as 3 and 4 form the negative enable testable D latch. In reversible logic fanout is not allowed so the Fredkin gate labeled as 1 is used to copy the input signal D. The Fredkin gate labeled as 6 works as the 2:1 MUX and transfer the output from one of these testable latches (negative enable D latch or the positive enable D latch) that is in the storage state (is holding its previous state) to the output Q. In the proposed design of testable reversible DET flip-flop pC1 and pC2 are the control signals of the testable positive enable D latch, while nC1 and nC2 are the control signals of the testable negative enable D latch. Depending on the values of the pC1, pC2, nC1 and nC2 the testable DET flip-flops works either in normal mode or in the testing mode.

1. *Normal Mode*: The normal mode of the DET flip-flop is illustrated in Fig.14(a) in which the pC1 = 0, pC2 = 1, nC1 = 0 and nC2 = 1. The pC1 = 0, pC2 = 1 helps in copying the output of the positive enable D latch thus avoiding the fanout while the nC1 = 0 and nC2 = 1 helps in copying the output of the negative enable D latch thus avoiding the fanout.
2. *Test Mode*: There will be two test modes:
   (a) *All 1s Test Vector*: This mode is illustrated in Fig. 14(c) in which control signals will have values as pC1 = 1, pC2 = 1, nC1 = 1 and nC2 = 1. The pC1 = 1 and pC2 = 1 help in breaking the feedback of the positive enable D latch, while the nC1 = 1 and nC2 = 1 help in breaking the feedback of the negative enable D latch. This makes the design testable by all 1s test vector for any stuck-at-0 fault.
   (b) *All 0s Test Vector*: This mode is illustrated in Fig. 14(b) in which the control signals will have values as pC1 = 0, pC2 = 0, nC1 = 0 and nC2 = 0. The pC1 = 0 and pC2 = 0 help in breaking the feedback of the positive enable D latch while the nC1 = 0 and nC2 = 0 help in breaking the feedback of the negative enable D latch. This makes the design testable by all 0s test vector for any stuck-at-1 fault.

6 Design of Testable Reversible Complex Sequential Circuits

The set of sequential building blocks proposed in this work can be used to build complex sequential circuits that provide the capability of testing a sequential circuit using two test vectors. The proposed sequential building blocks can be used to implement various types of sequential circuits such as shifters, sequence detectors, counters and systolic circuits etc. We have illustrated the method with examples of design of reversible ring counter and reversible Johnson counter using the proposed sequential building blocks.

6.1 Design of Testable Reversible Ring Counter

A testable reversible ring counter can be designed by cascading reversible master slave D flip-flops with asynchronous set/reset capability in which the output of

the last flip-flop is connected to the first flip-flop. The design of n bit testable reversible ring counter is illustrated in Fig. 15. In *normal working mode*, the ring counter needs to be initialized by setting the first master-slave flip flop to 1, while the remaining n-1 flip-flops need to be reset to 0. This will initialize the counter to a state $(10000\ldots0)_n$. In Fig. 15, this is performed as follows:

$$sc1_i \quad and \quad sc2_i = \begin{cases} 1 & \text{if } i = 0 \\ 0 & \text{if } 1 \leq i \leq \text{n} - 1 \end{cases}$$

The above values of $sc1_i$ and $sc2_i$ where $0 \leq i \leq n-1$ asynchronously sets the first testable reversible flip-flop to 1, while the n-1 testable reversible D flip-flops are asynchronously reset to 0. Once the counter is initialized, the values of $sc1_i$ and $sc2_i$ where $0 \leq i \leq n-1$ are changed:

$$sc1_i \quad and \quad sc2_i = \Big\{ 0 \quad and \quad 1 \quad \text{if } 0 \leq i \leq \text{n} - 1$$

Thus, in the proposed reversible n bit ring counter 1 bit is circulated so the state repeats every n clock cycles. For example, in a four bit testable reversible counter, the possible states for $Q_0Q_1Q_2Q_3$ will be 1000, 0100, 0010, and 0001.

The test mode of the reversible ring counter can be defined as follows:

(a) *All 1s Test Vector*: In this mode, all the inputs along with $sc1_i$ and $sc2_i$ where $0 \leq i \leq n-1$ are set to the value of 1 to detect any stuck-at-0 fault.
(b) *All 0s Test Vector*: In this mode, all the inputs along with $sc1_i$ and $sc2_i$ where $0 \leq i \leq n-1$ are reset to the value of 0 to detect any stuck-at-1 fault.

6.2 Design of Testable Reversible Johnson Counter

A testable reversible Johnson counter can be designed by cascading reversible master slave D flip-flops with asynchronous set/reset capability in which the Q' output of the last flip-flop instead of Q is connected to the first flip-flop. The design of n bit testable reversible Fig. 16. In *normal working mode*, the Johnson counter needs to be initialized by reseting all n flip-flops. This will initialize the counter to a state $(00000\ldots0)_n$. In Fig. 16, this is performed as follows:

$$sc1_i \quad and \quad sc2_i = \Big\{ 0 \quad and \quad 0 \quad \text{if} \quad 0 \leq i \leq \text{n} - 1$$

The above values of $sc1_i$ and $sc2_i$ where $0 \leq i \leq n-1$ asynchronously reset the n testable reversible D flip-flops to 0. Once the counter is initialized, the values of $sc1_i$ and $sc2_i$ where $0 \leq i \leq n-1$ are changed:

$$sc1_i \quad and \quad sc2_i = \Big\{ 0 \quad and \quad 1 \quad \text{if} \quad 0 \leq i \leq \text{n} - 1$$

Thus, the proposed reversible n bit Johnson counter produces a couting sequence so the state repeats every 2n clock cycles. For example, in a four bit testable reversible Johnson counter, the possible states for $Q_0Q_1Q_2Q_3$ will be 0000, 1000,

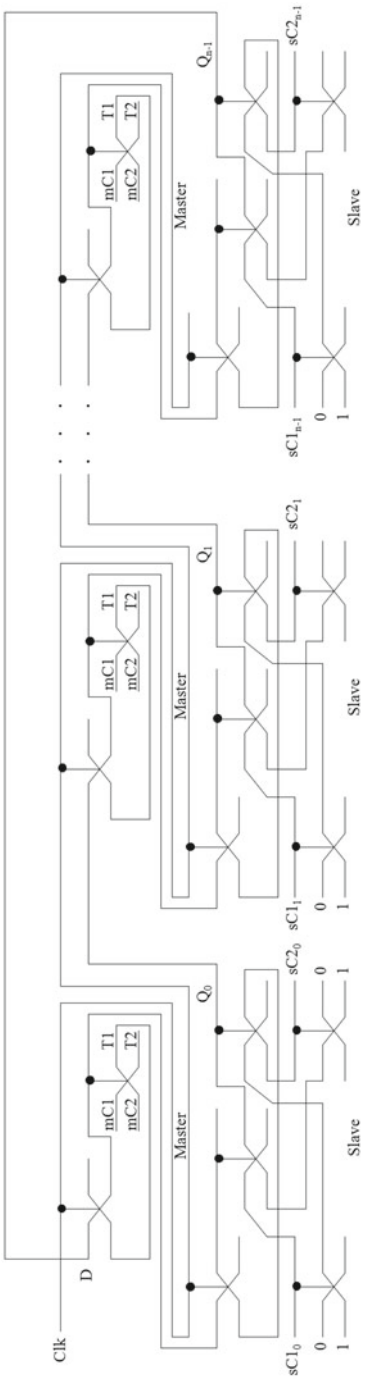


Fig. 15. Design of testable reversible ring counter

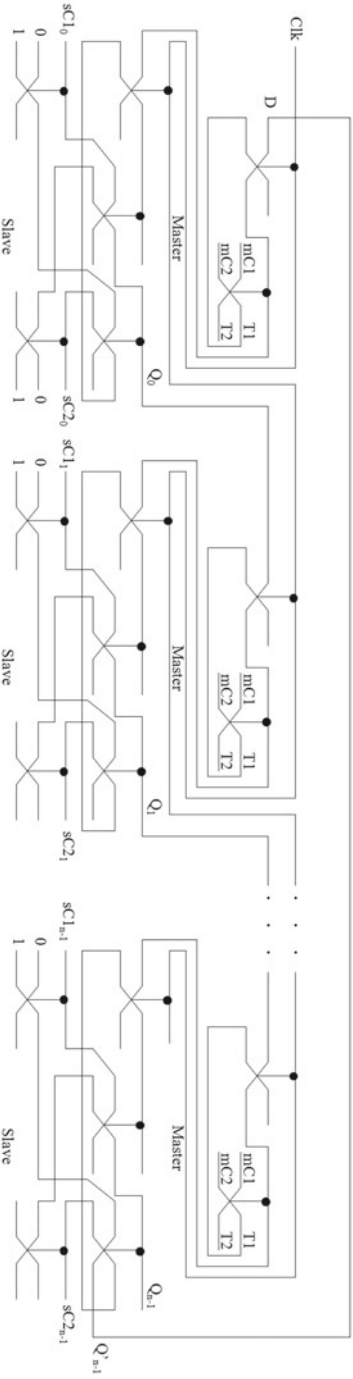


Fig. 16. Design of testable reversible Johnson counter

1100, 1110, 1111, 0111, 0011, 0001. The test mode of the reversible Johnson counter can be designed similar to the reversible ring counter as explained in the above section.

Similar design methodologies can be proposed based on the example study shown for testable reversible latches, flip-flops and counters to design other complex sequential circuits such as shift registers, memory based on proposed testable D flip-flop and reversible data path functional units.

7 Application of Two Vectors, All 0s and All 1s, Testing Approach to QCA Computing

QCA computing provides a promising technology to implement reversible logic gates. The QCA design of Fredkin gate is shown in Fig. 17 using the four-phase clocking scheme, in which the clocking zone is shown by the number next to z (z0 means clock 0 zone, z1 means clock 1 zone and so on). It can be seen that the Fredkin gate has two level majority voter (MV) implementation, and it requires 6 MVs and 4 clocking zones for implementation. The number of clocking zones in a QCA circuit represents the delay of the circuit (delay between the inputs and the outputs). Higher the number of clocking zones, lower the operating speed of the circuit [37].

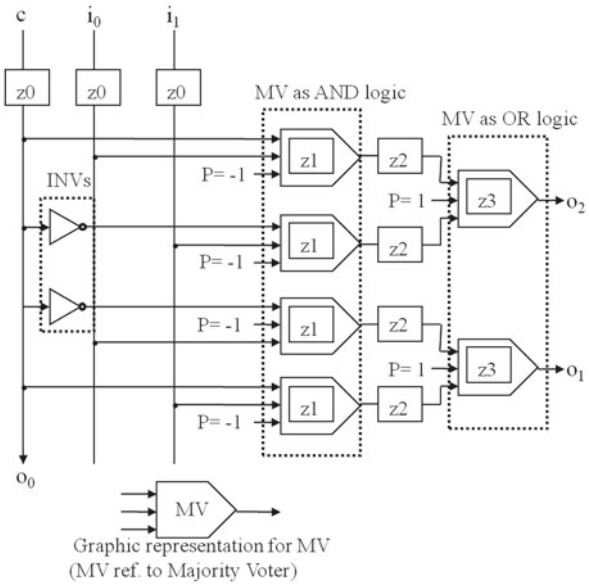


Fig. 17. QCA Design of Fredkin Gate using the four-phase clocking scheme, in which the clocking zone is shown by the number next to z (z0 means clock 0 zone, z1 means clock 1 zone and so on)

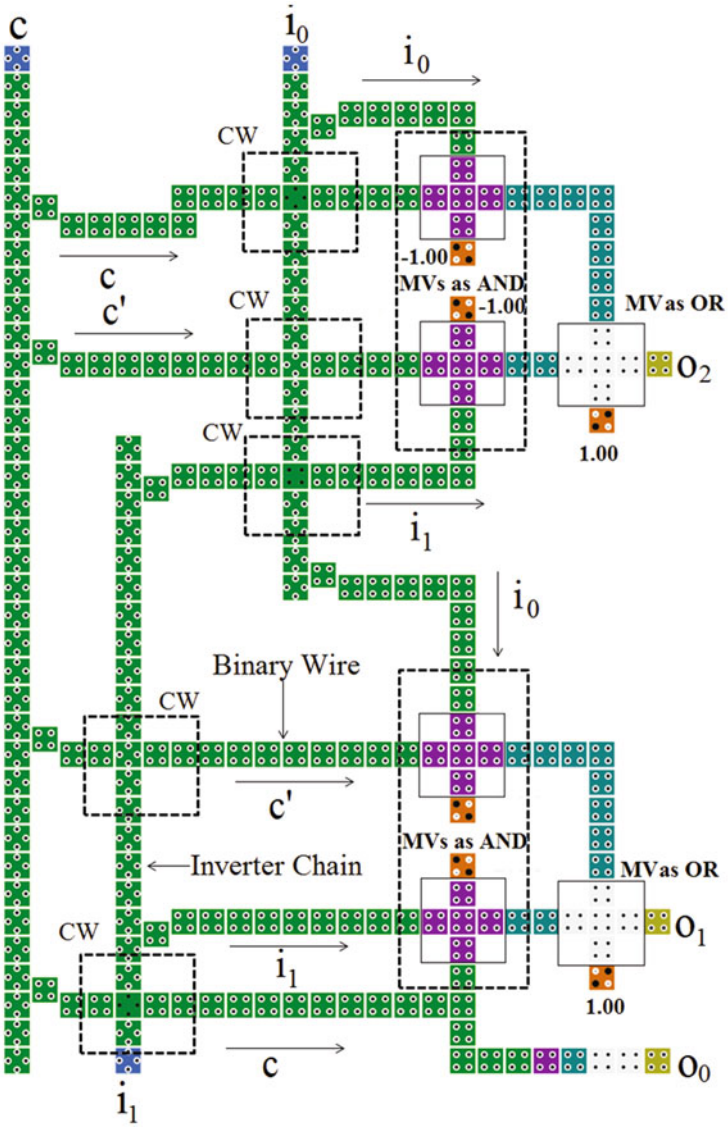


Fig. 18. QCA Layout of Fredkin Gate. CW represents cross wire resulted from intersection of binary wire and inverter chain, MV represents majority voter

In QCA manufacturing, defects can occur during the synthesis and deposition phases, although defects are most likely to take place during the deposition phase [62]. Researchers assume that QCA cells have no manufacturing defect; in metal QCA, faults occur due to cell misplacement. These defects can be characterized as cell displacement, cell misalignment and cell omission [23]. Researchers have

Table 2. Fault patterns in Fredkin gate

Input vector	Fault free	Fault patterns													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
a0	a0	a0	a0	a1	a1	a0	a0	a1	a2	a1	a1	a4	a2	a0	a2
a1	a1	a1	a1	a1	a1	a0	a1	a0	a1	a1	a1	a7	a3	a1	a3
a2	a2	a3	a2	a2	a3	a2	a0	a3	a2	a3	a3	a6	a0	a0	a2
a3	a3	a3	a3	a3	a3	a2	a1	a2	a1	a3	a3	a7	a1	a1	a3
a4	a4	a4	a5	a5	a5	a4	a4	a4	a4	a5	a4	a0	a4	a4	a4
a5	a6	a6	a7	a7	a7	a7	a6	a6	a6	a6	a6	a0	a6	a6	a6
a6	a5	a4	a4	a5	a4	a5	a5	a5	a5	a5	a5	a1	a5	a7	a5
a7	a7	a6	a6	a7	a6	a7	a7	a7	a7	a7	a7	a1	a7	a7	a7
Input vector	Fault free	Fault patterns													
		15	16	17	18	19	20	21	22	23	24	25	26	27	28
a0	a0	a0	a0	a1	a0	a2	a4	a4	a0	a0	a0	a0	a0	a0	a2
a1	a1	a1	a3	a0	a3	a3	a5	a1	a1	a1	a1	a0	a1	a1	a3
a2	a2	a2	a2	a3	a2	a0	a6	a6	a2	a2	a2	a2	a2	a2	a2
a3	a3	a3	a3	a2	a3	a1	a7	a3	a3	a1	a2	a2	a3	a3	a3
a4	a4	a5	a4	a6	a4	a6	a0	a4	a6	a4	a4	a4	a4	a6	a4
a5	a6	a7	a4	a4	a6	a6	a2	a0	a4	a6	a7	a6	a6	a6	a6
a6	a5	a5	a5	a7	a5	a7	a1	a5	a7	a7	a5	a5	a7	a7	a5
a7	a7	a7	a5	a5	a7	a7	a3	a1	a5	a7	a7	a7	a7	a7	a7

shown that molecular QCA cells are more susceptible to missing and additional QCA cell defects [42]. The additional cell defect is because of the deposition of an additional cell on the substrate. The missing cell defect is due to the missing of a particular cell. Researchers have been addressing the design and test of QCA circuits assuming the *single missing/additional cell defect model*. In [37], reversible logic was proposed as a means to detect single missing/additional cell defects. It was shown that reversible 1D array is C-testable. In [6], they address the robust coplanar crossing in QCA, proving that wires having rotated cells are thermally more stable.

In this section, we discuss how the QCA implementation of the Fredkin gate can be tested by only two test vectors, all 0s and all 1s, for the offline testing of any single missing/additional cell defect. The QCA layouts of the Fredkin gate is shown in Fig. 18. In the proposed work, the QCA layout of the Fredkin gate is converted into the corresponding hardware description language notations using the HDLQ Verilog library [46]. The HDLQ design tool consists of a Verilog HDL library of QCA devices, i.e., MV, INV, fan-out, Crosswire, L-shape wire with fault injection capability. The HDLQ model of the QCA layout of the Fredkin gate is shown in Fig. 19. In the Fig. 19, FO represents the fanout QCA device, LSW represents the L-shape wire, INV represents the QCA inverter, CW represents the crosswire, MV represents the majority voter. Thus it can be seen that modeled QCA layout of the Fredkin gate has 4 FOs, 2 INVs, 5 CWs, 9 LSWs and 6 MVs. The HDLQ modeled design of the Fredkin gate is

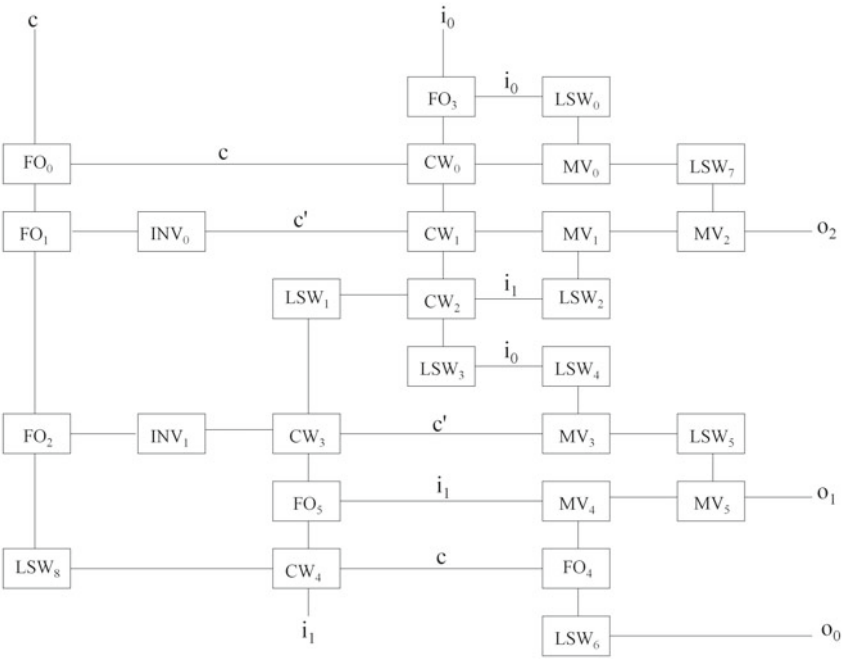


Fig. 19. Modeling of QCA Layout of Fredkin Gate. FO represents the fanout QCA device, LSW represents the L-shape wire, INV represents the QCA inverter, CW represents the crosswire and MV represents the majority voter

simulated for the presence of all possible single missing/additional cell defect in MVs (majority voters), INVs (Inverters), FOs (fanouts), Crosswires (CWs) and L-shape wires (LSWs). The design is simulated using the Verilog HDL simulator in the presence of faults to determine the corresponding outputs.

We conducted exhaustive testing of the HDLQ model of the Fredkin gate with 8 input patterns in the presence of all possible single missing/additional cell defect. Testing of the Fredkin gate generates 28 unique fault patterns at the output, as shown in Table 2. In the fault patterns study shown in the Table, ai is the 3 bit pattern with an equivalent decimal value of i. For example, a0 represents 000 (decimal 0) and a7 represents 111 (decimal 7). From fault tables we can see that there are 10 fault patterns 5, 6, 13, 15, 18, 23, 24, 25, 26, 27 that will produce the correct outputs for input vectors a0 (all 0s) and a7 (all 1s) even when there is a fault. Thus two test vectors a0 and a7 can only provide 64.28 % fault coverage. Thus in order to give the test vectors a0 and a7 100 % fault coverage we identified the logic devices in the HDLQ model of the Fredkin gate which can be replaced by their fault-tolerant counterpart. This will give the 100 % fault coverage for any single missing/additional cell defect to the two test vectors all 0s and all 1s. We observed that fanouts ($F0_2$ and $F0_3$), inverters (INV_1 and INV_2), crosswires (CW_4 and CW_2) and majority voters (MV_1, MV_3,

MV_4, MV_5 and MV_6) are devices in the QCA layout of the Fredkin gate that are making the design untestable by all 0s and all 1s test vectors. This work focuses on demonstrating the fault tolerant QCA circuits that provide ease of testability as the proposed sequential building blocks can be tested using only two test vectors. In the existing literature, several fault tolerant QCA components have been proposed such as Majority voters (MVs), Inverters (INVs), Fanouts (FOs), Crosswires (CWs) and L-shape wires (LSWs) [12,15,16,74]. Thus, these devices can be replaced by their fault tolerant counterparts in the QCA layout of the Fredkin gate to have the equivalent design that gives 100 % fault coverage to test vectors all 1s and all 0s. The HDLQ model of the Fredkin gate QCA layout having 100 % coverage for any single missing/additional cell defect to test vectors all 0s and all 1s is shown in Fig. 20. In Fig. 20, the shaded devices represent their fault tolerant counterparts. Thus, conservative logic QCA circuits based on our proposed QCA layout of the Fredkin gate show in Fig. 20, can be tested by all 0s and all 1s test vectors for presence single missing/additional cell defects.

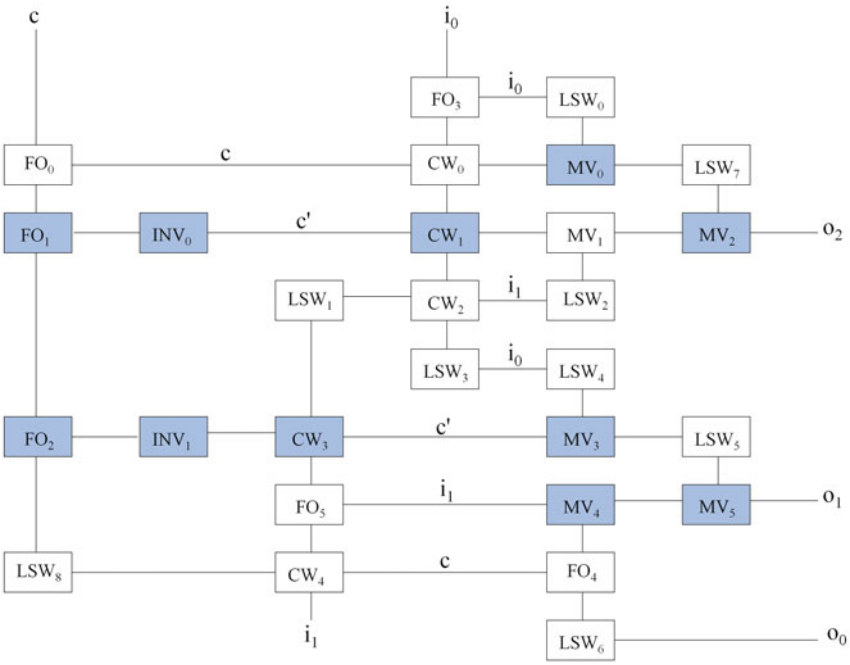


Fig. 20. QCA layout of the Fredkin gate testable with only all 0s and all 1s test vectors for any single missing/additional cell defect (the shaded devices represent their fault tolerant counterpart)

8 Proposed Multiplexer Conservative QCA Gate(MX-cqca)

For many of the designs, the designer could potentially be interested in using the testing advantages of conservative logic but saving the number of QCA cells. Thus, in this work we propose a new conservative logic gate that is conservative in nature but is not reversible. The proposed conservative logic gate is called multiplexer conservative QCA gate(MX-cqca) and has 3 inputs and 3 outputs. Mx-cqca has one of its outputs working as a multiplexer that will help in mapping the sequential circuits based on it, while the other two outputs work as AND and OR gates, respectively. The mapping of the inputs to outputs of the MX-cqca is: $o_0 = i_0 i_1$; $o_1 = i_0 i_1' + i_1 i_2$; $o_2 = i_2 + i_3$, where i_o, i_1 and i_2 are the inputs and o_o, o_1 and o_2 are the outputs, respectively. Figure 21 shows the block diagram of the MX-cqca gate. Table 3 shows the truth table of the MX-cqca gate. The table verifies the gate's conservative logic nature, i.e., the numbers of 1s in the inputs is equal to the number of 1s in the outputs. Figures 22 and 23 show the QCA design and layout of the proposed MX-cqca gate. From the QCA design, we can observe that the proposed MX-cqca gate requires 4 clocking zones and 5 majority gates for its QCA implementation. Table 4 shows the comparison between the proposed MX-cqca gate and the Fredkin gate in terms of area and number of QCA cells. The table illustrates that MX-cqca is better than the existing Fredkin gate for implementing multiplexer-based designs. The MX-cqca gate requires 5 majority voters and 218 QCA cells with an area of 0.71 um^2. Thus, it has 1 less majority gate, 1 less inverter, 11 % less QCA cells and 5.4 % less area compared to the Fredkin gate.

We also modeled the QCA layout of the MX-cqca gate using the HDLQ Verilog library for performing the fault testing. The HDLQ model of the QCA layout of the Fredkin gate is shown in Fig. 24. Thus it can be observe that modeled QCA layout have 4 FOs, 1 INV, 5 CWs, 8 LSWs and 5 MVs. We conducted exhaustive testing of the HDLQ model of the Mx-cqca gate with 8 input patterns in the presence of all possible single missing/additional cell defects. Testing of the Mx-cqca gate generates 24 unique fault patterns at the output, as shown in Table 5.

From fault tables we can see that there are 9 fault patterns 3, 7, 13, 17, 19, 20, 22, 23, 24 that will produce the correct outputs for test vectors a0 and a7 (all 0s and all 1s) even when there is fault. Thus two test vectors a0 and a7 can only provide 62.5 % fault coverage. Thus in order to give the test vectors 100 % fault

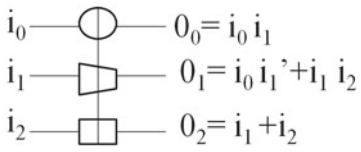


Fig. 21. Proposed MX-cqca gate

Table 3. Truth table of MX-cqca gate

i_0	i_1	i_2	o_0	o_1	o_2
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	0	1
0	1	1	0	1	1
1	0	0	0	1	0
1	0	1	0	1	1
1	1	0	1	0	1
1	1	1	1	1	1

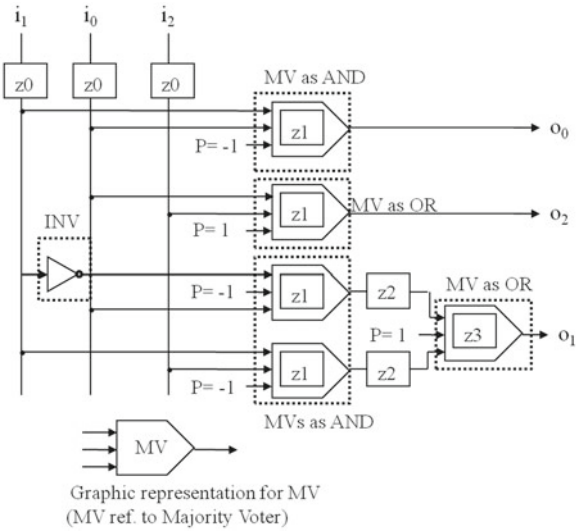


Fig. 22. QCA design of MX-cqca gate

coverage we identified the logic devices in the HDLQ model of the Mx-cqca gate which can be replaced by their fault-tolerant counterpart to give the 100 % fault coverage to two test vectors all 0s and all 1s, for any single missing/additional cell defect. We observed that fanout ($F0_3$), inverter (INV_1), crosswire (CW_4) and majority voters (MV_1, MV_2, MV_3, MV_4 and MV_5) are devices in the QCA layout of the Mx-cqca gate that are making the design untestable by all 0s and all 1s test vectors. In the existing literature, several fault tolerant QCA components have been proposed such as Majority voters (MVs), Inverters (INVs), Fanouts (FOs), Crosswires (CWs) and L-shape wires (LSWs) [12,15,16,74]. Thus, these devices can be replaced by their fault tolerant counterparts to have the equivalent design that gives 100 % fault coverage to test vectors, all 1s and all 0s, for any single missing/additional cell defect. The HDLQ model of the QCA layout having 100 % coverage for single missing/additional cell defect by all 0s and all 1s test

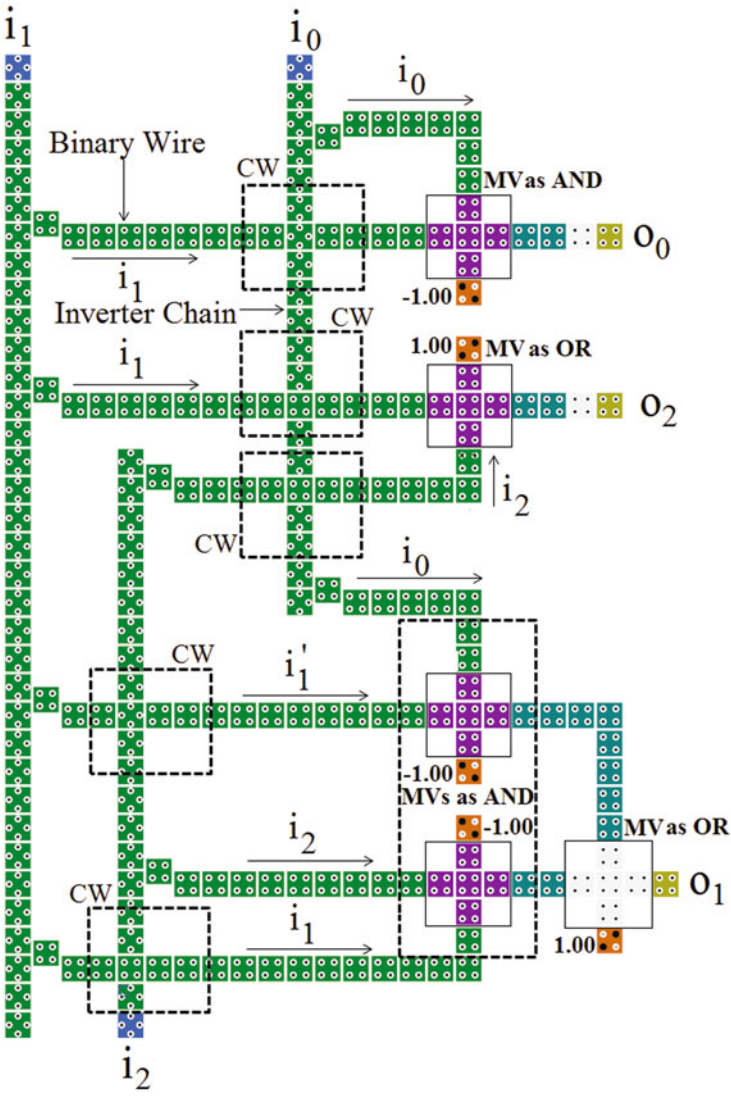


Fig. 23. QCA layout of MX-qca gate. CW represents cross wire resulted from intersection of binary wire and inverter chain, MV represents majority voter

vectors is shown in Fig. 25. The shaded devices in the Fig. 25 represent their fault tolerant counterparts. Thus, conservative logic QCA circuits based on the QCA layout of the Mx-cqca gate illustrated in Fig. 25, can be tested by all 0s and all 1s test vectors for presence of single missing/additional cell defect.

9 Design Methodology for Non-reversible Testable Design Based on MX-cqca Gate

The proposed conservative logic gate 'MX-cqca' is useful in designing any majority logic and multiplexer logic based testable non-reversible circuits. In the existing literature 13 standard functions are proposed to represent all three-variable Boolean functions [77]. These thirteen functions are widely used in

Table 4. A comparison of Fredkin and MX-cqca gates

	Fredkin	MX-cqca	Improvement % Mx-cqca to Fredkin
Majority voters	6	5	17
Inverters	2	1	50
Clk zones	4	4	–
Total cells	246	218	11.3
Area (L × W)	$0.4812\,\mu m \times 0.7698\,\mu m = 0.37\,\mu m^2$	$0.479\,\mu m \times 0.721\,\mu m = 0.35\,\mu m^2$	5.4

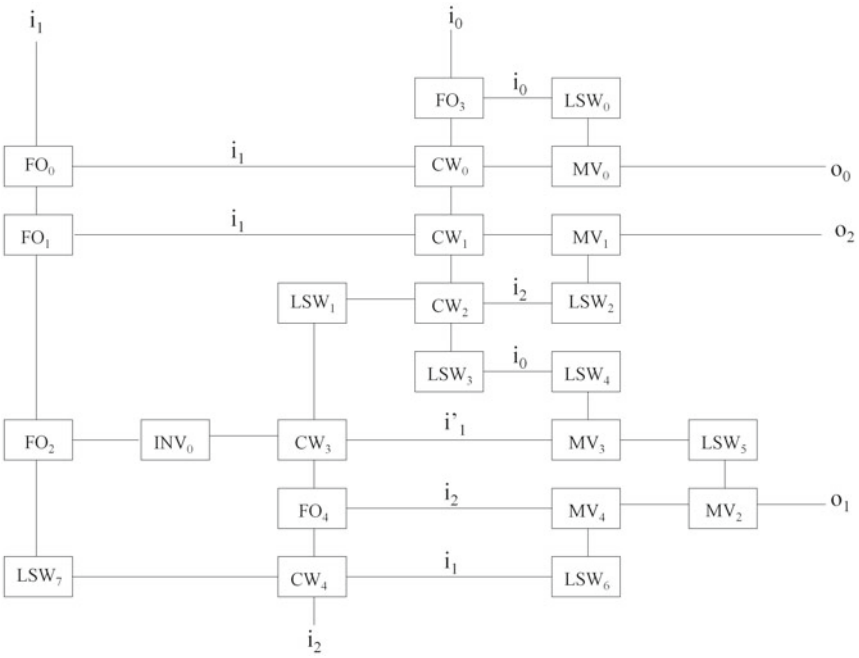


Fig. 24. Modeling of MX-qca gate QCA layout

Table 5. Fault patterns in Mx-cqca gate

Input vector	Fault free	Fault patterns											
		1	2	3	4	5	6	7	8	9	10	11	12
a0	a0	a0	a1	a0	a0	a0	a2	a0	a4	a1	a0	a0	a1
a1	a1	a1	a1	a1	a3	a1	a3	a1	a5	a0	a1	a1	a0
a2	a1	a1	a0	a1	a1	a5	a1	a5	a5	a3	a1	a3	a1
a3	a3	a3	a3	a3	a1	a7	a3	a7	a7	a1	a1	a1	a3
a4	a2	a6	a3	a0	a3	a2	a2	a2	a2	a3	a2	a2	a3
a5	a3	a7	a3	a1	a5	a3	a3	a3	a3	a2	a3	a3	a2
a6	a5	a1	a4	a7	a5	a1	a5	a5	a5	a7	a5	a7	a5
a7	a7	a3	a7	a7	a6	a3	a7	a7	a7	a3	a5	a5	a7
Input vector	Fault free	Fault patterns											
		13	14	15	16	17	18	19	20	21	22	23	24
a0	a0	a0	a2	a1	a2	a0	a0	a0	a0	a2	a0	a0	a0
a1	a1	a1	a3	a1	a1	a0	a0	a1	a1	a3	a3	a1	a1
a2	a1	a1	a1	a1	a1	a1	a1	a1	a3	a3	a1	a1	a1
a3	a3	a3	a3	a3	a3	a3	a2	a3	a3	a3	a3	a3	a3
a4	a2	a2	a0	a2	a2	a2	a2	a2	a2	a0	a2	a0	a0
a5	a3	a1	a1	a3	a1	a2	a2	a3	a3	a1	a3	a1	a1
a6	a5	a7	a5	a4	a5	a5	a5	a7	a7	a7	a5	a5	a5
a7	a7	a7	a7	a7	a7	a7	a6	a7	a7	a7	a7	a7	a7

QCA and majority logic based synthesis [37,38]. The 13 standard functions and their mapping based on MX-cqca gate is shown in Table 6. Thus based on Table 6 any three variable Boolean functions can be implemented based on MX-cqca gate. In Table 6, the complement of a input variable is used which can be easily generated using the MX-cqca as shown in Fig. 26. In order to design any complex function based on MX-cqca, the input function can be decomposed into the Boolean network in which every node has atmost three variables. Next, each node of three variables can be mapped to Mx-cqca gates based on Table 6. Finally, as fanout is not allowed in conservative logic, the nodes having fanout of more than one needs to be identified. At these identified nodes, MX-cqca gates need to be used to form the copy of the signals which have fanout of more than one. Figure 26 shows the use of the Mx-cqca gate to copy a signal (input i_1 is copied to outputs o_0 and o_2 of the MX-cqca gate). The proposed design methodology can be summarized in the following three steps:

– Step 1: The input function is decomposed into the Boolean network in which every node has atmost three variables. This step is similar to the design methodology proposed in [29].

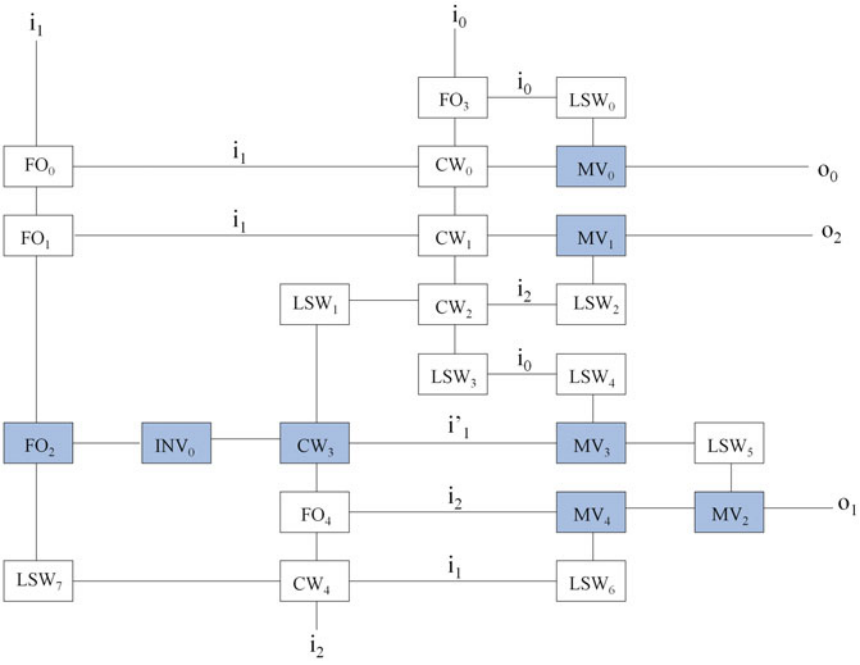


Fig. 25. QCA layout of Mx-cqca gate testable with only all 0s and all 1s test vectors for any single missing/additional cell defect (the shaded devices represent their fault tolerant counterpart)

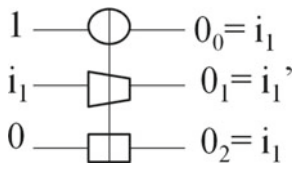


Fig. 26. MX-cqca gate for fanout and inversion of input B

– Step 2: The three variable functions generated at every node in Step 1 are mapped to their MX-cqca based implementation based on the thirteen standard functions. The MX-cqca based implementation of thirteen standard functions is shown in Table 6.
– Step 3: The nodes which have fanout of more than one are identified, and MX-cqca gates are used to form the copy of the signals which have fanout of more than one.

Table 6. MX-cqca based implementation of standard functions

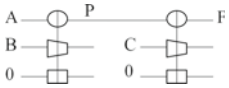
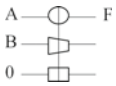
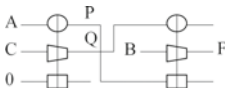
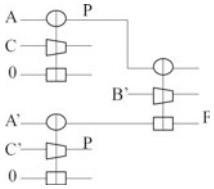
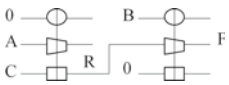
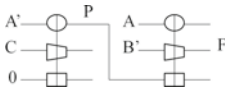
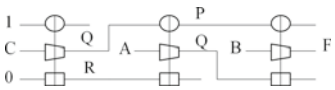
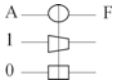
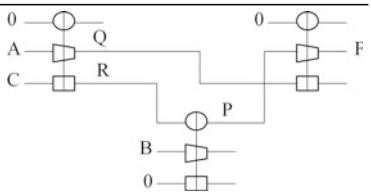
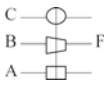
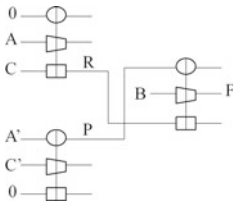
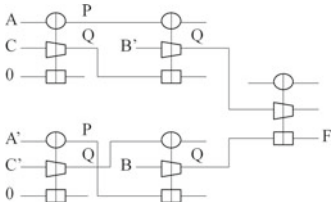
No.	Function	Majority expression	MX-cqca Implementation
1	$F = ABC$	$P = MXcqca(P = MXcqca(A, B, 0), C, 0)$	
2	$F = AB$	$P = MXcqca(A, B, 0)$	
3	$F = ABC + A\bar{B}\bar{C}$	$Q = MXcqca(Q = MXcqca(A, C, 0), B, P = MXcqca(A, C, 0))$	
4	$F = ABC + \bar{A}\bar{B}\bar{C}$	$Q = MXcqca(P = MXcqca(A, C, 0), \bar{B}, P = MXcqca(\bar{A}, \bar{C}, 0))$	
5	$F = AB + BC$	$P = MXcqca(B, R = MXcqca(0, A, C), 0)$	
6	$F = AB + \bar{A}\bar{B}C$	$Q = MXcqca(A, \bar{B}, P = MXcqca(\bar{A}, C, 0))$	
7	$F = ABC + \bar{A}B\bar{C} + A\bar{B}\bar{C}$	$Q = MXcqca(P = MXcqca(Q = MXcqca(1, C, 0), A, R = MXcqca(1, C, 0)), B, Q = MXcqca(Q = MXcqca(1, C, 0), A, R = MXcqca(1, C, 0)))$	
8	$F = A$	$P = MXcqca(A, 1, 0)$	

Table 6. *(Continued)*

9	$F = AB + BC + AC$	$R = MXcqca(0, P = MXcqca(R = MXcqca(0, A, C), B, 0), B, 0), Q = MXcqca(0, A, C))$	
10	$F = AB + \bar{B}C$	$Q = MXcqca(C, B, A)$	
11	$F = AB + BC + \bar{A}\bar{B}\bar{C}$	$Q = MXcqca(P = MXcqca(\bar{A}, \bar{C}, 0), B, P = MXcqca(0, A, C))$	
12	$F = AB + \bar{A}\bar{B}$	$Q = MXcqca(Q = MXcqca(1, B, 0), A, R = MXcqca(1, B, 0))$	
13	$F = ABC + \bar{A}\bar{B}C + A\bar{B}\bar{C} + \bar{A}B\bar{C}$	$R = MXcqca(0, Q = MXcqca(P = MXcqca(A, C, 0), \bar{B}, Q = MXcqca(A, C, 0)), Q = MXcqca(Q = MXcqca(\bar{A}, \bar{C}, 0), B, P = MXcqca(\bar{A}, \bar{C}, 0)))$	

10 Discussion and Conclusions

This work proposes testable reversible sequential circuits based on conservative logic. Conservative logic is testable for any unidirectional stuck-at faults using only two test vectors, all 0s and all 1s. The proposed conservative reversible sequential circuits have feedback that deters their testing by only two test vectors, thus a technique is demonstrated to disrupt the feedback in test mode. Experimental simulation on a single missing/additional cell defect has verified the application of the conservative logic towards fault testing in QCA computing. A new conservative gate (Mx-cqca gate) that is not reversible is also proposed especially suiting QCA computing. There are some major challenges associated with Mx-cqca based designs. Researchers have proposed several implementation techniques for QCA devices such as semiconductor, molecular QCA and magnetic QCA. However QCA devices are difficult to fabricate due to defects such as cell displacement, cell misalignment and cell omission Thus, all the fabricated and tested QCA designs are limited to small logic gates such as majority voter, fanout, and wire design. Thus, until a complex design such as an adder, multiplier and memory components are fabricated and tested in QCA computing; research community would have to wait to practically realize the Mx-cqca based designs. Further, the current literature lacks in the research about synthesis of QCA circuits. There is no synthesis tool for mapping HDL descriptions to QCA designs and to their corresponding QCA layouts to enable simulation using the QCADesigner tool. Thus, in order to implement Mx-CQCA based designs, a synthesis tool is needed that will be equivalent of an HDL description to layout generation tool in conventional CMOS computing is needed. There is also a need of a tool that can approximate the power dissipation in proposed Mx-cqca based design.

The proposed sequential circuits based on conservative logic gates outperform the sequential circuit implemented in classical gates in terms of testability. As the sequential circuits implemented using conventional classic gates does not provide inherited support for testability. A conventional sequential circuit needs modification in the original circuitry to provide the testing capability. Also as the complexity of a sequential circuit increases the number of test vector required to test the sequential circuit increases rapidly. For example, to test a general sequential circuit more than 2000 test vectors are required to test stuck at faults of the entire circuit, while if the same sequential circuit is build using proposed reversible sequential building blocks it can be tested using only two test vectors. The main advantage of proposed conservative logic gate based reversible sequential circuits compared to the conventional sequential circuit is that, the number of test vectors required to test the reversible sequential circuit is always two test vectors and the complexity of the circuit does not impact the number of test vectors. The proposed design of reversible sequential building blocks minimizes the overhead of test time for a reversible sequential circuit. A limitation of the proposed work is that it cannot detect multiple missing/additional cell defects. In conclusion, this work advances the state of the art of testing reversible

sequential circuits based on stuck-at-fault model, as well as, reversible circuits implemented in QCA circuits having single missing/additional cell defect.

References

1. QCADesigner. http://waluslab.ece.ubc.ca/qcadesigner/
2. IEEE. All rights reserved. Reprinted with permission from Thapliyal, H., Ranganathan, N., Kotiyal, S.: Design of testable reversible sequential circuits. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **21**(7), 1201–1209 (2013)
3. Alam, M.T., Kurtz, S.J., Siddiq, M.A.J., Niemier, M.T., Bernstein, G.H., Hu, X.S., Porod, W.: On-chip clocking of nanomagnet logic lines and gates. IEEE Trans. Nanotechnol. **11**(2), 273–286 (2012)
4. Anderson, N., Ercan, I., Ganesh, N.: Toward nanoprocessor thermodynamics. In: 2012 12th IEEE Conference on Nanotechnology (IEEE-NANO), pp. 1–6 (2012)
5. Bennett, C.H.: Logical reversibility of computation. IBM J. Res. Dev. **17**, 525–532 (1973)
6. Bhanja, S., Ottavi, M., Lombardi, F., Pontarelli, S.: QCA circuits for robust coplanar crossing. J. Electron. Test. **23**(2–3), 193–210 (2007)
7. Bhanja, S., Pulecio, J.: A review of magnetic cellular automata systems. In: 2011 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 2373–2376. IEEE (2011)
8. Bubna, M., Goyal, N., Sengupta, I.: A DFT methodology for detecting bridging faults in reversible logic circuits. In: Proceedings of 2007 IEEE Region 10 Conference, Tencon 2007, Taipei, pp. 1–4, Oct 2007
9. Cho, H., Swartzlander, E.: Adder designs and analyses for quantum-dot cellular automata. IEEE Trans. Nanotechnol. **6**(3), 374–383 (2007)
10. Cho, H., Swartzlander, E.: Serial parallel multiplier design in quantum-dot cellular automata. In: Proceedings of the IEEE Symposium Computer Arithmetic (ARITH), Montepellier, France, pp. 7–15 (2007)
11. Chuang, M.L., Wang, C.Y.: Synthesis of reversible sequential elements. J. Emerg. Technol. Comput. Syst. **3**(4), 1–19 (2008)
12. Dalui, M., Sen, B., Sikdar, B.K.: Fault tolerant QCA logic design with coupled majority-minority gate. Int. J. Comput. Appl. **1**(29), 81–87 (2010)
13. Ercan, I., Anderson, N.: Heat dissipation bounds for nanocomputing: theory and application to QCA. In: 2011 11th IEEE Conference on Nanotechnology (IEEE-NANO), pp. 1289–1294 (2011)
14. Farazmand, N., Zamani, M., Tahoori, M.B.: Online fault testing of reversible logic using dual rail coding. In: Proceedings of IEEE International On-Line Testing Symposium, pp. 204–205, May 2010
15. Fijany, A., Toomarian, B.N.: New design for quantum dots cellular automata to obtain fault tolerant logic gates. J. Nanopart. Res. **3**(1), 27–37 (2001)
16. Fijany, A., Toomarian, B.N.: New design for quantum dots cellular automata to obtain fault tolerant logic gates. J. Nanopart. Res. **3**, 27–37 (2001)
17. Frank, M.: Approaching the physical limits of computing. In: Proceedings of ISMVL 2005, The Thirty-Fifth International Symposium on Multiple-Valued Logic, Calgary, Canada, pp. 168–185, May 2005
18. Fredkin, E., Toffoli, T.: Conservative logic. Int. J. Theor. Phys. **21**, 219–253 (1982)
19. Frost-Murphy, S., Ottavi, M., Frank, M., DeBenedictis, E.: On the design of reversible QDCA systems. Technical Report SAND2006-5990, Sandia National Laboratories (2006)

20. Große, D., Wille, R., Dueck, G.W., Drechsler, R.: Exact synthesis of elementary quantum gate circuits for reversible functions with dont cares. In: Proceedings of the International Symposium on Multi-Valued Logic, Dallas, Texas, pp. 214–219, May 2008
21. Große, D., Wille, R., Dueck, G., Drechsler, R.: Exact multiple control toffoli network synthesis with SAT techniques. IEEE Trans. CAD **28**(5), 703175 (2009)
22. Gupta, P., Agarwal, A., Jha, N.K.: An algorithm for synthesis of reversible logic ciruits. IEEE Trans. Comput. Aided Des. **25**(11), 2317–2330 (2006)
23. Gupta, P., Jha, N.K., Lingappan, L.: A test generation framework for quantum cellular automata circuits. IEEE Trans. VLSI Sys. **15**(1), 24–36 (2007)
24. Hanninen, I., Takala, J.: Robust adders based on quantum-dot cellular automata, In: Proceedings of the IEEE International Conference Application-Specific Systems, Architectures and Processors (ASAP), Montreal, QC, Canada, pp. 391–396, Jul 2007
25. Huang, J., Momenzadeh, M., Lombardi, F.: Analysis of missing and additional cell defects in sequential quantum-dot cellular automata. Integr. VLSI J. **40**(1), 503–515 (2007)
26. Jin, Z.: Fabrication and measurement of molecular quantum cellular automata (QCA) device. Ph.D. thesis, University of Notre Dame (2006)
27. Kartschoke, P.: Implementation issues in conservative logic networks. In: M.S.E.E. Thesis, University of Virginia, Charlottesville, VA (1992)
28. Kim, K., Wu, K., Karri, R.: The robust QCA adder designs using composable QCA building blocks. IEEE Trans. Comput. Aided Des. **26**(1), 176–183 (2007)
29. Kong, K., Shang, Y., Lu, R.: An optimized majority logic synthesis methodology for quantum-dot cellular automata. IEEE Trans. Nanotechnol. **9**(2), 170–183 (2010)
30. Kostinski, N., Fok, M.P., Prucnal, P.R.: Experimental demonstration of an all-optical fiber-based Fredkin gate. Opt. Lett. **34**(18), 2766–2768 (2009)
31. Landauer, R.: Irreversibility and heat generation in the computational process. IBM J. Res. Dev. **5**, 183–191 (1961)
32. Chang, L., Frank, D.J., Montoye, R.K., Koester, S.J., Ji, B.L., Coteus, P.W., Dennard, R.H., Haensch, W.: Practical strategies for power-efficient computing technologies. Proc. IEEE **98**(2), 215–236 (2010)
33. Lent, C., Isaksen, B., Lieberman, M.: Molecular quantum-dot cellular automata. J. Am. Chem. Soc. **125**(4), 1056–1063 (2003)
34. Lent, C., Tougaw, P.: A device architecture for computing with quantum dots. Proc. IEEE **85**(4), 541–557 (1997)
35. Liu, W., Srivastava, S., Lu, L., O'Neill, M., Swartzlander, E.: Are QCA cryptographic circuits resistant to power analysis attack? IEEE Trans. Nanotechnol. **11**(6), 1239–1251 (2012)
36. Lu, Y., Liu, M., Lent, C.: Molecular quantum-dot cellular automata: from molecular structure to circuit dynamics. J. Appl. Phys. **102** (2007) (Article No. 034311)
37. Ma, X., Huang, J., Metra, C., Lombardi, F.: Reversible gates and testability of one dimensional arrays of molecular QCA. J. Elect. Test. **24**(1–3), 1244–1245 (2008)
38. Ma, X., Huang, J., Metra, C., Lombardi, F.: Detecting multiple faults in one-dimensional arrays of reversible qca gates. J. Elect. Test. **25**(1), 39–54 (2009)
39. Mahammad, S., Veezhinathan, K.: Constructing online testable circuits using reversible logic. IEEE Trans. Instrum. Meas. **59**, 101–109 (2010)
40. Maslov, D., Dueck, G.W.: Reversible cascades with minimal garbage. IEEE Trans. Comput. Aided Des. **23**(11), 1497–1509 (2004)

41. Mathew, J., Rahaman, H., Jose, B.R., Pradhan, D.K.: Design of reversible finite field arithmetic circuits with error detection. In: 21st International Conference on VLSI Design 2008, VLSID 2008, pp. 453–459. IEEE (2008)
42. Momenzadeh, M., Ottavi, M., Lombardi, F.: Modeling QCA defects at molecular level in combinational circuits. In: Proceedings of DFT in VLSI Systems, Monterey, CA, USA, pp. 208–216, Oct 2005
43. Morita, K.: Reversible computing and cellular automata-a survey. Theor. Comput. Sci. **395**(1), 101–131 (2008)
44. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, New York (2000)
45. Niemier, M.T., Rodrigues, A.F., Kogge, P.M.: A potentially implementable FPGA for quantum dot cellular automata. In: Proceedings of the 1st Workshop on Non-Silicon Computation (NSC-1), Boston, MS (2002)
46. Ottavi, M., Schiano, L., Lombardi, F., Tougaw, D.: HDLQ: a HDL environment for QCA design. ACM J. Emerg. Tech. **2**(4), 243–261 (2006)
47. Parhami, B.: Fault-tolerant reversible circuits. In: Proceedings of 40th Asilomar Conference Signals, Systems, and Computers, Pacific Grove, CA, pp. 1726–1729, Nov 2006
48. Patel, K.N., Hayes, J.P., Markov, I.L.: Fault testing for reversible circuits. IEEE Trans. CAD **23**, 410–416 (2004)
49. Pedram, M., Wu, Q., Wu, X.: A new design for double edge triggered flip-flops. In: Proceedings of the Asia South Pacific Design Automation Conference, Yokahama, pp. 417–421 (1998)
50. Polian, I., Fiehn, T., Becker, B., Hayes, J.P.: A family of logical fault models for reversible circuits. In: ATS '05: Proceedings of the 14th Asian Test Symposium on Asian Test Symposium, Kolkata, India, pp. 422–427 (2005)
51. Prasad, A.K., Shende, V., Markov, I., Hayes, J., Patel, K.N.: Data structures and algorithms for simplifying reversible circuits. ACM JETC **2**(4), 277–293 (2006)
52. Pulecio, J.F., Bhanja, S.: Magnetic cellular automata coplanar cross wire systems. J. Appl. Phys. **107**(3), 034308 (2010)
53. Pulecio, J., Pendru, P., Kumari, A., Bhanja, S.: Magnetic cellular automata wire architectures. IEEE Trans. Nanotechnol. **99**, 1 (2011)
54. Rahaman, H., Kole, D.K., Das, D.K., Bhattacharya, B.B.: On the detection of missing gate faults in reversible circuits by a universal test set. In: Proceedings VLSI Design 2008, 21st International Conference on VLSI Design, Hyderabad, India, pp. 163–168, Jan 2008
55. Ren, J., Semenov, V.K.: Progress with physically and logically reversible superconducting digital circuits. IEEE Trans. Appl. Supercond. **21**(3), 780–786 (2011)
56. Ren, J., Semenov, V.K., Polyakov, Y.A., Averin, D.V., Tsai, J.S.: Progress towards reversible computing with nSQUID arrays. IEEE Trans. Appl. Supercond. **19**, 961–967 (2009)
57. Rice, J.: A new look at reversible memory elements. In: Proceedings of International Symposium on Circuits and Systems (ISCAS) 2006, Kos, Greece, pp. 243–246, May 2006
58. Semenov, V.K., Danilov, G.V., Averin, D.V.: Classical and quantum operation modes of the reversible Josephson-junction logic circuits. IEEE Trans. Appl. Supercond. **17**, 455–461 (2007)
59. Shende, V.V., Prasad, A., Markov, I., Hayes, J.: Synthesis of reversible logic circuits. IEEE Trans. CAD **22**, 710–722 (2003)
60. Swaminathan, G.: Concurrent error detection techniques using parity. In: M.S.E.E. Thesis, University of Virginia, Charlottesville, VA (1989)

61. Swaminathan, G., Aylor, J., Johnson, B.: Concurrent testing of VLSI circuits using conservative logic. In: Proceedings of International Conference on Computer Design (ICCD), Cambridge, MA, pp. 60–65, Sep 1990
62. Tahoori, M.B., Huang, J., Momenzadeh, M., Lombardi, F.: Testing of quantum cellular automata. IEEE Trans. Nanotechnol. **3**(4), 432–442 (2004)
63. Taraphdara, C., Chattopadhyay, T., Roy, J.: Machzehnder interferometer-based all-optical reversible logic gate. Opt. Laser Technol. **42**(2), 249–259 (2010)
64. Taskin, B., Chiu, A., Salkind, J., Venutolo, D.: A shift-register-based QCA memory architecture. ACM J. Emerg. Tech. Comput. Sys. **5**(1) (2009) (Article No. 4)
65. Thapliyal, H.: Design, synthesis and test of reversible logic circuits for emerging nanotechnologies. Ph.D. thesis, University of South Florida, Tampa, Dec 2011
66. Thapliyal, H., Ranganathan, N.: Reversible logic-based concurrently testable latches for molecular QCA. IEEE Trans. Nanotechnol. **9**(1), 62–69 (2010)
67. Thapliyal, H., Srinivas, M.B., Zwolinski, M.: A beginning in the reversible logic synthesis of sequential circuits. In: Proceedings of the Military and Aerospace Programmable Logic Devices International Conference Washington, Sep 2005
68. Thapliyal, H., Vinod, A.P.: Design of reversible sequential elements with feasibility of transistor implementation. In: Proceedings of the 2007 IEEE International Symposium on Circuits and Systems, New Orleans, USA, pp. 625–628 (May 2007)
69. Thapliyal, H., Ranganathan, N.: Design of reversible sequential circuits optimizing quantum cost, delay and garbage outputs. ACM J. Emerg. Technol. Comput. Syst. **6**(4), 14:1–14:35 (2010). (Article No. 14)
70. Tougaw, P., Lent, C.: Logical devices implemented using quantum cellular automata. J. Appl. Phys. **75**(3), 1818–1825 (1994)
71. Tougaw, P., Lent, C.: Dynamic behavior of quantum cellular automata. J. Appl. Phys. **80**(8), 4722–4736 (1996)
72. Vasudevan, D.P., Lala, P.K., Parkerson, J.P.: Reversible-logic design with online testability. IEEE Trans. Instrum. Meas. **55**(2), 406–414 (2006)
73. Wang, P., Niamat, M., Vemuru, S.: Minimal majority gate mapping of 4-variable functions for quantum cellular automata. In: 2011 11th IEEE Conference on Nanotechnology (IEEE-NANO), pp. 1307–1312. IEEE (2011)
74. Wei, T., Wu, K., Karri, R., Orailoglu, A.: Fault tolerant quantum cellular array (QCA) design using triple modular redundancy with shifted operands. In: Proceedings of the 2005 Conference on Asia South Pacific Design Automation, Shanghai, China, pp. 1192–1195 (Jan 2005)
75. Yang, G., Song, X., Hung, W.N., Perkowski, M.A.: Bi-directional synthesis of 4-bit reversible circuits. Comput. J. **51**(2), 207–215 (2008)
76. Zhang, R., Walus, K., Wang, W., Jullien, G.: Performance comparison of quantum dot cellular automata adders. In: Proceedings of the IEEE International Symposium Circiuts and Systems, Kobe, Japan, pp. 2522–2526 (May 2005)
77. Zhang, R., Walus, K., Wang, W., Jullien, G.A.: A method of majority logic reduction for quantum cellular automata. IEEE Trans. Nanotechnol. **3**(4), 443–450 (2004)
78. Zhong, J., Muzio, J.: Analyzing fault models for reversible logic circuits. IEEE Congr. Evol. Comput., Vancouver, BC, pp. 2422–2427 (2006)

STT-Based Non-Volatile Logic-in-Memory Framework

Jayita Das[1(✉)], Syed M. Alam[2], and Sanjukta Bhanja[1]

[1] University of South Florida, Tampa, FL, USA
jayitadas@mail.usf.edu, bhanja@usf.edu
[2] Everspin Technologies Inc., Austin, TX, USA
salam@alum.mit.edu

Abstract. This work describes an integration of logic within the Spin Transfer Torque Magnetoresistive RAM (STT-MRAM) framework. For memory, a minimum separation between the cells is required to ensure bit-to-bit independency. For logic that relies on magnetostatic coupling, a maximum separation is allowed between magnetic cells for effective computation. Integration of the two functionalities therefore requires meeting the orthogonal spatial needs of separation. In this work the technological challenges of this integration are first described followed by the specifications of the new STT-MRAM based logic-in-memory architecture. How a spin transfer torque based control, also called clock, can tune the architecture between logic and memory modes is next described. A reference free variability tolerant differential read scheme leveraging the integration is presented. This logic-in-memory framework is also an integration between magnetic and CMOS planes. Finally, a logic partitioning between the two planes is described that can significantly improve the performance metrics.

Keywords: Spin transfer torque · MRAM · MTJ · NML · STT clock · Logic partitioning · Differential read · Variability tolerant

1 Introduction

For long, the semiconductor industry has relied on Dennard scaling and Moore's law to meet the ever increasing demands of the user. However, the downside of scaling (mainly leakage) has started becoming more apparent in newer technologies [1]. Therefore, it would not be long before the limits of application-specific power dissipation would end the benefits of scaling. The solution lies in alternative technologies to fulfill application specific demands. Two such post-CMOS technologies are the spin transfer torque magnetoresistive RAM (STT-MRAM) [2,3] used in building modern non-volatile memories and nanomagnetic logic (NML) used in building non-volatile zero leakage radiation hard logic. We will discuss further details of STT-MRAM in Sect. 2 and NML in Sect. 3.

Compared to CMOS memories like SRAM, STT-MRAM has a lower read latency. Its footprint is also smaller than DRAM. Unlike modern non-volatile

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 173–193, 2014.
DOI: 10.1007/978-3-662-43722-3_8, © Springer-Verlag Berlin Heidelberg 2014

memories like Flash, Ferroelectric RAM and Phase Change Memories, it has unlimited read/write endurance. In addition, STT-MRAM is thermally robust and radiation hard [4]. Freescale Semiconductor, Inc. is one of the pioneers in MRAM development. Its spin-off, Everspin Technologies, Inc. now offers fully functional 64-Mb double data rate type 3 (DDR3) STT-MRAM chips [4]. Everspin's MRAMs are already in use in Japanese research satellites and in BMW Motorrad Motorsport automotives [4]. STT-MRAM is also an active research area by semiconductor giants like Qualcomm Inc., Fujitsu Ltd., Grandis Inc., IBM Corp., Intel Corp. etc.

On the logic side, nanomagnetic logic (NML) is one of the experimentally demonstrated post-CMOS computing technologies. It uses single layered single domain nanomagnets to compute at room temperature [5,6]. However, NML has certain key issues. It has high dynamic power, which is also not scalable [7]. Moreover, there aren't any successful CMOS integration for electrically driven on-chip read and write from NML. These two key issues mars the benefits like zero leakage that we get from NML. In this chapter we will see how we can use STT-MRAM cells to compute in NML style and yet consume lesser dynamic power and have on-chip read/write techniques. Such thermally robust radiation hard non-volatile logic-in-memory architecture can provide a back-up to CPU under thermally challenged or radiation prone safety critical operations like automobile and space applications.

2 STT-MRAM

STT-MRAM is a modern non-volatile memory that uses magnetic tunnel junctions (MTJs) and their remanence to store information. MTJs are a class of multilayer devices. An insulating barrier separates two ferromagnetic layers (1) and (2) in a MTJ (see Fig. 1a). The magnetization of the ferromagnetic layer (2) remains fixed and is called the fixed layer or pinned layer or reference layer. It is pinned from underneath through antiferromagnetic coupling by another ferromagnetic layer (3). This layer is in turn pinned through exchange bias with antiferromagnetic seed layer (4) (see Fig. 1a).

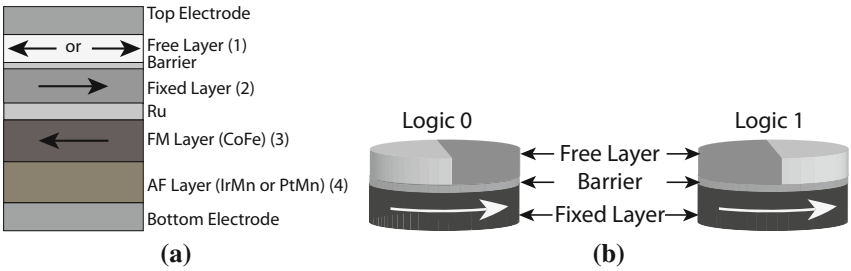


Fig. 1. (a) Main constituent layers of MTJ; (b) MTJ logic states;

The magnetization of the ferromagnetic layer (1) can be excited with the help of external magnetic fields or spin transfer torque generated by current through the device. The layer (1) is called the free layer. The magnetodynamics of the free layer under spin transfer torque is governed by the well known Landau-Lifshitz-Gilbert (LLG) equation given below (Eq. 1).

$$\frac{d\mathbf{m_1}}{dt} = -\gamma M_S \mathbf{m_1} \times (\mathbf{h_{eff}} - \frac{\alpha}{\gamma M_S}\frac{d\mathbf{m_1}}{dt} - \frac{J_e G}{J_p}\mathbf{m_2} \times \mathbf{m_1}) \tag{1}$$

where,

$$G = \left[-4 + (1+P)^3 \frac{(3 + \hat{s_1}.\hat{s_2})}{4P^{3/2}}\right]^{-1} \tag{2}$$

and

$$J_p = \mu_0 \cdot M_s^2 \frac{\mid e \mid t_F}{\hbar} \tag{3}$$

Table 1 defines the symbols. The first term on the right hand side of Eq. 1 relates to the precessional dynamics in the free layer resulting under the net magnetic field acting on it. The contributors to this field are the dipolar coupling from the fixed layer, shape anisotropy and demagnetization field. The second term relates to damping in the layer. The third term relates to the effect of spin transfer torque on the layer. Below are some of the key properties of MTJs that are crucial to understanding the operations of STT-MRAM.

Table 1. Symbol definitions.

Symbol	Description
$\mathbf{m_1}$, $\mathbf{m_2}$	Unit vectors in direction of magnetization of free and fixed layer
γ	Gyromagnetic ratio
M_s	Saturation magnetization
$\mathbf{h_{eff}}$	Unit vector along effective magnetic field on the free layer arising from crystalline and shape anisotropy, demagnetization field, exchange field and external field which also includes coupling from the fixed layer
α	Gilbert damping constant
P	Spin polarizing factor
$\hat{s_1}$, $\hat{s_2}$	Unit vectors along the global spin orientation of the free and fixed layers
t_F	Thickness of free layer
e	Electron charge
$\hbar$	Reduced Planck's constant
μ_0	Permeability of free space
H_K	Effective anisotropy field including magnetocrystalline anisotropy and shape anisotropy
H_{ext}	External magnetic field on free layer including coupling from underneath fixed layer

2.1 Magnetoresistance

The MTJs are conductive to electrical current. Interestingly their electrical resistance across the free and fixed layer is dependent on the relative magnetization angle θ between the two layers. The resistance, which is the inverse of the conductance $G(\theta)$, can be calculated from Eq. 4.

$$G(\theta) = \frac{1}{2}(G_P + G_{AP}) + \frac{1}{2}(G_P + G_{AP}) \cdot cos\ \theta \tag{4}$$

where, G_P and G_{AP} are the conductances of the MTJ for $\theta = 0°$ and $\theta = 180°$ respectively. The $\theta = 0°$ is also known as the parallel state and is commonly used in STT-MRAM to represent logic 0. The $\theta = 180°$ is known as the antiparallel state and is used to represent logic 1 (see Fig. 1b). Note, $G_p > G_{AP}$. For a quantum mechanical explanation of the angular dependence of MTJ resistance, readers are encouraged to refer [8,9]. In STT-MRAM the MTJ resistance is used to read its state.

It should be evident that greater the difference between the resistances of 0 and 1 state, the easier it is to read the MTJ. A parameter that represents this resistive difference is the Magnetoresistance (MR) defined by Eq. 5. This MR is an important parameter since it defines the ease and reliability for MTJ read. Higher the MR, higher is the read distinguishibility and lower are the read based errors for MTJs [3].

$$MR = \frac{G_P - G_{AP}}{G_{AP}} \tag{5}$$

Before we move on to the next property, we would like to briefly mention that the MR phenomenon was discovered as early as 1856 [10]. However, it was not until 1975 that Julliere in his seminal work defined the first MR model on a Fe/Ge/Co multilayer [11]. Since then various material combinations and fabrication steps has been explored to improve the MR at room temperature. Today a MR as high as 410 % is reported at room temperature with a Co(001)/MgO(001)/Co(001) material stack [12]. Interested readers may refer to [9] for a chronological description of MR developments.

2.2 Spin Transfer Torque

When spin conduction electrons interact with the localized magnetic moment of a ferromagnet, they transfer a part of their angular momentum to the magnet. This results in a torque on the magnet. When this torque is sufficiently strong, it can excite the magnetization in the material. This torque is called the spin transfer torque and is dependent on the magnitude and direction of the current. When electrons flow from the fixed to the free layer within the MTJ, the spin transfer torque and the damping act together to switch the free layer to logic 0 state (see Fig. 2). For electrons flowing in the reverse direction, the torque and the damping acts against each other. When the torque exceeds a critical magnitude, the MTJ switches to logic 1 state (see Fig. 2).

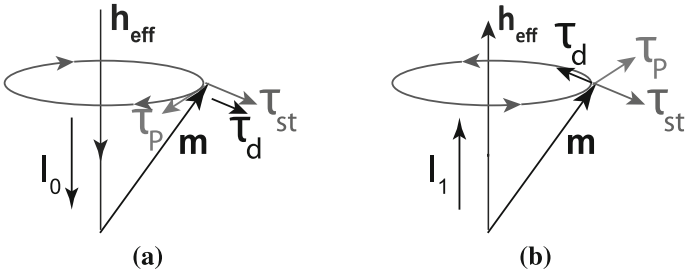


Fig. 2. (a) Switching to logic 0; (b) Switching to logic 1. τ_p refers to torque responsible for precessional motion, τ_d refers to torque responsible for damping and τ_{st} refers to spin transfer torque. I_0 and I_1 refers to the switching current through the MTJ.

The critical current density to switch the MTJs is given by Eq. 6

$$J_w = \frac{2e\alpha M_S t_F (H_K \pm H_{ext} + 2\pi M_S)}{\hbar\eta} \tag{6}$$

where, symbols are defined in Table 1. η is the spin transfer efficiency and is given by $\eta = p/(1 \pm p)$. The $\pm$ is for switching to logic 1 and 0 respectively. Interestingly, the critical switching current (Eq. 6) is proportional to the device dimensions. This means, that the current will scale with device dimensions unlike external fields required to switch the device. The STT-MRAM memory locations are written with the help of this spin transfer torque.

The spin transfer torque was first formulated by Slonczewski and was included in 1996 to the LLG equation Eq. 1. For the physics behind the spin transfer torque switching, readers are once again encouraged to consult Refs. [8,13].

2.3 STT-MRAM Architecture and CMOS Integration

The most popular STT-MRAM architecture is a crossbar array of 1T1MTJ cells (one transistor-one magnetic tunnel junction) (see Fig. 3a). Each MTJ in the STT-MRAM is accessed using their access transistors and bit (BL), source (SL) and word (WL) lines (see Fig. 3b). The thermal robustness of MTJs has helped to integrate them monolithically with CMOS using standard back-end-of-line (BEOL) manufacturing techniques [3]. In STT-MRAM, MTJs are typically fabricated between metal layers M3 and M4 [14]. Figure 3b shows a cross-section of the CMOS MTJ integration [15].

3 Nanomagnetic Logic

Use of magnets in the computing world was limited until recently to only information storage. Information propagation with the help of magnets was first demonstrated in 2000 by Cowburn and Welland [5]. In their work they used single layer single domain circular nanomagnetic dots. The magnetostatic interaction

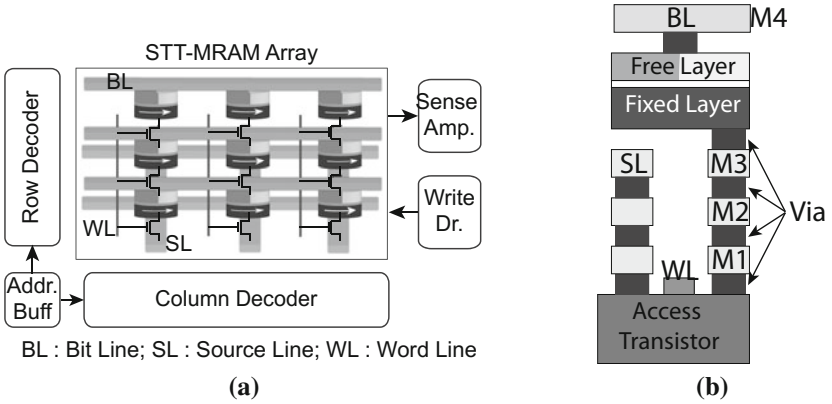


Fig. 3. (a) STT-MRAM architecture; (b) An example of STT-MRAM bit cross-section and CMOS integration. M1, M2, M3 and M4 are the metal layers 1, 2, 3 and 4 respectively.

between closely spaced nanomagnets was used to propagate information. The first logic computation using nanomagnets was first demonstrated by researchers from University of Notre Dame in 2006 [6]. In their work, the researchers once again used magnetostatic interaction between closely placed nanomagnets for computation. However, this time the nanomagnetic dots were elongated to give them a shape anisotropy. The function of shape anisotropy is to define two stable magnetic states at room temperature. This preferred magnetization direction is also called the easy axis and is used to represent the logic 0 and 1 states (see Fig. 4a, where the easy axis is along the y direction for the nanomagnets).

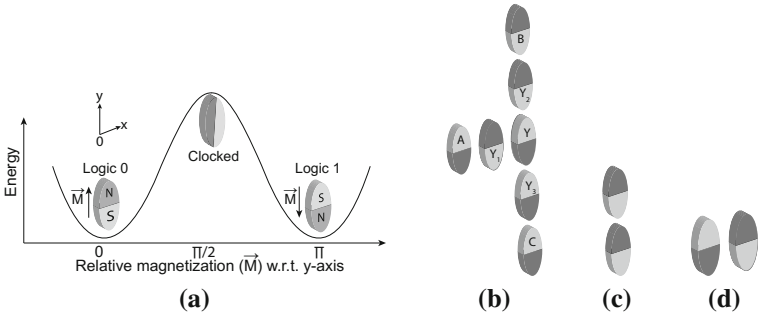


Fig. 4. (a) Logic 0 and 1 states along with clocked configuration in single layer single domain nanomagnets with shape anisotropy; (b) Majority voter. A, B and C are inputs and Y is the output. $Y = A \cdot B + B \cdot C + C \cdot A$; (c) Ferromagnetic coupling; (d) Antiferromagnetic coupling.

The basic logic construct in this computation is the 3-input majority voter [16] (Fig. 4b). In addition to the magnetic fields required to write the inputs into the nanomagnets, a clocking field was also required to be provided [17,18]. The purpose of the clock was to take the nanomagnets to their hard axis or orthogonal to the easy axis (see Fig. 4a). The hard axis is the energy maximum state for any nanomagnet. When released from this state, the nanomagnets can effectively respond to the weak magnetostatic coupling from the neighbors and compute any logic function. There are two types of magnetostatic coupling: ferromagnetic and antiferromagnetic (see Fig. 4c and d). The ferromagnetic coupling is equivalent to the logical buffer while the antiferromagnetic coupling is equivalent to the logical inverter. With clocking, the elongated nanomagnets can also be used to propagate information [19]. Clocking also helps in ordering the nanomagnets [20,21]. The remanence in the nanomagnets stores their states when no fields are applied to them. This means any NML function is non-volatile and has no leakage power.

The clocking and the writing fields are generated by current carrying wires placed underneath the nanomagnets [17,22]. The current required to generate the required fields ranges in mA for a $100 \times 50\,\mathrm{nm}^2$ nanomagnet. Moreoever, the current requirement scales inversely to the nanomagnet dimensions, making the overall technology non-scalable. Also, the fields cannot be precisely focussed only on the desired nanomagnets. This gives rise to stray fields that increase the chance of writing and clocking errors. There are also no well-defined on-chip read techniques. Till date NML mostly relies on magnetic sensors for reading. These issues diminishes the zero leakage computing advantage of NML.

Computing with the help of magnetostatic coupling between the free layers of MTJs can solve some of the above problems and still provide a zero leakage computing platform. Additionally, such logic would also benefit from the technological advancements already made in STT-MRAM and at the same time enjoy the properties characteristic of STT-MRAM like thermal robustness, radiation hardness and unlimited endurance. In this chapter we will discuss this STT-based computing platform [23–25] that can switch between logic and memory mode of operations with an acting clock as a classifier.

4 STT-Based Logic-in-Memory Architecture

In this architecture, the free layers of the MTJs operate both in computational and storage mode. The storage is through remanence in the free layer. The computation is through magnetostatic interaction between free layers of neighboring MTJs. The computation style is very similar to NML, differing only in the STT based write, clock and MR based read mechanisms. The architecture framework is again very similar to STT-MRAM. The MTJs are arranged in a crossbar fashion, which allows accessing them through rows and columns. The architecture is also very regular with bit, source and word lines running parallel to rows and columns of MTJs.

The main difference between the logic-in-memory architecture and MRAM is the bit density. In this architecture, the MTJs are patterned with a spacing of 20 nm to allow effective coupling between their free layers. On the other hand, in MRAM, the MTJs are placed much further apart to prevent any coupling between them. The close spacing between MTJs in this architecture has also its pros and cons. On one hand while it increases the bit density of the architecture, on the other it reduces the available room for the access transistors underneath the MTJs. This competing space requirement within the architecture generates a design challenge that can be resolved with proper choice of: (i) CMOS technology node; and (ii) integration between CMOS and MTJs. As a first step in this direction, we selected 22 nm CMOS technology node for integration with metal 1 pitch of 64 nm [26]. The second step was to redesign the MTJ-CMOS integration, which we have discussed in the next sections.

To best describe the architecture we will start with its salient features. We then intend to follow up with greater details of cell classification, logic blocks and write/clock/read techniques in the architecture.

4.1 Salient Features

1. *$100\times 50\,nm^2$ MTJ dimensions:* These dimensions ensure that the free layer of MTJs used in the architecture is single domain and is stable at room temperature (RT). The dimensions were arrived from both micromagnetic simulations in OOMMF [27] and LLG [28] software and our past fabricated results [29]. The aspect ratio ensures that there are two stable magnetization states for the free layer at RT. These states will be used to define 0 and 1. The dimensions also take care that a reasonably low critical current is required to switch the MTJs at RT.
2. *MTJ placement in regular 2D grid:* The MTJs in the architecture are laid out into well-defined rows and columns with a 120 nm row pitch and a 70 nm column pitch.
3. *MTJ integration with 22 nm CMOS technology node:* 22 nm CMOS technology node is selected for integration of access transistors.
4. *Access transistor integrated with every alternate MTJs in a row and column:* The smaller room for the access transistors together with the required drive strength and the metal pitch requirement of 22 nm CMOS node allows integration of only one access transistors for every alternate MTJs in rows and columns. Figure 5a shows a $3D$ view of the architecture. Figure 5b shows a cross-section of the architecture.
5. *Bit, source and word lines to address the MTJs in the architecture:* The word lines are used to turn on/off the access transistors. They are located in metal 3 and run parallel to the columns of MTJs. The bit and source lines are housed in metal 2 and 1 respectively and are used for row addressing as well as supplying the writing, clocking and reading current when required (see Fig. 5a).

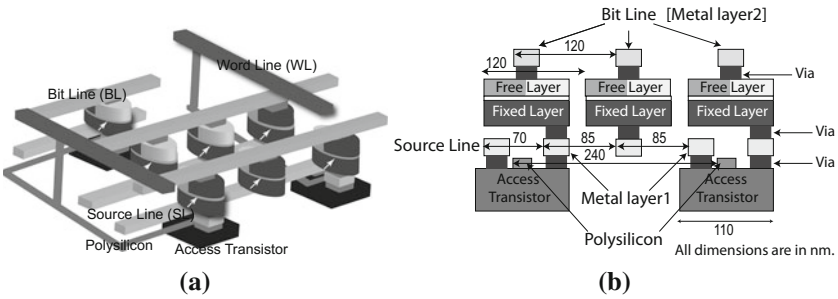


Fig. 5. (a) $3D$ view of the STT-based logic-in-memory architecture; (b) Cross-section of the architecture with the metal pitches.

4.2 Cell Classification

The selective integration of access transistors give rise to two types of cells in the architecture, which we will classify as:

1. *MTJs with access transistors:* To access these MTJs, an appropriate bias is applied across the bit and source lines of the row in which the MTJs are placed. Next the access transistors for the MTJs are turned on with the help of their word lines. This completes the path for the current through the MTJs.
2. *MTJs without access transistors:* These MTJs are accessed through their bit and source lines. Whenever a bias is applied across their bit and source lines, all of these MTJs conduct.

These cells are futher categorized into input, logic and output cells on the basis of their functionality and participation in logic. With the help of a 2-input exclusive-OR layout (Fig. 6), we will explain the different cell categories and the functionalities in the architecture.

1. *Input cells:* These are the MTJs with access transistors. They hold the inputs to any logic function. We will discuss the details of how to write the inputs into them in Sect. 4.3. In Fig. 6, X_1, X_2, X_3 and X_4 are the input cells. For 2-input exclusive-OR operation, $x_1 \oplus x_2 = x_1 \cdot x_2' + x_1' \cdot x_2$, the following values will be written into the input cells: $X_1 = x_1$, $X_2 = x_2'$, $X_3 = x_1'$ and $X_4 = x_2$.
2. *Logic cells:* These are the MTJs that compute the logic. They can be of either types and they form the body of the logic. These MTJs are to be clocked whenever a logic is to be computed. They get their values through neighbor interaction after the clock is released from them. These cells are further sub-categorized into the following:
   (a) *Gates:* Just as in NML, the basic logic gates in this architecture are the 3-input majority voter ($f(x_1, x_2, x_3) = x_1 \cdot x_2 + x_2 \cdot x_3 + x_3 \cdot x_1$) and the inverter ($f(x_1) = x_1'$). Together they form the universal minority logic function. 2-input AND/OR can be derived from this majority function

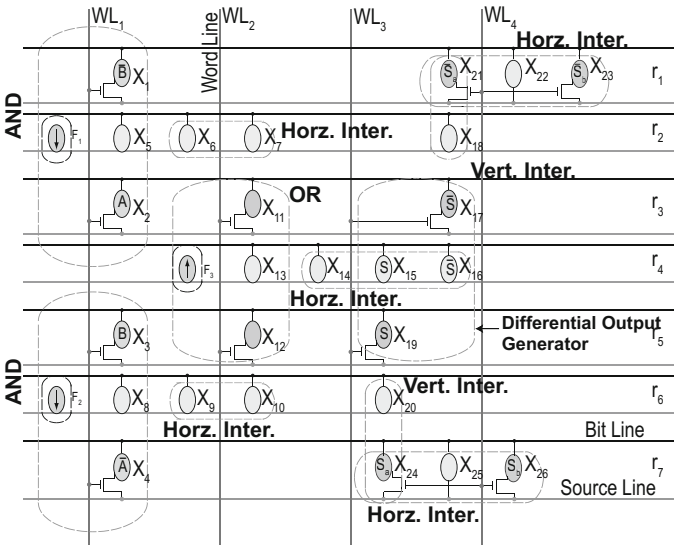


Fig. 6. Layout of 2-input exclusive-OR.

by fixing one of the inputs to 0/1. In Fig. 6, (X_1, X_2, X_5, F_1) forms a 2-input AND gate with inputs X_1 ($=x_1$) and X_2 ($=x_2'$) and output X_5 ($=x_1 \cdot x_2'$). F_1 is the fixed cell. Similarly, (X_3, X_4, X_6, F_2) forms a second 2-input AND gate with inputs X_3 ($=x_1'$) and X_4 ($=x_2$) and output X_6 ($=x_1' \cdot x_2$). (X_{11}, X_{12}, X_{13}, F_3) forms a 2-input OR gate. X_{11} ($=x_1 \cdot x_2'$) and X_{12} ($=x_1' \cdot x_2$) are the inputs and X_{13} ($=x_1 \cdot x_2' + x_1' \cdot x_2$) is the output. F_3 is the fixed cell.

(b) *Interconnects:* These are the cells that propagate information across the logic, just like wires in CMOS circuits. However unlike CMOS, the information here flows without the flow of any physical entity. These interconnects are of two types: vertical and horizontal. In a vertical interconnect, the information flows with the help of ferromagnetic coupling. In Fig. 6, (X_{18}, X_{21}), (X_{20}, X_{24}) are two vertical interconnects. In a horizontal interconnect, the information flows with the help of antiferromagnetic coupling between the MTJs. In Fig. 6, (X_6, X_7), (X_9, X_{10}), (X_{14}, X_{15}, X_{16}), (X_{21}, X_{22}, X_{23}), (X_{24}, X_{25}, X_{26}) are different horizontal interconnects. Note that cells like X_{21} and X_{24} are at the intersection of both types of interconnects.

(c) *Differential output generator:* They generate opposite output states with the help of antiferromagnetic coupling. In Sect. 4.3 we will see how these opposite states are used by the read circuit to read the output from the logic. In Fig. 6, (X_{15}, X_{16}, X_{17}, X_{19}) forms the differential output generator.

3. *Output cells:* These are MTJs with access transistors that store the final output and its complement. Again these cells are only clocked and are never

written into during the execution of a logic function. In Fig. 6, X_{21}, X_{23}, X_{24} and X_{26} are the output cells.

To summarize the execution of 2-input exclusive-OR, $x_1 \oplus x_2 = x_1 \cdot x_2' + x_1' \cdot x_2$, in relation to Fig. 6, there are two AND operations and one OR operation. The two ANDs are performed in the two AND gates with outputs $X_5 = x_1 \cdot x_2'$ and $X_8 = x_1' \cdot x_2$. Their outputs are ORed in the OR gate with $X_{13} = x_1 \oplus x_2$. The differential output generator generates the complementary output bit. Finally, X_{24} and X_{26} stores the output and X_{21} and X_{23} stores the output complement. Algorithm 1 compiles the steps till output generation in X_{13}.

Algorithm 1. 2-*Input* XOR *execution* $x_1 \oplus x_2$ *in STT-based logic-in-memory architecture*

1: **Inputs:** x_1 x_1', x_2, x_2'.
2: **Outputs:** $S_a = S_b = \mathrm{f}(x_1, x_2) = x_1 \oplus x_2$.
3: **Write** $X_1 = x_2'$, $X_2 = x_1$, $X_3 = x_2$, $X_4 = x_1'$.
4: **Clock** rows r_2 and r_6.
5: **Release clock.**
6: X_5, X_6, X_7 settle to $(x_1.x_2')$, $(x_1.x_2')$', $(x_1.x_2')$ resp.
7: X_8, X_9, X_{10} settle to $(x_1'.x_2)$, $(x_1'.x_2)$', $(x_1'.x_2)$ resp.
8: **Select** word line WL_2. **Clock** rows r_3 and r_5.
9: **Release** clock.
10: X_{11} and X_{12} **settle** to $(x_1.x_2')$, $(x_1'.x_2)$ resp.
11: **Clock** *row* r_4 *and then release the clock*
12: X_{13} **settles** to $(x_1 \oplus x_2)$.

4.3 Operational Details

As we mentioned earlier, one of the primary objectives behind replacing single layer NML cells with multilayer MTJs was to take advantage of the low power and on-chip write, clock and read mechanisms in MTJs. We will use the spin transfer torque to write and clock the MTJs and their MR to read them. Advantages of spin transfer torque over field induced write and clock are: (i) low power; and (ii) high precision without stray field generation. The advantages of MR over magnetic sensing include: (i) low read current; and (ii) on-chip read facility. To start with, we will first see how to write into the MTJs and thereafter follow with clocking and reading from them.

Writing. To write into the input cells we first need to apply the appropriate bias across their bit and source lines. With the help of their word lines, their access transistors are next turned on. This provides a path for the current from the bit to the source line or vice versa through the MTJ and its access transistor. The direction of current again depends on the potential applied across the bit and source lines. We will next see how the direction of electrons through the MTJ determines whether a 0 or a 1 is written into it.

When electrons flow from the fixed to the free layer inside the MTJ, their spins are first polarized by the fixed layer. These spin polarized electrons travel to the free layer and interacts with the local magnetization of the free layer. The electrons then transfer their angular momentum to the electrons of the free layer. This results in a spin transfer torque on the free layer that acts towards switching the free layer towards logic 0 state. When electrons flow in the opposite direction from the free to the fixed layer, they are first polarized by the free layer. The electrons that are polarized in the same direction of the fixed layer pass through it. Electrons with opposite polarization to the fixed layer are reflected back to the free layer. These reflected electrons exert a torque on the free layer that tends to switch it opposite to fixed layer, i.e. towards logic 1 state. The magnitude of the spin transfer torque is proportion to the current through the MTJ. Depending on the current direction, if the current magnitude exceeds a critical value (Eq. 6), the spin transfer torque succeeds in switching the free layer to logic 0 or 1 state. The LLG equation (Eq. 1) governs the magnetodynamics of the free layer from spin transfer torque. We will assume that cells on an average take t_w time to be written, the numerical value of which is mentioned in Table 2.

Clocking. For the chosen NML style computation in the free layers, we need to clock the free layers for effective coupling. The advantage is that unlike NML, we can now use spin transfer torque to clock the free layers, which is (i) low power; and (ii) highly confined to individual cells rather than to a group of cells. To elaborate the clocking we would first discuss the clocking mechanism and principle and then go on to describe if any necessary MTJ configurations are required to execute the spin transfer torque clock.

The intention of clocking is to take the MTJ free layers to a stationary state along their hard axis. Here, instead of external fields we will take the help of spin transfer torque to achieve the clocking mission. To reach the clocked state, the torques from precession, damping and spin transfer should cancel each when the magnetization of the free layer is along the hard axis. This condition can be achieved with a clocking current I_{clk} given by Eq. 7. The current magnitude is derived from Eq. 1 by taking the left hand side to a stationary state, i.e. $dm/dt = 0$. Note, the clocking current scales with the device dimensions. Moreover, the clocking current is also in the same order in magnitude as the switching current (Eq. 6). This ensures a clocking power of one order less than field induced clocking. For further details on the clocking, readers can refer to [30]. Clocking procedure for MTJs with access transistors is same as writing into them. First the required potential is applied across their bit and source lines. Next, their access transistors are turned on to complete the path for current. For MTJs without access transistors, clocking should be carried out by only applying the required potential across the bit and source lines. We will assume that cells on an average take t_{clk} time to be clocked and t_s time to settle down under neighbor coupling from the instant the clock is released. Again $t_s < t_{clk}$ is a condition to ensure that cells in one clocking zone have completely settled before the clock in the

next zone is released. This makes sure that information in the logic propagates only in one direction, which is from the input(s) to the output(s) and not other way. The numerical values of t_{clk} is mentioned in Table 2.

$$I_{clk} = \frac{\mu_0 M_S e Vol H_d}{\hbar G} \tag{7}$$

To make this STT clocking effective, the MTJs need to have a 45° tilt in the polarization of their fixed layer. The logic states for the tilted MTJs is shown in Fig. 7a. Interestingly, this type of tilted polarized fixed layer MTJs are also widely used in spin torque oscillators [14]. We will next see how this MTJ configuration suits computational needs as well. Remember for computation we are using interaction between free layers of neighboring MTJs. This interaction becomes most effective with minimum coupling from the fixed layer of MTJs. Ideally this means that the free and fixed layer in MTJs should have orthogonal polarizations. On the other hand, the MTJs also need to have good MR for reading their contents. As we have already seen from Eqs. 4 and 5, the MR is dependent on the relative angle between the magnetizations of free and fixed layers. The 45° polarization therefore provides a good optimal point between the orthogonal needs of computation and reading from MTJs.

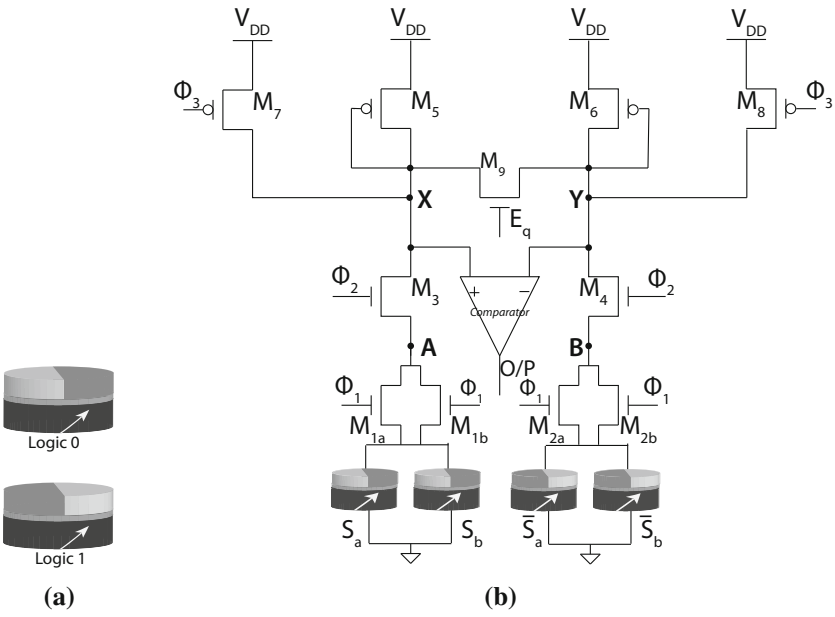


Fig. 7. (a) Logic states in MTJs with 45° polarized fixed layer; (b) Variability tolerant non-destructive differential read circuit.

Reading. One of the greatest advantages of using the MTJs over single layer nanomagnets is the ability to use their electrical resistance to read from the logic. The variable electrical resistance originates from the stacked composition and alignment of the two ferromagnetic layers (free and fixed) with their non-magnetic separation. Several studies have gone into improving the MR ratio for enhanced readibility at room temperature. Typically in memory, the bits are read by comparing their resistance against a reference value. This reference is typically maintained at the mid point of the resistances of the two logic states. In our logic, we can however take advantage of the bit to bit coupling to generate a reference free reading. Eliminating the need for reference reduces the cost of read circuit. In this section we will describe how we can read from the logic without using a reference value.

The read circuit is shown in Fig. 7b. It performs a differential read by comparing the output against its complementary value. The main advantage of this type of reading is its high sense margin. The read circuit is also non-destructive, which means that the values in the output cells are not disturbed during the read process. This is possible because the read current is much lower than the critical current for switching. Hence, the overall read circuit is also low power. Moreover, the read operation is made tolerant to the variations in the output MTJs by (i) regenerating a copy of both the output and its complement; and (ii) comparing a pair of the output states (M_{1a}, M_{1b}) against a pair of the complementary states (M_{2a}, M_{2b}). This averages out the variations and improves the overall sense margin.

The read circuit operates in two phases: the precharge and the sense phase. In the precharge phase, transistors M_3 and M_4 are turned off and transistors M_7, M_8 and M_9 are turned on. Together with M_5 and M_6 they help the nodes X and Y to rise to V_{DD} at the end of the precharge phase. At the start of the sense phase, transistors M_7, M_8 and M_9 are first turned off. A small potential V_{read} is then applied to transistors M_3 and M_4. The unequal resistances in the two arms, caused by the pair of complementary MTJ states, gives rise to a potential difference across X and Y. This potential difference is sensed by the comparator and a decision is taken on the output state. More in-depth analysis of the read circuit is available in [24,31].

4.4 Execution Time of 2-Input Exclusive-OR

The output generation in cell X_{13} (Fig. 6) can be subdivided into five different time zones.

(i) *Time zone* 1*:* Duration t_w, when the inputs are written into cells $X_1 \cdots X_4$.
(ii) *Time zone* 2*:* Duration t_{clk}, when the cells $X_5 \cdots X_7$ and $X_8 \cdots X_{10}$ are clocked.
(iii) *Time zone* 3*:* Duration t_{clk}, when the cells X_{11} and X_{12} are clocked and the cells $X_5 \cdots X_7$ and $X_8 \cdots X_{10}$ are released.

(iv) *Time zone* 4*:* Duration t_{clk}, when the cell X_{13} is clocked and the cells X_{11} and X_{12} are released.
(v) *Time zone* 5*:* Duration t_s, when the cell X_{13} is released.

The total time to compute the output at X_{13} is therefore $t_w + 3t_{clk} + t_s$.

4.5 Performance Analysis

To judge the performance of this architecture we will compare it to NML. As we discussed in Sect. 3, NML uses single layer single domain nanomagnets and their magnetostatic interaction to compute. It is non-volatile with zero leakage. However, as we mentioned, its main concerns are high power, non-scalability of write and clock current, lack of CMOS interface for read and write and lack of control over individual nanomagnets during computing. To alleviate these concerns, we have used magnetostatic coupling between multilayer STT-MRAM cells and STT current to realize the non-volatile logic-in-memory architecture. CMOS integrability in the architecture allows control over 2×2 array of cells. STT based write and clock used in the architecture are low power, scalable and CMOS driven when compared to field induced write and clock in NML. Finally, MR based reading practiced in the architecture is both low power and CMOS driven that can be integrated on-chip. Table 2 provides the necessary comparison. From the data, we can safely conclude that

1. the architecture has low power write, clock and read mechanisms;
2. it has more control over individual MTJs in logic than was possible in NML; and
3. it provides electrical interface with the logic through CMOS peripherals.

We will next see how the architecture can very interestingly share logic responsibility between its magnetic and CMOS plane for improved delay, energy and area performance.

5 Logic Partitioning

Till now we have seen the magnetic plane in the architecture to do the main computation. The role of the CMOS plane was to take care of the resource management for the magnetic plane. In this section we will see how to give some additional responsibility to the CMOS plane, in order to execute more efficiently while still enjoying the non-volatile property of the magnetic plane. We have named it "the logic partitioning". To do this partitioning, we will be taking the help of the well-known Shannon expansion of logic function. For better understanding, we will take the help of a 2-input exclusive-OR to explain the partitioning. Also, 2-input exclusive-OR is a fundamental component to any arithmetic and logical operation. This concept of logic partitioning can be extended for any n-input logic function as well [36,37].

Table 2. Comparison of logic behavior and logic performance between NML and the STT-based logic-in-memory architecture.

	Nanomagnetic logic	Logic-in-memory architecture
	Logic characteristics and signal types	
Cell types	Single layer nanomagnets	Multi-layer magnetic tunnel junctions [24]
Cell dimensions & spacing	$100 \times 50\,\text{nm}^2$, 20 nm [32]	$100 \times 50\,\text{nm}^2$, 20 nm [29]
Computing style	Dipolar magnetic coupling [6]	Dipolar magnetic coupling [29]
CMOS interface	None	Access transistors and metal lines
User control	Over groups of cells	Over single MTJ with access transistor [24]
Inputs	Oersted field or through explicit neighbor interaction [33]	Electric, STT current [24]
Clock	Oersted field [34]	STT current [30]
Outputs	Magnetic sensing [34]	MR based [24]
	Current & timing	
Input current (writing 1)	$> 2.29\,\text{mA}$, 3 ns [5,34]	$280\,\mu\text{A}$, < 0.5 ns [24,35]
Input current (writing 0)	$> 2.29\,\text{mA}$, 3 ns [5,34]	$216\,\mu\text{A}$, < 0.5 ns [24,35]
Clocking current	2.29 mA, 3 ns [34]	$170\,\mu\text{A}$, 3 ns [30,35]
Output current	–	$30\,\mu\text{A}$, 4 ns [24]

As a quick recap, Shannon expansion allows any logic function $f(x_1, x_2, \ldots, x_i, \ldots, x_n)$ to be expanded in terms of singled out variables x_i and their cofactor f_i as shown in Eq. 8.

$$f(x_1, x_2, \ldots, x_i, \ldots, x_n) = x_i f_i(x_1, x_2, \ldots, 1, \ldots, x_n) + x_i{}' f_i{}'(x_1, x_2, \ldots, 0, \ldots, x_n) \tag{8}$$

Instead of computing the entire logic, we will now compute the cofactors in the magnetic plane. The responsibility for the singled out variable(s) is transferred to the CMOS plane. For better understanding, let us once again look into the 2-input exclusive-OR function. It can be readily expressed into singled out variable x_i and cofactors $x_1 \oplus x_2 = x_1 \cdot x_2' + x_1' \cdot x_2$. Assuming the singled out variable is x_1, our approach would require us to compute (x_2 or x_2') in the magnetic plane. This is much simpler. The final choice between x_2 or x_2' in the final output is now made in the CMOS plane depending on the value of x_1. This approach significantly cuts down the cost of execution by avoiding some intermediary logic steps in the magnetic plane. At the same time, it also retains the flavor of non-volatile outputs. Some intermediary results are still left in the magnetic plane.

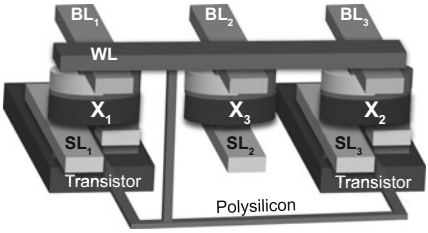


Fig. 8. Cell layout for 2-input exclusive-OR executed through logic partitioning.

5.1 Execution Principle for the 2-Input Exclusive-OR

The 2-input exclusive-OR is evaluated on the basis of the following logical deductions:

1. When $x_1 = 1$, $x_1 \oplus x_2 = x_2'$;
2. When $x_1 = 0$, $x_1 \oplus x_2 = x_2$.

The cell layout for the 2-input exclusive-OR is shown in Fig. 8. The logic execution takes place in the following four phases.
Phase I: The cofactors x_2 and x_2' are written into cells X_1 and X_2 respectively.
Phase II: X_3 is clocked followed by clocking X_1 (if $x_1 = 1$) or X_2 (if $x_1 = 0$).
Phase III: X_3 is released. The cell that was not clocked in Phase II, now acts as a driver for X_3. It eventually takes the value of $X_2 = x_2'$ (if $x_1 = 1$) or $X_1 = x_2$ (if $x_1 = 0$).
Phase IV: All clocked cells are released, i.e. X_1 (if $x_1 = 1$) and X_2 (if $x_1 = 0$).

The algorithm for the logic partitioning based 2-input exclusive-OR execution is provided in Algorithm 2.

5.2 Performance Analysis

We will investigate the cost of execution of the 2-input exclusive-OR by logic partitioning (Fig. 8) in terms of overall (i) area/cell count; (ii) delay; and (iii) energy and compare it to the standard mode of execution in the STT-based non-volatile logic-in-memory framework (Fig. 6).

Area/cell count: The total cell count of the logic is three (see Fig. 8). This is 75 % less than the sixteen cell standard 2-input exclusive-OR (see Fig. 6).

Delay: With logic partitioning, it takes $(t_w + t_{clk} + t_s)$ to generate the output. t_w is the MTJ writing time (Phase I), t_{clk} is the clocking time for the MTJ (Phase II) and t_s is the time for the MTJ to settle after the clock is released (Phase III). In the standard mode of operation, the time to compute the output is $(t_w + 3t_{clk} + t_s)$ (see Sect. 4.4). Logic partitioning therefore gives us a reduced delay of $2t_{clk}$.

Energy: With logic partitioning, the total energy required for computation is $(2I_w t_w + 2I_{clk} t_{clk})V_{dd}$ where, I_w and I_{clk} are the writing and clocking current

Algorithm 2. 2-Input xor execution $x_1 \oplus x_2$ with logic partitioning

1: **Input:** x_1, x_2. **Output:** $X_3 = f(x_1, x_2) = x_1 \oplus x_2$.
2: **Desired:** $f(0, x_2) = x_2$; $f(1, x_2) = x_2'$.
3: ***Logic execution sequence:***
4: **Phase I:** *Write* $X_1 = x_2$ and $X_2 = x_2'$.
5: **Phase II:** Clock cell X_3
6: **if** $A = 1$ **then**
7: *Clock cell* X_1
8: **else**
9: *Clock cell* X_2.
10: **end if**
11: **Phase III:** Release the clock for cell X_3.
12: **Phase IV:** Deactivate all metal lines.
13:
14: ***Cell Count N*** $= 3$.
15: ***Total Delay T***$= t_w + t_{clk}$, where t_w and t_{clk} are the durations writing and clocking cycles for a cell.
16: ***Total Energy E*** $= V_{dd} \times (2I_w t_w + 2I_{clk} t_{c)lk}$, where I_w and I_{clk} are the average currents to write and clock a cell.

for the MTJs. $2I_w t_w V_{dd}$ is the energy to write into X_1 and X_2, while $2I_{clk} t_{clk} V_{dd}$ is the energy to clock cells X_1 and X_3 or X_2 and X_3 depending on the input. Compare this to the standard mode of operation where the energy required is $(4I_w t_w + 9I_{clk} t_{clk}) V_{dd}$. $4I_w t_w V_{dd}$ is the energy to write into the four cells $X_1 \cdots X_4$ (Fig. 6) and $9I_{clk} t_{clk} V_{dd}$ is the energy to clock the nine cells $X_5 \cdots X_{13}$ in the body of logic. Logic partitioning therefore gives us an energy reduction by $(2I_w t_w + 7I_{clk} t_{clk})$.

For an average write current of $I_w = 250\,\mu\text{A}$ and an average clocking current $I_{clk} = 170\,\mu\text{A}$ and a conservative write time of 0.5 ns and clocking time of 3 ns, the logic partitioning provides a 48 % improvement in speed, over 65 % savings in

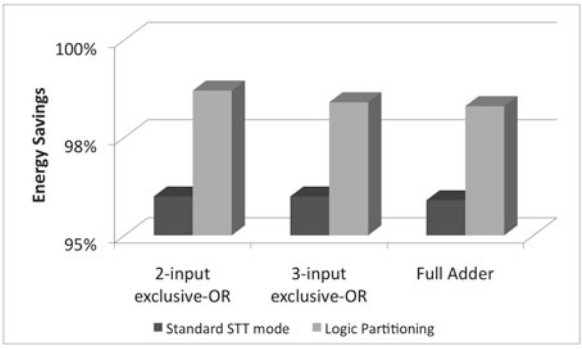


Fig. 9. Energy savings with STT-based logic-in-memory mode and logic partitioning mode over traditional NML.

energy and 81.25 % savings in area over the standard AND-OR implementation of the 2-input exclusive-OR in the logic-in-memory architecture.

Finally, Fig. 9 provides an estimate of the energy savings in some datapath components when executed in standard STT-based logic-in-memory mode and in logic partitioning mode over traditional NML.

6 Conclusion

In this chapter we have discussed a logic-in-memory architecture that is an integration of two different technologies, STT-MRAM and NML. With clock as control, the architecture switches between logic and memory mode of operations. The computation in the architecture is NML style using the magnetostatic coupling between the multilayer STT-MRAM cells. The writing and reading from the logic is STT-MRAM style using CMOS assisted spin transfer torque and magnetoresistance of their cells. This unique integration between the two technologies gives a logic platform that is non-volatile like NML but has a much lower power consumption and better user interface. Finally, we have shown a Shannon based logic partitioning between the magnetic plane and the CMOS plane of the architecture that provides significant performance boost to datapath elements like exclusive-OR in the architecture.

Acknowledgement. This work is partially supported by NSF Career Award CCF (0639624), NSF EMT/Nano CCF (0824838), NSF CRI (0551621) and USF Presidential Doctoral Fellowship.

References

1. Frank, D.J., Dennard, R.H., Nowak, E., Solomon, P.M., Taur, Y., Wong, H.-S.P.: Device scaling limits of Si MOSFET's and their application dependencies. Proc. IEEE **89**(3), 259–288 (2001)
2. Slonczewski, J.C.: Current-driven excitation of magnetic multilayers. J. Magn. Magn. Mater. **159**(1), L1–L7 (1996)
3. Parkin, S.S., Kaiser, C., Panchula, A., Rice, P.M., Hughes, B., Samant, M., Yang, S.-H.: Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers. Nat. Mater. **3**(12), 862–867 (2004)
4. http://www.everspin.com/
5. Cowburn, R., Welland, M.: Room temperature magnetic quantum cellular automata. Science **287**(5457), 1466–1468 (2000)
6. Imre, A., Csaba, G., Ji, L., Orlov, A., Bernstein, G.H., Porod, W.: Majority logic gate for magnetic quantum-dot cellular automata. Science **311**(5758), 205–208 (2006)
7. Salahuddin, S.: Current induced switching of ferromagnets for low-power memory applications. In: ISQED Symposium, Tutorial (2011)
8. Zhu, J.-G.J., Park, C.: Magnetic tunnel junctions. Mater. Today **9**(11), 36–45 (2006)

9. Yuasa, S., Djayaprawira, D.: Giant tunnel magnetoresistance in magnetic tunnel junctions with a crystalline MgO (0 0 1) barrier. J. Phys. D: Appl. Phys. **40**(21), R337 (2007)
10. Thomson, W.: Tunneling between ferromagnetic films. Proc. R. Soc. Lond. **8**, 546 (1856)
11. Julliere, M.: Tunneling between ferromagnetic films. Phys. Lett. A **54**(3), 225–226 (1975)
12. Yuasa, S., Fukushima, A., Kubota, H., Suzuki, Y., Ando, K., et al.: Giant tunneling magnetoresistance up to 410% at room temperature in fully epitaxial Co/MgO/Co magnetic tunnel junctions with bcc Co (001) electrodes. Appl. Phys. Lett. **89**(4), 42505–42505 (2006)
13. Maekawa, S. (ed.): Concepts in Spin Electronics. Oxford Science Publications, Oxford (2006)
14. Lin, C., Kang, S., Wang, Y., Lee, K., Zhu, X., Chen, W., Li, X., Hsu, W., Kao, Y., Liu, M., et al.: 45 nm low power CMOS logic compatible embedded STT MRAM utilizing a reverse-connection 1T/1MTJ cell. In: 2009 IEEE International Electron Devices Meeting (IEDM), pp. 1–4. IEEE (2009)
15. Chen, E., Apalkov, D., Diao, Z., Driskill-Smith, A., Druist, D., Lottis, D., Nikitin, V., Tang, X., Watts, S., Wang, S., et al.: Advances and future prospects of spin-transfer torque random access memory. IEEE Trans. Magn. **46**(6), 1873–1878 (2010)
16. Vacca, M., Graziano, M., Zamboni, M.: Majority voter full characterization for nanomagnet logic circuits. IEEE Trans. Nanotechnol. **11**(5), 940–947 (2012)
17. Niemier, M., Alam, M., Hu, X., Bernstein, G., Porod, W., Putney, M., DeAngelis, J.: Clocking structures and power analysis for nanomagnet-based logic devices. In: ISLPED, pp. 26–31. ACM, New York (2007)
18. Graziano, M., Vacca, M., Chiolerio, A., Zamboni, M.: An NCL-HDL snake-clock-based magnetic QCA architecture. IEEE Trans. Nanotechnol. **10**(5), 1141–1149 (2011)
19. Pulecio, J.F., Bhanja, S.: Magnetic cellular automata coplanar cross wire systems. J. Appl. Phys. **107**(3), 034308–034308-5 (2010)
20. Cowburn, R., Adeyeye, A., Welland, M.: Controlling magnetic ordering in coupled nanomagnet arrays. New J. Phys. **1**(1), 16 (1999)
21. Kumari, A., Bhanja, S.: Landauer clocking for magnetic cellular automata (MCA) arrays. IEEE Trans. Very Large Scale Integr. VLSI Syst. **19**(4), 714–717 (2011)
22. Csaba, G., Lugli, P., Csurgay, A., Porod, W.: Simulation of power gain and dissipation in field-coupled nanomagnets. J. Comput. Electron. **4**(1), 105–110 (2005)
23. Das, J., Alam, S.M., Bhanja, S.: Low power CMOS-magnetic nano-logic with increased bit controllability. In: 2011 11th IEEE Conference on Nanotechnology (IEEE-NANO), pp. 1261–1266. IEEE (2011)
24. Das, J., Alam, S., Bhanja, S.: Ultra-low power hybrid CMOS-magnetic logic architecture. IEEE TCAS-I **59**, 2008–2016 (2012)
25. Das, J., Alam, S.M., Bhanja, S.: Nanoelectronic Device Applications Handbook, Chap. 60, vol. 16, 1st edn. CRC Press Llc, Boca Raton (2013)
26. International technology roadmap for semiconductor (2009)
27. Donahue, M., Porter, D.: Oommf user's guide, version 1.0, interagency report nistir 6376. NIST, Gaithersburg, MD (1999)
28. Scheinfein, M.R.: LLG micromagnetics simulator
29. Karunaratne, D.K., Bhanja, S.: Study of single layer and multilayer nano-magnetic logic architectures. J. Appl. Phys. **111**(7), 07A928–07A928-3 (2012)

30. Das, J., Alam, S., Bhanja, S.: Low power magnetic quantum cellular automata realization using magnetic multi-layer structures. IEEE JETCAS **1**, 267–276 (2011)
31. Das, J., Alam, S.M., Bhanja, S.: Non-destructive variability tolerant differential read for non-volatile logic. In: Proceedings of the 55th International Midwest Symposium on Circuits and Systems (2012)
32. Pulecio, J., Pendru, P., Kumari, A., Bhanja, S.: Magnetic cellular automata wire architectures. IEEE Trans. Nanotechnol. **10**(6), 1243–1248 (2011)
33. Bhanja, S., Pulecio, J.: A review of magnetic cellular automata systems. In: 2011 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 2373–2376 (2011)
34. Dingler, A., Niemier, M.T., Hu, X.S., Lent, E.: Performance and energy impact of locally controlled NML circuits. J. Emerg. Technol. Comput. Syst. **7**, 2:1–2:24 (2011)
35. Zhao, H., Lyle, A., Zhang, Y., Amiri, P., Rowlands, G., Zeng, Z., Katine, J., Jiang, H., Galatsis, K., Wang, K., et al.: Low writing energy and sub nanosecond spin torque transfer switching of in-plane magnetic tunnel junction for spin torque transfer random access memory. J. Appl. Phys. **109**(7), 07C720–07C720-3 (2011)
36. Das, J., Alam, S., Bhanja, S.: Nano magnetic STT-logic partitioning for optimum performance. IEEE Trans. Very Large Scale Integr. VLSI Syst. **22**, 90–98 (2014)
37. Das, J., Alam, S., Bhanja, S.: A novel design concept for high density hybrid cmos-nanomagnetic circuits. In: 2012 12th IEEE Conference on Nanotechnology (IEEE-NANO), pp. 1–6 (2012)

Security Issues in QCA Circuit Design - Power Analysis Attacks

Weiqiang Liu[1(✉)], Saket Srivastava[2], Máire O'Neill[3], and Earl E. Swartzlander Jr.[4]

[1] College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
liuweiqiang@nuaa.edu.cn

[2] School of Engineering, University of Lincoln, Lincoln LN6 7TS, UK
ssrivastava@lincoln.ac.uk

[3] ECIT, Queen's University Belfast, Belfast BT3 9DT, UK
m.oneill@ecit.qub.ac.uk

[4] Department of ECE, University of Texas at Austin, Austin 78712, USA
eswartzla@aol.com

Abstract. Quantum-dot cellular automata (QCA) technology has advantages of fast computation performance, high density and low power consumption. Thus, it is believed that QCA is attractive for designing future digital systems. Side channel attacks including power analysis attacks have become a significant threat to the security of cryptographic circuits using CMOS technology. A power analysis attack can reveal the secret key of a cryptographic cipher by measuring the power consumption of the cipher's hardware platform while it is encrypting or decrypting data. As the power consumption of QCA circuits is extremely low when compared to their CMOS counterparts, it may be possible to build cryptographic circuits that are immune to power analysis attacks by using QCA technology. Therefore, in this chapter an investigation into both the best and worst case scenarios for attackers is carried out to ascertain if QCA circuits have such an advantage. A more efficient QCA design of a sub-module of the Serpent cipher is proposed and compared to a previous design. By using an upper bound power model, the first power analysis attack of a QCA cryptographic circuit (Serpent sub-module) is presented. The results show that in the best case scenario for attackers, QCA cryptographic circuits would be vulnerable to power analysis attack. However, the security of practical QCA circuits can be greatly improved by applying a smoother clock. Moreover, in the worst case scenario, reversible QCA circuits with Bennett clocking could be used as a natural countermeasure to power analysis attack. Therefore, it is believed that QCA could be a niche technology in the future for the implementation of security architectures resistant to power analysis attack.

Keywords: Quantum-dot Cellular Automata (QCA) · QCA power models · Power analysis attack · Cryptography · Serpent cipher · S-box · Security architectures

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 194–222, 2014.
DOI: 10.1007/978-3-662-43722-3_9, © Springer-Verlag Berlin Heidelberg 2014

1 Introduction

CMOS technology is approaching its scaling limitation [1]. New nanotechnologies are explored to continue Moore's Law. Among the possible alternatives to CMOS technology, Quantum-dot cellular automata (QCA) [2,3] is a promising one. QCA can offer significant advantages including fast speed, high density and low power consumption. There are four possible implementations of QCA technology [4]: metal-island QCA, semiconductor QCA, molecular QCA and magnetic QCA. The use of semiconductor QCAs is assumed in this chapter, as most prototypes of QCA circuits were demonstrated by metal-island and semiconductor QCA so far. Semiconductor QCAs have been manufactured from standard semiconductive materials [5]. However, they can only work in very low temperature.

The elementary units of QCA technology are cells. The cells consist of square nano-structures with a quantum-dot in each corner as shown schematically in Fig. 1(a). Two electrons that can tunnel among the four quantum dots are resident in a QCA cell. These two electrons tend to occupy antipodal sites within the cell due to Coulombic repulsion. The tunneling action only occurs within cells. Therefore, two ground states with polarisations of '−1' and '1' can be used to represent binary '0' and '1', respectively. There is no physical wire in QCA. Instead, a chain of coupled cells is used as a QCA "wire," as shown in Fig. 1(b). There are two basic logic components in QCA: inverters and three-input majority gates. Three-input majority gates realize the following logic function:

$$M(a, b, c) = ab + ac + bc, \tag{1}$$

where a, b and c are inputs. Two-input AND or OR gates can be implemented by fixing one of majority gate's input to '0' or '1', respectively. A four-phase clocking scheme is typically used for semiconductor QCAs [3]. QCA circuits are divided into four clocking zones and each zone contains four phases with a 90° phase shift between adjacent zones. A latched clocking zone is used as the input to the subsequent zone and cells in the other two zones do not affect these computing zones as they are in inactive states. Since the cells in one clocking zone become latched and remain in this state until the cells are latched in the next clocking zone, information is transferred and processed in a pipelined manner.

QCA technology not only provides a fundamentally novel physical structure, but also offers a new kind of computing architecture for digital design. Simple QCA components have been fabricated and demonstrated [6–8] and more complex circuits [9–15] have been designed and verified by simulation tools. Digital design methods for QCA circuits [16–21] have also been explored to achieve more efficient designs.

The power consumption of QCA designs is not expected to be significant [22]. However, side channel analysis (SCA) attacks [23,24] have emerged as a significant threat to CMOS cryptographic circuits in the past decade. These attacks exploit the key information leaked by the physical implementation of a cryptographic cipher. Typically the amount of side-channel information required to break the cipher is very small [24]. One of the most powerful SCA techniques is

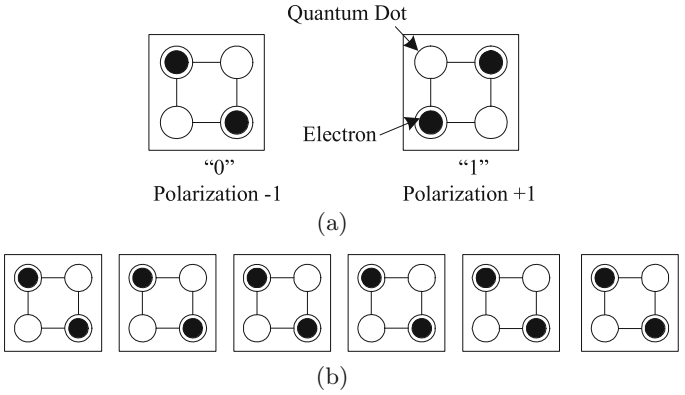


Fig. 1. Schematics of QCA cell and wire: (a) Binary QCA cells, (b) A QCA wire composed of coupled cells.

a power analysis attack [25–27], which can extract the secret key of an electronic security device by measuring the power consumption of the device while it is performing cryptographic operations. The power consumption of a CMOS device is highly dependant on processed data. As QCA is a field coupled computing paradigm which only involves changing the position of electrons in the cells, its power consumption is ultra-low. Therefore, it is interesting and appropriate to ask if QCA cryptographic circuits are immune to power analysis attack.

The power dissipation of QCA circuits has been studied. A quantum mechanical power model based on the density matrix formalism was proposed by Timler and Lent [22,28]. This model can estimate the power dissipation of QCA circuits quite accurately, however, it is computationally expensive to use for large scale designs. The energy dissipated in a 2-cell chain was investigated using an RC model by Bond and Macucci [29]. Based on the work by Timler and Lent [22], a upper bound power model to compute the upper bound of the power consumed in QCA circuits was proposed by Srivastava *et al.* [30]. This power model can provide a tight upper bound of the power consumption in QCA circuits for quasi-adiabatic (smooth) switching. The lower bound power dissipation for QCA circuits was also studied by Lent *et al.* [31]. The lower bound power dissipation was studied by applying a Bennett clock to logically reversible QCA circuits.

Both the best case and the worst case scenarios for an attacker are considered to evaluate the security of QCA cryptographic circuits in terms of power analysis attacks. Power dissipation during quasi-adiabatic switching is very low [22]. However, a more realistic scenario of switching is more abrupt and hence will introduce more power dissipation. Generally, the greater the power dissipation, the easier it is to perform a power analysis attack [25,32]. The best case scenario for attackers involves upper bound power dissipation with non-adiabatic clocking. Therefore, the upper bound power model [30] can be used to assess the performance of QCA cryptographic circuits under power analysis attack. To fully address the security issues related to power analysis attack, the worst

case scenario (for attackers) should also be considered with a lower bound power model that is related to Bennett-clocked QCA circuits. The power dissipation under Bennett clocking can be arbitrarily low, even lower than the bit erasure energy $k_B T ln(2)$ [31]. The lower bound power model proposed by Lent *et al.* [31] is used to study this worst case scenario.

First the best case scenario is considered. The power dependence on the Hamming distance of inputs in QCA gates is demonstrated as a fundamental step in the power analysis attack. As a case study, a sub-module of the Serpent cryptographic block cipher with its 4-bit × 4-bit S_0-box is designed in QCA. It is verified using QCAPro, which is a fast power estimation tool for QCA circuit design [33]. The power consumption of the Serpent sub-module is then simulated in QCAPro based on the upper bound power model. A power analysis attack procedure for QCA is proposed to reveal the secret key by statistically comparing the simulated power consumption and all hypothetical key guesses. The first power analysis attack results show that QCA cryptographic circuits could be at risk of being attacked under typical quasi-adiabatic clocking. Then a more realistic scenario is discussed for practical QCA devices. Finally, the worst case scenario for attackers is also studied based on a Bennett-clocked QCA gate.

This chapter is organized as follows: Sect. 2 presents an overview of the upper bound power model for QCA circuits. The power dissipation that is dependent on the Hamming distance (HD) in basic QCA gates is also shown in this section. The design of the Serpent sub-module with its 4-bit × 4-bit S_0-box is presented and compared with a previous design in Sect. 3. Section 4 proposes a procedure of performing a power analysis attack on QCA circuits using the upper bound power model and provides a power analysis attack results of QCA circuits. A discussion of practical scenario of QCA devices is given in Sect. 5. The vulnerability of Bennett-clocked QCA circuits to power analysis attack is also studied in Sect. 6. Section 7 concludes this chapter.

2 Upper Bound Power Model and Power Dependence on Hamming Distance

The power flow of a QCA cell is shown in Fig. 2(a), where P_{in} is the signal power from the neigbouring cell on its left, P_{out} is the signal power transferred to the cell on its right, P_{clock} is the energy provided by the clock and P_{diss} is the power dissipated. The signal powers P_{in} and P_{out} are generally equal as demonstrated in [22]. A considerable amount of energy is drawn into the cell from the clock as the barriers are being raised. Most of that energy is returned to the clock as the barriers are lowered. The difference between the powers in and out of the clock is P_{diss}. Since P_{diss} is of interest in this research, it is not necessary to include other parts of the cell's power flow in the simulation of power dissipation. Power is only dissipated in a QCA cell when actual computation is performed.

As shown in Fig. 2(b), the power dissipation of a CMOS circuit can be measured by monitoring the power supply, i.e., P_{supply}. Similarly, since the power

of a QCA circuit is provided by its clock [22], the dissipated power can be measured by the power provided by the clock, i.e., P_{clock}. Previous research [34] has shown that the power dissipated in the clocking wires is fairly small and that dissipation in the QCA devices themselves will dominate the power dissipation. Even if the power losses in the clock wire are large, these power losses are not data dependent, thus they will not affect the results of a power analysis attack [32]. Therefore, only the power dissipated by the QCA cells is considered.

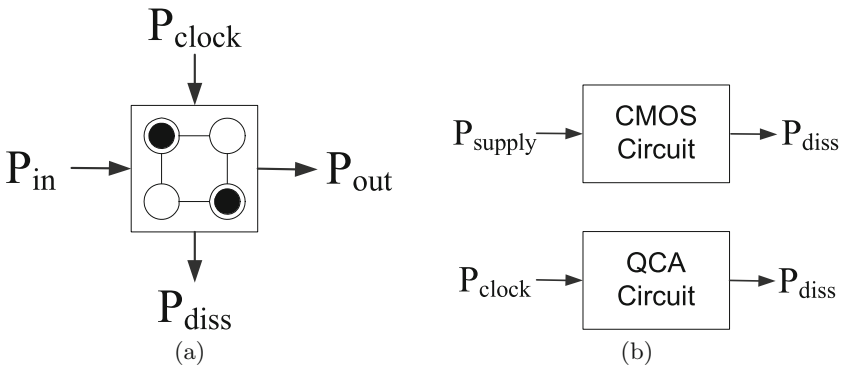


Fig. 2. Power dissipation in QCA: (a) Power flow in a QCA cell; (b) Power dissipation in CMOS and QCA circuits.

As typical switching in QCA is not purely adiabatic, power consumption is unavoidable. The upper bound of quasi-adiabatic power dissipation is reached with non-adiabatic clocking. The best case scenario for attackers is considered by using the upper bound power of QCA circuits to determine if they are vulnerable to power analysis attack. An upper bound power dissipation model [30] is used, which has been derived from a quasi-adiabatic model [22]. In this section, an overview of the upper bound power model is presented. The validity of the upper bound power is confirmed by calculating the actual power dissipated for various levels of clock smoothness in a QCA cell. Then the power dependence on the Hamming distance in QCA cells and basic components under different tunnelling energy levels and temperatures is illustrated.

2.1 Upper Bound Power Model

A Hamiltonian matrix can be used to describe the total energy of a QCA cell. Using the Hartree-Fock approximation [35] and assuming that only Coulombic interactions apply between cells, the matrix representation of the Hamiltonian for an array of cells is [22,30]:

$$H = \begin{bmatrix} -\frac{1}{2}\sum_i E_k x_i f_i & -\gamma \\ -\gamma & \frac{1}{2}\sum_i E_k x_i f_i \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}G & -\gamma \\ -\gamma & \frac{1}{2}G \end{bmatrix}, \tag{2}$$

where
$\sum_i$ is the sum over the cells,
E_k is the kink energy,
f_i is the geometric factor that specifies the electronic falloff with distance between cells,
x_i is the polarisation of the ith neighbour cell,
γ is the tunnelling energy between two logic states of a cell which is controlled by the clock,
G is the total kink energy caused by neighbouring polarized cells.

The energy of a QCA cell at each clock cycle is the expected value of the Hamiltonian, which is given by [22,30]:

$$E =< H >= \frac{\hbar}{2}\boldsymbol{\Gamma} \cdot \boldsymbol{\lambda}, \tag{3}$$

where $\hbar$ is the reduced Planck constant, $\boldsymbol{\lambda}$ is the coherence vector and $\boldsymbol{\Gamma}$ is the three dimensional energy vector:

$$\boldsymbol{\Gamma} = \frac{1}{\hbar}\left[-2\gamma, 0, G\right]. \tag{4}$$

The equation for the instantaneous power can then be derived as follows:

$$P_{total} = \frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\hbar}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{\Gamma} \cdot \boldsymbol{\lambda}\right) = \frac{\hbar}{2}\left(\frac{\mathrm{d}\boldsymbol{\Gamma}}{\mathrm{d}t}\right) \cdot \boldsymbol{\lambda} + \frac{\hbar}{2}\boldsymbol{\Gamma} \cdot \left(\frac{\mathrm{d}\boldsymbol{\lambda}}{\mathrm{d}t}\right). \tag{5}$$

The first term represents the power in and out of the clock and the inter-cell power flow. It is the second term, namely P_{diss}, that refers to the dissipated power. Therefore, the power dissipation of a QCA cell can be calculated as:

$$P_{diss} = \frac{\hbar}{2}\boldsymbol{\Gamma} \cdot \left(\frac{\mathrm{d}\boldsymbol{\lambda}}{\mathrm{d}t}\right). \tag{6}$$

Energy dissipated in one clock cycle $T_c = [-D, D]$ can be computed by [30]:

$$\begin{aligned} E_{diss} &= \frac{\hbar}{2}\int_{-D}^{D} \boldsymbol{\Gamma} \cdot \frac{\mathrm{d}\boldsymbol{\lambda}}{\mathrm{d}t}\mathrm{d}t = \frac{\hbar}{2}\left(\left[\boldsymbol{\Gamma} \cdot \boldsymbol{\lambda}\right]_{-D}^{D} - \int_{-D}^{D} \boldsymbol{\lambda} \cdot \frac{\mathrm{d}\boldsymbol{\Gamma}}{\mathrm{d}t}\mathrm{d}t\right) \\ &= \frac{\hbar}{2}\left(\boldsymbol{\Gamma}_+ \cdot \boldsymbol{\lambda}_+ - \boldsymbol{\Gamma}_- \cdot \boldsymbol{\lambda}_- - \int_{-D}^{D} \boldsymbol{\lambda} \cdot \frac{\mathrm{d}\boldsymbol{\Gamma}}{\mathrm{d}t}\mathrm{d}t\right), \end{aligned} \tag{7}$$

where $\boldsymbol{\Gamma}_-$ and $\boldsymbol{\Gamma}_+$ are used to denote $\boldsymbol{\Gamma}\left(-D\right)$ and $\boldsymbol{\Gamma}\left(D\right)$ and the same notation is used for $\boldsymbol{\lambda}$. The maximum power will be dissipated when the change of $\boldsymbol{\Gamma}$ is a maximum under non-adiabatic switching. By modelling a step change with a delta function, the upper bound of power dissipation for a QCA cell is derived as follows [30]:

$$P_{diss} = \frac{E_{diss}}{T_c} < \frac{\hbar}{2T_c}\boldsymbol{\Gamma}_+ \times \left[-\frac{\boldsymbol{\Gamma}_+}{|\boldsymbol{\Gamma}_+|}tanh\left(\frac{\hbar\,|\boldsymbol{\Gamma}_+|}{kT}\right) + \frac{\boldsymbol{\Gamma}_-}{|\boldsymbol{\Gamma}_-|}tanh\left(\frac{\hbar\,|\boldsymbol{\Gamma}_-|}{kT}\right)\right], \tag{8}$$

where $\boldsymbol{\Gamma}_+$ and $\boldsymbol{\Gamma}_-$ are the values of the Hamiltonian before and after the transition, k is the Boltzmann constant and T is the temperature. Once the pre-transition and post-transition Hamiltonians are known, the upper bound of power dissipation for a QCA cell can be calculated. The power model is derived by including the effects of dissipative coupling to a heat bath.

The power dissipation model for each QCA cell is similar, therefore the total power for one clock cycle can be computed by summing up the power consumed by each cell. The cells interact via the electronic kink energy between them. Power can be computed by tracking the polarisation of the cells before and after the switching of the input cells, where next-near-neighbour coupling is ignored since the interaction between cells falls off by the fifth power of the distance [22]. Using the values of cell polarisation and clock energy, the total power consumption in a QCA circuit for the transition from a current input to the next input can be calculated.

To perform a power analysis attack of a QCA circuit, a power tool for large circuits is required. Based on the upper bound power model discussed earlier, a power tool called QCAPro was developed [33]. QCAPro also offers a quick design check to verify the correct polarisation of a design. The input required for the QCAPro tool is the layout file generated by QCADesigner [36] which is a widely used design and simulation tool for QCA. The current version of QCAPro [33] only provides the average, maximum and minimum power consumption of a QCA circuit during the input switching. For this research, it was necessary to modify the tool to compute and provide the power consumption in one clock cycle according to every input change.

2.2 Power Consumption in a Cell

The power dissipated in each cell is a function of the rate of change of the clock and the tunneling energy. The adiabaticity of the system is directly proportional to the amount of clock smoothing which can be implemented by a Gaussian function [30]. Clock signals with smoothing factors 0.6, 1.2, 1.8, 2.4 and 3.0, which correspond to the clock smoothing of 3.98×10^{-4}, 1.58×10^{-3}, 6.31×10^{-3}, 2.51×10^{-2} and 1.0×10^{-1}, are shown in Fig. 3. The actual power dissipated using the quantum model for various values of these parameters was calculated and compared with the upper bound power, which is shown in Fig. 4. The results show that the upper bounds do indeed hold and are reached when the clock smoothing is zero, i.e., non-adiabatic switching. Therefore, it confirms the validity of the upper bound power assumption in this research. Note that, as can be seen from Fig. 4, for a smoother clock, for higher tunneling energies, less power is dissipated, while for a less smooth clock, for higher tunneling energies, more power is dissipated.

The power consumption of a QCA cell is computed under a non-adiabatic clock. Two cases are shown in Fig. 5, including cell's polarisation keeps unchanged as 1 and changes from -1 to 1. There are three parts to the power consumption of a cell whenever the Hamiltonian changes, as shown in Fig. 5(b). The first part occurs when the clock goes from low to high (γ_L *to* γ_H) to depolarise a cell.

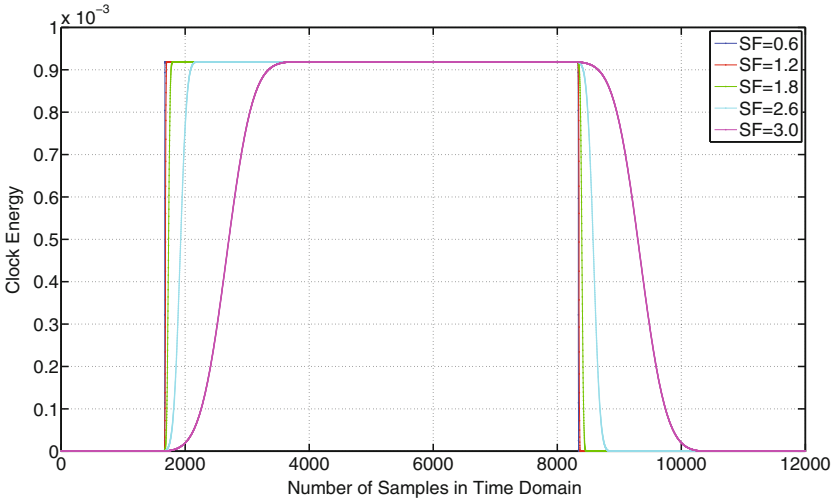


Fig. 3. Clock signals under smoothing factors (SF) of 0.6, 1.2, 1.8, 2.6 and 3.0.

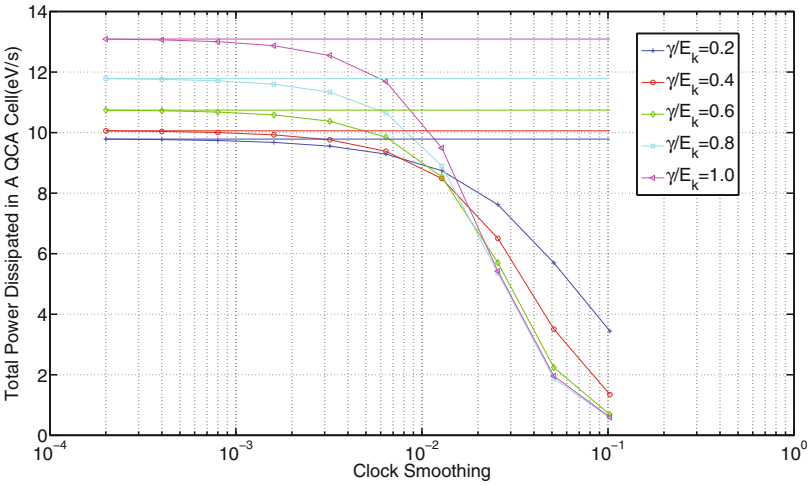


Fig. 4. Total power consumption in a QCA cell under various clock smoothing levels for different tunelling energy γ values ($T = 5.0\,\mathrm{K}$). Adiabaticity of the switching process is controlled by smoothness of the clock. The horizontal line show the upper bounds for each case.

The second part happens when the driven cell changes state (G_- *to* G_+), and the third part is produced by changing the clock from high to low (γ_H *to* γ_L), which latches the cell to the new state. The first and third parts of the power consumption occur even if there are no changes to the cell's polarisation as shown

in Fig. 5(a). Thus, these can be considered as static power. The second part of the power consumption is dependent on the changes to the cell's polarisation, which can be considered as dynamic power. It is clear from Fig. 5(b) that the dynamic power is significantly larger than the static power in a QCA cell. Since the dynamic power is dependent on the polarisation changes, the total power loss in a QCA cell can possibly provide information about its inputs.

Note that the absolute value of the power consumption is on the order of a few hundredths of an electron volt, i.e., eV, which is extremely low. However, a power analysis attack exploits the difference between the static power and dynamic power rather than the absolute value of the power.

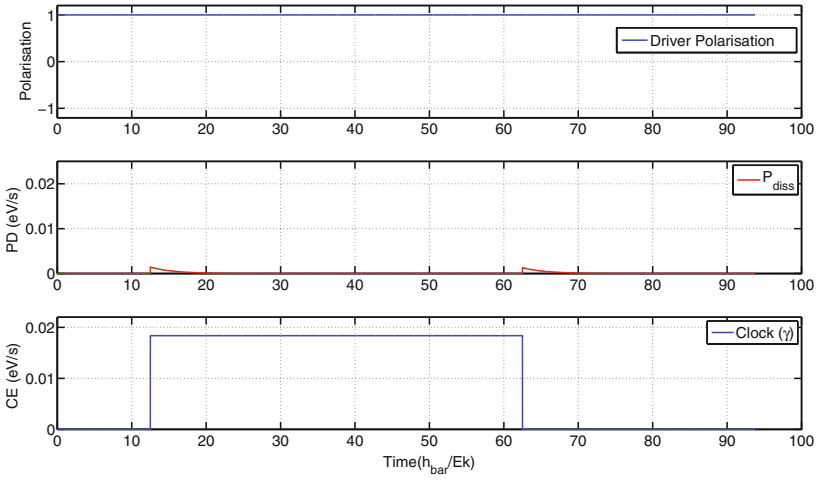
2.3 Power Dependence on Hamming Distance in CMOS and QCA Gates

The fundamental of power analysis attack is based on the fact that the power consumption of cryptographic circuits is dependent on the input data. The power consumption of CMOS circuits is mainly divided into two parts: static power consumption and dynamic power consumption [32]. The N-type and P-type MOS transistors are used to build the CMOS cells so that the pull-up network and pull-down network are not conducting at the same time. When there is no change from the input signal, there is only a little leakage current flowing through the turned-off transistor, which is termed the static power consumption. The case when only static power is consumed in CMOS is shown in Fig. 6 for an inverter that has a constant input.

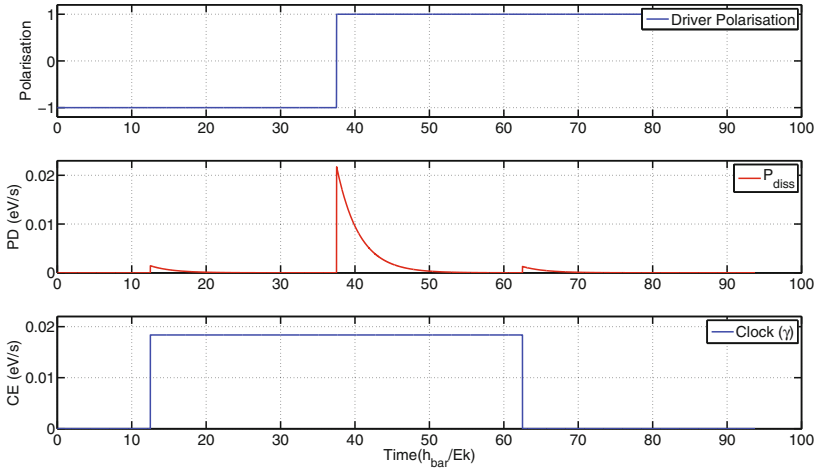
While the input of the inverter is changing from '1' to '0' or from '0' to '1', dynamic power is also consumed. The dynamic power is from two parts: the first one is the load capacitance which needed to be charged and the second one is a short circuit current occurs due to switched output signal. As shown in Fig. 7(a), when the input signal changed from '1' to '0', the inverter draw a charging current from the power supply to the load capacitance. In the other case as shown in Fig. 7(b) when input is changing from '0' to '1', the previously charged load capacitance dissipates energy. The power consumption of a CMOS inverter under four cases are summarized in Table 1. It is clear that the power consumption of an CMOS inverter is dependent on the Hamming Distance (HD) of its input. The HD between two inputs X_1 and X_2 is the Hamming Weight (HW) of $X_1 \oplus X_2$, where HW is the number of binary '1's and $\oplus$ represents XOR:

$$HD(X_1, X_2) = HW(X_1 \oplus X_2). \tag{9}$$

The power dependence on processed data of QCA circuits is studied here with an inverter and a majority gate. The power consumption of a QCA inverter, as shown in Fig. 8, is provided in Table 2 under different tunneling energy levels. The power consumption does not vary as the temperature range is small, from 1.0 K to 8.0 K. It can be seen from the table that in the case of transitions from $0 \rightarrow 0$ and $1 \rightarrow 1$, little static power is dissipated, while for transitions from $0 \rightarrow 1$ and $1 \rightarrow 0$ significantly more power is consumed. This is similar to that of



(a)



(b)

Fig. 5. Power consumption ($T = 2.0\,\mathrm{K}, T_c = 75\frac{\hbar}{E_k}$) in a QCA cell under non-adiabatic switching during driver polarisation: (a) keep unchanged as 1; (b) changing from -1 to 1 (PD: Power Dissipation, CE: Clock Energy).

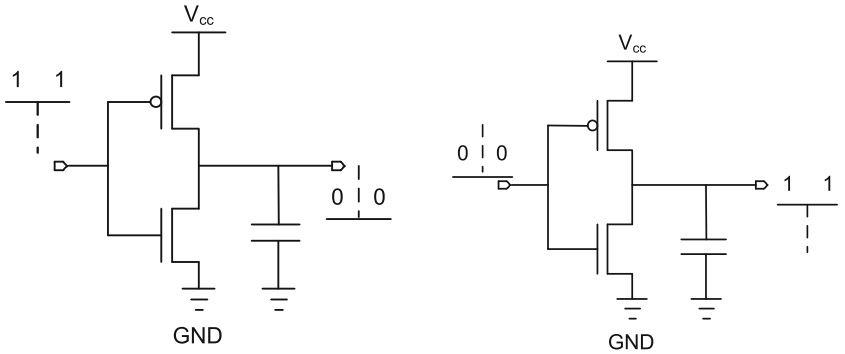


Fig. 6. Static power consumption in a CMOS inverter when input keeps unchanged as: (a) logic '1'; and (b) logic '0'.

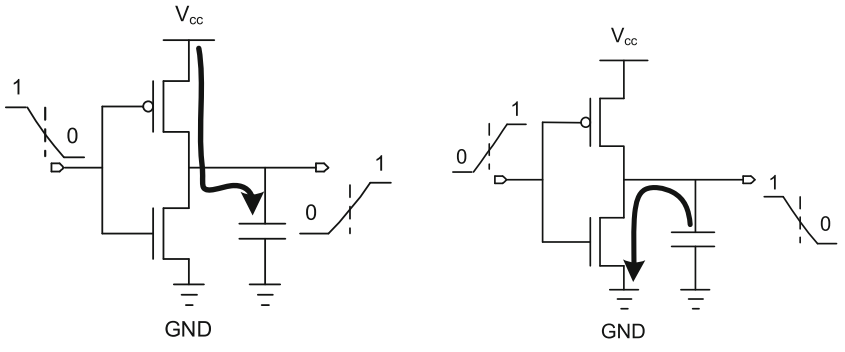


Fig. 7. Dynamic power consumption in a CMOS inverter when input is changing: (a) from '1' to '0'; and (b) from '0' to '1'.

Table 1. Power consumption of an CMOS inverter with different input changes

Inputs	Hamming distance	Types of power consumption
$0 \rightarrow 0$	0	Static power
$1 \rightarrow 1$		
$1 \rightarrow 0$	1	Static power + dynamic power
$0 \rightarrow 1$		

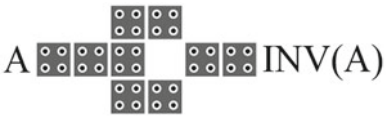


Fig. 8. The QCA layout of an inverter.

Table 2. Power consumption of an inverter with different tunnelling energy levels ($T = 2.0\,\mathrm{K}$)

Input switching	Hamming distance	Power dissipated			
		$\gamma = 0.25E_k$ (meV)	$\gamma = 0.5E_k$ (meV)	$\gamma = 0.75E_k$ (meV)	$\gamma = 1.0E_k$ (meV)
$0 \rightarrow 0$	0	0.8	2.7	5.2	8.0
$1 \rightarrow 1$		0.8	2.7	5.0	8.0
$0 \rightarrow 1$	1	28.4	28.6	29.3	30.2
$1 \rightarrow 0$		28.4	28.6	29.3	30.2

a CMOS inverter. Therefore, the power consumption of a QCA inverter is also dependent on the HD of its inputs.

The other basic gate in QCA is the majority gate, the layout of which is shown in Fig. 9. Its power consumption for input switching from 0 ("000") to 7 ("111") is provided in Table 3 under different tunneling energy levels. As with the inverters, the power consumption does not vary as the temperature range is small. For $HD = 1$, only one input of the majority gate changes its polarisation. Thus, only a little power is consumed. In the case of $HD > 1$, more cell inputs change and the output polarisation changes, which increases the power consumption. It is apparent that switching with a higher HD results in higher power consumption. Figure 10 clearly illustrates the strong dependence between the HD and the power consumption under various tunneling energy levels. Since the functional

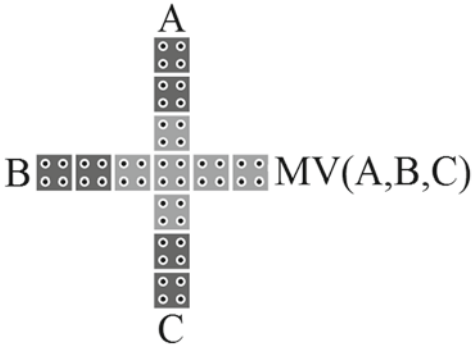


Fig. 9. The QCA layout of a majority gate.

Table 3. Power consumption of a majority gate with different tunnelling energy levels ($T = 2.0\,\mathrm{K}$)

Switching (A, B, C)	Hamming distance	Power dissipated			
		$\gamma = 0.25E_k$ (meV)	$\gamma = 0.5E_k$ (meV)	$\gamma = 0.75E_k$ (meV)	$\gamma = 1.0E_k$ (meV)
000 → 000 (0)	0	0.8	2.8	5.3	8.1
000 → 001 (1)	1	2.3	4.2	6.8	9.8
000 → 010 (2)		2.3	4.2	6.8	9.8
000 → 100 (4)		2.3	4.2	6.8	9.7
000 → 011 (3)	2	25.3	26.4	27.9	29.8
000 → 101 (5)		25.3	26.4	28.0	29.8
000 → 110 (6)		25.3	26.4	28.0	29.8
000 → 111 (7)	3	41.0	41.2	41.9	42.9

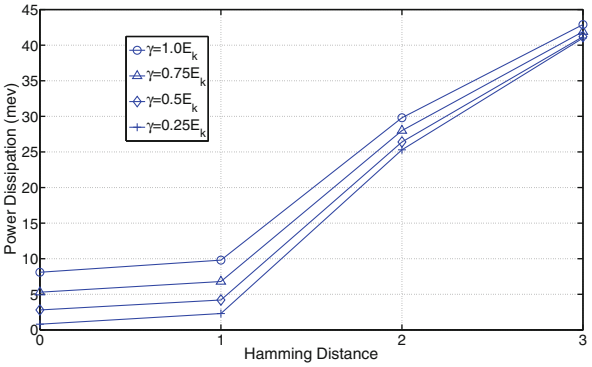


Fig. 10. Power consumption ($T = 2.0\,\mathrm{K}$) in majority gate versus Hamming distance with various tunneling energy γ levels.

temperature range of current semiconductor QCA devices is limited, there is little power consumption difference for different temperatures.

As power analysis attacks exploit the fact that the power consumption of a cryptographic circuit is correlated with the processed data, such as the HD of inputs [32], from the analysis outlined above, QCA cryptographic circuits may be vulnerable to attack. However, since QCA architectures are very different from equivalent CMOS architectures, the vulnerability is further studied by performing a power analysis attack on a sub-module of the Serpent cryptographic algorithm designed in QCA.

3 Design of the QCA Cryptographic Circuit: Serpent Sub-Module

Cryptographic block ciphers including AES and DES are widely used to encrypt confidential information. However, they are vulnerable to power analysis attack

which typically targets the substitution boxes (S-boxes) [26,37–39]. S-boxes are at the heart of block ciphers and are the only non-linear parts of the ciphers. They are typically used to hide the relationship between the key and the ciphertext following Shannon's property of confusion to resist mathematical cryptanalysis. In general, an S-box takes an n-bit input and transforms it into an m-bit output, namely an $n \times m$ S-box.

In this work, a sub-module of the Serpent cipher [40] including its S-box, is designed in QCA. The Serpent cipher is chosen because the presently available QCA design and simulation tools are currently limited in their support for large circuit designs. Serpent is a well-designed modern block cipher that offers a large security margin. It was a finalist in the AES contest [41]. The 32 rounds in the Serpent cipher provide an even higher security margin than the Rijndael cipher which is the current AES standard. It uses eight 4×4 S-boxes which are strongly secure against all known mathematical attacks [40]. However, similar to AES and DES, Serpent can be attacked using a power analysis attack. The essential idea of a power analysis attack is to attack a small part of the whole key to reduce the computational complexity. More specifically, it can be used to target a key-dependent sub-module of a cipher (usually the S-boxes for block ciphers) to uncover a small subkey [32]. The whole key can then be revealed by attacking each of the subkeys. A successful power analysis attack of the 4×4 Serpent S-box has been demonstrated on 65 nm CMOS cryptographic circuits [42]. Therefore, Serpent is an appropriate example for demonstrating a power analysis attack on QCA cryptographic circuits.

3.1 Sub-Module of Serpent Cipher

Serpent [40] is a 32-round substitution-permutation network that operates on four 32-bit words. It encrypts a 128-bit plaintext to a 128-bit ciphertext in 32 rounds under the control of 33 128-bit subkeys $\hat{K}_0, \ldots, \hat{K}_{32}$. An initial permutation is applied to a plaintext before the first round. A set of eight 4×4 S-boxes is used four times. In each round, only a single replicated S-box is used with a subkey. The last round is slightly different from the others and uses two subkeys, $\hat{K}_{31}$ and $\hat{K}_{32}$.

The sub-module implemented in this work is expressed as follows:

$$V = \hat{S}_0 \left(\hat{B}_0 \oplus \hat{K}_0 \right), \tag{10}$$

where
V is an intermediate value,
$\hat{S}_0$ is the first Serpent S-box, namely the S_0-box,
$\hat{B}_0$ is the permutated plaintext used as the input to the first round,
$\hat{K}_0$ is the first subkey.

Thus the function of the sub-module as shown in Fig. 11 is to produce the first four bits of an intermediate vector by taking bits $0, 1, 2, 3$ of $\hat{B}_0 \oplus \hat{K}_0$ as the input to the S_0-box, which is chosen without loss of generality. As a replicated S-box is used in each round, the next S_0-box takes bits $4, 5, 6, 7$ of $\hat{B}_0 \oplus \hat{K}_0$ and

returns the next four bits of the intermediate vector, and so on. Every four bits of the subkey related to the S-box will be attacked. In this way, the subkey can be revealed by the power analysis attack. If each subkey is revealed, the whole secret key of the Serpent cipher can be obtained.

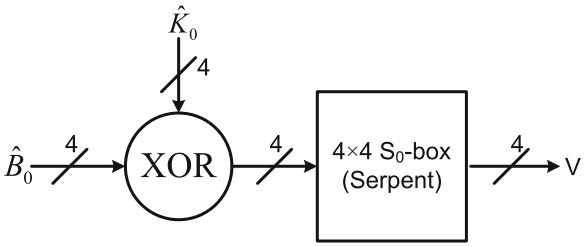


Fig. 11. The sub-module of Serpent cipher.

3.2 QCA Implementation of the Serpent Sub-Module

The sub-module of the Serpent cipher contains two parts: one is the 4-bit XOR operation and the other is the S_0-box. An XOR gate is designed in QCA and its schematic and layout are shown in Fig. 12.

The S_0-box is the first Serpent S-box and comprises four inputs and four outputs. Its truth table is shown in Table 4. There are two methods to implement the S-box. One is to emulate a lookup table (LUT); however, memory architectures in QCA require many cells and introduce high latency. The second technique is to design a logic-based S-box. This method will occupy less area and introduce low latency. In this work, the logic-based QCA S_0-box is designed by using majority logic reduction.

Let X denote the input and Y denote the output of the S_0-box. First, by using the Karnaugh map, the minimized logic expressions for $Y = y_4y_3y_2y_1$ can be obtained with AND and OR gates as follows:

$$\begin{aligned} y_1 &= \bar{x}_4\bar{x}_3\bar{x}_1 + \bar{x}_4x_2 + x_4x_3\bar{x}_1 + x_4\bar{x}_3\bar{x}_2x_1 \\ &= \bar{x}_4\left(\bar{x}_3\bar{x}_1 + x_2\right) + x_4\left(x_3\bar{x}_1 + \bar{x}_3x_1\bar{x}_2\right), \end{aligned} \tag{11}$$

$$\begin{aligned} y_2 &= \bar{x}_4\bar{x}_3\bar{x}_1 + \bar{x}_4x_3x_1 + x_4\bar{x}_3x_2x_1 + \bar{x}_2\bar{x}_1 \\ &= \bar{x}_4\left(\bar{x}_3\bar{x}_1 + x_3x_1\right) + x_4\left(\bar{x}_3x_2x_1\right) + \bar{x}_2\bar{x}_1, \end{aligned} \tag{12}$$

$$\begin{aligned} y_3 &= \bar{x}_4x_2\bar{x}_1 + \bar{x}_4x_3\bar{x}_2x_1 + x_4\bar{x}_3\bar{x}_1 + x_4\bar{x}_2\bar{x}_1 + x_4\bar{x}_3\bar{x}_2 + x_4x_3x_2x_1 \\ &= \bar{x}_4\left(x_2\bar{x}_1 + \bar{x}_2x_1x_3\right) + x_4\left[\bar{x}_3\bar{x}_1 + \left(\bar{x}_1\bar{x}_2 + \bar{x}_2\bar{x}_3 + x_1x_2x_3\right)\right], \end{aligned} \tag{13}$$

$$\begin{aligned} y_4 &= \bar{x}_4x_3\bar{x}_2\bar{x}_1 + \bar{x}_4\bar{x}_3x_2\bar{x}_1 + x_4x_3x_2 + x_4\bar{x}_3\bar{x}_2 + \bar{x}_3\bar{x}_2x_1 + x_3x_2x_1 \\ &= \bar{x}_4\bar{x}_1\left(x_3\bar{x}_2 + \bar{x}_3x_2\right) + \left(x_4 + x_1\right)\left(x_3x_2 + \bar{x}_3\bar{x}_2\right). \end{aligned} \tag{14}$$

Table 4. Truth table of S_0-box in form of Karnaugh map

x_4, x_3 \ x_2, x_1	00	01	11	10
00	3 (0011)	8 (1000)	1 (0001)	15 (1111)
01	10 (1010)	6 (0110)	11 (1011)	5 (0101)
11	7 (0111)	0 (0000)	12 (1100)	9 (1001)
10	14 (1110)	13 (1101)	2 (0010)	4 (0100)

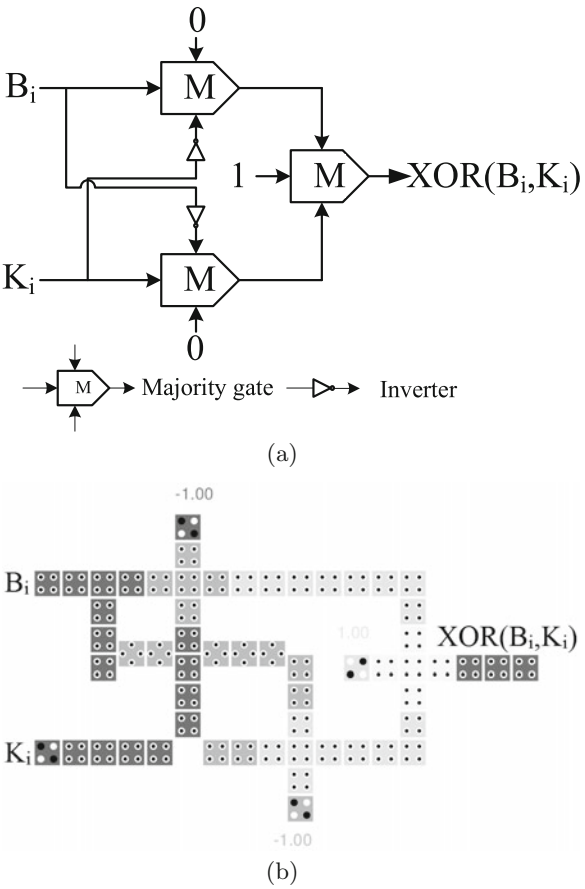


Fig. 12. Design of a QCA XOR gate: (a) schematic; (b) QCA layout.

By using the majority logic reduction method [43], the above expressions can be further minimized by the following logic reduction expressions:

$$\begin{aligned} y_1 :& x_3\bar{x}_1 + \bar{x}_3 x_1 \bar{x}_2 \\ &= M(M(x_3, 0, \bar{x}_1), \bar{M}(x_3, 1, \bar{x}_1), M(x_3, 1, \bar{x}_2)). \qquad (15) \\ y_2 :& \bar{x}_4 \left(\bar{x}_3\bar{x}_1 + x_3x_1\right) + x_4 \left(\bar{x}_3 x_2 x_1\right) \\ &= M(M((\bar{x}_3\bar{x}_1 + x_3x_1), 0, \bar{x}_4), 1, M(x_4, 0, (\bar{x}_3 x_2 x_1))). \qquad (16) \\ y_3 :& x_2\bar{x}_1 + \bar{x}_2 x_1 x_3 \\ &= M(M(x_2, 0, \bar{x}_1), \bar{M}(x_2, 1, \bar{x}_1), M(x_2, 1, x_3)), \qquad (17) \\ & (\bar{x}_1\bar{x}_2 + \bar{x}_2\bar{x}_3 + x_1x_2x_3) \\ &= M(M(x_1, 1, \bar{x}_2), \bar{M}(\bar{x}_2, 1, \bar{x}_3), M(\bar{x}_1, \bar{x}_2, \bar{x}_3)). \qquad (18) \end{aligned}$$

The majority gate based schematic of the Serpent sub-module is shown in Fig. 13. By integrating the XOR gates and S_0-box, the QCA design of the Serpent sub-module has been implemented in QCADesigner with appropriately assigned clocking zones as shown in Fig. 14 (different shades indicate the different clocking zones). The design has been checked with QCAPro, which verified that the outputs are correct for all given inputs.

A QCA design of the Serpent S_0-box based on majority logic was previously proposed [44]. In that design, each term of the logic expressions is implemented by AND and OR gates which are directly mapped from the CMOS design. However, since majority based logic reduction was not applied, the design incurs a much higher cell count and delay [45]. A comparison between the previous work and this design is presented in Table 5. In terms of area, the QCA cell size used in both designs is 18 nm with a centre-to-centre cell distance of 20 nm. It can be seen from the comparison table that a much more efficient design of the Serpent S_0-box is achieved in this work, with a reduction of 45 %, 42 % and 59 % in terms of the number of cells, area and latency, respectively.

Table 5. Comparison of two designs of the Serpent S_0-box

Compared items	Previous design [44]	This design	Improvement (%)
Complexity	3965 cells	2186 cells	45
Area	5.75 μm^2	3.35 μm^2	42
Latency	8 cycles	3.25 cycles	59

4 Power Analysis Attack of QCA Circuits

Differential Power Analysis (DPA) [25] is the most effective power analysis attack. DPA attacks do not require a detailed knowledge of the target device or the circuit architecture. The adversary does not need to know when a particular operation is actually computed by the cryptographic circuit. Power data

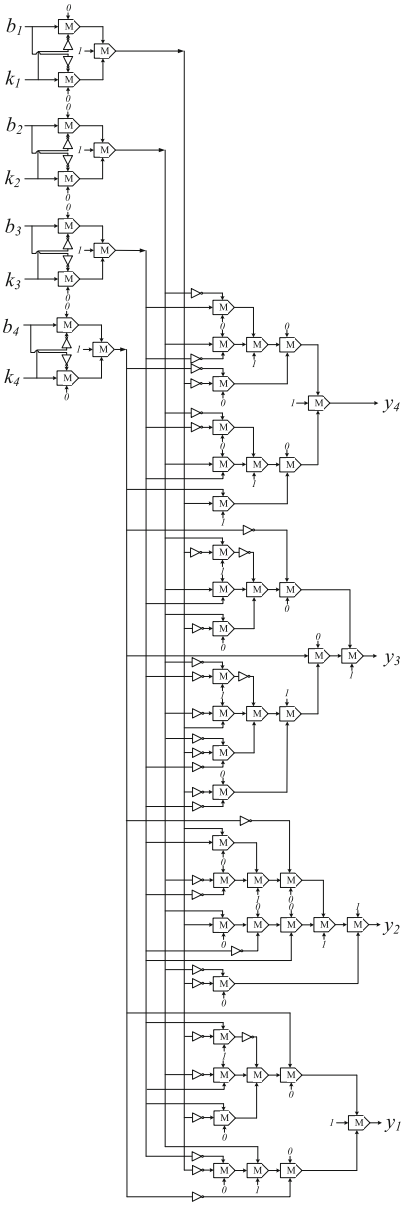


Fig. 13. Schematic of Serpent sub-module based on majority gates and inverters.

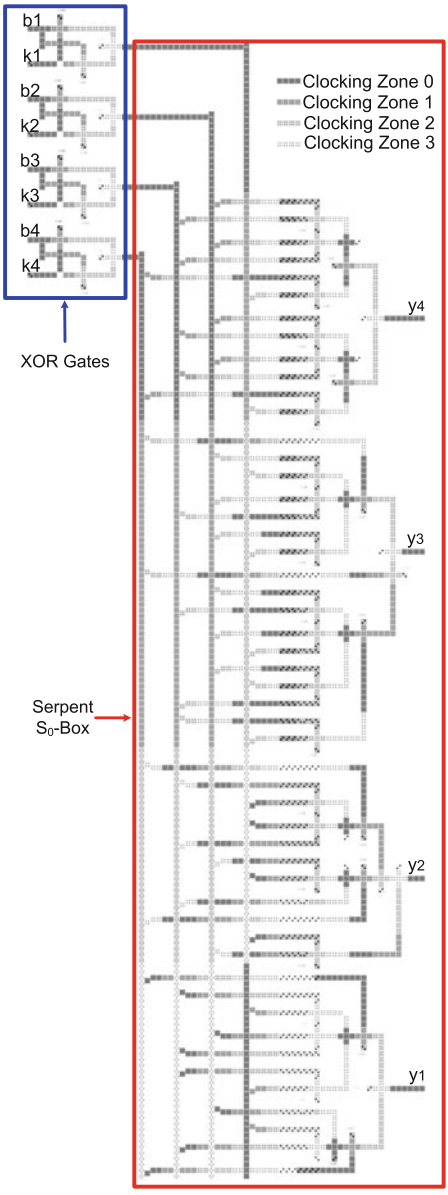


Fig. 14. QCA layout of the Serpent sub-module

that have been measured while a circuit encrypts or decrypts data can be statistically analysed to reveal the secret key. A large number of power measurements are used to analyse the power consumption at a fixed moment in time. For a block cipher under DPA, the S-boxes are the main target and they have been shown to be vulnerable to this kind of attack in CMOS technology. In this section, the first DPA attack results of a QCA cryptographic circuit are presented with a proposed DPA attack procedure for QCA circuits.

4.1 DPA Attack Procedure for QCA Circuits

As discussed in Sect. 2.3, the power consumption in basic QCA circuits depends on the Hamming Distance (HD) of the input switching. Therefore, the power data of QCA circuits could also be used to uncover the secret key. In a DPA attack, the power consumption of a circuit is correlated with a hypothetical power model that targets a key-dependent cryptographic operation. By comparing the hypothetical power values with the real power consumption of cryptogrpahic circuits by statistical tests, the key can be revealed. As the power consumed at a fixed moment of time is used in DPA attacks, the total power consumed by a QCA circuit during a full clock cycle is considered and simulated in this work.

A DPA procedure for QCA circuits is proposed, which has been adapted from the process used for CMOS-based devices. The proposed DPA procedure consists of the following five steps:

Step 1: Choose a target key-dependent intermediate result which is generated by the cryptographic circuit. This intermediate signal should be a function of a known non-constant data value (the plaintext or the ciphertext) and a small part of the key. A sub-module including S-boxes is usually chosen in a block cipher. In this work, the intermediate result under the attack is a sub-module of the Serpent cipher with its S_0-box:

$$Intermediate\ Result = \hat{S}_0\left(\hat{B}_0 \oplus \hat{K}_0\right). \tag{19}$$

Step 2: Measure the power consumption while the cryptographic circuit is encrypting the plaintext or decrypting the ciphertext. To perform this step in this work, $n+1$ random inputs, $\boldsymbol{B}$, are applied to the QCA Serpent sub-module:

$$\boldsymbol{B} = (b_0, b_1, \ldots, b_n)', \tag{20}$$

where b_i denotes the data value in the i^{th} encryption or decryption process. For each switching between two inputs, the power consumption, $\boldsymbol{P}$, is measured by QCAPro and stored as:

$$\boldsymbol{P} = (p_0, p_1, \ldots, p_{n-1})'. \tag{21}$$

Step 3: Calculate a hypothetical intermediate value for every possible choice of the key vector $\boldsymbol{K}$. In this work, all possible keys in the key vector are represented as $\boldsymbol{K} = (0, 1, 2, \ldots, 15)$. Each possible key element in the key vector is referred to as a **key guess**. An attacker can calculate hypothetical values

for the intermediate result chosen in **Step 1** for all random inputs and for all key guesses. The calculation results in a matrix $\boldsymbol{V}_{(n+1)\times(16)}$ with each element calculated as:

$$v_{i,j} = \hat{S}_0\left(b_i \oplus k_j\right),\ i = 0, 1, \ldots, n;\ j = 0, 1, \ldots, 15. \tag{22}$$

The j^{th} column of $\boldsymbol{V}$ includes the intermediate results calculated with the key guess k_j. As $\boldsymbol{K}$ contains all possible keys, the real key used in the cryptographic circuit is one element of the key vector $\boldsymbol{K}$. The aim of DPA is to find out which column of the matrix $\boldsymbol{V}$ has actually been processed during the encryption or decryption. Therefore, one column of $\boldsymbol{V}$ will be most highly correlated with the correct key guess (i.e., the real key).

Step 4: Map the hypothetical intermediate values in $\boldsymbol{V}$ to a matrix $\boldsymbol{H}_{(n)\times(16)}$ of hypothetical power consumption values. The elements in the hypothetical power consumption matrix $\boldsymbol{H}_{(n)\times(16)}$ are usually the HD or HW of the hypothetical intermediate values in matrix $\boldsymbol{V}$. The HD is used in this work, as the total power of QCA circuits heavily depends on the HD between two consecutive inputs, which has been clearly shown in Sect. 2.3. The HD model requires both previous and current values of the target intermediate result. Therefore, the calculation of the hypothetical power from hypothetical intermediate values is as follows:

$$h_{i,j} = HW\left(v_{i,j} \oplus v_{i+1,j}\right),\ i = 0, 1, \ldots, n-1;\ j = 0, 1, \ldots, 15. \tag{23}$$

A larger HD usually leads to higher power consumption.

Step 5: The hypothetical power values are compared with the measured power data. The attacker compares the hypothetical power values of each key guess with the measured power data at every position by comparing each column $\boldsymbol{h}_j$ of matrix $\boldsymbol{H}$ with the measured power $\boldsymbol{P}$. Pearson's correlation function [46] is applied to calculate the correlation coefficient between the hypothesis and the simulated power data. The comparison result is a vector $\boldsymbol{R}$ of correlation coefficients with each element calculated as follows:

$$r_j = \frac{\sum_{i=0}^{n-1}(h_{i,j} - \bar{h_j}) \times (p_i - \bar{p})}{\sqrt{\sum_{i=0}^{n-1}(h_{i,j} - \bar{h_j})^2 \times \sum_{i=0}^{n-1}(p_i - \bar{p})^2}},\ j = 0, 1, \ldots, 15, \tag{24}$$

where $\bar{h_j}$ and $\bar{p}$ denote the mean values of column $\boldsymbol{h}_j$ and the power vector $\boldsymbol{P}$, respectively. In a successful attack, the highest value of correlation coefficient corresponds to the correct key guess.

A diagram of the DPA procedure for QCA circuits is shown in Fig. 15. For more information on differential power analysis, please refer to [32].

4.2 DPA Attack of the Serpent Sub-Module

The first power analysis attack results of a QCA cryptographic circuit using the upper bound power model are presented here. The Serpent sub-module inputs

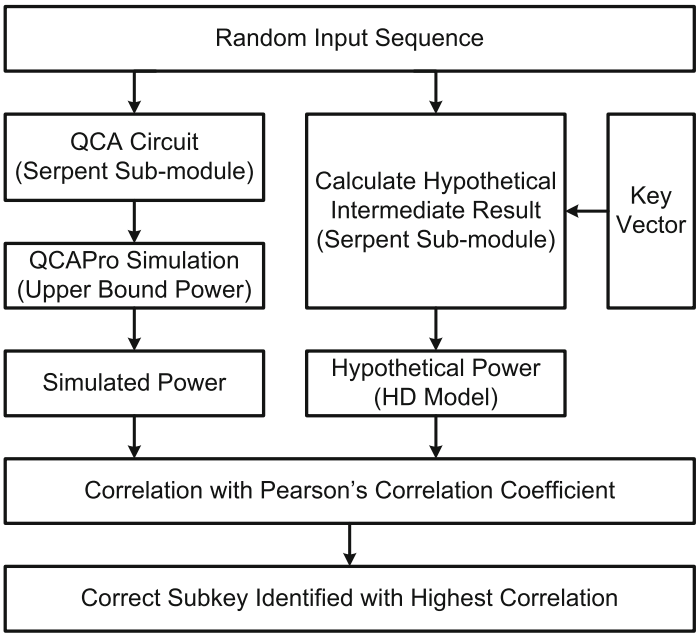


Fig. 15. A outline of the DPA procedure for QCA circuits.

and a subkey are XORed before being passed to the S-box transformation. All possible combinations of S_0-box inputs are simulated by QCAPro and the corresponding power consumption values are provided in Table 6, where the left vertical column of the table represents the current input and the top horizontal row represents the next input.

Following the DPA attack steps, the Serpent sub-module is attacked and two correlation results are shown in Figs. 16 and 17. It can be seen that the correct key guess in each case is clearly distinguishable from wrong key guesses after 1000 inputs. Since one subkey can be revealed, the whole Serpent key set can also be revealed using this power analysis attack.

To illustrate that all of the possible subkeys from "0000" to "1111" will result in a distinguishable correlation coefficient, they were applied to the Serpent sub-module and DPA attacks were carried out. The results are shown in Fig. 18. All subkey correlation coefficients vary between 0.16 and 0.25 and the number of power points required to reveal the subkey from the incorrect key guesses varies between 200 and 2000. Therefore, with just 2000 power points all of the Serpent subkeys can be revealed. In Fig. 18, for each subkey applied to the sub-module, the correlation coefficients of all key guesses are shown in the vertical direction, which are chosen when the number of power points is 2000. For example, the correlation coefficients for all key guesses when the correct subkey is 4 (SK = 4) or 13 (SK = 13) in Fig. 18 is obtained by a cross section of Figs. 16 or 17 taken at 2000 power points. For every case, each correct key guess may have

Table 6. Power consumption of S_0-box for all possible input switching. ($T = 2.0\,\mathrm{K}, \gamma = 1.0E_k$, unit of power consumption: eV)

Input	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	3.55	6.15	4.93	6.02	5.44	7.31	6.47	6.06	4.92	7.00	6.16	7.29	6.50	8.36	7.52
2	6.15	3.55	4.97	6.83	7.30	5.47	6.34	6.68	7.54	5.26	6.40	14.40	13.79	12.09	11.45
3	13.79	12.09	11.45	7.25	6.49	6.13	5.23	7.34	6.27	6.03	5.01	14.08	13.58	14.19	13.73
4	13.58	14.19	13.73	3.55	4.90	5.20	5.95	6.83	7.42	7.93	8.44	11.62	11.48	13.71	13.54
5	11.48	13.71	13.54	4.90	3.54	5.86	4.91	7.77	6.84	8.54	7.63	13.19	12.36	12.63	12.33
6	12.36	12.63	12.33	5.20	5.88	3.55	4.82	8.37	8.69	6.95	7.61	13.95	13.65	11.48	11.37
7	13.65	11.48	11.37	5.94	4.92	4.82	3.55	8.74	7.81	7.59	6.61	12.37	12.29	12.71	12.37
8	12.29	12.71	12.37	6.84	7.76	8.34	8.76	3.55	4.87	5.04	5.93	11.93	11.29	14.38	13.56
9	11.29	14.38	13.56	7.40	6.84	8.63	7.80	4.88	3.55	5.82	4.97	12.53	12.32	12.26	12.38
10	12.32	12.26	12.38	7.94	8.60	6.96	7.64	5.08	5.82	3.56	4.72	14.14	13.29	11.83	11.15
11	13.29	11.83	11.15	8.46	7.66	7.61	6.61	5.97	4.98	4.71	3.55	14.19	13.52	14.37	13.64
12	13.52	14.37	13.64	5.03	6.20	6.62	7.19	5.55	5.91	6.66	7.12	11.47	11.23	13.76	13.30
13	11.23	13.76	13.30	6.02	5.04	7.14	6.20	6.37	5.31	7.05	6.19	13.14	11.93	12.77	12.04
14	11.93	12.77	12.04	6.46	7.17	5.18	6.05	6.92	7.01	5.47	6.07	13.82	13.17	11.38	11.10
15	13.17	11.38	11.10	7.06	6.21	6.10	5.05	7.29	6.33	6.10	5.18	5.77	4.71	4.60	3.55

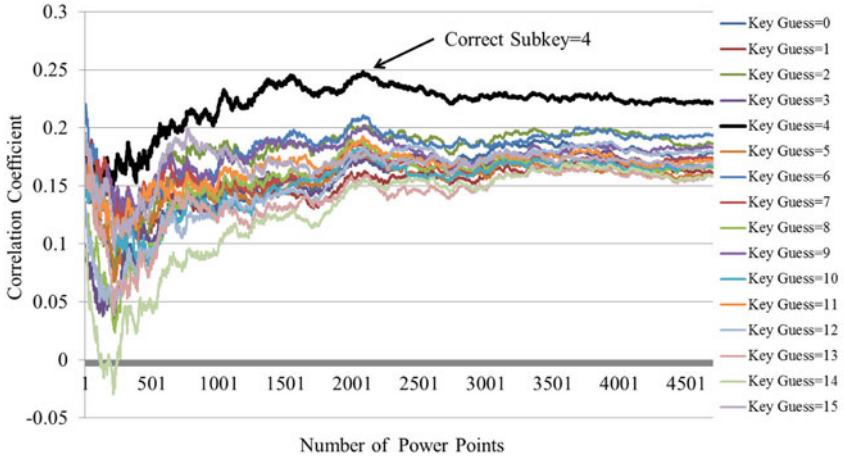


Fig. 16. Correlation results for DPA attack of Serpent sub-module when the real subkey is 4 ($T = 2.0\,\mathrm{K}, \gamma = 1.0\ E_k$).

a different correlation coefficient value, but it will be significantly higher than the correlation coefficients of the wrong key guesses.

An observation is that all of the correlation coefficients are much lower than 1. The reason is probably that the power data simulated by QCAPro is the total power consumed by the sub-module over one clock cycle. As a result, some power information that is not related to the target intermediate value is also included

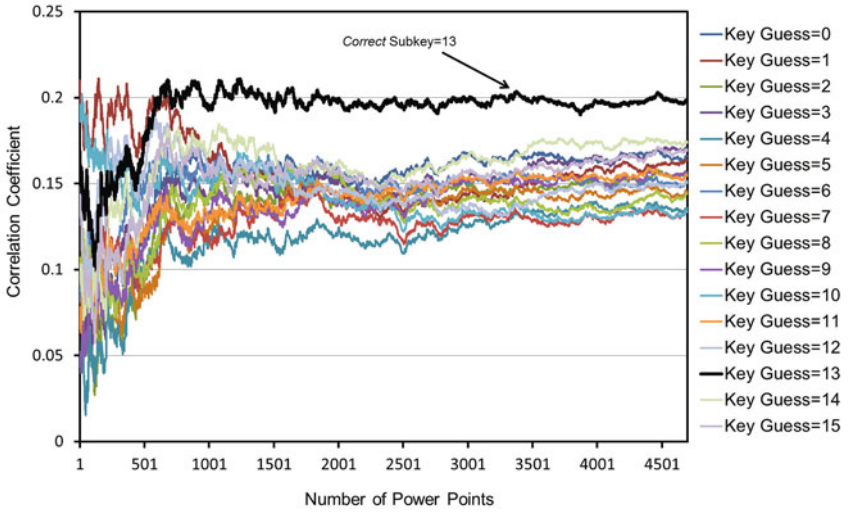


Fig. 17. Correlation results for DPA attack of Serpent sub-module when the real subkey is 13 ($T = 2.0\,\mathrm{K}, \gamma = 1.0\ E_k$).

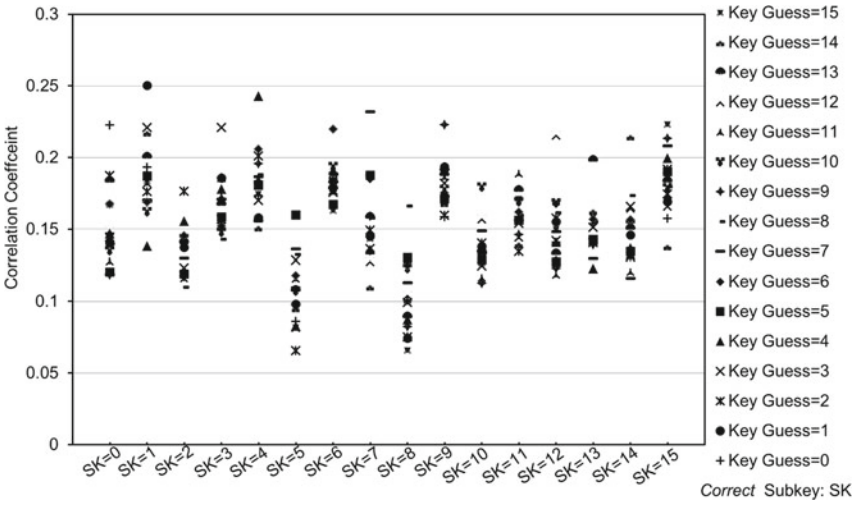


Fig. 18. Correlation of *Correct* Subkey (SK) with all possible key guesses applied to the Serpent sub-module ($T = 2.0\,\mathrm{K}, \gamma = 1.0\ E_k$). The results in each column can be visualized as a cross section of results equivalent to that shown in Figs. 16 or 17 for each *correct* subkey at 2000 power points.

in the total power. The vulnerability of the circuit is due to the close relationship between the processed key-related information and the corresponding power dissipation. Generally, if more cells or gates are involved in this processing, it is easier for an attacker to perform a power analysis attack.

5 Discussion of Practical QCA Devices

The upper bound power model was derived for semiconductor QCAs. However, it can also be extended for molecular and magnetic QCAs by using the appropriate Hamiltonian in the derivation [30]. Therefore, the power analysis attack results obtained in this chapter can also be applied to other QCA implementations under best case scenario. Although the power consumption in QCA is very low, for power analysis attacks, it is the power difference between transitions from '0' → '0' (or '1' → '1') and '0' → '1' (or '1' → '0') that is important since it indicates a power dependence on the processed data. It has been shown that cryptographic algorithms implemented in QCA with typical four-phase clocking could be vulnerable to power analysis attack. However, this result is for a best case scenario for attackers which assumes non-adiabatic switching.

In a more typical scenario, a smoother clock would be used, which would reduce the power consumption of transitions from 0 → 1 (or 1 → 0) significantly [30]. As shown in Fig. 4, the real power consumption will be lower than the upper bound. The difference between the dynamic power and static power is less and the power dependence on the Hamming distance of the inputs is also reduced. This will require more power data to reveal the key, which would enhance the security. Also, no noise is considered in this work. In a practical situation, noise will be present when measuring the power consumption, which will affect the power analysis attack. Therefore, in practice using a smoother clock, the security of QCA cryptographic circuits would be greatly enhanced. Also, even if the power dependence on processed data is still measurable under a smoother clock, the measurement of power consumption with a magnitude of eV is very difficult and expensive. Power analysis attacks were proposed as a cheap attack technique using readily accessible equipment. This may not be the case for future QCA devices.

6 The Worst Case Scenario for Attackers-Bennett Clocking

QCA circuits using typical quasi-adiabatic clocking perform irreversible computing. Landauer has shown that any logically irreversible operation must dissipate at least $K_B T ln(2)$ per bit, independent of the operation speed [47]. However, if a copy of the bit to be erased is reserved, the operation can dissipate an arbitrarily small amount of energy [48]. Furthermore, Bennett extended Landauer's theory by showing that any computation could be implemented as a logically reversible operation [49]. In this research, the logically irreversible clocking is referred to as Landauer clocking which is the typical quasi-adiabatic four phase clocking and the logically reversible clocking is referred to as Bennett clocking. Although the reversible computing theory has been proposed for almost four decades, no concrete circuits exist as they are infeasible to implement in CMOS due to the complexity involved in their design. However, QCA technology provides a practical platform for the realization of reversible computing without the requirement

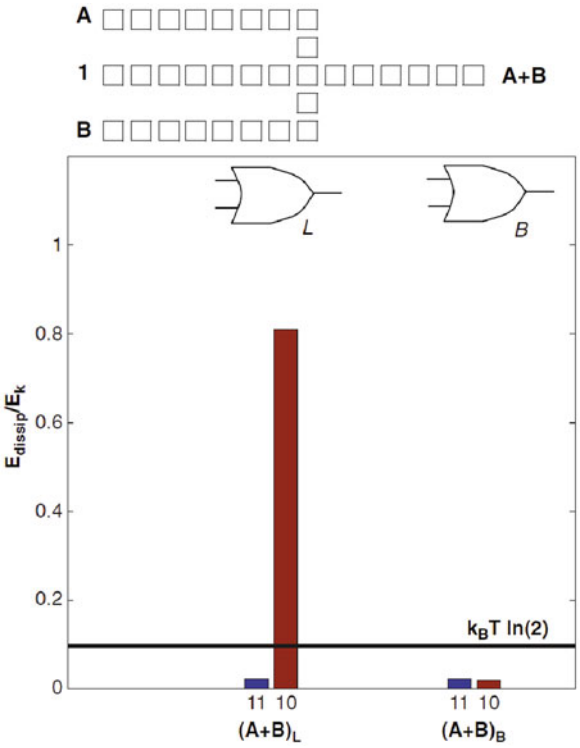


Fig. 19. Calculated power dissipation of QCA OR gate for Landauer-clocked (L) and Bennett-clocked (B) OR gates [31].

for additional circuit complexity [31,50]. Here the vulnerability of QCA circuits under Bennett clocking [49] to power analysis attack is considered as the worst case scenario for attackers.

The wave of Bennett clocking used in this study is described as follows [31]:

$$E_c(x,t) = E_C^0 min\left[\left(1 - \frac{x}{\lambda_c}\right) + sin\left(\frac{t}{T_c}\right), 1\right], \tag{25}$$

where x is the position of the bit information, t is the time, λ_c is the spatial width of the Bennett-clocked region and T_c is the temporal clocking period. The lower bound power is modelled by describing the relevant physics of QCA switching in a thermal environment, which is illustrated in detail by Lent *et al.* [31]. To implement QCA circuits with Bennett clocking as shown in Eq. (25), only the timing of the clocking signals needs to be altered, in order to keep the bit information in place using the clock until a computational block is finished; then the information is erased during the reverse order of computation.

It has already been demonstrated that a QCA OR gate using Bennett clocking produces very low and similar power consumption values for inputs with

different Hamming distance [31], as shown in Fig. 19. From this figure, it is clear that the power dissipated in the Bennett-clocked QCA OR gate during input changes is extremely low, even lower than the bit erasure energy, i.e., $K_B T ln(2)$. Compared with the Landauer-clocked OR gate, the most important point is that there is almost no power difference for a Bennett-clocked OR gate between the input changes from '1' to '1' and '1' to '0'. Therefore, by using Bennett clocking, the power dependence of basic gates on the inputs is effectively removed making it impossible to perform a power analysis attack albeit at the cost of speed. As a result, cryptographic circuit design using Bennett clocking in QCA would act as a natural countermeasure to power analysis attacks.

7 Conclusion

An investigation into the resistance of QCA cryptographic circuits to power analysis attacks is presented under both the best case and the worst case scenarios from an attackers' point of view.

Based on the upper bound power model, the power dependence on Hamming distance in basic QCA gates including majority gates and inverters has been shown. An efficient design of the Serpent S_0-box is presented with a reduction of 45 %, 42 % and 59 % in terms of number of cells, area and latency, respectively over previous work. The power consumption of the Serpent sub-module is simulated using QCAPro based on the upper bound power model. A power analysis attack procedure for QCA is proposed to reveal the secret key by statistically comparing the power consumption and all hypothetical key guesses. The first power analysis attack of a QCA cryptographic circuit is presented under a best case scenario (for the attacker). A DPA attack of the Serpent sub-module is performed and the results show that all possible subkeys of Serpent can be revealed by power analysis attacks. Therefore, QCA cryptographic circuits under quasi-adiabatic switching could be vulnerable to power analysis attack. However, the security of QCA circuits can be improved greatly by applying a smoother clock.

In the worst case scenario for the attacker, QCA cryptographic circuits can be designed with reversible computing using Bennett clocking, which removes the power dependence on the basic gates, making power analysis attack impossible. Therefore, a Bennett-clocked QCA circuit design could be used to prevent power analysis attacks. It is believed that QCA could be a niche technology to implement security architectures resistant to power analysis attack in the future.

References

1. ITRS: International Technology Roadmap for Semiconductors (ITRS), website (2011). http://www.itrs.net/Links/2011ITRS/Home2011.htm
2. Lent, C.S., Tougaw, P.D., Porod, W., Bernstein, G.H.: Quantum cellular automata. Nanotechnology **4**(1), 49–57 (1993)
3. Lent, C.S., Tougaw, P.D.: A device architecture for computing with quantum dots. Proc. IEEE **85**, 541–557 (1997)

4. Walus, K., Jullien, G.A.: Design tools for an emerging soc technology: quantum-dot cellular automata. Proc. IEEE **94**(6), 1225–1244 (2006)
5. Smith, C., Gardelis, S., Rushforth, A., Crook, R., Cooper, J., Ritchie, D., Linfield, E., Jin, Y., Pepper, M.: Realization of quantum-dot cellular automata using semiconductor quantum dots. Superlattices Microstruct. **34**(3), 195–203 (2003)
6. Orlov, A.O., Amlani, I., Bernstein, G.H., Lent, C.S., Snider, G.L.: Realization of a functional cell for quantum-dot cellular automata. Science **277**(5328), 928–930 (1997)
7. Amlani, I., Orlov, A.O., Toth, G., Bernstein, G.H., Lent, C.S., Snider, G.L.: Digital logic gate using quantum-dot cellular automata. Science **284**(5412), 289–291 (1999)
8. Arima, V., Iurlo, M., Zoli, L., Kumar, S., Piacenza, M., Matino, F., Maruccio, G., Rinaldi, R., Paolucci, F., et al.: Toward quantum-dot cellular automata units: thiolated-carbazole linked bisferrocenes. Nanoscale **4**(3), 813–823 (2012)
9. Frost, S.E., Rodrigues, A.F., Janiszewski, A.W., Rausch, R.T., Kogge, P.M.: Memory in motion: a study of storage structures in QCA. In: Proceedings of the 1st Workshop on Non-Silicon Computing, vol. 2, pp. 30–37 (2002)
10. Vankamamidi, V., Ottavi, M., Lombardi, F.: A line-based parallel memory for QCA implementation. IEEE Trans. Nanotechnol. **4**, 690–698 (2005)
11. Walus, K., Mazur, M., Schulhof, G., Jullien, G.A.: Simple 4-bit processor based on quantum-dot cellular automata (QCA). In: Proceedings of the 16th IEEE International Conference on Application-Specific Systems, Architecture Processors, pp. 288–293 (2005)
12. Hanninen, I., Takala, J.: Pipelined array multiplier based on quantum-dot cellular automata. In: Proceedings of the 18th European Conference on Circuit Theory and Design, pp. 938–941 (2007)
13. Cho, H., Swartzlander Jr, E.E.: Adder and multiplier design in quantum-dot cellular automata. IEEE Trans. Comput. **58**, 721–727 (2009)
14. Swartzlander Jr, E.E., Cho, H., Kong, I., Kim, S.W.: Computer arithmetic implemented with QCA: a progress report. In: Conference Record of the 44th Asilomar Conference on Signals, Systems and Computers, pp. 1392–1398 (2010)
15. Lu, L., Liu, W., O'Neill, M., Swartzlander Jr, E.E.: QCA systolic matrix multiplier. In: Proceedins of the IEEE Annual Symposium on VLSI, pp. 149–154 (2010)
16. Niemier, M.T., Kogge, P.M.: Problems in designing with QCAs: layout = timing. Int. J. Circuit Theory Appl. **29**(1), 49–62 (2001)
17. Zhang, R., Walus, K., Wang, W., Jullien, G.A.: A method of majority logic reduction for quantum cellular automata. IEEE Trans. Nanotechnol. **3**, 443–450 (2004)
18. Srivastava, S., Bhanja, S.: Hierarchical probabilistic macromodeling for QCA circuits. IEEE Trans. Comput. **56**, 174–190 (2007)
19. Choi, M., Patitz, Z., Jin, B., Tao, F., Park, N., Choi, M.: Designing layout-timing independent quantum-dot cellular automata (QCA) circuits by global asynchrony. J. Syst. Architect. **53**, 551–567 (2007)
20. Liu, W., Lu, L., O'Neill, M., Swartzlander Jr, E.E., Woods, R.: Design of quantum-dot cellular automata circuits using cut-set retiming. IEEE Trans. Nanotechnol. **10**(5), 1150–1160 (2011)
21. Lu, L., Liu, W., O'Neill, M., Swartzlander Jr, E.E.: QCA systolic array design. IEEE Trans. Comput. **62**, 548–560 (2013)
22. Timler, J., Lent, C.S.: Power gain and dissipation in quantum-dot cellular automata. J. Appl. Phys. **91**(2), 823–830 (2002)

23. Kocher, P.C.: Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 104–113. Springer, Heidelberg (1996)
24. Kelsey, J., Schneier, B., Wagner, D., Hall, C.: Side channel cryptanalysis of product ciphers. In: Quisquater, J.-J., Deswarte, Y., Meadows, C., Gollmann, D. (eds.) ESORICS 1998. LNCS, vol. 1485, pp. 97–110. Springer, Heidelberg (1998)
25. Kocher, P., Jaffe, J., Jun, B.: Differential power analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 388–397. Springer, Heidelberg (1999)
26. Messerges, T.S., Dabbish, E.A., Sloan, R.H.: Examining smart-card security under the threat of power analysis attacks. IEEE Trans. Comput. **51**, 541–552 (2002)
27. Standaert, O.X., Peeters, E., Rouvroy, G., Quisquater, J.J.: An overview of power analysis attacks against field programmable gate arrays. Proc. IEEE **94**, 383–394 (2006)
28. Timler, J., Lent, C.S.: Maxwell's demon and quantum-dot cellular automata. J. Appl. Phys. **94**(2), 1050–1060 (2003)
29. Bond, L., Macucci, M.: Analysis of power dissipation in clocked quantum cellular automaton circuits. In: Proceeding of the 36th European Solid-State Device Research Conference, pp. 57–60 (2006)
30. Srivastava, S., Sarkar, S., Bhanja, S.: Estimation of upper bound of power dissipation in QCA circuits. IEEE Trans. Nanotechnol. **8**(1), 116–127 (2009)
31. Lent, C.S., Liu, M., Lu, Y.: Bennett clocking of quantum-dot cellular automata and the limits to binary logic scaling. Nanotechnology **17**, 4240–4251 (2006)
32. Mangard, S., Oswald, E., Popp, T.: Power Analysis Attacks: Revealing the Secrets of Smart Cards. Springer, New York (2007)
33. Srivastava, S., Asthana, A., Bhanja, S., Sarkar, S.: QCAPro - an error-power estimation tool for QCA circuit design. In: Proceedings of the IEEE International Symposium on Circuits and Systems, pp. 2377–2380 (2011)
34. Blair, E.P., Yost, E., Lent, C.S.: Power dissipation in clocking wires for clocked molecular quantum-dot cellular automata. J. Comput. Electron. **9**(1), 49–55 (2010)
35. Lent, C.S., Tougaw, P.D.: Lines of interacting quantum-dot cells: a binary wire. J. Appl. Phys. **74**(10), 6227–6233 (1993)
36. Walus, K., Dysart, T.J., Jullien, G.A., Budiman, R.A.: QCADesigner: A rapid design and simulation tool for quantum-dot cellular automata. IEEE Trans. Nanotechnol. **3**(1), 26–31 (2004)
37. Morioka, S., Satoh, A.: An optimized S-box circuit architecture for low power AES design. In: Kaliski Jr, B.S., Koç, Ç.K., Paar, C. (eds.) CHES 2002. LNCS, vol. 2523, pp. 172–186. Springer, Heidelberg (2003)
38. Prouff, E.: DPA attacks and S-boxes. In: Gilbert, H., Handschuh, H. (eds.) FSE 2005. LNCS, vol. 3557, pp. 424–441. Springer, Heidelberg (2005)
39. Boey, K.H., Hodgers, P., Lu, Y., O'Neill, M., Woods, R.: Security of AES S-box designs to power analysis. In: Proceedings of the 17th IEEE International Conference on Electronics, Circuits, and Systems, pp. 1232–1235 (2010)
40. Anderson, R., Biham, E., Knudsen, L.: Serpent: a proposal for the advanced encryption standard. NIST AES Proposal (1998)
41. Nechvatal, J., Barker, E., Bassham, L., Burr, W., Dworkin, M., Foti, J., Roback, E.: Report on the development of the advanced encryption standard (AES). J. Res. Nat. Inst. Stand. Technol. **106**(3), 511–576 (2001)
42. Alioto, M., Giancane, L., Scotti, G., Trifiletti, A.: Leakage power analysis attacks: a novel class of attacks to nanometer cryptographic circuits. IEEE Trans. Circuits Syst. I: Reg. Pap. **57**, 355–367 (2010)

43. Kong, K., Shang, Y., Lu, R.: An optimized majority logic synthesis methodology for quantum-dot cellular automata. IEEE Trans. Nanotechnol. **9**, 170–183 (2010)
44. Amiri, M., Mahdavi, M., Mirzakuchaki, S.: Logic-based QCA realization of a 4×4 S-box. In: Proceedings of International Conference on Computer Applications and Industrial Electronics, pp. 415–420 (2010)
45. Liu, W., Lu, L., O'Neill, M., Swartzlander Jr, E.E.: Design rules for quantum-dot cellular automata. In: Proceedings of the IEEE International Symposium on Circuits and Systems, pp. 2361–2364 (2011)
46. Rodgers, J.L., Nicewander, W.A.: Thirteen ways to look at the correlation coefficient. Am. Stat. **42**(1), 59–66 (1988)
47. Landauer, R.: Irreversibility and heat generation in the computing process. IBM J. Res. Dev. **5**(3), 183–191 (1961)
48. Keyes, R.W., Landauer, R.: Minimal energy dissipation in logic. IBM J. Res. Dev. **14**(2), 152–157 (1970)
49. Bennett, C.H.: Logical reversibility of computation. IBM J. Res. Dev. **17**(6), 525–532 (1973)
50. Ottavi, M., Pontarelli, S., DeBenedictis, E., Salsano, A., Frost-Murphy, S., Kogge, P., Lombardi, F.: Partially reversible pipelined QCA circuits: combining low power with high throughput. IEEE Trans. Nanotechnol. **10**, 1383–1393 (2011)

NanoMagnet Logic: An Architectural Level Overview

Marco Vacca[1], Mariagrazia Graziano[1(✉)], Juanchi Wang[1], Fabrizio Cairo[1], Giovanni Causapruno[1], Gianvito Urgese[2], Andrea Biroli[1], and Maurizio Zamboni[1]

[1] Dipartimento di elettronica e telecomunicazioni, Politecnico di Torino, Turin, Italy
mariagrazia.graziano@polito.it
[2] Dipartimento di Ingegneria informatica e dei sistemi, Politecnico di Torino, Turin, Italy

Abstract. In recent years Field-Coupled devices, like Quantum dot Cellular Automata, are gaining an ever increasing attention from the scientific community. The computational paradigm beyond this device topology is based on the interaction among neighbor cells to propagate information through circuits. Among the various implementations of this theoretical principle, NanoMagnet Logic (NML) is one of the most studied. The reason lies to some interesting features, like the possibility to combine memory and logic in the same device and the possible low power consumption. Since the working principle of Field-Coupled devices is completely different from CMOS technology, it is important to understand all the implications that this new computational paradigm has on complex circuit architectures.

In this chapter we deeply analyze the major issues encountered in the design of complex circuits using Field-Coupled devices. Problems are analyzed and techniques to solve them and to improve performance are presented. Finally, a realistic analysis of the applications best suited for this technology is presented. While the analysis is performed using NanoMagnet Logic as target, the results can be applied to all Field-Coupled devices. This chapter therefore supplies researchers and designers with the essential guidelines necessary to design complex circuits using NanoMagnet Logic and, more in general, Field-Coupled devices.

1 Introduction

With the conception of Quantum dot Cellular Automata (QCA) [1] technology, a completely new computational paradigm to process and propagate information was presented to the scientific community. Signals were no more represented by voltage or current levels, but by charge configurations. Circuits are made by many identical cells and information processing is driven by electrostatic interaction among neighbor cells [2]. Different practical implementations of this principle were proposed, the two most promising are Molecular QCA [3] and Magnetic QCA also called NanoMagnet Logic [4]. Molecular QCA [5–8] use complex molecules as basic cells. The main interest for this QCA implementation resides in

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 223–256, 2014.
DOI: 10.1007/978-3-662-43722-3_10, © Springer-Verlag Berlin Heidelberg 2014

the amazing clock speed (1 THz) that can be theoretically reached [9]. However, the experimental fabrication is extremely challenging with current technological processes. NanoMagnet Logic uses single domain nanomagnets (Fig. 1(A)) as basic cells [10]. Due to magnetic properties, like shape or magnetocrystalline anisotropy, magnets can have only two stable states that are used to represent logic values '0' and '1'. The main interest around this technology lies in its magnetic nature, that allows to potentially mix logic and memory in the same device. For the same reason circuits built with this technology have an high resistance to radiations and heat and a power consumption potentially lower than that of state-of-the-art CMOS circuits [11].

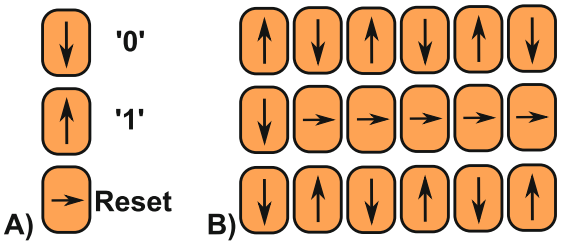


Fig. 1. (A) Single domain nanomagnets are used as basic cell in NML technology. Only two stable states are possible and are therefore used to represent digital values '0' and '1'. A third intermediate state is possible but it is unstable and can therefore be forced only by external means. (B) To switch magnets from one state to the other a RESET mechanism is required. Magnets are driven in the RESET state by an external mechanism, a magnetic field in the most classical implementation. When the magnetic field is removed, magnets realign following the input element.

Information propagation depends on the magnetostatic interaction among neighbor magnets. Nonetheless, the magnetic field generated by one magnet is not strong enough to switch a neighbor element from one stable state to the other. As a consequence a mechanism called clock must be used to help magnets switching, forcing them in an intermediate unstable state (RESET in Fig. 1(A)). An external mean (a magnetic field in the most classical implementation [12,13]) must be used to force magnets in the RESET state (Fig. 1(B)). When the magnetic field is removed magnets realign following the input element, thereby propagating the information through the circuit.

Signals within the circuit propagate with a Domino-like effect. Magnets switch one by one in a sort of chain reaction. Unfortunately the number of magnets that switch correctly in a chain is limited by the intrinsic instability of the RESET state. The RESET state is unstable, so external stimuli (like thermal noise [14]) can generate unwanted switches, therefore leading to errors during signals propagation. To solve this problem a multiphase clock system must be used [15–17]. Circuits are divided in small areas called clock zones. Every clock zone is made of a limited number (less than 5 in the ideal case [14]) of cascaded

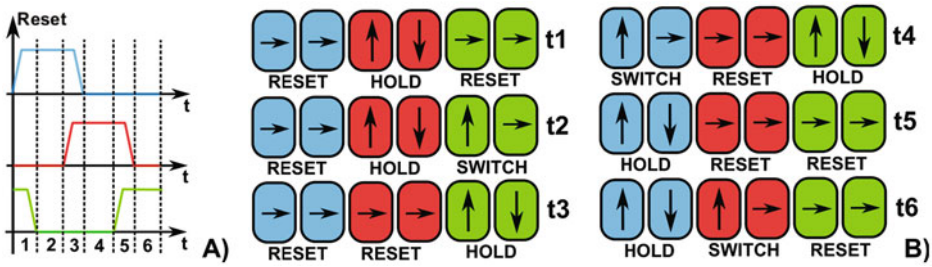


Fig. 2. 3-phase overlapped clock system. (A) Clock signal waveforms. 3 clock signals with a phase difference of 120°, partially overlapped, are used to assure a correct signals propagation. (B) Detailed signal propagation through a simple NML wire. 6 different time step can be identified thanks to this clock system. When magnets of a clock zone are switching (SWITCH) magnets on their left are in the HOLD state and act as an input, while magnets on their right are in the RESET state so they have no influence.

elements. A different clock signal is applied to every clock zone. Different clock schemes can be used. One of the most simple clock scheme uses three overlapped clock signals [18], with a phase difference of 120° (Fig. 2(A)). Figure 2(B) shows an example of NML wire and how its state changes when the clock signals are applied. Three different states can be identified: HOLD when no clock signal is applied, RESET when the clock signal is applied and SWITCH when the clock signal is slowly removed. When magnets of a clock zone are switching (SWITCH state), they see on their left magnets that are in the HOLD state and therefore act as an input. Magnets on the right are in the RESET state and they have no influence on the signals propagation. To obtain an errorless signal propagation magnets of a clock zone must be forced in the RESET state before neighbor magnets start to switch (see Fig. 2(B)). This is obtained by overlapping the clock waveforms [18].

1.1 Designing Circuits with Clock Zones Constraints

The clock mechanism is central to the entire NML (and QCA) technology and it is the source of the main problems encountered at architectural level. Particularly in case of NML implementation four possible clock mechanisms can be identified, as shown in Fig. 3. The first mechanism uses a magnetic field generated by a current flowing through a wire placed under the magnets' plane [13,19], as shown in Fig. 3(A). The second mechanism uses Spin-Torque coupling with a current flowing through the magnets themselves [20–22], as can be seen from Fig. 3(B). In this case the basic element is no more a simple magnet but it is a Magneto-Tunnel Junction (MTJ). The third clock mechanism, shown in Fig. 3(C), uses the mechanical stress induced by the strain of a piezoelectric layer to force magnets in the RESET state [23,24]. In this case the clock mechanism is no more based on a current but on a voltage. Finally, a 4th clock topology is available as shown in Fig. 3(D). In this case, magnets are made by Cobalt/Platinum multilayered

structures where the magnetization lies out of plane [25–27]. The clock is an oscillating magnetic field applied perpendicular to the plane. No clock zones are used because the magnetic field is applied globally to the circuit.

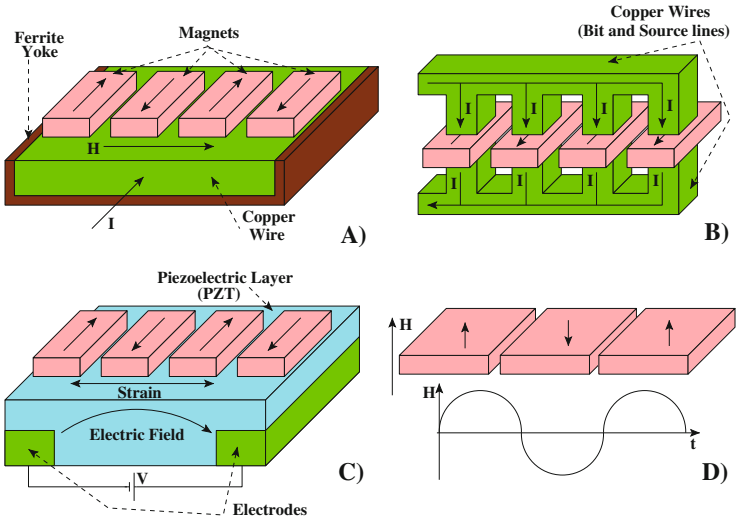


Fig. 3. NML main clock systems. (A) Magnetic field clock. A magnetic field is generated by a current flowing through a wire placed under the magnets plane. (B) STT clock. Spin-Torque coupling due to a current flowing through the magnets, which in this case are MTJ junctions. (C) Magnetoelastic clock. The mechanical stress induced by the strain of a piezoelectric layer at which a voltage is applied forces magnets in the RESET state. (D) Out-of-plane NML clock. In this particular implementation of NML, the magnetization of each dots points out-of-plane. The clock is an oscillating magnetic field applied perpendicularly to the plane. No clock zones are present because clock field is applied uniformly to the entire circuit.

The clock generation system is important from the circuits architecture point of view because it has a strong impact on the layout of the clock zones. Particularly, to drive the currents and the voltages required by different clock systems, wires must be properly routed through the circuit. Thanks to the size of these wires and the limitations of the technological processes available, not every clock zones layout is possible. Different clock systems have different constraints. For example, relatively to the magnetic field clock we have developed the "snake-clock" system [15,18]. The clock zones layout is outlined in Fig. 4(A). Clock zones are made by parallel strips that represent the physical wires used to generate the magnetic field. To propagate correctly, signals must be routed through a precise sequence of clock zones, from zone 1 to zone 2 and finally to zone 3. As can be seen from Fig. 4(A), with this clock zone layout, magnets can propagate in only one direction, from left to right. Moreover, signals cannot propagate vertically in the same clock zone because the number of magnets that can be cascaded is

quite limited [14,28]. As a consequence vertical signals follow a stair-like propagation (see Fig. 4(A)). This clock system is called "snake clock" because the vertical signals propagation recalls the movement of a snake. To propagate signals in the opposite direction, from right to left it is instead necessary to switch the order of phase 2 and 3. This is possible locally twisting the correspondent wires as shown in Fig. 4(B). Wires are physically located on different planes [13] so they can be twisted freely without interferences. Magnets cannot be placed in the area correspondent to the clock wires crossing. We have demonstrated in [29,30] that this clock system is immune to possible crosstalk between neighbor clock wires.

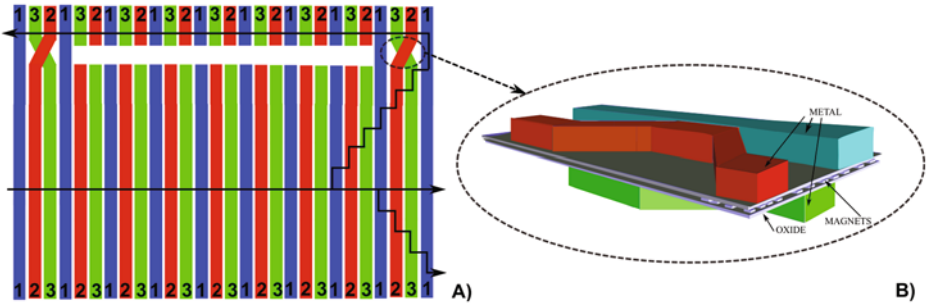


Fig. 4. Snake clock system. (A) Clock zones are made by parallel strips that correspond to the clock wires used to generate the magnetic field. (B) To allows signals propagation wires are twisted when signals need to propagate in different directions. The wire twisting is possible because wires are alternatively placed over and under magnets plane.

Other clock solutions are possible, for example in [31] a more simple 2-phases clock was presented. Moreover other clock systems will have different clock constraints. However two important things must be understood: first, the clock zones layout must be chosen carefully according to the technological processes used to fabricate the clock generation network and to the requirements of the NML signals propagation. The second important fact to be considered are the constraints and limitations generated by the chosen clock system, that must be carefully taken into account in the circuits design. More in general this is true also for other QCA implementations.

1.2 Problems and Solutions

While the clock mechanism influences the circuits layout differently, depending on the particular system chosen, it has a second more important consequence that is shared by all kind of NML and QCA implementations. Considering a N-phase clock system, every group of N consecutive clock zones has a delay of exactly one clock cycles and it is therefore equivalent to a CMOS register (Fig. 5). A QCA wire can therefore be seen as a shift register [18,32], leading

therefore to an intrinsic pipelined behavior. Pipelining is a common feature also in CMOS technology but in this case there are two important differences. In CMOS the level of pipelining is generally small and moreover it is a parameter chosen by the designer. In this case the level of pipelining is extremely high and it is not a parameter chosen by the designer but it depends on the circuit layout [33]. This is valid for all the clock implementations, also in case of the Out-of-plane NML clock (Fig. 3(D)). While with this clock system no clock zones exists, in a wire signals needs exactly one clock cycle to propagate through two magnets [27]. As a consequence a block of two magnets in a wire has a delay of one clock cycles.

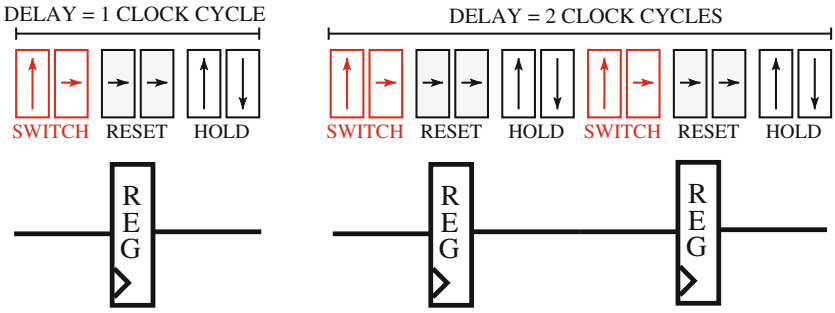


Fig. 5. NML intrinsic pipelining. Every group of 3 consecutive clock zones has a delay of 1 clock cycle and is therefore equivalent to a CMOS register.

In Field-Coupled technologies, like QCA and NML, three important problems can be identified at the architectural level. Some of them are already known. Here we assess them and we discuss thoroughly several alternative solutions, some of them previously mentioned and some newly proposed here.

- *LAYOUT=TIMING*. This is the first problem that arises thanks to the intrinsic circuits pipelining. The propagation delay of a signal in terms of clock cycles depends on the length of its correspondent wire in terms of clock zones. As a consequence mismatches in the wires length generates errors in the logic operations performed by the circuit. This problem is analyzed in detail in Sect. 2.
- *LOOPS*. This is the second problem due to the intrinsic pipelining. If a loop with a length of N clock cycles is present inside the circuit, the throughput is reduced of N times. Moreover serious synchronization problems arise. This problem is thoroughly studied in Sect. 3.
- *INTERCONNECTIONS*. This further issue is connected to both the Field-Coupling principle that represents the base of this technology, and to its intrinsic pipelining. In complex circuits interconnections occupy a lot of area, moreover long interconnections generate a long delay in signals propagation. This problem is described in Sect. 4.

It is important to further underline that while the results here presented are obtained using NanoMagnet Logic as a target technology, they can be applied to any Field-Coupled device. This chapter shows therefore the guidelines that a designer should follow to effectively develop circuits with this technology. These guidelines are also used in ToPoliNano [34–36] a dedicated synthesis/simulation tool for NML that we are developing. As part of ToPoliNano we are working on an automatic layout generator [37] that is necessary based on the rules presented in this chapter.

2 Layout=Timing

The first important problem that a designer must face during the development of any NML circuit is called "layout=timing" [33]. Thanks to the multiphase clock system, every group of N consecutive clock zones has a delay of exactly one clock cycle. The main consequence is that the propagation delay is no more a designer choice but it depends on the circuit layout. However, synchronization issues can arise due to mismatch in the wires length and therefore in the propagation delay of signals. The situation is explained in Fig. 6. Two input wires are connected to a NML AND gate. AND/OR gates can be obtained changing the shape of one magnets as shown in [31], and altogether with the majority voters [30,38] and inverter they represent the logic gates set available with this technology. In Fig. 6(A) input wires have a different length in terms of clock zones. The two input signals arrive at the gate inputs with the difference of one clock cycle and the output of the gate is consequently wrong. To solve this problem the wires length must equalized, as shown in Fig. 6(B). Signals are therefore perfectly synchronized and the gate output is correct.

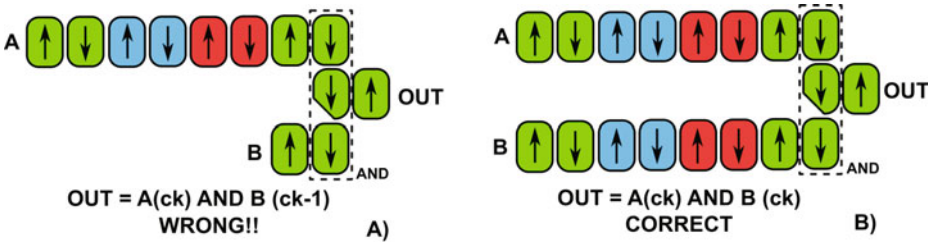


Fig. 6. Layout=Timing problem. Since the propagation delay of a signal depends on the wires length, if input wires of one gate have different lengths, their signals will arrive with different delays and the result will be wrong. Wires length must be equalized to synchronize signals and obtain the correct result.

Nonetheless, considering a complex circuit based on hundreds of thousands or millions of gates, the equalization of the length of every signal in it can be a nearly impossible task. In the following we discuss two different solutions, one

at logic level and one at layout level. The first solution, described in Sect. 2.1, is based on the use of delay insensitive asynchronous logic to obtain automatic signals synchronization. The second solution, described in Sect. 2.2, is based on the introduction of constraints on the clock zones layout making it very regular and automatically equalizing the length of every wire in the circuit.

2.1 Asynchronous Logic

The first solution that can be adopted to solve the "layout=timing" problem is to use a different kind of logic, particularly a specific type of asynchronous logic called Null Convention Logic (NCL). In this logic [39,40] signals are coded using two bits (Fig. 7(A)). The logic '0' corresponds to the coding '01' while logic '1' corresponds to the coding '10'. '00' corresponds to a particular state called NULL. '11' is a not allowed state. Figure 7(B) shows an example of NCL gate, called TH22, and its logic equation. The particularity of this kind of logic is that it is completely delay insensitive. The circuit behavior is depicted in Fig. 7(C). Circuits switch periodically from DATA to NULL and from NULL to DATA, however the transition happens only when all inputs switch. This means that circuit will switch from NULL to DATA only when all inputs switch from NULL to DATA. Circuits will remain in the DATA state until at least one of the inputs is in the DATA state. The transition from DATA to NULL will happen only if and when all inputs will switch from DATA to NULL. Then the cycle will restart. As a consequence a complete delay insensitivity is reached, because it does not matter if there is a difference in the propagation delay of signals. The transition from one state to the other will occur only when the last signal arrives at the circuit inputs.

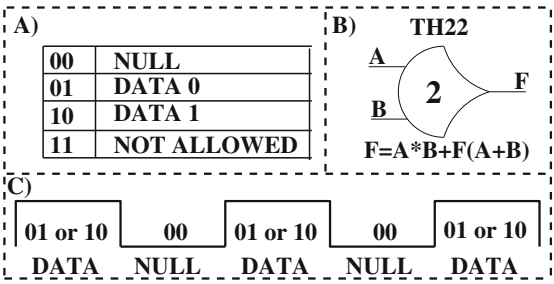


Fig. 7. Null Convention Logic (NCL). (A) Signals are coded using two bits. (B) Example of NCL gate and its logic equation. (C) NCL circuits switch periodically from DATA to NULL state and then from NULL to DATA.

The delay insensitivity of NCL logic apparently makes it perfect for NML circuits. This solution works correctly as demonstrated in [32,41], but it greatly reduces circuits performance. This also happens if CMOS technology is used as a target, because the high robustness of this logic comes at the price of

a big area overhead. In case of NML circuits the price that must be paid to achieve complete delay independence is much higher. There are two reasons behind this fact: the signal coding and the communication protocol. With this logic, signals are coded using two bits, that means that circuit area is at least two times higher with respect to a not-coded circuit, but in NML more area means more delay. NCL logic is also an asynchronous logic and as a consequence it uses a communication protocol, which plays an important role in the overall performance losses. Figure 8 shows the communication protocol used. Circuits can be generically represented by blocks of combinational logic embraced by two asynchronous registers which implement the communication protocol. The communication protocol works following the step explained in Fig. 8. First a new DATA is sent to the circuit (Fig. 8(A)). When the DATA reaches the second register, an acknowledgment signal is sent back to the first register (Fig. 8(B)). The first register then sends a NULL to the circuit (Fig. 8(C)). When the NULL is received by the second register, an acknowledgment is sent back to the first register (Fig. 8(D)) and a new cycle can start. While this communication protocol works very well, it requires a signal to traverse the circuit four times. As a consequence every DATA cycle requires a time equal to four times the signal propagation delay of the circuit. This explains the high losses of performance due to the use of NCL logic.

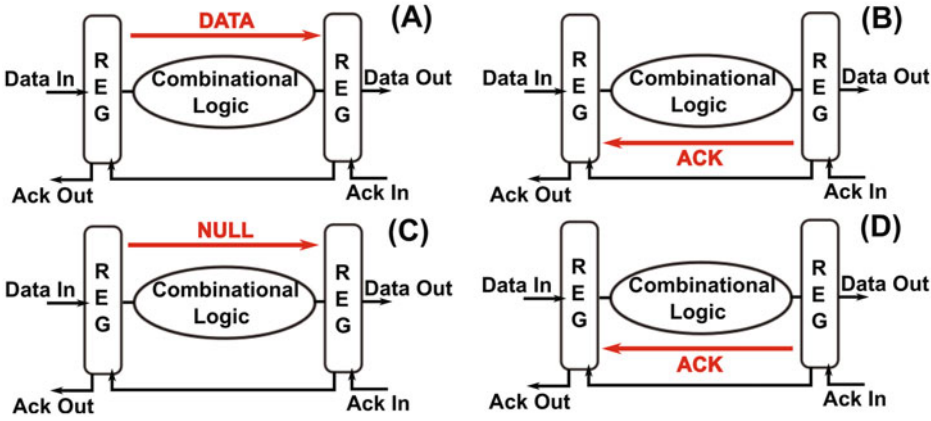


Fig. 8. NCL asynchronous protocol. (A) A DATA is sent to the circuit. (B) An acknowledgment that states the reception of the DATA is sent back. (C) A NULL is sent to the circuit. (D) An acknowledgment that states the reception of the NULL is sent back.

Other types of asynchronous logic with a simplified communication protocol can be studied and possibly adopted to improve performance. However asynchronous logic is based by definition on a communication protocol which uses messages between units to assure signals synchronization. The problem is that in this kind of technology the delay penalty along interconnections is very high.

As a consequence in [41] we have developed a new mixed synchronous-asynchronous protocol to exploit the maximum potential from this technology. The protocol is described in Fig. 9. It is based on simple boolean logic gates with no signals encoding, greatly reducing the area overhead. Asynchronous registers are substituted by a synchronization block shown in Fig. 9, which is simply a multiplexer with the output connected to one of its inputs. Normally the Selection bit (*Sel* in Fig. 9) is at logic '0', that means that the multiplexer is in memory state and its output is constant. Every N clock cycles a new data is sent to the multiplexer input followed by an enable signal which is connected to the *Sel* pin of the multiplexer. As a consequence the multiplexer samples the new data and after it returns in the memory state. The consequence is that a new data is sent to the circuit every N clock cycles and in the middle the output value of the multiplexer is kept constant. The value of N is chosen according to the longest propagation delay in the combinational path (in case of Fig. 9 it is equal to 2 clock cycles) in this way all signals have time to propagate through the circuit, granting a correct operation of the circuit.

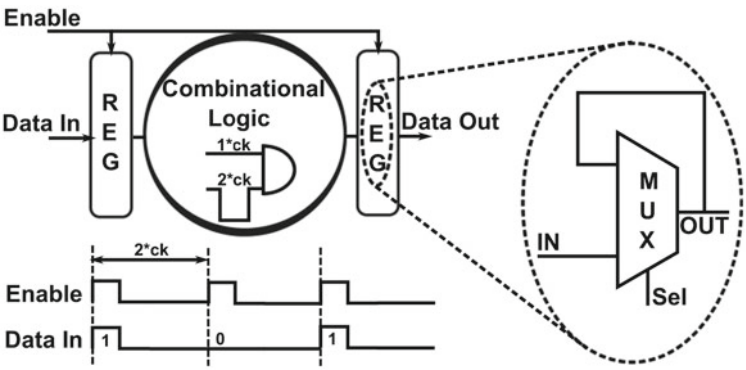


Fig. 9. New asynchronous protocol. No signal coding is used and asynchronous registers are substituted by a multiplexer with the output connected to one of its inputs. The output of the multiplexer is kept constant and only every N clock cycles a new data is sampled. The value of N is chosen according to the maximum propagation delay in the combinational path, in order to ensure that all signals had time to propagate correctly through the circuit.

To evaluate the performance of different logic solutions a simple but complete microprocessor was developed in [32,41]. The microprocessor is schematically reported in Fig. 10(A). The architecture is based on four main components, a program counter to run programs, two memories, one for instructions and one for data, and an arithmetic-logic unit to execute the selected instruction. Asynchronous registers are exploited to implement the communication protocol. The microprocessor architecture is simple but it allows to execute all kind of instructions commonly found in the instruction set of commercially available

microprocessors. As a consequence it represents a good test to evaluate the performance of different logic solutions applied to NML logic. The microprocessor is described using a RTL model of NML technology written using VHDL language. The model is reported in Fig. 10(B). The basic idea behind the high level modeling of NML technology is to exploit its intrinsic pipelining. Figure 10(B) shows an example of NML circuit and its equivalent RTL model. Registers are used to emulate the signals propagation delay while ideal logic gates without delay are used to model the circuit logic behavior. The result is a circuit that behaves exactly like its NML counterpart, but the advantage is that this circuit can be described and simulated using powerful design tool already available for CMOS technology, like Modelsim [42]. In this way complex NML circuits can be easily described and simulated. More information on the model itself can be found in [29].

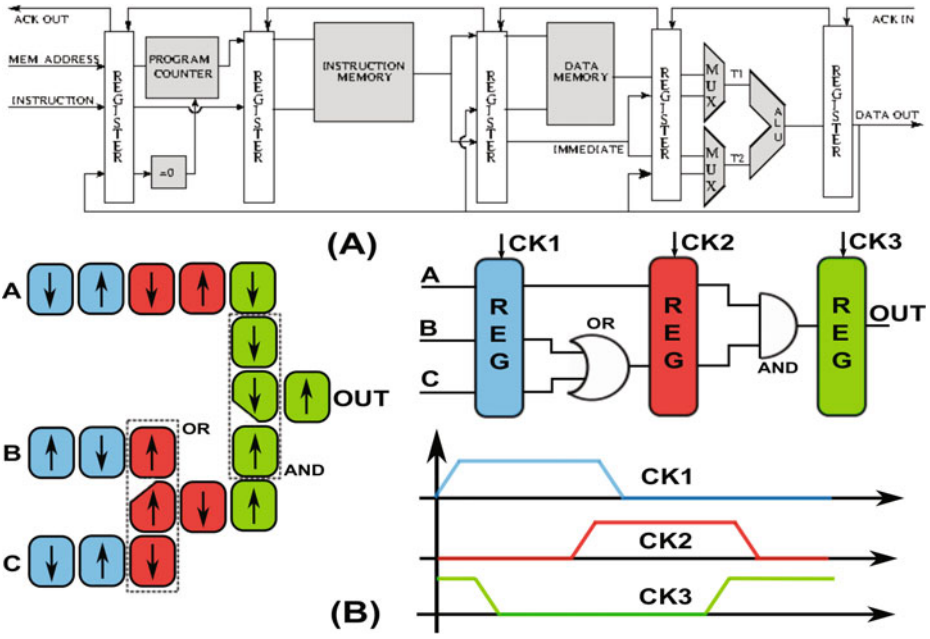


Fig. 10. NML microprocessor. (A) The microprocessor architecture is made by a program counter, to execute programs, two memories, one for data and one for instructions and an ALU to execute the instructions. The architecture is simple but it allows the execution of all common microprocessors instructions. (B) NML equivalent RTL model used to describe and simulate the microprocessor.

Detailed simulations of the microprocessor are not reported here, and can be found together with a thorough description in [32,41]. The most important result obtained from simulations is indeed the time needed to execute one instruction. Using NCL logic one instruction takes 5.35 µs to be executed, while with the

improved asynchronous protocol only 0.546 µs are required. As a consequence a speed up of 10 times can be observed. This result demonstrates the validity of the logic solution here proposed, which allows to solve the layout=timing problem without speed penalties. The clock frequency used in the simulation is about 100 MHz, that means a clock period of about 10 ns. The execution of one instruction with the improved logic requires therefore 53 clock cycles. This slowdown is not due to the use of asynchronous logic but to the presence of a long feedback signals inside the circuit. This aspect is described in detail in Sect. 3.

2.2 Clock Zones Layout for Automatic Signals Synchronization

In order to synchronize signals in a complex circuit it is however possible to work on the clock zones layout instead that on the logic type. The idea is to exploit the potential of circuit layout shown in Fig. 4(A). A circuit detailed example is reported in Fig. 11. Clock zones are made by parallel strips and inputs arrive from the left side, all of them starting from the first clock zone. Output signals are generated at the circuit right side. With these constraints, perfect signals synchronization is achieved. At any point inside the circuits all signals are perfectly synchronous. This result is obtained without the need of using asynchronous logic keeping the circuit as simple as possible, minimizing the area overhead and maximizing performance. Moreover this solution allows to gain two further advantages. Signals synchronization is automatically achieved independently of circuit complexity and this approach can be successfully applied for hierarchical circuit descriptions.

While the clock zones layout of Fig. 11 was built upon the constraints of NML circuits based on magnetic field clock, it allows other considerable results to be achieved, like automatic signals synchronization and easy circuits description. As a consequence this same layout can be exploited also for other clock systems. Its regularity also greatly helps the fabrication process.

3 Loops

If the pipelined nature of these technologies causes troubles to achieve signals synchronization in combinational circuits, the situation worsens when loops are present inside the circuit. Loops are required to build any complex circuit. An example of such a system is reported in Fig. 12, where the detailed layout of a 2 bits multiply and accumulate unit (MAC) is shown. In Fig. 12 the simplified circuit schematic is reported in detail. Two incoming signals are multiplied and then the result is added to the result of the previous operation. MAC units are a fundamental block in digital signals processors (DSP).

The first and most important problem due to the presence of loops is the loss of performance. In pure combinational circuits thanks to the pipelining a new data can be sent to the circuit every clock cycle. If a loop of length N clock cycles is present, a new input can be sent only every N clock cycles, thus reducing the

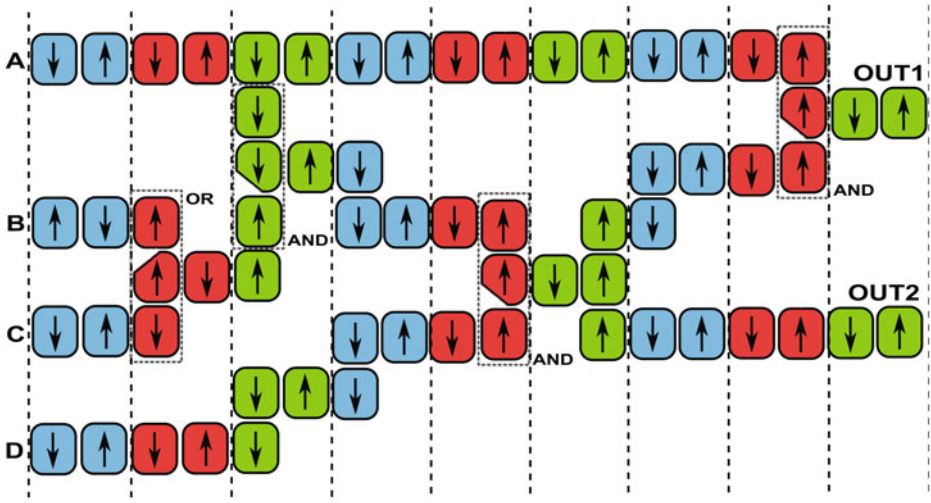


Fig. 11. Regular clock zones layout to obtain automatic signals synchronization. Clock zones are made by parallel strips, input comes from left direction while outputs are generated at the circuit right side.

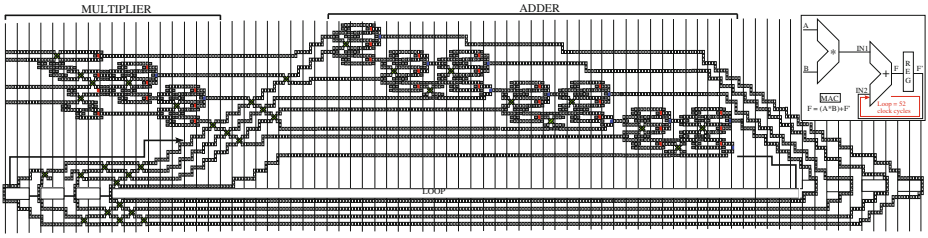


Fig. 12. NML 2 bits Multiply and accumulate unit layout. (A) Detailed circuit layout. The loop length is 52 clock cycles. In the upper detail

throughput of N times. To demonstrate this fact the MAC unit was described using the VHDL model outlined in Sect. 2.1. Simulation results are reported in Fig. 13. The loop length is 52 clock cycles, so if inputs are fed to the circuit one every 52 clock cycles the MAC output is correct (Fig. 13(A)). If the delay between subsequent data is lower, for example 50 clock cycles like in Fig. 13(B), errors are generated. This behavior can be easily explained by the fact that once the multiplication output is generated it needs 52 clock cycles to travel back and to reach the adder inputs. Therefore to synchronize signals the next output of the multiplier must be generated exactly with a delay of 52 clock cycles, in this way it will reach the adder input together with the previously generated multiplier output. This is a common problem also in CMOS technology, for example in superscalar microprocessors. However in this case the level of pipelining is much

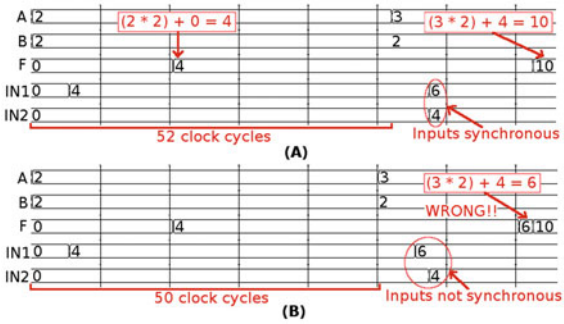


Fig. 13. MAC simulations, with inputs sent one every 52 clock cycles (A) and every 50 clock cycles (B). If the delay between new inputs is lower than the loop length errors are generated.

deeper, it depends on the circuit layout and it is not a pure designers choice. Solutions to improve performance in circuits with loops are presented in Sect. 3.1.

The second problem encountered with loops in intrinsically pipelined circuits, is related to signals synchronization if more loops are present inside the circuit. Synchronization problems and related solutions are presented in Sect. 3.2.

3.1 Throughput Maximization

To maximize performance two possible approaches can be used. It is possible to work at algorithmic level, rearranging data avoiding data dependencies, or it is possible to work on the circuit architecture with the aim of reducing the loops length. In the following both solutions are presented.

Interleaving. As seen above, if a loop is present, two consecutive data operands can only be sent to the system after a number of clock cycles equal to the loop length degrading system performance. The reason for this problem lies in the fact that there is data dependencies between two consecutive data, i.e. one data depends on the previously sent data. It is however possible to work at algorithmic level rearranging data to avoid data dependencies. To do so, some techniques that are commonly adopted in CMOS technology can also be applied here, like dynamic data dependency rearrangements and predictive techniques used in superscalar microprocessors. These solutions may potentially improve performance. However they aggravate the system with additional calculations, and require careful effectiveness analysis from application to application. A more simple technique, called “interleaving”, can be adopted. It does not require significant modifications to the original algorithm. The basic concept of interleaving is to parallelize relatively independent operations by interleaving data sequence at the inputs. As an example two independent sequences of data A, B and C, D

are fed to the circuit. Data from different sequences are completely independent. The aim is to calculate (Eqs. 1 and 2)

$$\sum_{i=0}^{3} A * B \tag{1}$$

$$\sum_{i=0}^{3} C * D \tag{2}$$

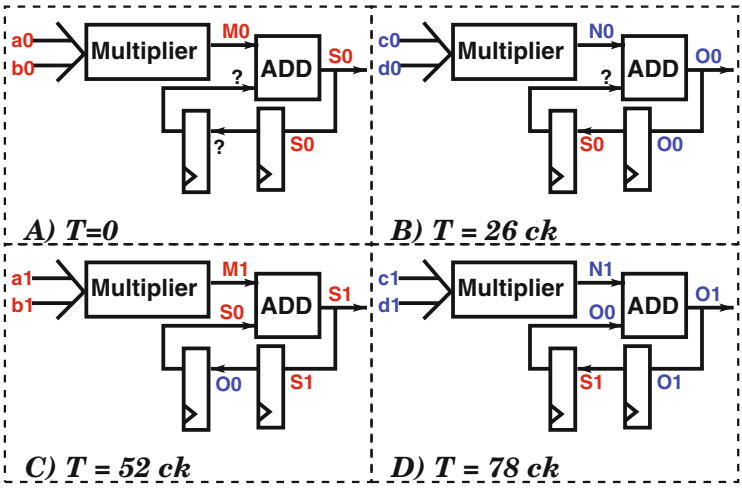


Fig. 14. Example of interleaving applied to a MAC unit. Two input sequences are sent to the circuit, both send one value every 52 clock cycles, with 26 clock cycles between data of different sequences. The left column shows the calculation of sequence A, B, while the right one represents the calculation of sequence C, D. As can be observed, the two sequences are executed in parallel but they do not interfere with each other. (A) *a0* and *b0* are sent to the circuit. (B) After 26 clock cycles *c0* and *d0* are sent to the circuit. (C) At a time correspondent to 52 clock cycles *a1* and *b1* are sent to the circuit and they reach the adder input exactly with *S0*, the result of the previous operation. (D) At 78 clock cycles *c1* and *d1* are sent to the circuit.

At the beginning *a0* and *b0*, the first two data of the first sequence are sent to the circuit (Fig. 14(A)). Just for this example, to better clarify the interleaving principle, the multiplier is considered ideal without delay, so data propagate directly from the general MAC inputs to the adder inputs. After a time equal to half the loop length (26 clock cycles in this case), *c0* and *d0* are sent to the inputs (Fig. 14(B)). This operation is correct because there is no data dependency between them and *a0*, *b0*. At the 52nd clock cycle, *a1* and *b1* are then sent to the circuit and they arrive at the adder inputs together with *S0* (Fig. 14(C)), the result of the previous operation. After other 26 clock cycles *c1*

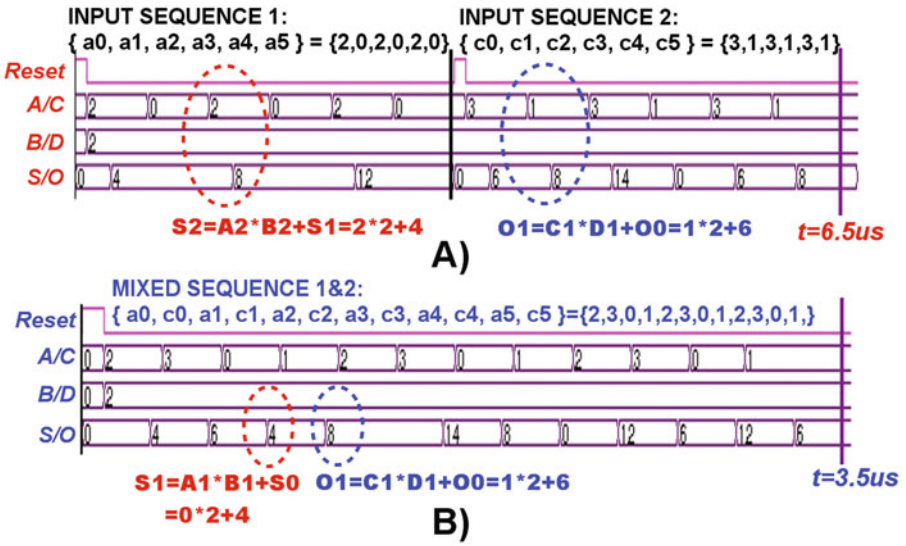


Fig. 15. Simulation comparison between the original circuit behavior (A) and the circuit with data interleaving (B). The only modification done is mixing input sequence 1 and input sequence 2. As can be observed, the calculation results are the same, only are also mixed, but the total simulation time is significantly reduced.

and *d1* are sent to the circuit. They arrive at the adder inputs together with *O1* (Fig. 14(D)). Data from the same sequence are always sent every 52 clock cycles granting perfect signals synchronization, but two sequences are executed in the same time required to execute one sequence alone, effectively doubling the throughput.

Figure 15 shows a MAC simulation with and without interleaving. The sequences, originally executed serially as A, B and then C, D (Fig. 15(A)), are parallelized and interleaved (Fig. 15(B)). The original time interval between two consecutive input values (52 clock cycles) is halved. Therefore, the total execution time of these two independent operations is halved, and the throughput has been doubled. This principle can be expanded executing 52 operations in parallel, sending therefore effectively one data every clock cycle. The throughput is therefore maximized as in a pure combinational circuit. This clearly demonstrates that interleaving is the perfect technique to be adopted in case of NML (and QCA) circuits. To fully exploit the potential of this technology, it requires a large number of independent data sequences to process in parallel. Only applications where a large number of data to process is available are therefore best adapted to this technology.

Architecture Redesign for Loops Length Reduction. In addition to algorithm rearrangement techniques like interleaving, it is also possible to modify architectures with the aim of reducing the loops length. Since loops are the

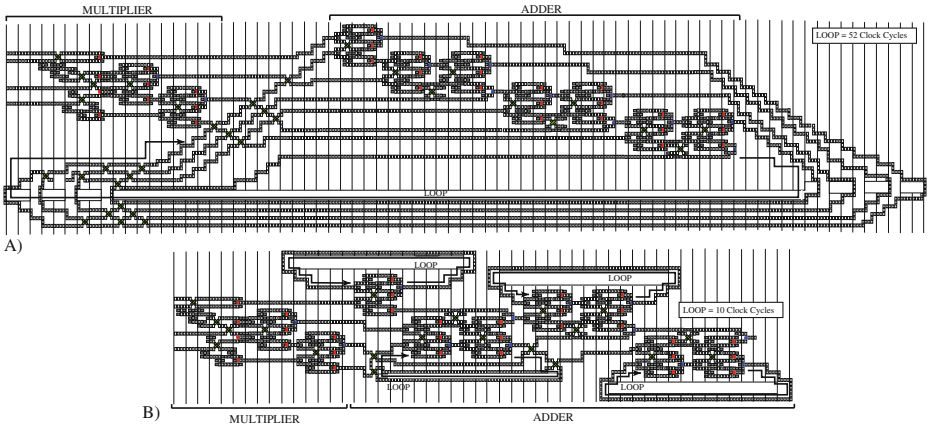


Fig. 16. Redesign of MAC to reduce the loop length. With this new design the loop length is only 10 clock cycles, moreover it is independent from the bit number.

major sources of performance degradation in this technology, the shorter the loop, the better the performance. With a better analysis of the circuit layout, the MAC circuit can be modified as shown in Fig. 16, so the original 52 clock cycle long loop has become only 10 clock cycle long, without changing the system algorithm. The delay is also independent from the MAC bit number, and overall the area is smaller. Without interleaving the delay between one data and the next is therefore of only 10 clock cycles.

Figure 17 shows a simulation comparison (without interleaving) of the original architecture (Fig. 17(A)) and the optimized one (Fig. 17(B)). The execution time, and therefore the throughput, is improved of 5 times. While this technique shows good results, it can be only used as a complementary technique, because it does not allow to completely eliminate the loop. Interleaving is still necessary, however the number of operations required to reach maximum performance is lower.

3.2 Signals Synchronization

While the loss of performance is probably the more relevant problem due to the presence of loops, some important issues also arise for signals synchronization. When more loops are present inside the circuit, they must be carefully designed to achieve perfect signals synchronization. Another issue is related to the necessity of adding a specific delay on a particular signal. This is a common requirement in many circuits, where the algorithm mapping implies the necessity of delaying some signals of a specific amount of clock cycles. In case of NML and all the intrinsic pipelined technologies particular rules must be followed to add delays on specific signals. Both problems and the related solutions are described in the following.

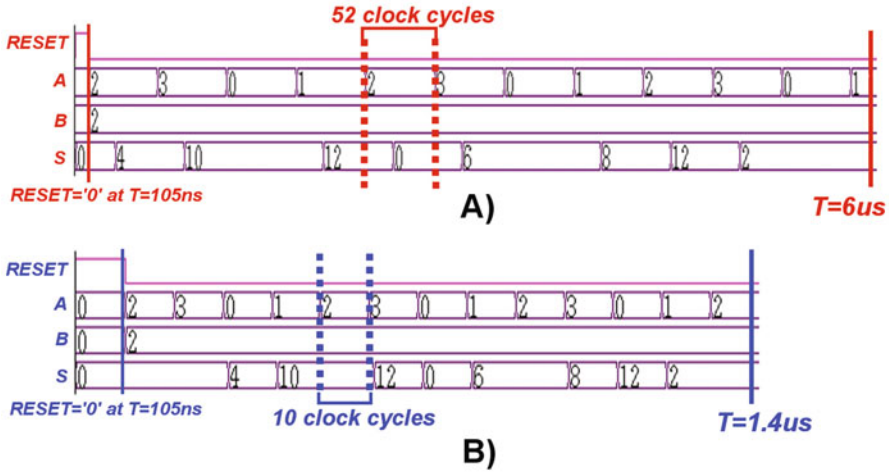


Fig. 17. Simulation comparison between the original architecture (A) and the redesigned architecture (B).

Nested Loops. Generally, it is quite common to find multiple loops in a relatively complex system, and they can be independent from each other or they can be nested. In this second case, signal synchronization problems arise. A circuit example with two nested loops is represented in Fig. 18. The algorithm is simply $SUB\_OUT = SUM\_OUT - SUB\_IN = A + ADD\_IN - SUB\_IN$. Figure 19 shows instead the circuit simulation. Since $ADD\_IN = SUB\_IN$, the result is simply the value of A.

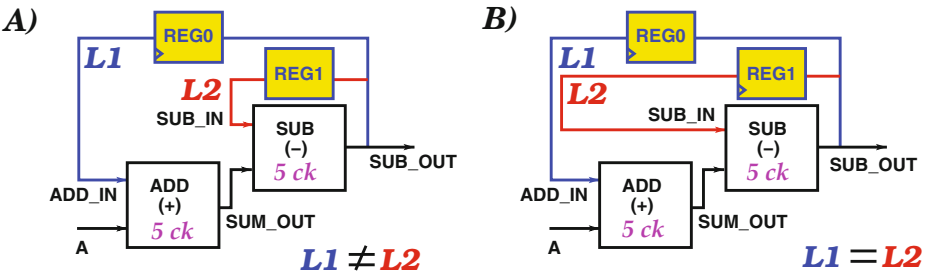


Fig. 18. Circuit representation with two nested loops (A). Originally, the two loops L1 and L2 do not have the same length (B). For signal synchronization, loop L2 is extended to the same length as loop L1.

If the loop lengths are different (Fig. 18(A)) signals are not perfectly synchronized and the output result is wrong as can be observed from Fig. 19(A). If the loop lengths are equal (Fig. 18(B)) the circuit output is correct, as the simulation of Fig. 19(B) clearly demonstrates.

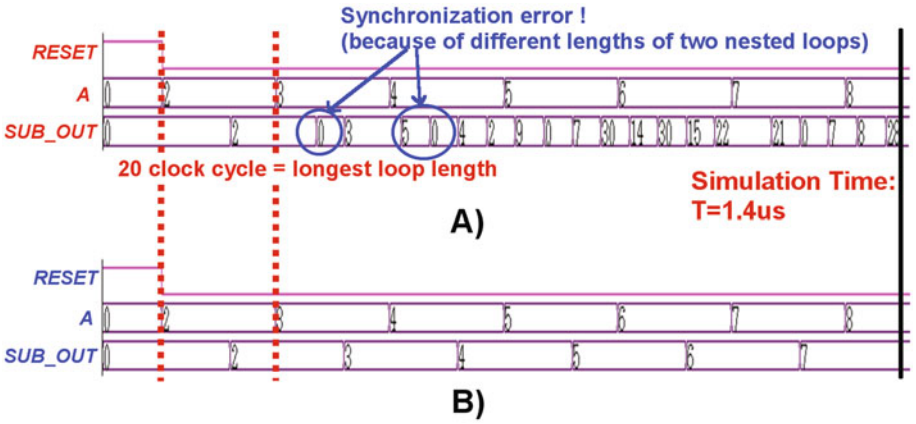


Fig. 19. Simulation comparison with nested loops that have different (A) and equal length (B). As can be observed, synchronization error will appear if the two nested loop lengths are not equal.

The consequence is simple: if there are more loops inside the circuit and these loops are nested inside each other, they must have the same length to obtain perfect signals synchronization.

Additional Delay Loops. The presence of loops introduces another problem related to signal synchronization. Sometimes mapping an algorithm to an electronic circuit requires the insertion of a delay on specific signals. An example is shown in Fig. 20(A). An adder calculates the sum between a signal A, the same signal delayed of an additional clock cycle $A(t-1)$ and the output of the previous operation $SUM(t-1)$. When trying to map the same circuit onto NML (or QCA) technology, one important problem arises. In the CMOS implementation the loop has a delay of exactly 1 clock cycle, a new input A is sent every clock cycle. All signals arrive at the adder input with perfect synchronization, respecting the algorithm.

However in NML (and QCA) loops normally have a delay of N clock cycles. The result of the previous operation arrives at the adder input after N clock cycles. As explained in Sect. 3 a new data must be sent every N clock cycles. This is true also in case of interleaving, because N operations are interleaved, but the delay between one data and the subsequent data of the same operation is always N clock cycles. Therefore to map the same algorithm it is not sufficient to delay the input A by 1 clock cycle. It must be delayed by N clock cycles (Fig. 20(B)). Considering an example, an input ($a0$) is sent to the first adder input. The output is calculated and after N clock cycles it reaches the adder inputs. After N clock cycles a new data ($a1$) is sent to first adder

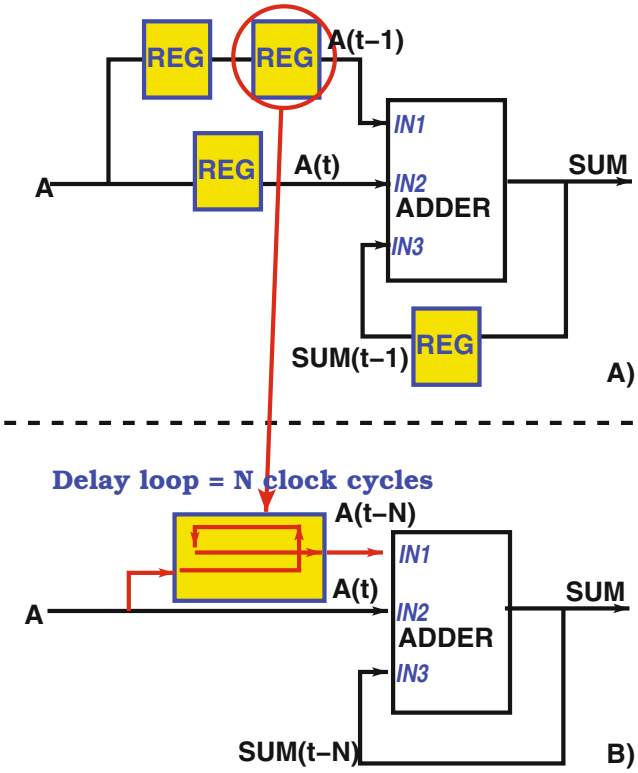


Fig. 20. Circuit example. The algorithm is $SUM(t) = A(t) + A(t-1) + SUM(t-1)$. (A) CMOS implementation. (B) NML implementation. An additional delay, under the form of a wire loop, is used to map the additional delay on signal A.

input, but at the same time the previous input ($a0$) reaches the second adder input because it is delayed of exactly N clock cycles. As a general rule, for each additional delay of one clock cycle in the original circuit, an additional delay of N clock cycles must be added to the correspondent wire in the NML implementation.

Figure 21 shows the circuit simulation. In Fig. 21(A) the CMOS implementation is depicted, while in Fig. 21(B) the NML simulation, with the use of the additional synchronization delay, is shown. The time scale is different because in the CMOS implementation, one data is sent every clock cycle while in the NML implementation one data is sent every N clock cycles, but the behavior is the same. If the signal A is not correctly delayed the result is wrong, as can be seen from Fig. 21(C).

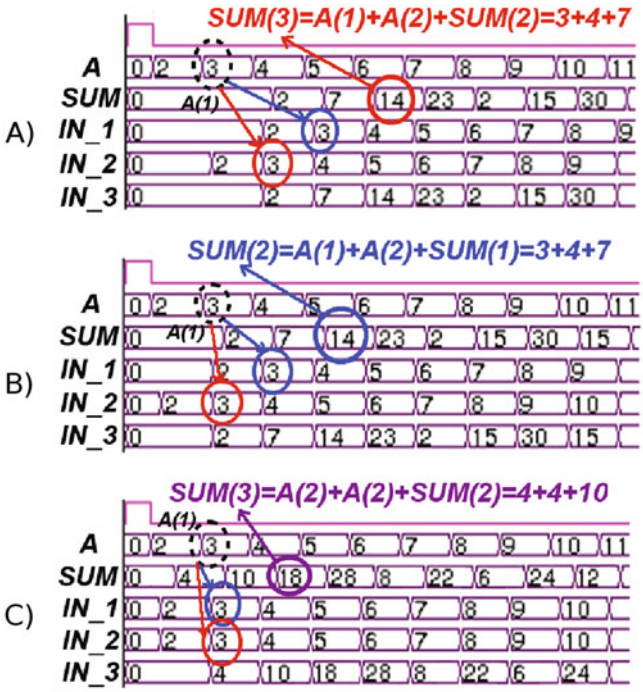


Fig. 21. Simulation comparison between the original circuit (A), the correct NML circuit with additional loop (B), and the NML circuit without additional loop (C), which gave the wrong result.

4 Interconnections

The final problem encountered in the design of complex circuits with Field-Coupled technologies, is the strong impact of interconnections both on the propagation delay and on the circuit area. Figure 22 shows the detailed layout of a NML 8 bits comparator, that is part of an LDPC decoder for wireless signals [43–45]. The circuit schematic is shown in Fig. 22, while the decoder description is not reported here because it is not relevant for this discussion. Details on the architecture itself can be found in [43]. What it is important to understand is that most of the area (up to 99 %) is occupied by interconnection area. This layout is based on the technological constraints described in Sect. 1.1, so in other types of NML or QCA the impact of interconnections can be different. However interconnections will always represent an important part of the circuit area. It is worth noting again that in Field-Coupled devices more area means more delay and more power consumption.

The reason of this problem comes from the nature of the technology itself, which favors local interactions over neighbor elements, penalizing long interconnect wires. A possible solutions is to use particular architectures, called systolic arrays, that can greatly reduce the interconnections overhead. Systolic arrays

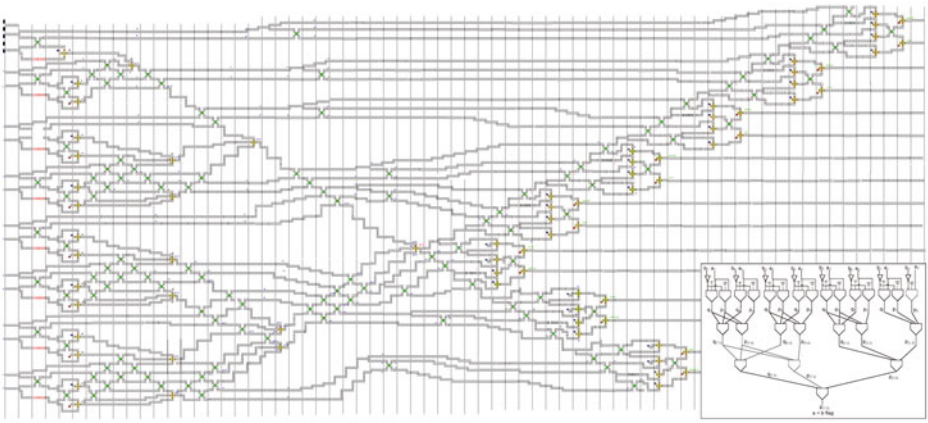


Fig. 22. Example of interconnection impact in complex NML circuits. The circuit, reported schematically in the detail, is a 8 bit comparator used inside a wireless decoder.

are briefly described in Sect. 4.1 alongside techniques to optimize the design and to improve performance.

4.1 Systolic Arrays

A Systolic Array (SA) is a network of Processing Elements (PEs), also called "Cells", that are locally interconnected and can work in parallel. SAs were first introduced by Kung and Leiserson in 1978, who stated: "a systolic system is a network of processors which rhythmically compute and pass data through the system" [46]. Each PE receives data from neighboring cells or from outside and outputs result to the outside or to near PEs. Two are the main concepts at the basis of SAs: parallel computation (i.e. all PEs work in the same way on different data) and local transmission of data (i.e. there are not global signals). In Fig. 23 three examples of SAs arrangements are shown: a bi-dimensional matrix-like shape SA (a), a linear SA (b), and an SA with signals flowing in three different directions (c). The shape is determined by a three design process steps: starting from the algorithm description, the Dependence Graph (DG) is derived; this is shrunk along one axis to obtain the Signal Flow Graph (SFG) which represents the real shape of the SA. Finally, PE internal structure is derived.

In last decades, increasing operating frequency has been enough to sustain the request of higher computational efficiency of digital circuits, reducing the need of parallel architectures; for this reason SAs have not represented an attractive solution. However, approaching the boundary limit for CMOS scaling [47], and thus for frequency increase, parallel solutions are being exploited and SAs are back in the limelight. They have been designed for image processing [48–50], signal processing [51–53] and video algorithms (such as those for MPEG compression). Recently, automatic tools that translate algorithms to SAs for FPGAs have been explored [54]. Also, reconfigurable arrays, that are not application-specific, have

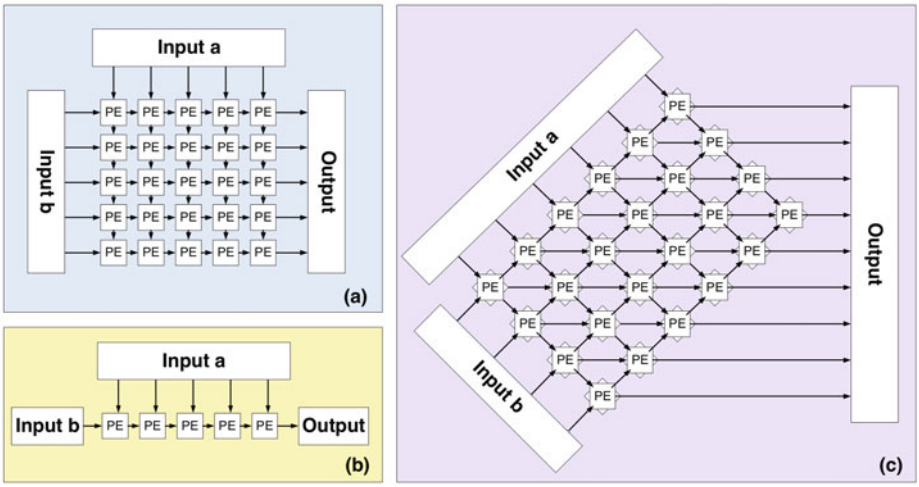


Fig. 23. Three examples of Systolic Arrays. (a) Matrix-like SA. (b) Linear SA. (c) SA with signals flowing in three different directions.

been introduced in [55]. SAs are a good architectural solution for new technologies (beyond-CMOS), where benefits in terms of speed and required area can be achieved only avoiding global interconnections, and SAs verify intrinsically this requirement. SA paradigm allows to design local small circuits (the PEs) and replicate them to organize the whole array structure. In particular, in QCA several circuits have been designed and simulated; as an example, in [56,57] SAs for matrix multiplication and Galois field multiplication have been proposed. Also, NML implementations for convolution filters have been proposed in [58].

While SAs help to solve the interconnection problem of this technology, they still suffer from a loss of performance in presence of loops. Interleaving must therefore be used in conjunction with SAs to maximize performance [59]. As far as optimization with pipeline interleaving is concerned, SAs can be distinguished between those that have PEs WithOut Internal Loops (WOIL), and those with PEs With Internal Loops. The latter can then be further divided in SAs that Store result in cells (WIL-S) and SAs that evaluate the final results passing partial results through lines (WIL-PT). WOIL SAs are composed of PEs without loops; for this reason, pipelining is enough to achieve the required performance increase. Pipeline interleaving for this reason is applied to WIL SAs only.

The generic PE of a WIL SA is shown in Fig. 24. It is composed of 4 parts: an entry section, made of blocks numbered from 1 to i; the forward part of the loop, made of blocks from $i+1$ to j; the feedback part of the loop, made of blocks from $j+1$ to $k-1$; the output block, called k. Each of these blocks has a delay d_n, $n = 1, 2, \ldots, k$ and cannot be internally pipelined. Call Z_e the total delay of the entry block, Z_{fo} the delay of the forward side of the loop, Z_{fb} the delay of the feedback in the loop and Z_o the output delay:

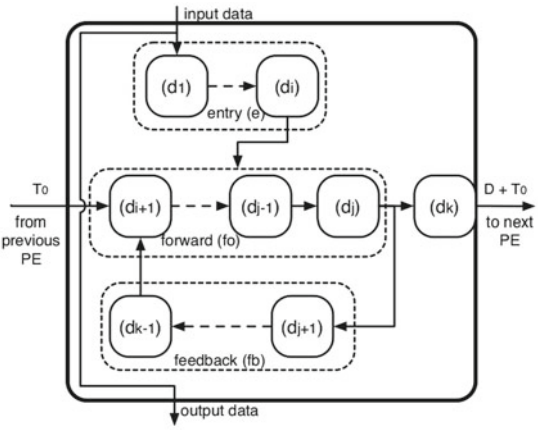


Fig. 24. WIL Systolic Array generic processing element structure.

$$Z_e = \sum_{n=1}^{i} d_n \tag{3}$$

$$Z_{fo} = \sum_{n=i+1}^{j} d_n \tag{4}$$

$$Z_{fb} = \sum_{n=j+1}^{k-1} d_n \tag{5}$$

$$Z_o = d_k \tag{6}$$

Input data coming from outside enter in the first block, while data coming from the neighbor processing element can enter at any stage of the cell. Inputs must be provided every $Z_{loop} = Z_{fo} + Z_{fb}$ cycles, that is the total time of the feedback loop, in order to match them with data coming from the feedback. To apply interleaving the intrinsic pipelined nature of the structure can be exploited. Thus we can improve performance and usage of the cell giving inputs every $J = max\{d_n\}$. Every J cycles a new operation can be started, and in this way M different operations can be interleaved, $M = Z_{loop}/J$ (integer division). After Z_{loop} cycles, the second set of inputs is fed. When Z_{loop} is not a perfect multiple of J, the remainder of the division, called R, must be taken into account. After M successive inputs have been fed, the following one must provided with a delay of $J + R$, so to have synchronization with the value coming from the loop. R represents a number of "stalls" that must be inserted between the one set of M inputs and the following set.

Applying pipeline interleaving it is possible to evaluate M different operations in parallel, having an increase in performance of M. If the number of stalls is high,

it would be favorable to increase the delay of the feedback loop in order to achieve a deeper level of interleave. For example, consider $Z_{fo} = 20, Z_{fb} = 6, J = 9$; in this case $M = 2$ and $R = 8$. If it is possible to increase of 1 cycle the delay of the feedback, then 3 operations can be interleaved and no stalls will be present. As far as the global array is concerned, in WIL-S SAs processing elements work independently each from the others, and the only connections are the wires to pass inputs from one cell to another. Being S the delay in clock cycles of connection wires, inputs must be given with the same rule for all cells (number of interleaved operations and stall cycles), but starting at cycle $S_n = m_n \times S$. m_n is the Manhattan distance between cell n and the top-left one (if we consider data moving from top to bottom and left to right). This delay must then be reported to the actual inputs of the array at boundary cells. In WIL-PT SAs instead, results of a cell are re-used in the same cell but also passed to the neighboring cell. In this case the order of inputs is given by the mismatch between the feedback loop delay and the one to transmit data to next cell.

In order to evaluate the benefits of pipeline interleaving the Cell Updates Per Second (CUPS) parameter can be computed. Consider a square SA as the one in Fig. 23(a): the top-left PE is the one that starts operating first, while the bottom-right one is the last. Given a finite number of inputs, the bottom-right PE is also the one that finishes later computation; hence, total time will be given by the time at which this PE will finish to compute the last result. Total time, Z_{end}, is given by the time for last inputs to reach last PE, plus the time to execute the operation inside the PE itself, called Z_{cell}. S was previously defined as the number of cycles needed for an input to pass through a cell, that is the number of registers of the shift chain. It is possible to assume that this value is equal to transmit one value from top input to bottom output of the PE, or from left input to right output. First inputs will be available at last cell after $2(M-1)S$, where $2(M-1)$ is the Manhattan distance between first cell and last one in an array of $M \times M$ PEs. Call $Z_{end}^{(m)}$ the time at which computation of m-th result is available. Then: $Z_{end}^{(1)} = 2(M-1)S + Z_{cell}$. J is the delay between one input and the following one; if l successive inputs are fed into the array, then we obtain Eq. 7:

$$Z_{end} = Z_{end}^{(l)} = 2(M-1)S + (l-1)J + Z_{cell} \tag{7}$$

Formula (7) expresses the total time needed for executing operations on a $M \times M$ SA that receives l successive data from each input path. During this period of time each cell will execute l operation (one every time a new input is received). The total cell updates are lM^2, and CUPS can be evaluated as:

$$CUPS = f_{clk} \frac{lM^2}{2(M-1)S + (l-1)J + Z_{cell}} \tag{8}$$

This formula must be adapted in two cases: when the array is used without interleaving operations, and when interleaving is exploited to achieve an improvement in performance. In case of no-interleaving formula (8) can be adapted considering $l = M$. In case of n-interleaving instead, each cell will update n times

more than the previous case, hence $l = nM$; this increase is also reflected in the denominator of formula (8) as an increase in total time. Moreover, Z_{cell} and J must be re-evaluated considering interleaving. SAs performance are often evaluated computing peak CUPS $\max(CUPS) = f_{clk} \times \#PEs$ where $\#PEs$ is the number of PEs in the array. Equation 8 is more precise and takes into account the effects of interconnections and delays inside PEs. This equation is upper bounded by $\max(CUPS)$.

In order to demonstrate the effectiveness of the proposed mechanism of pipeline interleaving to increase performance, we evaluated CUPS according to (8) for different levels of interleaving in a practical case. Hereinafter values are given in number of clock cycles. Consider a WIL-S SA with $Z_e = 8$, $Z_{fo} = 7$, $Z_{fb} = 5$, with Z_e and Z_{fo} that cannot be further pipelined, while the feedback can be represented as a shift register. Without interleaving $Z_{cell} = Z_e + Z_{fo} = 15$ and $J = Z_{loop} = 12$. Moreover, let $S = 10$ whatever the level of interleaving, since this value is not influenced by interleaving. In this case Eq. 8 reduces to: $CUPS = M^3/(32M - 17)$. It is possible to achieve interleave 2 considering that $J = \max\{Z_e, Z_{fo}\} = 8$ so it is required to design the feedback loop to have a delay of 12 clock cycles. In this case Eq. 8 must be adapted considering $J = 8$ and $P = 2M$. Figure 25 shows the trend of CUPS in function of M for different levels of interleaving. It can be immediately noticed that applying interleaving results in significant benefits in terms of CUPS. Moreover, it is evident that benefits of applying interleaving saturate with deeper levels of interleaving.

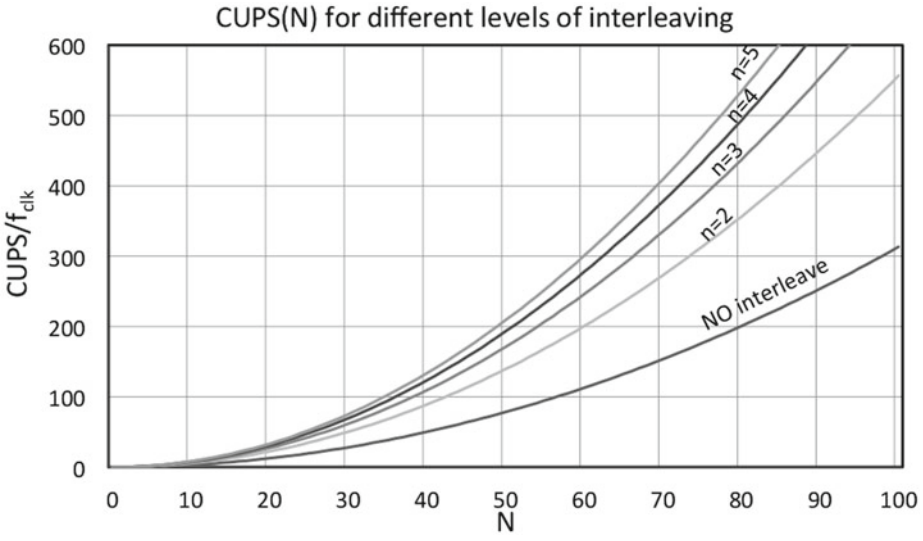


Fig. 25. The effect of *pipeline interleaving* on CUPS: given an array of $M \times M$ cells, increasing the level of interleaving CUPS increase as well.

5 Analysis of Applications Suited for NML Technology

Since the nature of Field-Coupled devices is quite different from traditional CMOS circuits, it is important to identify in which kind of applications the true potential of this technology can be fully exploited. The architectural overview presented in this chapter represents therefore the ideal starting point for such analysis. In the following the most important features of Field-Coupled devices are summarized.

1. **Parallelism.** To maximize performance parallel computation is required. As seen in Sect. 3 the interleaving of a big number of operations is required to maximize throughput. As a consequence all applications that are intrinsically parallel, or where it is necessary to process a huge amount of data are the ideal target of these devices.
2. **Locality.** A huge delay penalty is acquired by signals traveling over long interconnection wires. The consequence is simple: Only local interconnections should be used (see Sect. 4). Therefore all algorithms that require only a local exchange of data are favored in this technology.

Along with these two features that are common to all Field-Coupled devices, NML circuits, due to their magnetic nature, have some more specific advantage.

3. **Logic-in-memory.** Magnets are intrinsically memory devices. They are used here for logical computation, but since they maintain their states also without external power supply, they also provide a natural memory ability. This possibility can be exploited to build logic-in-memory devices, where memory and logic functions are combined together.
4. **No stand-by power consumption.** Since magnets maintain their state without power external supply, there is no power consumption at all when circuits are in stand-by mode. This fact can be exploited to reduce power consumption, activating the circuit only when it is necessary and shutting it down in the meantime.
5. **Radiation hard.** NML circuits are extremely resistant to radiations. Circuits still work correctly after absorbing a considerable amount of high energy radiation.
6. **Low dynamic power consumption.** The dynamic power consumption can be very low, depending on the clock system used.

Every application that can benefit from one of more of these characteristics can therefore be an ideal target for this technology. Clearly the more features are exploited, the more benefit it is possible to gain from this technology. A short (but necessary not complete) list of possible applications for NML technology is presented below.

Aerospace and Military. The near-immunity to radiations is a property that gives to NML technology and undeniable advantage over CMOS circuits.

Aerospace and military applications are therefore the ideal target for NML technology. In these fields there is an always increasing demand for radiation hard, heat resistant processing units that must work in hostile environments. With CMOS technology it is very difficult to match all of aforementioned characteristics, because circuits can be damaged by radiations. NML devices are the perfect solution to match these fields necessities. Moreover, the NML logic-in-memory ability paves the way for "on the flight" circuit reconfigurability that could enhance space and military applications with re-programmable hardware and run-time adaptive functionality. This feature enables the possibility to change the function or to repair part of the system if something went wrong.

Embedded Sensors in Remote Zones. One of the main characteristics offered by NML technology, the zero stand-by power consumption, can be exploited to built intelligent sensor stations in remote zones. In this kind of systems it is quite common to have circuits that need to operate for a limited period of time, then they have to spent the rest of the time in stand-by mode. These kind of systems are actually used in many applications; one of these is the environmental monitoring in remote zones such mountains peaks or buoys placed in the ocean, which detect the rate of pollution. The use of NML technology can give a great advantage compared to the devices actually used, mainly thanks to the zero stand-by power consumption, but also thanks to its resistance to extreme environmental conditions.

Automotive. The resistance to extreme environments can also be useful in automotive applications. It is well known that electronic components that must be placed near the mechanical parts of a car, like the engine, are exposed to a great deal of stress. This stress comes under many forms, heat, vibrations, voltage spikes, straining the resistance of CMOS circuits to their limits. Clearly the use of NML logic can represent a substantial improvement to the reliability of electronic circuits applied to the automotive field.

Integration with Other Magnetic Systems. While here magnets are used to build logic circuits, magnetic devices are commonly employed in many fields as sensors and actuators, not to mention as memory devices. Hard-disks are the leading devices for mass data storage, and in recent years Magnetic RAMs were also developed. These memories are gaining an increasing interest for their prominent features [60,61], like radiation resistance, high speed, non-volatile nature, heat resistance and nearly-unlimited endurance. Since all these devices share the same magnetic nature, it is possible to imagine the merging logic devices and MRAM together. This is particularly interesting because memory performance is fast becoming the key bottleneck that limits system performance, and critical applications are becoming more data-dependent, and less computation-dependent. Instant-on will be a requirement for a lot of applications. Merging NML circuits and MRAM can be a solution to this problem, a further way to exploit the logic-in-memory principle.

Nonetheless it is also good to consider that many embedded applications implicate the acquisition of signals from sensors, the processing of acquired signal, the storage of the elaborated data, the command of actuators and the communication with other system. A lot of sensors, actuators and memory devices are based on magnetism. This consideration can be exploited to obtain a deeper level of embedding. If, for example, magnetic sensors are constructed on the same chip as NML devices using the same fabrication processes, new scenarios will become possible. Indeed, one possibility is a new kind of a system on a chip with potential advantages in terms of both performance and power.

Dynamic Circuit Reconfigurability. The ability to mix logic and memory in the same device can be used to enhance the concept of reconfigurable logic circuit. For example it is possible to imagine computational algorithms that dynamically manage the logic and the memory according to the necessaries resources. As an example it is possible to consider an algorithm that works in two phases: the first one needs a large amount of computational power and a little memory; the second phase requires less computation but an increased quantity of memory. A dynamic reconfigurability coupled with the fact that both logic and memory are available in the same device can revolutionize the design of algorithms and their implementations in hardware. Another possibility is to build logic devices that can be reconfigured run-time, leading to a totally new class of "smart" circuit, obtaining considerable advantages in terms of circuit area, power consumption and speed.

Image Processing. The possible applications presented so far are particularly suited for NML logic. However any application requiring huge amounts of data may be the ideal target for all QCA technologies, not only for the magnetic version. An example is the image processing field where digital images, made by a big matrix of pixel, must be elaborated. The amount of data to process is very high, secondly the algorithms involved in the elaboration of an image are often applied to a pixel and its neighbors. Image processing is therefore an application particularly suited for QCA technology because it can exploit both the principle of *Parallelism* and the principle of *Locality*. Among all QCA implementations, NML logic is favored due to its low power consumption. It is possible to think, for example, of a dedicated co-processor coupled with the sensor of a mobile camera. In this case, NML logic would present a huge advantage over CMOS technology.

Informatics for Biotechnology. Similarly to image processing also the biotechnology field can gain a huge advantage from the use of QCA technology. The massive parallelism offered by QCA could be successfully exploited to develop hardware accelerators or specialized co-processors to execute specific heavy-computation tasks. One of these task that recently has obtained great interest is the modeling of molecular dynamics in the cellular membrane.

This kind of simulation is useful for the investigation of cellular effects of molecular alterations due to pathological conditions or to extend the characterization of other sub-cellular structures [62]. Recently a highly parallel implementation of molecular dynamics simulation was developed for general purpose GPUs in order to exploit the massive parallelism offered by such devices [63]. As shown in this GPU implementation, the massive parallelism is one of the major feature that can improve the computation in the biotechnology field.

Another biotechnology task that has recently attracted interest from many interdisciplinary research groups is the study of biological sequences and the relative development of tools. Biologists study the similarities between proteins to reconstruct phylogenetic trees and to assess the presence of mutations that lead to genetic diseases or tumors. Since proteins consist of long sequences of amino acids, the fastest way to performs a first analysis is to align the studied protein with the protein coming from huge databases searching for regions of similarity. After that, the short list of proteins that share a sufficient level of similarity are investigated in detail. There are many alignment algorithms that are chosen according to the type of analysis that the biologist wants to perform. One of the most used is the method developed by Smith and Waterman that performs, in dynamic programming, the exhaustive local alignment between two sequences [64]. We have designed a systolic architecture that accelerate the Smith-Waterman algorithm execution [65–67] mapping it to NML technology [68,69] demonstrating how much gain can be obtained from the use of NML technology.

6 Conclusions

In this chapter we have presented a thorough analysis of the main problems that arise at architectural level in Field-Coupled devices. The analysis is mainly based on NML logic but most of the results here presented are valid for all QCA implementations. For every problem presented a solution is proposed. The solutions described in this chapter are designed around both logic and technology constraints. Finally, a brief overview of applications that can fully exploit the true potential of this technology is presented. This chapter provides researchers and designers guidelines for the design of complex circuits with this technology, and provides directions for future development of this technology.

Now it is important to continue the architectural analysis focusing on the system level integration and all related problems. Important problems must be investigated, like interfacing with the outside world and the generation of clock signals. Particularly, in case of NML logic it is important to investigate how to reduce the necessity of CMOS transistors in the support circuits, like the input/output interfaces and clock waveforms generators. The necessity to use CMOS transistors in the external circuits represents a weakness because it reduces one of the most important advantages of this technology, the immunity to radiations.

We will therefore continue to investigate NML (and QCA) architectures following these guidelines, always keeping an eye at the technological constraints in order to get the most realistic results.

References

1. Lent, C., Tougaw, P., Porod, W., Bernstein, G.: Quantum cellular automata. Nanotechnology **4**, 49–57 (1993)
2. Csurgay, A., Porod, W., Lent, C.: Signal processing with near-neighborcoupled time-varying quantum-dot arrays. IEEE Trans. Circ. Syst. **47**(8), 1212–1223 (2000)
3. Lu, U., Lent, C.: Theoretical study of molecular quantum-dot cellular automata. J. Comput. Electron. (Springer) **4**, 115–118 (2005)
4. Imre, A., Ji, L., Csaba, G., Orlov, A.O., Bernstein, G., Porod, W.: Magnetic logic devices based on field-coupled nanomagnets. In: 2005 International Semiconductor Device Research Symposium, p. 25, December 2005
5. Pulimeno, A., Graziano, M., Abrardi, C., Demarchi, D., Piccinini, G.: A write-in system based on electric fields for molecular QCA. In: 2011 IEEE International NanoElectronics Conference (INEC), Tao-Yuan, Taiwan, pp. 1–2. IEEE, June 2011
6. Pulimeno, A., Graziano, M., Piccinini, G.: Molecule interaction for QCA computation. In: IEEE International Conference on Nanotechnology, pp. 1–5 (2012)
7. Pulimeno, A., Graziano, M., Demarchi, D., Piccinini, G.: Towards a molecular QCA wire: simulation of write-in and read-out systems. Solid State Electron. **77**, 101–107 (2012)
8. Pulimeno, A., Graziano, M., Saginario, A., Cauda, V., Demarchi, D., Piccinini, G.: Bis-ferrocene molecular QCA wire: ab-initio simulations of fabrication driven fault tolerance. IEEE Trans. Nanotechnol. **12**(4), 498–507 (2013)
9. Lent, C., Isaksen, B.: Clocked molecular quantum-dot cellular automata. IEEE Trans. Electron devices **50**(9), 1890–1896 (2003)
10. Porod, W.: Magnetic logic devices based on field-coupled nanomagnets. Nano & Giga (2007)
11. Pulimeno, A., Graziano, M., Piccinini, G.: UDSM trends comparison: from technology roadmap to UltraSparc Niagara2. IEEE Trans. VLSI Syst. **20**(7), 1341–1346 (2012)
12. Niemier, M., Hu, X., Alam, M., Bernstein, G., Porod, W., Putney, M., DeAngelis, J.: Clocking structures and power analysis for nanomagnet-based logic devices. In: International Symposium on Low Power Electronics and Design, Portland-Oregon, USA, pp. 26–31. IEEE (2007)
13. Niemier, M., et al.: Nanomagnet logic: progress toward system-level integration. J. Phys. Condens. Matter **23**, 34 (2011)
14. Csaba, G., Porod, W.: Behavior of nanomagnet logic in the presence of thermal noise. In: International Workshop on Computational Electronics, Pisa, Italy, pp. 1–4. IEEE (2010)
15. Graziano, M., Chiolerio, A., Zamboni, M.: A Technology Aware Magnetic QCA NCL-HDL Architecture, Italy, Genova, pp. 763–766. IEEE (2009)
16. Chiolerio, A., Allia, P., Graziano, M.: Magnetic dipolar coupling and collective effects for binary information codification in cost-effective logic devices. J. Magn. Magn. Mater. **324**(19), 3006–3012 (2012)

17. Graziano, M., Vacca, M., Zamboni, M.: Magnetic QCA design: modeling, simulation and circuits. Cellular Automata - Innovative Modelling for Science and Engineering, InTech (2011). http://www.intechopen.com/articles/show/title/magnetic-qca-design-modeling-simulation-and-circuits
18. Graziano, M., Vacca, M., Chiolerio, A., Zamboni, M.: A NCL-HDL snake-clock based magnetic QCA architecture. IEEE Trans. Nanotechnol. **10**(5), 1141–1149 (2011)
19. Alam, M., Siddiq, M., Bernstein, G., Niemier, M., Porod, W., Hu, X.: On-chip clocking for nanomagnet logic devices. IEEE Trans. Nanotechnol. **9**(3), 348–351 (2009)
20. Das, J., Alam, S., Bhanja, S.: Low power magnetic quantum cellular automata realization using magnetic multi-layer structures. J. Emerg. Sel. Top. Circ. Syst. **1**(3), 267–276 (2011)
21. Das, J., Alam, S., Bhanja, S.: Ultra-low power hybrid CMOS-magnetic logic architecture. Trans. Comput. Syst. **59**(9), 2008–2016 (2011)
22. Karunaratne, D., Bhanja, S.: Study of single layer and multilayer nano-magnetic logic architectures. J. Appl. Phys. **111**, 07A928 (2012)
23. Vacca, M., Crescenzo, L., Graziano, M., Zamboni, M., Chiolerio, A., Lamberti, A., Enrico, E., Celegato, F., Tiberto, P., Boarino, L.: Electric clock for NanoMagnet logic circuits. In: Field Couple Computing Workshop (FCN) (2013)
24. Fashami, M.S., Atulasimha, J., Bandyopadhyay, S.: Magnetization dynamics, throughput and energy dissipation in a universal multiferroic nanomagnetic logic gate with fan-in and fan-out. Nanotechnology **23**(10), 105201 (2012)
25. Rizos, N., Omar, M., Lugli, P., Csaba, G., Becherer, M., Schmitt-Landsiedel, D.: Clocking schemes for field coupled devices from magnetic multilayers. In: International Workshop on Computational Electronics, Beijin, China, pp. 1–4. IEEE (2009)
26. Becherer, M., Kiermaier, J., Csaba, G., Rezgani, J., Yilmaz, C., Osswald, P., Lugli, P., Schmitt-Landsiedel, D.: Characterizing magnetic field-coupled computing devices by the Extraordinary Hall-effect. In: Proceedings European Solid State Device Research Conference, Athens, Greece, pp. 105–108. IEEE (2009)
27. Ju, X., Niemier, M., Becherer, M., Porod, W., Putney, M., Lugli, P., Csaba, G.: Systolic pattern matching hardware with out-of-plane nanomagnet logic devices. IEEE Trans. Nanotechnol. **12**(3), 376–386 (2013)
28. Ercan, I., Anderson, N.: Heat Dissipation Bounds for Nanocomputing: Theory and Application to QCA (2011)
29. Vacca, M., Graziano, M., Zamboni, M.: Nanomagnetic logic microprocessor: hierarchical power model. IEEE Trans. VLSI Syst. **21**(8), 1410–1420 (2012)
30. Vacca, M., Graziano, M., Zamboni, M.: Majority voter full characterization for nanomagnet logic circuits. IEEE Trans. Nanotechnol. **11**(5), 940–947 (2012)
31. Niemier, M., Varga, E., Bernstein, G., Porod, W., Alam, M., Dingler, A., Orlov, A., Hu, X.: Shape engineering for controlled switching with nanomagnet logic. IEEE Trans. Nanotechnol. **11**(2), 220–230 (2012)
32. Graziano, M., Vacca, M., Blua, D., Zamboni, M.: Asynchrony in quantum-dot cellular automata nanocomputation: elixir or poison? IEEE Des. Test Comput. **28**(5), 72–83 (2011)
33. Niemier, M., Kogge, P.: Problems in designing with QCAs: Layout = Timing. Int. J. Circ. Theor. Appl. **29**, 49–62 (2001)
34. Vacca, M., Frache, S., Graziano, M., Zamboni, M.: ToPoliNano: a synthesis and simulation tool for NML circuits. In: IEEE International Conference on Nanotechnology, pp. 1–6, August 2012

35. Frache, S., Chiabrando, D., Graziano, M., Riente, F., Turvani, G., Zamboni, M.: ToPoliNano: nanoarchitectures design made real. In: IEEE International Symposium on Nanoscale Architectures (NANOARCH), pp. 160–167 (2012)
36. Frache, S., Chiabrando, D., Graziano, M., Vacca, M., Boarino, L., Zamboni, M.: Enabling design and simulation of massive parallel nanoarchitectures. J. Parallel Distrib. Comput. **74**(6), 2530–2541 (2013)
37. Vacca, M., Frache, S., Graziano, M., Crescenzo, L., Cairo, F., Zamboni, M.: Automatic Place&Route of Nano-magnetic Logic circuits, pp. 58–63, July 2013
38. Vacca, M., Vighetti, D., Mascarino, M., Amaru, L., Graziano, M., Zamboni, M.: Magnetic QCA majority voter feasibility analysis. In: 2011 7th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME), pp. 229–232 (2011)
39. Fant, K., Brandt., S.: NULL convention logic^{TM}, a complete and consistent logic for asynchronous digital circuit synthesis. In: International Conference on Application Specific Systems, Chicago-Illinois, USA, pp. 261–273. IEEE (1996)
40. Choi, M., Patitz, Z., Jin, B., Tao, F., Park, N.: Designing layout-timing independent quantum-dot cellular automata (QCA) circuits by global asynchrony. J. Syst. Architect. Elsevier **53**, 551–567 (2007)
41. Vacca, M., Graziano, M., Zamboni, M.: Asynchronous solutions for nano-magnetic logic circuits. ACM J. Emerg. Tech. Comp. Syst. **7**(4), 15 (2011)
42. Mentor Graphics. http://www.modelsim.com
43. Awais, M., Vacca, M., Graziano, M., Masera, G.: Quantum dot cellular automata check node implementation for LDPC decoders. IEEE Trans. Nanotechnol. **12**(3), 368–377 (2013)
44. Awais, M., Vacca, M., Graziano, M., Masera, G.: FFT implementation using QCA. In: 2012 19th IEEE International Conference on Electronics, Circuits and Systems (ICECS), pp. 741–744 (2012)
45. Martina, M., Masera, G.: Turbo NOC: a framework for the design of network-on-chip-based turbo decoder architectures. IEEE Trans. Circ. Syst. I **57**(10), 2776–2789 (2010)
46. Kung, H., Leiserson, C., Science, C.M.U.D.o.C.: Systolic Arrays for (VLSI). CMU-CS. Carnegie-Mellon University, Department of Computer Science (1978)
47. Haron, N., Hamdioui, S.: Why is CMOS scaling coming to an END? In: 3rd International Design and Test Workshop, IDT 2008, pp. 98–103, December 2008
48. Pan, S.B., Park, R.H.: Unified systolic arrays for computation of the DCT/DST/DHT. IEEE Trans. Circ. Syst. Video Technol. **7**(2), 413–419 (1997)
49. Panchanathan, S., Goldberg, M.: A systolic array architecture for image coding using adaptive vector quantization. IEEE Trans. Circ. Syst. Video Technol. **1**(2), 222–229 (1991)
50. Iyengar, G., Panchanathan, S.: Systolic array architecture for Gabor decomposition. IEEE Trans. Circ. Syst. Video Technol. **5**(4), 355–359 (1995)
51. Lim, H., Swartzlander, E.J.: Multidimensional systolic arrays for the implementation of discrete Fourier transforms. IEEE Trans. Signal Process. **47**(5), 1359–1370 (1999)
52. Herzberg, H., Haimi-Cohen, R.: A systolic array realization of an LMS adaptive filter and the effects of delayed adaptation. IEEE Trans. Signal Process. **40**(11), 2799–2803 (1992)
53. Chang, L.W., Wu, M.C.: A unified systolic array for discrete cosine and sine transforms. IEEE Trans. Signal Process. **39**(1), 192–194 (1991)
54. Buyukkurt, B., Najj, W.: Compiler generated systolic arrays for wavefront algorithm acceleration on FPGAs. In: International Conference on Field Programmable Logic and Applications, FPL 2008, pp. 655–658, September 2008

55. Jin, W., Zhang, C., Li, H.: Mapping multiple algorithms into a reconfigurable systolic array. In: Canadian Conference on Electrical and Computer Engineering, CCECE 2008, pp. 001187–001192, May 2008
56. Lu, L., Liu, W., O'Neill, M., Swartzlander, E.: QCA systolic matrix multiplier. In: 2010 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), pp. 149–154, July 2010
57. Lu, L., Liu, W., O'Neill, M., Swartzlander, J.E.: QCA systolic array design. IEEE Trans. Comput. **62**(3), 548–560 (2013)
58. Crocker, M., Hu, X., Niemier, M.: Design and Comparison of NML Systolic Architectures. Nanoarch (2010)
59. Causapruno, G.: Analysis and optimization of parallel processing architectures for nanotechnologies. Master's thesis, Politecnico di Torino, November 2012
60. Rajaram, S., Karunaratne, D., Sarkar, S., Bhanja, S.: Study of dipolar neighbor interaction on magnetization states of nano-magnetic disks. IEEE Trans. Magn. **49**(7), 3129–3132 (2013)
61. Panchumarthy, R., Karunaratne, D., Sarkar, S., Bhanja, S.: Magnetic state estimator to characterize the magnetic states of nano-magnetic disks. IEEE Trans. Magn. **49**(7), 3545–3548 (2013)
62. Deriu, M.A., Shkurti, A., Paciello, G., Bidone, T.C., Morbiducci, U., Ficarra, E., Audenino, A., Acquaviva, A.: Multiscale modeling of cellular actin filaments: from atomistic molecular to coarse-grained dynamics. Proteins Struct. Funct. Bioinf. **80**(6), 1598–1609 (2012)
63. Shkurti, A., Orsi, M., Macii, E., Ficarra, E., Acquaviva, A.: Acceleration of coarse grain molecular dynamics on GPU architectures. J. Comput. Chem. **34**(10), 803–818 (2013)
64. Smith, T., Waterman, M.: Identification of common molecular subsequences. Mol. Biol. **147**, 195–197 (1981)
65. Urgese, G.: Analysis and design of an optimized HW accelerator for protein alignment. Master's thesis, Politecnico di Torino (2012)
66. Graziano, M., Frache, S., Zamboni, M.: A hardware viewpoint on biosequence analysis: what's next? ACM J. Emerg. Technol. Comput. Syst. **9**(4), 29 (2013)
67. Urgese, G., Graziano, M., Vacca, M., Awais, M., Frache, S., Zamboni, M.: Protein alignment HW/SW optimizations. In: The IEEE International Conference on Electronics, Circuits, and Systems (ICECS), pp. 145–148 (2012)
68. Wang, J.: Emerging technologies for biosequence analysis. Master's thesis, Politecnico di Torino (2012)
69. Wang, J., Vacca, M., Graziano, M., Zamboni, M.: Biosequences analysis on Nano-Magnet Logic. In: International Conference on IC Design and Technology, pp. 131–134, May 2013

Modeling and Simulation

Modelling Techniques for Simulating Large QCA Circuits

Faizal Karim(✉) and Konrad Walus

The University of British Columbia, Vancouver, BC, Canada
faizalk@ece.ubc.ca

1 Introduction

In the past several years, incredible advances in the availability of nano fabrication processes have been witnessed, and have demonstrated molecular-scale production beyond the usable limit for CMOS process technology [1–3]. This has led to the research and early development of a wide-range of novel computing paradigms at the nanoscale; amongst them, quantum dot cellular automata (QCA) [4]. QCA is a nanoelectronic computing paradigm in which an array of cells, each electrostatically interacting with its neighbors, is employed in a locally interconnected manner to implement general purpose digital circuits [4]. Several proof-of-concept QCA devices have been fabricated using silicon-on-insulator (SOI) [5], metallic island devices operating in the Coulomb blockade regime [6–12], and nano-magnetics [13–18]. In recent years, research into implementing these devices using single molecules has also begun to generate significant interest [1,19–25], and most recently, it was demonstrated that silicon atom dangling bonds (DBs), on an otherwise hydrogen terminated silicon crystal surface, can serve as quantum dots [1].

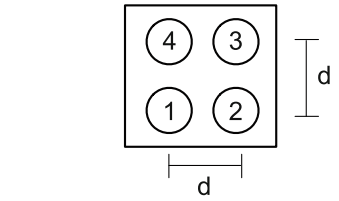
Research into QCA has been motivated by the potential of reaching the limits to physical scaling, as well as the predictions of ultra low power consumption, a major problem in the industry [4,11,20,26–28]. It has been reported that molecular and atomic implementations can achieve device densities on the order of $10^{14}/\text{cm}^2$ (for $1\,\text{nm}^2$ devices) [29]. The power dissipated from current-switched devices such as FETs operating at GHz speeds could melt the chip at those densities [30], however, molecular QCA has been predicted to reduce the power dissipation by several orders of magnitude [26,31]. Additionally, as the size of the cells gets smaller, the interaction energies between cells increases. At the molecular (and atomic) scale, these energies are expected to be in the 0.2–0.5 eV range [1,4,32,33] which allow for room temperature operation since these energies are greater than the thermal ambient energy (i.e, thermal noise), k_BT ($\sim$25 meV at room temperature), where k_B is Boltzmann's constant and T is the temperature in kelvin ($T = 293\,\text{K}$ at room temperature). However, while the development of such devices is promising, a number of technical challenges, including choice of molecules [19,25], the design of proper interfacing mechanisms [34], and clocking technology [35,36] - amongst other things - remain to

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 259–273, 2014.
DOI: 10.1007/978-3-662-43722-3_11, © Springer-Verlag Berlin Heidelberg 2014

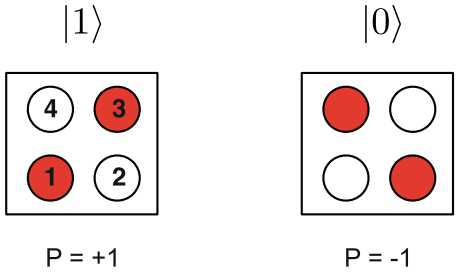
be solved before QCA devices can be implemented. Thus, studying the dynamics of QCA systems has remained mainly theoretical up until now.

A great deal of theoretical and modelling work has been done on the topic of QCA [37–41]. This research has aimed to capture the qualitative and quantitative characteristics of QCA cells and of arrays of cells. Because of the difficulties inherent in solving the complete quantum mechanical problem, a number of simplifying assumptions are typically made. These include a reduction of the Hilbert space to two states per cell [39], treatment of intercell interactions via a mean-field approach [37,38], and finally an assumption of exponential energy relaxation [26,40,42]. In this chapter, we provide a thorough review of these approximations and discuss the current methods for modelling the dynamics of QCA circuits.

Simulation of QCA Systems. In the proposed QCA implementations, a QCA cell will feature 4–6 quantum dots and two mobile electrons [1,20,43]. A schematic diagram of a single QCA cell is shown in Fig. 1(a). This figure shows a cell consisting of four quantum dots arranged in a square pattern. The two basis states of the QCA cell considered in this work is also shown. While cells containing five or six dots have shown to improve the behaviour of QCA devices [20,28,35], it greatly increases the numerical complexity of the cell model, and thus we limit the scope of this chapter to four-dot cells.



(a) 4-dot QCA Cell



(b) Bistable QCA states

Fig. 1. Schematic of the basic four-site cell. (a) The geometry of the cell. The inter-dot distance is designated by d. (b) Coulombic repulsion causes the electrons to align along the diagonals of the cell. These two bistable states result in cell polarizations of $P = +1$ and $P = -1$.

The two mobile electrons are free to tunnel between adjacent dots, however, tunneling out of the cell is assumed to be completely suppressed.

1.1 Full-Basis Quantum Mechanical Treatment

For the single four-dot cell considered in this chapter (with two electrons of opposite spin), there are a total of sixteen underlying basis vectors. Modelling a single cell within the full sixteen-dimensional Hilbert space makes it possible to capture the full dynamics of the cell, and represents the most complete way of modelling a QCA cell [39]. To maintain the full many-electron degrees of freedom of an N-cell system however (including the correlation and exchange effects between cells), one must treat the entire N-cell system as a single quantum system by constructing a new basis set consisting of the direct-product of all combinations of the single-cell basis vectors. Thus, a complete basis set for an N-cell system would require 16^N basis vectors. A $16^N \times 16^N$ Hubbard-type Hamiltonian for this system can be constructed using the standard second-quantized notations [39],

$$\begin{aligned}\hat{H} = & \sum_{i,\sigma,m} E_0 \hat{n}_{i,\sigma}(m) - t_{i,j} \sum_{i>j,\sigma,m} (\hat{a}_i^{\dagger}(m)\hat{a}_j(m) + \hat{a}_j^{\dagger}(m)\hat{a}_i(m)) \\ & + \sum_{i,m} E_Q \hat{n}_{i,\uparrow}(m)\hat{n}_{i,\downarrow}(m) + \sum_{i>j,\sigma,\sigma',m} V_Q \frac{\hat{n}_{i,\sigma}(m)\hat{n}_{j,\sigma'}(m)}{|r_i(m) - r_j(m)|} \\ & + \sum_{i>j,\sigma,\sigma',k>m} V_Q \frac{\hat{n}_{i,\sigma}(m)\hat{n}_{j,\sigma'}(k)}{|r_i(m) - r_j(k)|}, \end{aligned} \tag{1}$$

where the operator $\hat{a}_{i,\sigma}(m)$ ($\hat{a}_{i,\sigma}^{\dagger}(m)$) annihilates (creates) an electron in the i^{th} site of cell m with spin σ, the operator $\hat{n}_{i,\sigma}(m) \equiv \hat{a}_{i,\sigma}^{\dagger}(m)\hat{a}_{i,\sigma}(m)$ is the number operator for the i^{th} site of cell m, and $V_Q = q_e^2/(4\pi\epsilon)$ is a constant, where q_e is the charge of the electron and ϵ the electrical permittivity of the medium. The first term in Eq. (1) represents the on-site energy of a dot. The second term describes the electron tunnelling between neighbouring sites i and j within a cell, m. The third term in Eq. (1) accounts for the energetic cost of putting two electrons of opposite spin at the same site, and the final two terms are related to the Coulombic interactions between electrons in the same cell and in neighbouring cells, respectively. The polarization of each cell can be found by evaluating,

$$P_m = \frac{(\rho_1^m + \rho_3^m) - (\rho_2^m - \rho_4^m)}{\rho_1^m + \rho_2^m + \rho_3^m + \rho_4^m}, \tag{2}$$

where ρ_i^m is the expectation value of the number operator on the i^{th} site of cell m, i.e., $\rho_i^m = \langle \hat{n}_i(m) \rangle$. Using this treatment, it has been shown that the dynamics of small arrays can indeed be solved for directly while retaining the full many-electron degrees of freedom [39]. However, the exponential growth in

the basis set makes it computationally prohibitive to model any device larger than just a few cells. Thus, further approximations are typically required for modelling larger arrays of QCA cells.

1.2 Two-State Approximation

The first of such approximations is the two-state approximation. As shown in [39], the ground state of a single cell remains almost completely contained within a two-dimensional subspace of the full sixteen-dimensional Hilbert space. For simplicity, we will consider idealized cells with saturation polarizations of ± 1. A more realistic treatment would have the polarizations spanning a slightly smaller range, however the basic argument that follows would be unchanged. We therefore refer to our reduced basis as the polarization basis, and denote the two states as $|0\rangle$ and $|1\rangle$, as shown in Fig. 1. The Hamiltonian in the polarization basis for a system of N interacting QCA cells, under the influence of driver cells, is described by a $2^N \times 2^N$ Ising-like Hamiltonian:

$$\hat{H} = -\sum_{i=1}^{N} \gamma_i \hat{\sigma}_x(i) - \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} E_k^{i,j} \hat{\sigma}_z(i) \hat{\sigma}_z(j) + \frac{1}{2} \sum_{i=1}^{N} \sum_{D \in \Omega_i} E_k^{i,D} P_D \hat{\sigma}_z(i), \quad (3)$$

where Ω_i represents the neighbourhood of driver cells that have appreciable effect on cell i, and P_D is the polarization of those cells. The driver cells have polarizations that can range from -1 to $+1$. The tunnelling energy, γ, takes the place of the transverse magnetic field in the Ising spin model, and is determined from the nature of the potential barriers between the dots in the cell. When the potential barriers are raised, γ is small, and the cells are allowed to become polarized (i.e., take on one of the antipodal configurations). When the potential barriers are lowered, the opposite is true, and the electrons become delocalized. $E_k^{i,j}$ is the kink energy between cells, and is the energetic cost of two cells having opposite polarization. In general, the interaction between cells can be described by a quadrupole-quadrupole interaction in which the kink energy decreases as the fifth power of the distance between cells. Thus it is typical to only consider nearest-neighbour coupling within the Hamiltonian, as the kink energy between next-to-nearest neighbours is reduced by a factor of $\frac{1}{32}$. The kink energy is analogous to the coupling constant, J, in the Ising spin model. The Pauli operators, $\hat{\sigma}_a(i)$; $a = x, y, z$, represent the tensor product of N 2×2 identity operators, with the i^{th} identity operator replaced by the usual Pauli matrix shown below. For example, $\hat{\sigma}_y(2) \equiv \mathbb{1} \otimes \hat{\sigma}_y \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1}$. The polarization of a cell, i, is defined as $P_i = -\langle \hat{\sigma}_z(i) \rangle$.

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{\sigma}_y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \hat{\sigma}_z = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

As shown in [39], the ground state of a single cell within the full sixteen-dimensional Hilbert space remains almost completely contained within the

two-dimensional subspace of the two-state basis vectors. For sufficiently low temperatures, we can therefore expect the state of a cell or a line to be well described within the two-state approximation. However, while the two-state approximation offers significant computational savings over the full basis quantum mechanical treatment described in the previous section, the exponential growth in the basis set still limits its application to systems containing only a small number of interacting cells. A thirty-cell system, for instance, would yield over a billion basis vectors. In practice, QCA circuits will likely require millions of cells. A simple 4-bit processor, for example, was designed using almost 30,000 cells [44]. Thus, further approximations are required in order to reduce the computational complexity of modeling large QCA systems.

1.3 Intercellular Hartree Approximation

In the full quantum mechanical treatments discussed above, arrays of QCA cells are being solved for directly while retaining the full many-electron degrees of freedom. The inclusion of all of these additional variables contributes to the exponential growth in the basis set as the system grows with the number of cells, N. Thus, in order to mitigate this growth, some quantum degrees of freedom must be removed during calculations. One such approach is to use the intercellular Hartree approximation (ICHA), for which the effects arising from quantum correlations and exchange are fully neglected [37,38]. In this treatment, an N-cell system is decoupled into a set of N single-cell subsystems that are assumed to interact classically through the expectation values of their polarization without any quantum mechanical coherence between neighboring cells. Using this approximation in conjunction with the two-state approximation, it is only necessary to diagonalize N 2×2 Hamiltonians as opposed to one $2^N \times 2^N$ Hamiltonian. This allows large circuits, which would otherwise be intractable using a full quantum mechanical model, to be simulated on a classical computer. The Hamiltonian (in the polarization basis) for a single cell, i, is simply,

$$\hat{H}_i = -\gamma_i \hat{\sigma}_x + \frac{1}{2} \sum_{j \in \Omega} E_k^{i,j} P_j \hat{\sigma}_z, \tag{4}$$

where P_j is the polarization of cell j, and is found by evaluating, $P_j = -\langle \sigma_z \rangle$. Because the solutions of one cell's Hamiltonian define parameters that enter its neighbours' Hamiltonians, the system of coupled Schrödinger equations must be solved iteratively to obtain self-consistency. If a circuit evolves adiabatically and cells remain at the ground state, then this problem simplifies even further by recognizing that the polarization of any cell, i, can be evaluated numerically using [45],

$$P_i = \frac{\frac{1}{2\gamma} \sum\limits_j E_k^{i,j} P_j}{\sqrt{1 + \left(\frac{1}{2\gamma} \sum\limits_j E_k^{i,j} P_j \right)^2}}, \tag{5}$$

which results in the well-known nonlinear cell-to-cell response function shown in Fig. 2. Thus the ICHA carries with it extremely low computational complexity, making it very desirable when solving large arrays of QCA cells. However, while the ICHA is generally capable of arriving at the correct ground state of an array of QCA cells in a single clocking zone with a fixed driver, it has been shown that it does not always arrive at the correct ground state for many other types of circuits [40,41]. Furthermore, the ICHA is not capable of predicting the dynamic behavior of an array, as acknowledged in [39]. In order to calculate the dynamic response, or finite-temperature behavior, of a QCA array, the inclusion of the many-cell excited states is required. Thus, while the ICHA offers a great deal of simplicity and low computational overhead, it represents, in a sense, the minimum inclusion of quantum mechanical effects in that correlations are completely ignored, except within the Hilbert space of each cell. It should be noted however, that for such cases where the tunnelling rates are very slow ($\gtrsim 100\,\mu s$), we witness a crossover from a regime where quantum tunnelling is important to one where electrons are completely localized. The normal relaxation dynamics are suppressed because the system is strongly coupled to the environment through σ_z, which has the effect of localizing charge [46]. In such cases, the ICHA is capable of predicting the correct ground state of a QCA system. However, in molecular and atomic implementations, it is likely that the range of tunnelling rates that allows this crossover from quantum dynamics to classical dynamics will not be achievable. For these reasons, it is useful to develop other techniques if we aim to accurately model the time-dependance of many-cell devices.

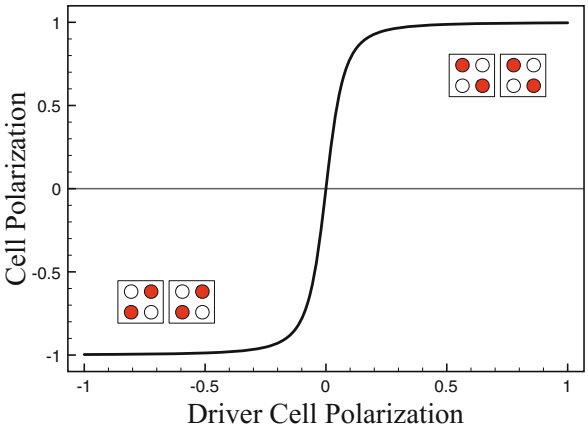


Fig. 2. Nonlinear cell-to-cell response function. The output cell is almost completely polarized for even a small polarization.

2 Coherent Cell-Cell Dynamics

By ignoring correlations and coherence, the ICHA can often fail to predict the correct ground state and dynamics of QCA systems. However, as previously discussed, including both coherence and correlation results in a non-polynomial expansion of the basis set - making solving large systems computationally intractable. The problem lies in the fact that the ICHA uses only the single cell operators (the Pauli spin matrices) to derive its solutions, which on their own, do not carry enough information to accurately predict the ground state of most QCA systems. The full-basis two-state approximation, on the other hand, uses all N cell operators when arriving at its solution, which provides us with more information than we need. Thus, what we seek is an intermediate treatment which uses only the bare-minimum number of operators required to obtain an accurate ground state solution for a QCA system. Attempting to develop such a treatment in the Schrödinger picture will prove cumbersome, and thus it will benefit us to formulate our problem in the Heisenberg picture, in which the operators incorporate a dependency on time, but the state vectors are time-independent. In the Heisenberg picture, we can separate the state variables of our system into groups corresponding to the state of the individual cells (single-cell operators) and into groups corresponding to the two-cell, three-cell, etc., operators [40]. We can then remove any higher-order operators deemed negligible to the system's evolution, and thus reduce the overall computational requirements for solving the ground state of a QCA system. This method is known as the Coherence Vector formulation, and allows us to easily compute the dynamics of a QCA system as well.

The coherence vector can be constructed by first recognizing that the density operator in Hilbert space of dimension n, like any Hermitian operator, can be represented by the $s = n^2 - 1$ generating (basis) operators of $SU(n)$ [47]. For an N cell system (i.e., a system with dimension $n = 2^N$), the density operator $\hat{\rho}$ can be represented as [47]:

$$\hat{\rho} = \frac{1}{2^N}\hat{1} + \frac{1}{2^N}\sum_{i=1}^{s} \lambda_i \hat{\lambda}_i, \tag{6}$$

where $\hat{\lambda}_i$ represents the i^{th} generating operator of $SU(2^N)$, and the elements λ_i are defined by

$$\lambda_i = \langle \hat{\lambda}_i \rangle = \mathrm{Tr}\{\hat{\rho}\hat{\lambda}_i\}, \tag{7}$$

and are the components that form the coherence vector. For any system of dimension $n = 2^N$, the $\hat{\lambda}_i$ basis operators take the form

$$\hat{\lambda}_i = \hat{\Lambda}(1) \otimes \hat{\Lambda}(2) \otimes \cdots \otimes \hat{\Lambda}(N), \tag{8}$$

where each term of the direct product can be any of the generators of the $SU(2)$ group - namely the $\hat{\sigma}_x$, $\hat{\sigma}_y$, and $\hat{\sigma}_z$ Pauli spin matrices - and the unit matrix:

$$\hat{\Lambda}(i) = \begin{cases} \hat{1} \\ \hat{\sigma}_x \\ \hat{\sigma}_y \\ \hat{\sigma}_z \end{cases} \tag{9}$$

The special case for which $\hat{\lambda}_i = \hat{1} \otimes \hat{1} \otimes \cdots \otimes \hat{1}$, is not allowed. Thus, there are a total of $s = 4^N - 1$ $(= 2^{2N} - 1)$ possible combinations, which is consistent with our earlier definition of s.

For a simple two-cell system, i.e., $N = 2$, the $\hat{\lambda}_i$ basis operators are shown in Fig. 3.

$\hat{\sigma}_x(1)$	$\hat{\sigma}_y(1)$	$\hat{\sigma}_z(1)$	Single-Cell Operators
$\hat{\sigma}_x(2)$	$\hat{\sigma}_y(2)$	$\hat{\sigma}_z(2)$	
$\hat{\sigma}_x(1)\hat{\sigma}_x(2)$	$\hat{\sigma}_x(1)\hat{\sigma}_y(2)$	$\hat{\sigma}_x(1)\hat{\sigma}_z(2)$	
$\hat{\sigma}_y(1)\hat{\sigma}_x(2)$	$\hat{\sigma}_y(1)\hat{\sigma}_y(2)$	$\hat{\sigma}_y(1)\hat{\sigma}_z(2)$	Two-Cell Operators
$\hat{\sigma}_z(1)\hat{\sigma}_x(2)$	$\hat{\sigma}_z(1)\hat{\sigma}_y(2)$	$\hat{\sigma}_z(1)\hat{\sigma}_z(2)$	

Fig. 3. The $s = 15$ basis operators for a two-cell system.

There are three single-cell operators for each of the two cells, and nine two-cell operators. Again, the expectation values of the $s = 15$ basis operators are the components that make up the coherence vector:

$$\boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}(1) \\ \boldsymbol{\lambda}(2) \\ \boldsymbol{K}(1,2) \end{bmatrix}, \tag{10}$$

where $\boldsymbol{\lambda}(1)$ and $\boldsymbol{\lambda}(2)$ are single-cell coherence vectors containing the expectation values of the single-cell operators, and $\boldsymbol{K}(1,2)$ is the two-cell coherence vector containing the two-point correlation terms (i.e., the expectation values of the two-cell operators). The single- and two-cell coherence vectors are shown below,

$$\begin{aligned} \boldsymbol{\lambda}(1) &= [\lambda_x(1)\ \lambda_y(1)\ \lambda_z(1)]^T, \\ \boldsymbol{\lambda}(2) &= [\lambda_x(2)\ \lambda_y(2)\ \lambda_z(2)]^T, \\ \boldsymbol{K}(1,2) &= [K_{xx}(1,2)\ K_{xy}(1,2)\ K_{xz}(1,2) \\ &\quad K_{yx}(1,2)\ K_{yy}(1,2)\ K_{yz}(1,2) \\ &\quad K_{zx}(1,2)\ K_{zy}(1,2)\ K_{zz}(1,2)]^T. \end{aligned} \tag{11}$$

Once the coherence vector has been computed, the polarization of the i^{th} cell can be found by evaluating, $P_i = -\lambda_z(i)$. It is worthwhile to point out that $K_{ab}(i,j) = K_{ba}(j,i)$ $(a,b = x,y,z)$, and as such, in the above example, it is not necessary to compute the two-cell coherence vector $\boldsymbol{K}(2,1)$. If we increase the size of our system to three cells, i.e., $N = 3$, then our coherence vector would expand to include the expectation values for the additional basis operators of our system, i.e.,

$$\boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}(1) \\ \boldsymbol{\lambda}(2) \\ \boldsymbol{\lambda}(3) \\ \boldsymbol{K}(1,2) \\ \boldsymbol{K}(1,3) \\ \boldsymbol{K}(2,3) \\ \boldsymbol{K}(1,2,3) \end{bmatrix}, \tag{12}$$

where $\boldsymbol{K}(1,2,3)$ is the three-cell coherence vector containing the 27 three-cell correlation terms (i.e., the expectation values of the three-cell operators). Note that there are now three two-cell coherence vectors corresponding to the three unique two-cell groupings possible within a three cell circuit. The increasing number of unique two-cell, three-cell, etc, groupings as the circuit grows is what leads to the exponential growth in the coherence vector. While this exponential growth makes solving larger systems intractable, it has been shown in [40] that the contribution of the higher-order (i.e., three-cell, four-cell, etc.) correlation terms in the coherence vector are negligible in calculating the dynamics of QCA systems. Thus, if we remove these higher-order correlation terms, then our coherence vector grows according to $s = 4.5N^2 - 1.5N$, which offers a much more manageable growth in state variables, and is far more suitable for solving QCA systems with a large number of cells. If we limit the scope of our Hamiltonian to simply nearest-neighbour (NN) coupling, then the number of state variables in our system can be reduced even further to $s = 12N - 9$. Comparatively, the number of state variables when modelling a QCA system using the ICHA grows according to $s = 3N$. Thus, by considering only two-cell correlations and nearest-neighbour coupling, we can achieve the accuracy of the full-basis two-state approximation, but with the computational complexity of the ICHA.

A comparison of the number of state variables being solved for in each approximation is shown in Fig. 4.

2.1 The Liouville Equation

In the Heisenberg picture of quantum mechanics, the state vector, $|\psi\rangle$, does not change with time, and an observable, A, satisfies,

$$\frac{d}{dt}\hat{A}(t) = \frac{i}{\hbar}\left[\hat{H}, \hat{A}\right], \tag{13}$$

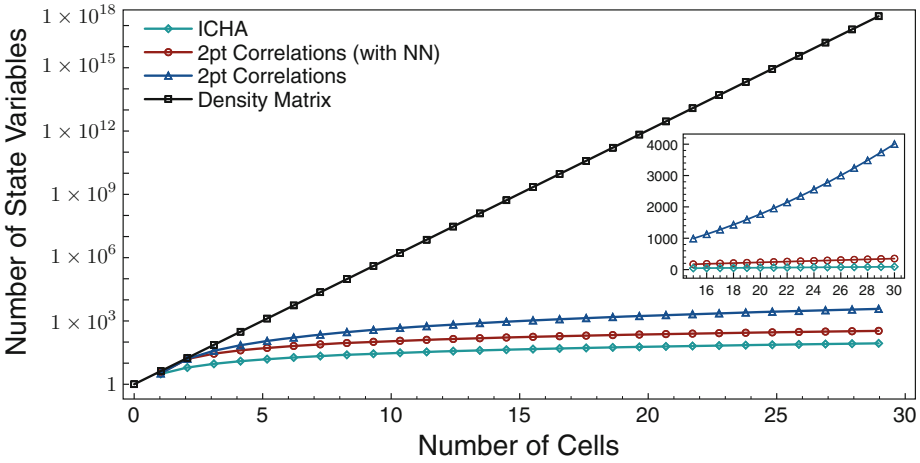


Fig. 4. The number of state variables that need to be solved as a function of the number of cells in the circuit. Plots are shown for the different quantum mechanical treatments discussed.

where $\hat{H}$ is the Hamiltonian and $[\cdot,\cdot]$ is the commutator of $\hat{H}$ and $\hat{A}(t)$. Equation (13) is known as the Liouville equation of motion. By using this equation and substituting in the Hamiltonian defined in Eq. (3), we can derive a set of ordinary differential equations (ODE) describing the evolution of each of the basis operators of our system, i.e.,

$$\begin{aligned}
\frac{d}{dt}\hat{\sigma}_x(j) &= \frac{i}{\hbar}\left[\hat{H}, \hat{\sigma}_x(j)\right], \\
\frac{d}{dt}\hat{\sigma}_y(j) &= \frac{i}{\hbar}\left[\hat{H}, \hat{\sigma}_y(j)\right], \\
&\vdots \\
\frac{d}{dt}\hat{\sigma}_x(j)\hat{\sigma}_y(k) &= \frac{i}{\hbar}\left[\hat{H}, \hat{\sigma}_x(j)\hat{\sigma}_y(k)\right], \\
&\vdots \\
\frac{d}{dt}\hat{\sigma}_z(j)\hat{\sigma}_z(k) &= \frac{i}{\hbar}\left[\hat{H}, \hat{\sigma}_z(j)\hat{\sigma}_z(k)\right].
\end{aligned} \tag{14}$$

The dynamics of the coherence vector components can be obtained by simply taking the expectation values of both sides of the equations listed above, i.e.,

$$\begin{aligned}
\frac{d}{dt}\lambda_x(j) &= \frac{i}{\hbar}\left\langle\left[\hat{H},\hat{\sigma}_x(j)\right]\right\rangle,\\
\frac{d}{dt}\lambda_y(j) &= \frac{i}{\hbar}\left\langle\left[\hat{H},\hat{\sigma}_y(j)\right]\right\rangle,\\
&\vdots\\
\frac{d}{dt}K_{xy}(j,k) &= \frac{i}{\hbar}\left\langle\left[\hat{H},\hat{\sigma}_x(j)\hat{\sigma}_z(k)\right]\right\rangle,\\
&\vdots\\
\frac{d}{dt}K_{zz}(j,k) &= \frac{i}{\hbar}\left\langle\left[\hat{H},\hat{\sigma}_z(j)\hat{\sigma}_z(k)\right]\right\rangle.
\end{aligned} \tag{15}$$

As an example, let us consider the dynamics of the $\hat{\sigma}_y(j)$ basis operator. In doing so, the following commutation property for Pauli matrices will prove useful:

$$[\hat{\sigma}_a,\hat{\sigma}_b] = 2i\epsilon_{abc}\hat{\sigma}_c, \tag{16}$$

where ϵ_{abc} is the Levi-Civita function and is defined as,

$$\epsilon_{abc} = \begin{cases} +1 \text{ if } (a,b,c) \text{ is } (x,y,z),(y,z,x) \text{ or } (z,x,y)\\ -1 \text{ if } (a,b,c) \text{ is } (z,y,x),(x,z,y) \text{ or } (y,x,z)\\ 0 \text{ if } a=b \text{ or } b=c \text{ or } a=c \end{cases}. \tag{17}$$

Using this property and Eq. (14), the time dependance of $\hat{\sigma}_y(j)$ is then,

$$\begin{aligned}
\frac{d}{dt}\hat{\sigma}_y(j) &= \frac{i}{\hbar}\left[\hat{H},\hat{\sigma}_y(j)\right],\\
&= \frac{1}{\hbar}\left[2\gamma_j\hat{\sigma}_z(j) - \sum_m^N E_k^{j,m}\hat{\sigma}_x(j)\hat{\sigma}_z(m)\right].
\end{aligned} \tag{18}$$

Taking the expectation values of both sides, we get:

$$\hbar\frac{d}{dt}\lambda_y(j) = 2\gamma_i\lambda_z(j) - \sum_m^N E_k^{j,m}K_{xz}(j,m). \tag{19}$$

In addition to λ_z, we notice that K_{xz} also appears on the right-hand side of Eq. (19), and therefore its dynamics must also be computed in order to capture the full dynamic behaviour of λ_y. Similarly, when we derive the dynamical equation for K_{xz}, various two-point correlation terms will appear in that equation, and thus we will need to compute the dynamics for those terms as well, and so on. At the end, we will be left with a linear system of coupled ODEs that must all be solved in order to compute the system's dynamics.

3 Dissipative Coupling to the Heat Bath

A final approximation to be discussed here is the relaxation time approximation, which is commonly used in simulations of QCA dynamics. In the absence of energy dissipation and other decohering effects, a QCA array will evolve coherently according to the Liouville equation shown in Eq. (14). Over fairly short time scales, quantum mechanical systems often fall to a thermal steady state [26]. If the QCA system is weakly-coupled to the environment, and the energy transfer between the system and environment is well-described by a Markov process, then at low temperatures, the simplest way to incorporate energy dissipation into a model of QCA dynamics is via the relaxation time approximation [26,40,42,46,48]. This is done by adding a dissipation term to Eq. (15):

$$\frac{d}{dt}\lambda_i = \frac{i}{\hbar}\left\langle \left[\hat{H}, \hat{\lambda}_i\right]\right\rangle - \frac{1}{\tau}(\lambda_i - \lambda_{ss}^i), \tag{20}$$

where τ is a phenomenological time constant and describes the relaxation of the coherence vector towards the steady-state coherence vector element, λ_{ss}^i. The value of τ depends on the specific implementation details of the QCA system as well as the nature of the coupling to the thermal environment, and would need to be determined experimentally. The λ_{ss}^i term can be found by evaluating,

$$\lambda_i^{ss} = \mathrm{Tr}\left\{\hat{\rho}_{ss}\hat{\lambda}_i\right\}, \tag{21}$$

where $\hat{\rho}_{ss}$ is the steady-state density matrix and is defined as,

$$\hat{\rho}_{ss} \equiv \frac{e^{-\hat{H}(t)/k_BT}}{\mathrm{Tr}\left\{e^{-\hat{H}(t)/k_BT}\right\}}. \tag{22}$$

Determining the steady state density matrix exactly is critical to solving the complete Schrödinger equation for the system, and thus comes up against all the difficulties mentioned above. It is therefore usually calculated using the ICHA and the two-state approximation.

4 Conclusions

In this chapter, we reviewed the most common methods for modelling the dynamics of large QCA circuits. Due to the inherent difficulties of solving complete quantum mechanical problems, a number of simplifying assumptions are typically made when attempting to model the dynamic behaviour of large QCA circuits. These include a reduction of the Hilbert space to two states per cell (2-state approximation), treatment of intercell interactions via a mean-field approach (the ICHA), and an assumption of exponential energy relaxation (relaxation-time approximation). Collectively, these assumptions form the basis of almost

all QCA simulations performed to date, and are currently used to treat QCA layouts in QCADesigner. While the use of the 2-state and relaxation-time approximations have been validated in previous literature, the ICHA, which neglects any correlations or entanglement between cells, is known to arrive at the incorrect ground state for many circuits. In spite of its shortcomings, the ICHA is widely-used due to its relative simplicity and low computational overhead.

Fortunately, other methods for solving the dynamical behaviour of QCA circuits exist, whose computational overhead exists somewhere in the solution space spanned by the ICHA and a full quantum mechanical treatment. Using the so-called coherence-vector formalism, the state variables of a QCA system are separated into groups corresponding to the state of individual cells, and into groups corresponding to the two-cell, three-cell, etc., operators. In doing so, we are able to remove those groups from our calculations that do not contribute to the system's dynamics in any meaningful way, and thus reduce the overall computational requirements of treating a QCA circuit. As it turns out, only the 2^{nd}-order operators (i.e., two-cell correlations) were large enough to play a significant role in the dynamics, and thus allowing all higher-order correlations terms to be either neglected or approximated using 2^{nd}-order operators. In doing so, the complexity of solving for the dynamics of an N-cell QCA circuit reduces from $O(4^N)$ to $O(N^2)$. Further reductions in the complexity are also possible if only nearest-neighbour interactions are considered. In such cases, the complexity drops to $O(N)$, which is on par with that of the ICHA. The Liouville equation of motion is used to calculate the time-dependance for each of the system's remaining operators. The collection of these equations form a co-dependent set of ODEs which can be solved using standard ODE solvers. And lastly, interactions between the cells and the environment are accounted for via the use of the relaxation-time approximation which drives the QCA system towards its steady-state. These methods for solving large QCA systems have been shown to offer accurate results with relatively low computational overhead.

References

1. Haider, M.B., et al.: Controlled coupling and occupation of silicon atomic quantum dots at room temperature. Phys. Rev. Lett. **102**, 046805 (2009)
2. Wang, L., Krapek, V., Ding, F., Horton, F., Schliwa, A., Bimberg, D., Rastelli, A., Schmidt, O.G.: Self-assembled quantum dots with tunable thickness of the wetting layer: role of vertical confinement on interlevel spacing. Phys. Rev. B **80**(8), 9 (2009)
3. Keyser, U.F., Schumacher, H.W., Zeitler, U., Haug, R.J., Eberl, K.: Fabrication of quantum dots with scanning probe nanolithography. Phys. Status Solidi B **224**(3), 681–684 (2001)
4. Lent, C.S.: Quantum cellular automata. Nanotechnology **4**, 49–57 (1993)
5. Macucci, M., Gattobigio, M., Bonci, L., Iannaccone, G., Prins, F.E., Single, C., Wetekam, G., Kern, D.P.: A QCA cell in silicon-on-insulator technology: theory and experiment. Superlattices Microstruct. **34**, 205–211 (2004)

6. Orlov, A.O., Kummamuru, R.K., Ramasubramaniam, R., Lent, C.S., Berstein, G.H., Snider, G.L.: Clocked quantum-dot cellular automata shift register. Surf. Sci. **532–535**, 1193–1198 (2003)
7. Amlani, I., Orlov, A.O., Toth, G., Bernstein, G.H., Lent, C.S., Snider, G.L.: Digital logic gate using quantum-dot cellular automata. Science **284**, 289–291 (1999)
8. Amlani, I., Orlov, A., Snider, G.L., Lent, C.S.: Demonstration of a functional quantum-dot cellular automata cell. J. Vac. Sci. Technol. B **16**, 3795–3799 (1998)
9. Lent, C.S., Snider, G.L., Bernstein, G.H., Porod, W., Orlov, A., Lieberman, M., Fehlner, T., Niemier, M.T., Kogge, P.: Quantum-Dot Cellular Automata. Kluwer Academic Publishers, Boca Raton (2003)
10. Snider, G.L., Amlani, I., Orlov, A., Toth, G., Bernstein, G., Lent, C.S., Merz, J.L., Porod, W.: Quantum-dot cellular automata: line and majority gate logic. Jpn. J. Appl. Phys. **38**, 7227–7229 (1999)
11. Kummamuru, R.V., Timler, J., Toth, G., Lent, C.S., Ramasubramaniam, R., Orlov, A., Bernstein, G.H.: Power gain and dissipation in a quantum-dot cellular automata latch. Appl. Phys. Lett. **81**, 1332–1334 (2002)
12. Toth, G., Lent, C.S.: Quasiadiabatic switching of metal-island quantum-dot cellular automata. J. Appl. Phys. **85**(5), 2977–2984 (1999)
13. Imre, A., Csaba, G., Ji, L., Orlov, A., Bernstein, G.H., Porod, W.: Majority Logic gate for magnetic quantum-dot cellular automata. Science **311**(5758), 205–208 (2006)
14. Csaba, G., et al.: Nanocomputing by field-coupled nanomagnets. IEEE Trans. Nanotechnol. **1**(4), 209–213 (2002)
15. György, C., Porod, W.: Simulation of field coupled computing architectures based on magnetic dot arrays. J. Comput. Electron. **1**(1), 87–91 (2002)
16. Parish, M.C.B.: Modeling of physical constraints on bistable magnetic quantum cellular automata. Ph.D. thesis, University of London (2003)
17. Cowburn, R.P., Welland, M.E.: Room temperature magnetic quantum cellular automata. Science **287**, 1466–1468 (2000)
18. Bernstein, G.H., Imre, A., Metlushko, V., Ji, L., Orlov, A., Csaba, G., Porod, W.: Magnetic QCA systems. Microelectron. J. **36**, 619–624 (2005)
19. Lu, Y., Lent, C.S.: Theoretical study of molecular quantum-dot cellular automata. J. Comput. Electron. **4**, 115–118 (2005)
20. Lent, C.S., Isaksen, B., Lieberman, M.: Molecular quantum-dot cellular automata. J. Am. Chem. Soc. **125**, 1056–1063 (2003)
21. Jiao, J., Long, G.J., Grandjean, F., Beatty, A.M., Fehlner, T.P.: Building blocks for the molecular expression of quantum cellular automata. Isolation and characterization of a covalently bonded square array of two ferrocenium and two ferrocene complexes. J. Am. Chem. Soc. **125**(25), 7522–7523 (2003)
22. Lent, C.S., Isaksen, B.: Clocked molecular quantum-dot cellular automata. IEEE Trans. Electron Devices **50**(9), 1890–1896 (2003)
23. Li, Z., Fehlner, T.P.: Molecular QCA cells. 2. Characterization of an unsymmetrical dinuclear mixed-valence complex bound to a au surface by an organic linker. Inorg. Chem. **42**(18), 5715–5721 (2003)
24. Li, Z., Beatty, A.M., Fehlner, T.P.: Molecular QCA cells. 1. Structure and functionalization of an unsymmetrical dinuclear mixed-valence complex for surface binding. Inorg. Chem. **42**(18), 5707–5714 (2003)
25. Qi, H., Sharma, S., Li, Z., Snider, G.L., Orlov, A.O., Lent, C.S., Fehlner, T.P.: Molecular quantum cellular automata cells. Electric field driven switching of a silicon surface bound array of vertically oriented two-dot molecular quantum cellular automata. J. Am. Chem. Soc. **125**(49), 15250–15259 (2003)

26. Timler, J., Lent, C.S.: Power gain and dissipation in quantum-dot cellular automata. J. Appl. Phys. **91**, 823–831 (2002)
27. Timler, J., Lent, C.S.: Maxwell's demon and quantum-dot cellular automata. J. Appl. Phys. **94**, 1050–1060 (2003)
28. Lent, C.S., Isaksen, B.: Clocked molecular quantum-dot cellular automata. IEEE Trans. Electron Dev. **50**, 1890–1896 (2003)
29. Jin, Z.: Fabrication and measurement of molecular quantum cellular automata (QCA) device. Master's thesis, University of Notre Dame, Notre Dame, IN 46556 (2006)
30. Data flow in molecular QCA: Logic can "sprint," but the memory wall can still be a "hurdle" (2005)
31. Lent, C., Liu, M., Lu, Y.: Bennett clocking of quantum-dot cellular automata and the limits to binary logic scaling. Nanotechnology **17**(16), 4240–4251 (2006)
32. Walus, K., Budiman, R.A., Jullien, G.A.: Impurity charging in semiconductor quantum-dot cellular automata. Nanotechnology **16**(11), 2525–2529 (2005)
33. Walus, K.: Design and simulation of quantum-dot cellular automata devices and circuits. Ph.D. thesis, University of Alberta, September (2005)
34. Walus, K., Karim, F., Ivanov, A.: Architecture for an external input into a molecular QCA circuit. J. Comput. Electron. **8**, 35–42 (2009)
35. Karim, F., Walus, K., Ivanov, A.: Analysis of field-driven clocking for molecular quantum-dot cellular automata. J. Comput. Electron. **9**, 16–30 (2010)
36. Hennessy, K., Lent, C.S.: Clocking of molecular quantum-dot cellular automata. J. Vac. Sci. Technol. B **19**(5), 1752–1755 (2001)
37. Lent, C.S., Tougaw, P.D.: Lines of interacting quantum-dot cells: a binary wire. J. Appl. Phys. **74**, 6227–6233 (1993)
38. Tougaw, P.D., Lent, C.S.: Logical devices implemented using quantum cellular automata. J. Appl. Phys. **75**, 1818–1825 (1994)
39. Tougaw, P.D., Lent, C.S.: Dynamic behavior of quantum cellular automata. J. Appl. Phys. **80**, 4722–4735 (1996)
40. Tóth, G., Lent, C.S.: Role of correlation in the operation of quantum-dot cellular automata. J. Appl. Phys. **89**, 7943–7953 (2001)
41. Karim, F., Navabi, A., Walus, K., Ivanov, A.: Quantum mechanical simulation of QCA with a reduced hamiltonian. In: Proceedings of the 8th IEEE Conference on Nanotechnolgy (2008)
42. Srivastava, S., Sarkar, S., Bhanja, S.: Estimation of upper bound of power dissipation in QCA circuits. IEEE Trans. Nanotechnol. **8**(1), 116–127 (2009)
43. Lieberman, M., Chellamma, S., Varughese, B., Wang, Y.L., Lent, C., Bernstein, G.H., Snider, G., Peiris, F.C.: Quantum-dot cellular automata at a molecular scale. Mol. Electron. II **960**, 225–239 (2002)
44. Walus, K., Mazur, M., Schulhof, G., Jullien, G.A.: Simple 4-bit processor based on quantum-dot cellular automata (QCA). In: Proceedings of Application Specific Architectures, and Processors Conference, pp. 288–293, July 2005
45. Lent, C.S., Tougaw, P.D., Porod, W.: Bistable saturation in coupled quantum dots for quantum cellular automata. Appl. Phys. Lett. **62**, 714–716 (1993)
46. Leggett, A.J., Chakravarty, S., Dorsey, A.T., Fisher, M.P.A., Garg, A., Zwerger, W.: Dynamics of the dissipative two-state system. Rev. Mod. Phys. **59**, 1–85 (1987)
47. Mahler, G., Weberruß, V.A.: Quantum Networks: Dynamics of Open Nanostructures. Springer, Berlin (1998)
48. Weiss, U.: Quantum Dissipative Systems. World Scientific, Stuttgart (2008)

ToPoliNano: NanoMagnet Logic Circuits Design and Simulation

Marco Vacca, Stefano Frache, Mariagrazia Graziano(✉), Fabrizio Riente, Giovanna Turvani, Massimo Ruo Roch, and Maurizio Zamboni

Dipartimento di elettronica e telecomunicazioni, Politecnico di Torino, Turin, Italy
mariagrazia.graziano@polito.it

Abstract. Among the emerging technologies Field-Coupled devices like Quantum dot Cellular Automata are one of the most interesting. Of all the practical implementations of this principle NanoMagnet Logic shows many important features, such like a very low power consumption and the feasibility with up-to-date technology. However its working principle, based on the interaction among neighbor cells, is quite different from CMOS circuits. Dedicated design and simulation tools for this technology are necessary to further study this technology, but at the moment there are no such tools available in the scientific scenario.

In this chapter we present ToPoliNano, a software developed as a design and simulation tool for NanoMagnet Logic, that can be easily adapted to many other emerging technologies, particularly to any kind of Field-Coupled devices. ToPoliNano allows to design circuits following a top-down approach similar to the ones used in CMOS and to simulate them using a switch model specifically targeted for high complexity circuits. This tool greatly enhances the ability to analyze and optimize the design of Field-Coupled circuits.

1 Introduction on Simulation of Complex NML Circuits

Among the emerging technologies NanoMagnet Logic (NML) is one of the most intriguing. In this technology single domain nanomagnets with only two stable states are used to represent the logic values '0' and '1' [1,2], as shown in Fig. 1. They represent a particular application of the Quantum dot Cellular Automata [3] idea, and more generally of the Field-Coupled principle, where the computation is performed by the interaction of neighbor cells [4–7]. Molecular QCA is the other main implementation of the Quantum dot Cellular Automata principle [8,9], which relies on complex molecules to represent the digital values [10]. The specific advantages of NanoMagnet Logic are represented by low power consumption [11], the possibility of combining memory and logic in the same devices, high radiation resistance and, not less important, the possibility to fabricate circuits with up-to-date technology [12,13].

In this technology logic circuits can be fabricated placing cells on a plane [14]. Signal propagation and logic computation are obtained through magnetic coupling among neighbor cells [15,16], because magnets align themselves in order

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 274–306, 2014.
DOI: 10.1007/978-3-662-43722-3_12, © Springer-Verlag Berlin Heidelberg 2014

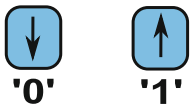


Fig. 1. Single domain nanomagnets are used as basic cells. At the equilibrium only two stable states are possible.

to reach the minimum energy state. The alignment is antiferromagnetic (every element is in the opposite state of its neighbors) if magnets are aligned horizontally, while the alignment is ferromagnetic (every element is in the same state of its neighbors) if magnets are aligned vertically [17]. However, the magnetic field generated by a magnet is not sufficient to cause a state alteration in its neighbors. To switch magnets from one state to the other it is necessary to use a mechanism called clock [18]. The behavior is depicted in Fig. 2. Magnets are forced in an unstable (RESET) state through an external mean, like a magnetic field [19,20]. When the magnetic field is removed magnets realign with a domino-like effect following the input element. With this mechanism signals propagate correctly through the circuit. As well as a magnetic field, other systems can be used to clock circuits, like STT-current coupling [21] or an electric field [22].

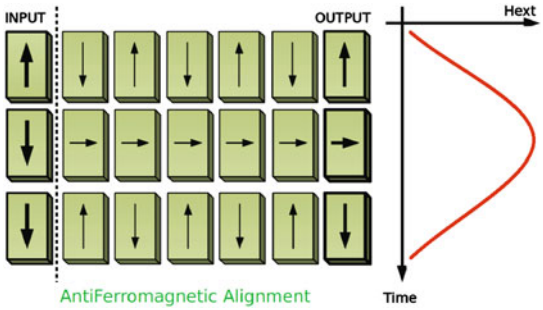


Fig. 2. Clocking mechanism for NML logic. Magnets are forced in an intermediate unstable state through an external mean, like a magnetic field that in a particular portion of time reaches a maximum appropriate value.

The RESET state is unstable. If too many magnets are cascaded some of them along the chain will switch in the wrong state due to external influences, like thermal noise [23,24]. To have a safe signal propagation no more than 5 magnets should be cascaded [23]. As a consequence a multiphase clock system is required. Circuits are divided in small areas, called clock zones. Each zone is made by a limited number of magnets. Every clock zone requires the application of a different clock signal, like shown in Fig. 3(A) where three clock signals with a phase difference of 120° are used. As depicted in Fig. 3(B), when magnets of a clock zone are in the SWITCH state (the magnetic field is slowly removed)

they see from one side magnets that are in the HOLD state (no magnetic field applied). Magnets in the HOLD state can assume a value of logic '0' and '1' so they are seen as an input by magnets that are switching. At the same magnets near the other side of the switching zone are in the RESET state (magnetic field is applied) and therefore they have no influence on the switching ones. Figure 3(B) shows the circuit time evolution when a multiphase clock system is used.

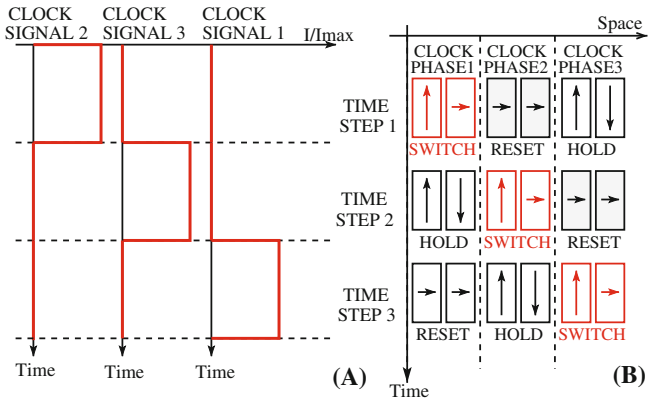


Fig. 3. 3-phase clock. Circuits are divided in areas, called clock zones, made by a limited number of magnets. (A) Three clock signals with a phase difference of 120° are selectively applied to clock zones. (B) When magnets of a clock zone are in the SWITCH state, magnets on the left are in the HOLD state and act like an input, while magnets on the right are in the RESET state and have no influence on the switching magnets.

To design circuits, clock zones must be arranged following a proper layout. Moreover the layout must take into account the constraints related to the technological fabrication of the clock generation network. For example the magnetic field is normally generated by a current flowing through a wire placed under the magnets plane [25] (Fig. 4(A)). With this clock mechanism the clock zones layout is made by parallel stripes (Fig. 4(B)). Each stripe corresponds to one of the clock wires used to generate the various clock signals [13]. While this layout was developed for the magnetic field clock, and other clock systems can have different layouts, it has the advantage to synchronize signals propagation. Thanks to the multiphase clock the circuit is intrinsically pipelined, that means every group of 3 consecutive clock zones has a delay of 1 clock cycle. As a consequence, if the length of input wires of a logic gate is not the same, signals will have different propagation delay and propagation errors will occur. This problem is called "layout=timing" [26,27]. With the clock zones layout shown in Fig. 4(B) the length of every input wire of any gate inside the circuit is always equalized, solving therefore the "layout=timing" problem. For this reason this clock zones layout is chosen as a reference regardless to the clock mechanism used.

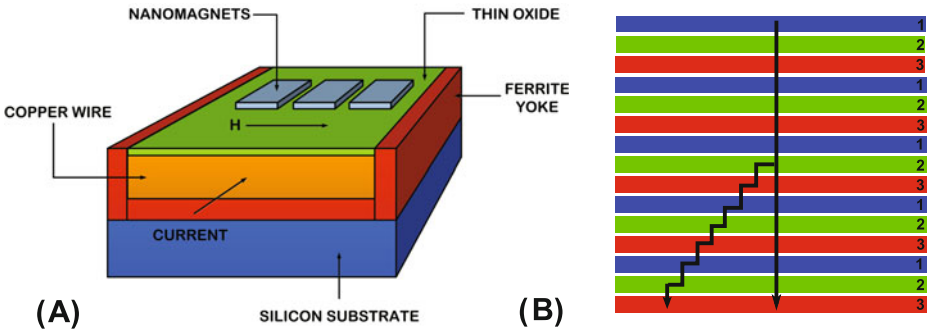


Fig. 4. (A) A magnetic field can be generated by a current flowing through a wire placed under the magnets plane. (B) Clock zones layout is made by parallel strips that follows the wires used to generate the magnetic fields.

To simulate QCA circuits a dedicated simulator, called QCADesigner, was developed [28]. Unfortunately QCADesigner does not support magnetic circuits. To simulate NML circuits two paths can be followed. First of all low level magnetic simulators, like OOMMF [29], NMAG [30] or [31], can be used. Low level simulators allow to obtain the most accurate simulation, but they are very slow and only small circuits can be simulated due to the high memory usage.

As a second option circuits can be studied using a RTL model [32]. The idea is to describe using the VHDL language a CMOS circuit that behaves exactly like its NML counterpart. For example, starting from the NML circuit of Fig. 5(A), its RTL model can be built as shown in Fig. 5(B) and then described with VHDL language. Registers are used to model the propagation delay. This is possible due to the intrinsic pipelined behavior of the technology. Ideal logic gates without delay are used to model the logic function. The logic gate set available in this technology is based on majority voters [16], AND/OR gates [33] and crosswires [13], that are particular blocks that allows to cross two wires on the same plane. At every register one of the three clock signals shown in Fig. 5(C) is then applied. In this way complex circuits can be easily described using VHDL and simulated using the powerful CAD tools available in CMOS technology, like Modelsim [34]. We applied this model successfully designing complex NML circuits in [26,27,35].

2 ToPoliNano

Both simulation mechanisms available have their flaws. The situation is summarized in Fig. 6.

To have the most accurate results physical simulators are required, but they can be used only on very small circuits due to their computational requirements. At the same time, while the RTL model is a powerful tool that allows fast description and simulation of complex circuits, it gives only estimations of real circuits performance, because a lot of informations on the circuit layout are

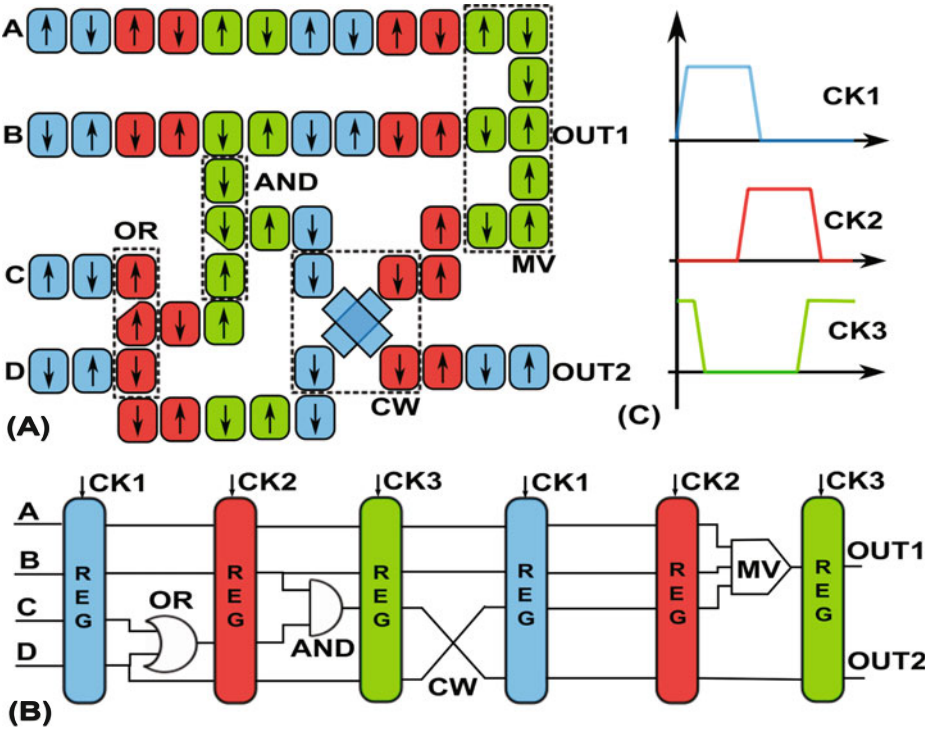


Fig. 5. VHDL model for NML circuits. (A) Example of NML circuit. (B) RTL model described with VHDL. Registers are used to simulate the propagation delay while ideal logic gates are used to model the logic function. (C) Clock signals applied to each register.

missed. We have therefore created our own tool, ToPoliNano, Torino Politecnico Nanotechnology tool [36], a tool targeted to design and simulation of Nano-Magnet Logic circuits. ToPoliNano emulates the top-down approach used in CMOS design, where circuits are described using the VHDL language and the layout is automatically generated. Circuits can be simulated and important information on the circuit behavior and the power consumption can be extracted, knowing exactly the circuit area and the precise placement of every element [37]. Mostly important the open and modular structure of the software allows to easily integrated others emerging technologies, like we have done with Silicon Nanowires NanoPLA [38,39], making it the ideal platform for the study of emerging technologies.

2.1 Tool Overview

ToPoliNano has been developed in C++ and is built around the idea to give to researcher the possibility to design NML circuits following the same top-down

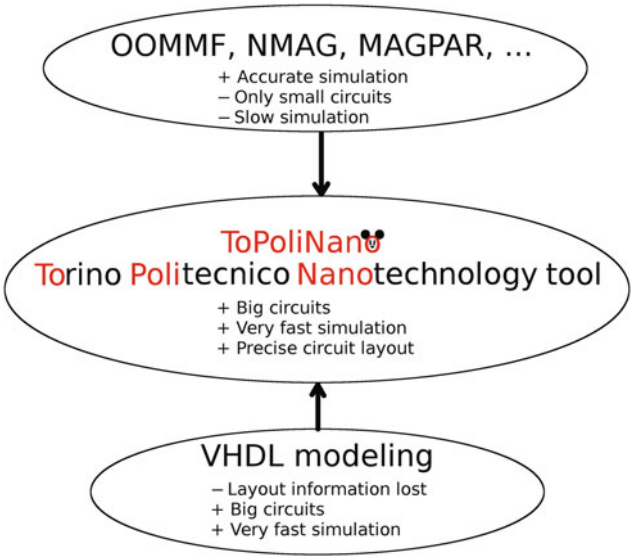


Fig. 6. Simulation of NML circuits. While physical simulators provide the most accurate simulation they can be used only on small circuits. At the same time, VHDL model gives inaccurate results for the loss of information regarding circuits layout. ToPoliNano was created to overcome these problems and to provide a tool that allows to have both accurate results and fast simulation of complex circuits.

methodology used in CMOS circuits. This means to describe circuits of any kind of complexity using VHDL language, to automatically generate the layout and to fast simulate the obtained circuit. For this emerging technology there are no tools available to perform these analysis, therefore it has been necessary to design a completely new system. The structure of Topolinano is shown in Fig. 7.

- The **Logic Synthesizer** is the first block encountered in the traditional CMOS design flow. Starting from a generic VHDL description it translates it on a specific logic gates set. In this case it takes an entry VHDL file and it generates another VHDL file with a structural description, that means the circuit is described only using the set of gates available in this technology (majority voter, and, or, inverter). The logic synthesizer is still partially in development.
- The **Parser** takes the structural VHDL file generated by the logic synthesizer and creates an in-memory representation of the circuit itself. The internal description is based on a hierarchical graph to efficiently handle the circuits in terms of both time and memory occupation on the host computation machine. The parser is complete and it is throughly described in Sect. 3.
- The **Place & Route** takes the graph generated by the parser and automatically creates the circuit layout. This block is still in development, as up to now it can handle only combinational circuits of any complexity but not

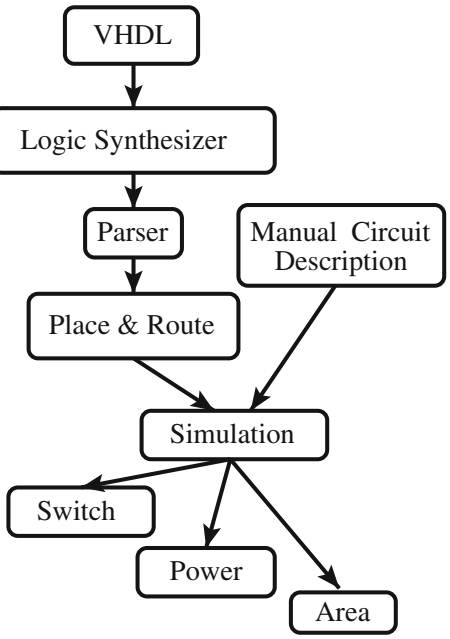


Fig. 7. ToPoliNano structure. Circuits are described through VHDL, a logic synthesizer maps the circuit on the technology library available and a parser generates the in-memory description. The layout can be generated automatically or manually, circuits are then simulated obtaining data on the circuit behavior, the area occupied and the power consumption estimation.

sequential. The circuit layout is based on the clock zones layout described in Sect. 1 and shown in Fig. 4(B). Section 4 provides more details on the Place & route in its current state.

– It is also possible to **Manually Describe** circuits with a full custom approach. This possibility is granted for two important reasons, because the Place & Route block is still in development and because, no matter what level of development the Place & Route block will reach, in certain cases the hand of a designer is requested to reach the maximum level of optimization. Up to now circuits can be described either directly writing the code that describes the circuit or using external vectorial graphic editors and then importing the circuit in ToPoliNano. Further details are given in Sect. 6.
– Once the circuit layout is generated, a **Simulator** is used to verify the correct behavior of the circuit. The algorithm used is based on a behavioral model extracted from low level simulations. This tool is designed for high complexity circuits (million of magnets), so only a behavioral algorithm allows a fast enough simulation. However, since the model is based on physical simulations it still gives accurate results. More details on the simulation algorithm are provided in Sect. 5.

– After the simulation the **Output Generator** allows to calculate the circuit area and to estimate the power consumption considering a magnetic field clock. More details can be found in Sect. 7.

3 Parser

One of the most common uses of a parser in computing is as a component of a compiler or an interpreter. This component usually parses the source code of a programming language in order to create an internal representation. The parsing process (i.e. syntax analysis) consists in the analysis of an input sequence to determine its grammatical structure with respect to a given formal grammar. The parsing process operates a transformation of the input text into a data structure (a tree in the present context), suitable for later processing. The data structure must be such to capture the implied hierarchy of the input, and a tree certainly is suitable for this purpose. The typical operation of parsers is in two stages: first, it identifies the meaningful tokens in the input. Then, it builds a data structure out of the tokens (Fig. 8).

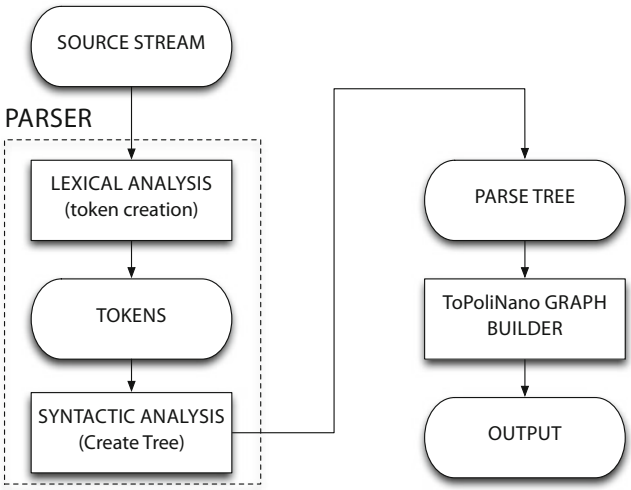


Fig. 8. Overview of the parsing process.

In ToPoliNano the use of the parser is related to the need to input circuit descriptions by means of a Hardware Description Language (HDL). In particular, the VHDL language is currently supported.

3.1 Lexical Analysis

Lexical analysis is the process of converting a sequence of characters into a sequence of tokens. A program which performs the lexical analysis is called lexer

or scanner. The input characters stream is split into meaningful symbols defined by a grammar. They define the set of possible character sequences used to form individual tokens. The term token, in the present context, refers to an abstraction for the smallest unit of VHDL code that is convenient when describing the syntax of the VHDL language. A token is a string of characters, categorized according to the rules as a symbol (e.g. identifier, number, comma, etc.). Starting from this time on, the interpreted data may be loaded into data structures for general use: in ToPoliNano data is used to build an internal representation of the circuit by means of a graph.

3.2 Syntactic Analysis

Syntactic analysis has the objective to determine the structure of the input stream and to build the data structure. Token are fed to the syntactic analyzer and, as output, in case of a tree-based data structure, one would get a node for each element. Basic elements are represented by leaf nodes, and other elements by composite nodes. This stage basically checks that the tokens form an allowed expression.

3.3 The Parse Tree

According to [40], a parse tree is an ordered and rooted tree that represents the syntactic structure of a string, text file, source code written in a given programming language according to some formal grammar. A parse tree for a source code is called Abstract Syntax Tree (AST). The syntax is 'abstract' in the sense that it does not represent every detail that appears in the real syntax.

3.4 Parsing Expression Grammar Definition

Parsing Expression Grammars (PEGs) are formal grammars that allow to describe a formal language in terms of a set of rules for recognizing tokens. The grammar encapsulates a set of rules, primitive parsers and sub-grammars. After being defined, rules can be used as parser components in more complex expressions in order to form a grammar. Grammar is basically a container for one or more rules allowing to encapsulate more complex parsers. A grammar has the same template arguments as a rule.

When the Parsing Expression Grammar to be defined is complex and nested, as in ToPoliNano, where we must define a PEG for VHDL93 language specifications, it may be useful to build user-defined parser components. The grammar for VHDL structural descriptions can be built starting from the grammar definition of the unique constructs, i.e. constructs that do not include any other.

3.5 VHDL Grammar

A VHDL file for structural description includes an Entity Declaration and an Architecture Body (in what follows, the libraries included at the begin of every VHDL file are not considered):

```
VHDL = EntityDeclaration
       > ArchitectureBody;
```

The Entity Declaration grammar could be written as:

```
EntityDeclaration = EntityKeyword > Identifier > IsKeyword
                    > ( GenericClause | *ascii::space )
                    >  PortClause
                    > EndKeyword > Identifier > lit(';');
```

Such a grammar may match a Generic Clause, if it is defined in the VHDL source file, otherwise zero or more white spaces, and a Port Clause every time an Entity Declaration is found.

The grammar for an Architecture Body is divided into two blocks:

- Architecture Declarative Part Grammar
- Architecture Statement Part Grammar

The custom grammar for the Architecture body could be written as:

```
ArchitectureBody = ArchitectureKeyword > ArchitectureIdentifier > OfKeyword
                   > EntityIdentifier > IsKeyword
                   > ArchitectureDeclarativePart
                   > BeginKeyword
                   > ArchitectureStatementPart
                   > EndKeyword >  ArchitectureIdentifier > lit(";");
```

In an Architecture Declarative Part there are one or more Component interfaces and zero or more Signal Declarations:

```
ArchitectureDeclarativePart = (+ ComponentDeclaration)
                              > (* SignalDeclaration);
```

The Component Declaration Grammar is very similar to the Entity Declaration, since the Component interface is the copy of the Entity interface. In an Architecture Statement there may be instantiated components with Generic Map and Port Map, assigned values to signals and generate well-patterned structures. These three statements can be found in any order and in any number within the Architecture Statement Part (here Generic Map and Port Map are joined for the sake of simplicity):

```
ArchitectureStatementPart = + ( GenericMapPortMap
                                | SignalAssignment
                                | GenerateStatement
                              );
```

In Fig. 9 the parser components have been defined, the grammars and the other parser components used are summarized.

Once the VHDL grammar has been defined, the need to store the information parsed from the VHDL projects arises. This information must be stored at parse time, and elaborated at a later time, possibly by other modules of the ToPoliNano

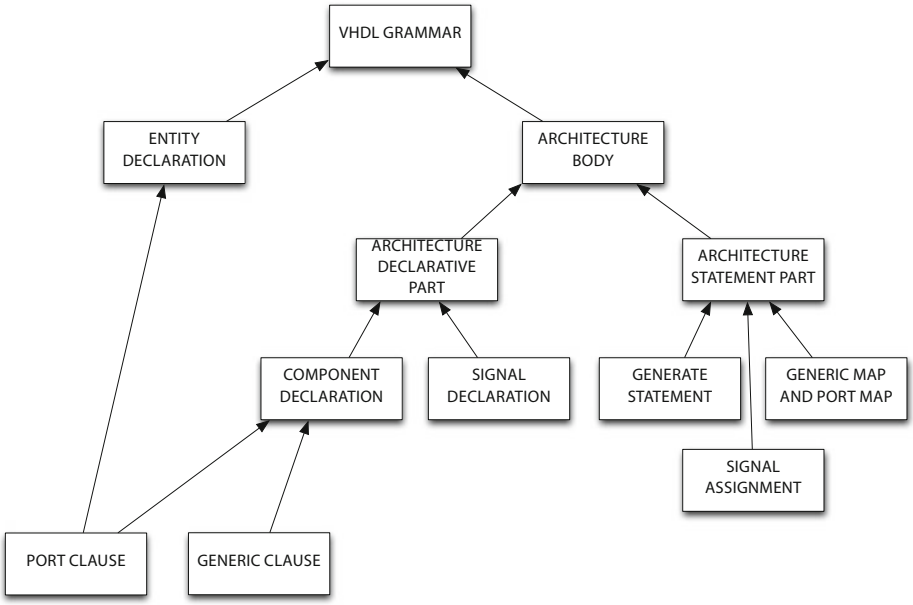


Fig. 9. Hierarchy of VHDL Grammar.

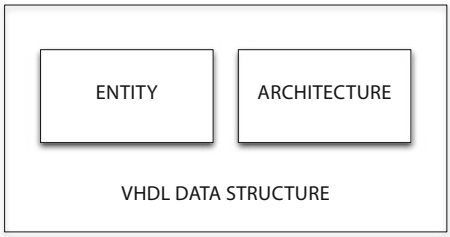


Fig. 10. VHDL Design Entity Data structure.

tool. Up to now, the Parser just recognizes data, but does nothing with it. A dynamic data structure is built and instantiate at runtime dynamic objects of certain classes just when they are needed, and populate them at parse time. The data structure used to store the parsed information from a VHDL design unit is a dynamic object of a main class called VhdlClass, which comprises pointers to an Entity object and to an Architecture object. It is shown in Fig. 10.

3.6 Intermediate Form Representation

Once all the VHDL design units have been parsed, a data structure that represents the digital circuit, component by component, has to be implemented. The idea is to create a hierarchical graph of nodes that represents the circuit and all

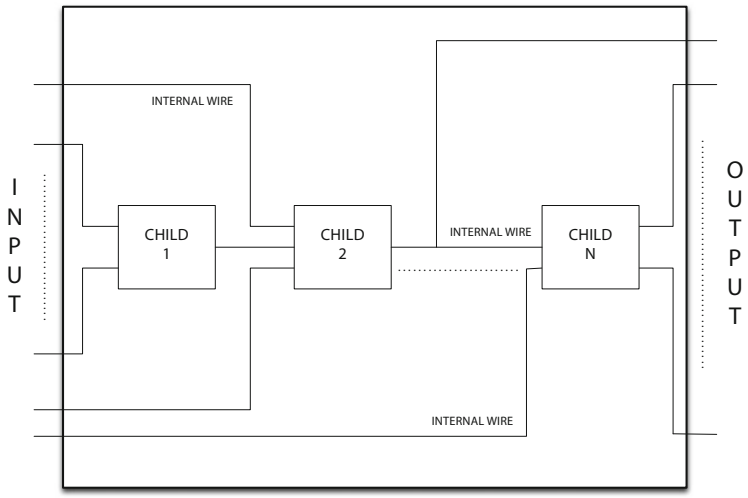


Fig. 11. Composite component.

its components; moreover, it has to capture how they are connected together. This graph is generated starting from the temporary data structure created during the VHDL code parsing. The graph must have as many nodes as the number of basic blocks plus the number of composite components in the circuit. All the nodes of the same grade represent the whole circuit seen at a certain level of abstraction. The lower the level, the higher the abstraction. The root has the maximum level of abstraction, because it depicts the circuit in just one block; all the graph leaves represent the basic blocks out of which the circuit is built. Every node contains all the information regarding a component (Input, Output, Bidirectional ports, internal Signals, Interconnections), and all of its children (if any) represent the components it is made of.

The basic element of the graph is an object of the abstract class Node. Two concrete classes are derived from the Node class: CompositeNode and LeafNode. Every node contains a set of Input, a set of Output, a potential set of Bidirectional I/O, a potential set of internal Signals, a potential set of child Nodes and the Interconnections. A Composite Node is a node that has at least one child. Composite nodes are used to represent the top level and all the intermediate level components. Composite nodes represent components like that shown in Fig. 11.

A Leaf Node is a node that does not have children. They are used to represent all the basic blocks in the circuit. All Inputs, Outputs, Bidirectional I/O and internal Signals can be represented by objects of the same class, i.e. the Wire class, because all of them are merely wires that carry the same kind of logic data. The class members are capable to differentiate from other Wires by means of a name within a design Entity and by means of a serial number within a set. The data structure built during the parsing phase is the starting point of the graph creation. It contains one object for each VHDL design unity. Starting from the

output of the Parser, one has to calculate the hierarchy among the components in the circuit, then, for each component, to instantiate a Node and populate it.

The first step is to detect every instance of a component. For each VHDL design unity, we have to create as many objects as the number of components instantiated in the Port Map clauses. The objects in the VHDL design unities vector are copied and in case modified (if the component is instantiated with a generic interface different from its declaration) every time they are instantiated in a Port Map aspect. For this purpose a new object type is created in order to facilitate the work of the stage that instantiates the graph, that is an object of class Hierarchy. A vector containing all the composite and all the basic design unities is temporarily built.

Each object of class Hierarchy contains the following information:

- A pointer to the VHDL design unity object.
- The name of the Design Entity Declaration.
- The instantiation name, which is unambiguous among nodes with the same parent.
- A flag indicating if it is a basic block (Leaf Node).
- The index of the parent object in the vector.

The first item of the Hierarchy vector (position 0) contains the information about the top level (the root); it is the first item to be pushed into the vector. Then the population cycle starts (with only the top level in the vector). All the other items are pushed back in the vector during the cycle. For every design unity composed of some other components (i.e. for every composite), it takes the Port Map aspects, and, for each component instantiated, a new Hierarchy item is created, in which the VHDL design unity object with the correspondent Entity identifier, taken from the VHDL design unities vector, is copied and possibly modified, depending if it has a generic interface that differs from its declaration. Then, also the information about the name of the Entity, the instantiation name, the index of the parent object (which is merely the index of the FOR cycle) and the flag indicating if it is a basic block are inserted. After its population, the object is pushed back in the Hierarchy vector. The cycle ends when all the components in all the composite design unities are instantiated, i.e. when all the components instantiated in all the composite entities Port Map aspects are created. The graph is then instantiated and populated through an object of a concrete class, following the Visitor Pattern.

4 Place and Route

In recent years the rapid advances in fabrication technology have dramatically increased the complexity in VLSI circuits. Indeed, major focus is on development of tools for the design and analysis of heuristic algorithms for partitioning, placement, routing and layout compaction. The developed place and route engine [41] uses different algorithms from traditional technology, adapted in order to solve

NML technology issues. What is new is how these algorithms are used together in order to obtain a final layout compliant with the technology constrains and limitations. The proposed NML design flow can be divided in two main parts: graph elaboration and physical mapping, which will be discussed in the following sections. The algorithm is divided in two parts, *Graph Elaboration* where the graph generated by the parser is optimized (see Sect. 4.1) and *Physical Mapping* where the circuit layout is generated (described in Sect. 4.2).

4.1 Graph Elaboration

The flow diagram of the graph elaboration phase is shown in Fig. 12. Starting from a structural description of the circuit, mapped on a set of cells available (majority voter, and, or, inverter), the HDL parser generates a graph which is the primary input of the place and route engine.

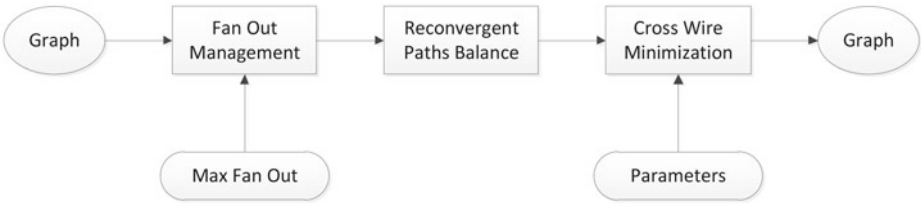


Fig. 12. Graph Elaboration flow diagram. The entry point is the graph generated by the parser while the output is an optimized graph used to create the circuit layout. Several optimizations are performed on the input graph, each part of the algorithm can be customized using appropriate parameters.

This data structure is then handled according to NML characteristics and the clock zones layout. The operations performed by these algorithms can be summarized as:

- *Fan-out management*, the graph is modified to take into account a limited fan-out, given as a parameter, for each graph node.
- *Reconvergent paths balance*, paths inside the graph are balanced to avoid "layout=timing" problems.
- *Wire crossing minimization*, the graph is elaborated using various algorithms (customizable with several input parameters) to reduce the number of wires-crossing.

At the end of this process the output graph is used as input for the physical mapping phase. In the following a detailed description of each algorithm is reported.

Fan-Out Management. As in the case of CMOS technology, also NML has a limitation on the fan-out that each cell can support. The main reasons are related to NML clock zones dimensions, to the physical space occupied by the wires and to the number of magnets that can be cascaded inside a clock zone. With a parallel clock zones layout organization, the vertical magnetic wires length should be limited in order to avoid propagation errors. Besides, considering that each wire is made by magnets there must be enough space to allow their physical placement. Therefore, the input graph is iteratively analyzed and mapped into a new one where the fan-out limit is satisfied.

Reconvergent Paths Balance. The graph generated starting from a circuit netlist can present many *reconvergent paths*. Two paths, in a direct acyclic graph (DAG), are called *reconvergent* if they diverge from and reconverge to the same blocks. This situation is common with traditional technology, but with NML it can generate some problems due to its intrinsic pipelined behavior that originates the so called "layout = timing" problem. In order to guarantee signals synchronization at the input of each logic gate, additional intermediate nodes must be added, so that all the reconvergent paths are composed by the same number of nodes.

Cross-Wires Minimization. Since NML is a planar technology, up to now just one layer can be used to build a circuit. A particular component, called *cross-wire*, is available to cross two wires on the same plane without interferences. Even if this block allows physical wires intersection a specific optimization is needed to reduce the number of cross-wire, therefore the wasted area. Different techniques are here implemented, such as *Barycenter*, *Fan-out Tolerance Duplication*, *Simulated Annealing* and *Kernighan-Lin*. They can be used alone or combining one or more of them together.

Barycenter. The basic idea behind this method is to rearrange nodes in order to place them directly above the nodes to which they are connected. The algorithm explores each rows of the graph two by two from inputs to outputs. For each couple of rows analyzed one is kept frozen while nodes in the other row are changed in position to reduce the number of cross-wires. This algorithm is quite simple and fast but leads to an unoptimized result. This is due because there can be situation in which multiple solutions satisfy the requisites of the algorithm. This solutions have however a different number of cross-wires, so the efficiency of the algorithm heavily depends on the policy chosen to solve this conflicting situations.

Fan-out Duplication. The job of the fan-out duplication algorithm is to integrate the *Barycenter* method in order to reduce wire crosses. As can be gathered from the name of this technique, graph nodes are duplicated trying to reduce the cross-wires number. The number of cross-wires can be theoretically reduced to 0, however the circuit area grows exponentially.

Kernighan-Lin. This algorithm is one of the most commonly used in the class of partition based methods. It heuristically divides the graph into sub-regions, trying to minimize the cut, i.e. the number of edges that connect one sub-region from the others.

Simulated Annealing. It is a stochastic algorithm that iteratively swap the position of nodes inside the graph trying to find the global minimum through consecutive solutions. The purpose of this technique is to minimize the number of cross-wires during the global placement. The algorithm needs three data as input:

- The graph nodes V.
- The minimum temperature that must be reached.
- The number of iteration that are necessary to obtain the final state.

The parameters used determine the efficiency of the simulated annealing algorithm, and these parameters must be chosen according to experience. While simulated annealing can lead to very good results, its a stochastic method that heavily relies on the parameters used, on the function used to generate random numbers and generally requires a huge time to converge.

References. Since the cross-wires minimization techniques here proposed are based on algorithm taken from the literature, here some references are included for interested readers. For the barycenter technique users can refer to [42,43] where these algorithms were already applied to QCA technology. For Kernighan-Lin technique the readers may refer to [44,45] while for simulated annealing readers may refer to [46–48]. These were the starting point, the algorithms have been reworked and adapted in order to be used on nanomagnet logic circuits.

4.2 Physical Mapping

During the physical mapping phase, the previous elaborated graph is transformed into the final circuit layout. Due to the high complexity of today ICs, the physical placement cannot be obtained in a single routing phase. For this reason, a three steps approach is followed:

- Placement, where graph nodes are mapped to their correspondent logic gate and they are initially placed on the layout.
- Global routing, where an iterative approach is used to find the optimum position for logic gates.
- Detailed routing, where interconnection wires among gates are routed.

In Fig. 13 a general flow diagram is shown.

Every node of the graph is mapped into its corresponding logic gate and it is placed into the circuit. Then, a global routing is performed. This process try to place each logic gate in order to minimize the occupied area. After the blocks positioning phase a detailed routing is performed among gates.

Fig. 13. Physical Mapping flow diagram. The circuit layout is generated starting from its graph following several optimization steps.

Placement. During the placement stage each node of the graph is mapped into its logic gate. As shown in Fig. 14, nodes are placed row by row starting from the top of the graph without any optimization. Thus, blocks are aligned with a minimum spacing equal to one equivalent magnet height.

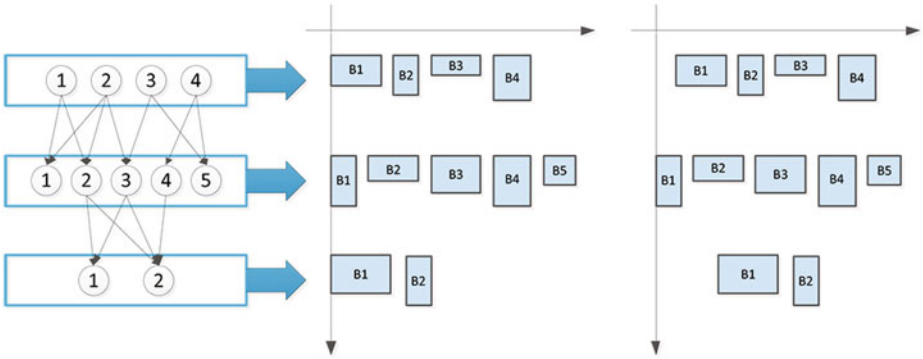


Fig. 14. *Center*: seed row placement for maximum with evaluation; *Right*: barycentered placement

This technique is used to evaluate the maximum width of the circuit, i.e the width of the largest row. At this point, a barycenter alignment is performed to shift placed blocks in order to reduce the overall wire length.

Global Routing. The final position of each gate is obtained with a fine shift performed during the global routing phase. The idea is to maximize the circuit compaction, therefore reducing the length of interconnection wires. Figure 15 shows the flow diagram of the global routing phase.

The implemented procedure is composed by the following main steps iteratively applied to each couple of rows:

- Logic gates of row i are shifted.
- Wires among row i and row $i+1$ are routed.
- Interconnection area is evaluated.

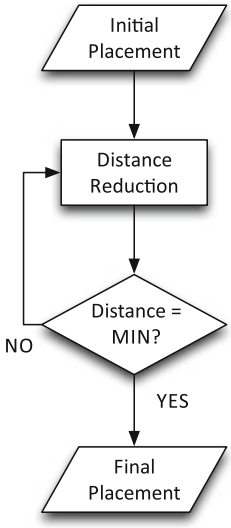


Fig. 15. Global routing flow diagram. It is an iterative process where the position of each block for every row is shifted with the aim of minimizing the interconnections length.

The goal is to find the global minimum area of the circuit, so an iterative approach is needed.

Channel Routing. When the final position of logic gates and pins is defined, the channel routing routine can start in order to obtain the final layout. In NML there is a limit to the maximum number of element that can be cascaded inside a clock zone. This feature coupled with the clock zones layout makes the signals propagation follow a "stair-like" path, as shown in Fig. 4(B). To model this behavior the channel routing algorithm is based on the mini-swap [49] method, which uses diagonal interconnections. Again though based on these algorithm, the method is a mix and largely adapted to the NML case.

Circuit Example. Figure 16 shows an example of circuit obtained after the whole process. It is a 6 bits ripple carry adder where the zoomed element represents a cross wire block.

References. The literature about placement and routing of VLSI circuits is quite wide. For interested readers an overview of the automatic placement and routing of CMOS VLSI circuits can be found in [50,51]. More detail on the wire and channel routing can instead be found in [49,52–54].

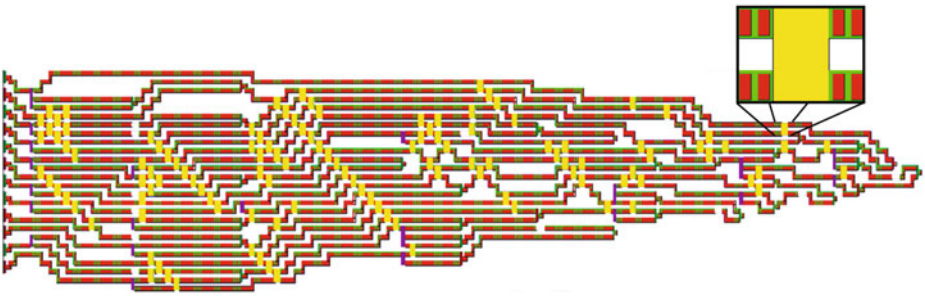


Fig. 16. Layout of a 6 bit Ripple Carry Adder.

5 Simulation Engine

After obtaining the detailed circuit layout it is important to simulate it to verify its behavior and to evaluate its performance. For this purpose a behavioral simulation engine was developed to be used in ToPoliNano. In order to simulate the circuit the graph used to represent the circuit must be modified. The hierarchical graph structure, which describes the circuit, is flattened by the simulation engine. A completely new data structure based on a dynamical matrix is created, in this way all physical information are kept (magnets positions) and the structural ones are discarded.

An example of simulation matrix is shown in Fig. 17. It represents a majority voter but the structural information are lost so it is seen simply as a matrix containing or not containing magnets in each cell. Three different type of magnets can be identified: input, output and normal. Every magnet is seen as a tristate device, because it can assume only three values, '0', '1' and RESET.

5.1 Simulation Algorithm

The behavior of the simulation algorithm has been captured and encapsulated in a finite-state machine (FSM). The clock waveforms Fig. 18(A) can be divided in 6 defined states and are therefore mapped to the FSM shown in Fig. 18(B).

The initial state is represented by the 0, while the first transition depends on the clock signals behavior. For example, the transition to the state 1 takes place when the first clock signal is equal to one and the second and the third are equal to 0. The FSM cycles through the states because of the periodic behavior of the clock waveforms. In order to manage transitions, the future step S_f can be calculated starting from the present state S_p (Eq. 1):

$$S_f = (S_p + 1) mod 6 \tag{1}$$

The general simulation algorithm follows the FSM switching periodically through its states. In each state, the value of each magnet must be calculated depending on the clock zone in which it is located. In particular, all magnets belonging to

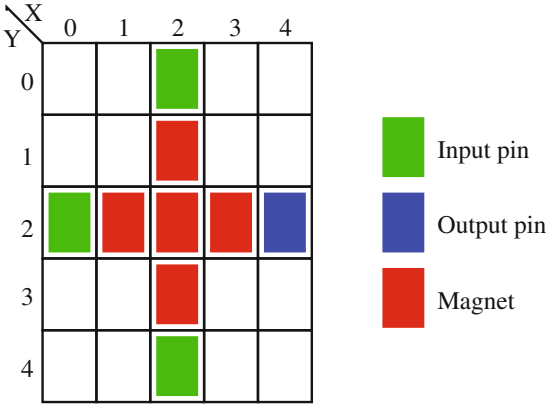


Fig. 17. Simulation matrix. Three categories of magnets can be identified: normal, input and output magnets. Every magnets is represented with a tristate approximation.

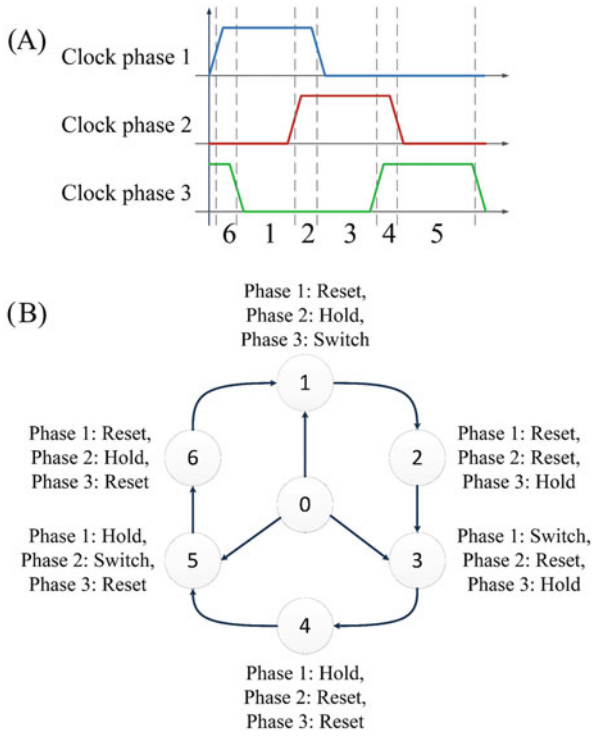


Fig. 18. (A) Clock signals. (B) Behavioral simulation finite state machine.

HOLD state are left untouched, while all magnets in RESET state are reset. In case of magnets belonging to a clock zone that is in the SWITCH state, the situation is more complex because the calculation of the state of each magnet requires a specific algorithm.

5.2 Matrix Exploration

For each clock zone where the SWITCH state is active, the simulation matrix must be scanned and the new value of magnetization of each magnet must be calculated. The matrix exploration algorithm is based on two nested loops which scan every column of a zone and for each of them scan every row.

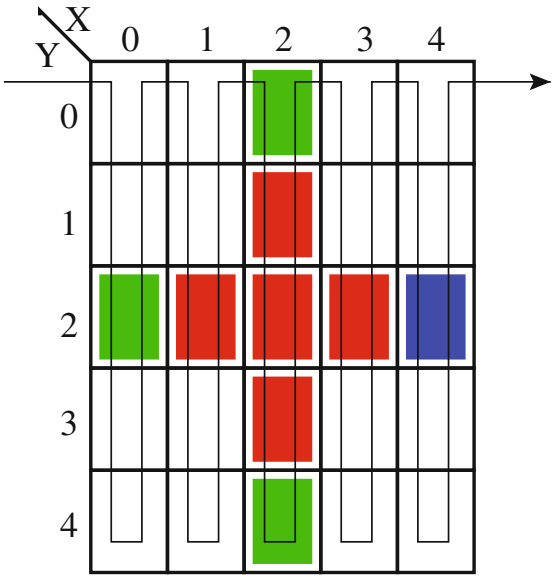


Fig. 19. Exploration matrix algorithm. Matrix columns are scanned one by one from left to right, each row of the column is scanned from up to down and then from down to up.

The behavior of the matrix exploration algorithm is depicted in Fig. 19. Each column of the matrix is scanned starting from the left border going toward the right border. For each column rows are scanned one by one from up to down and then from down to up. During the matrix visit for the new magnet state is evaluated following the magnetization algorithm (Sect. 5.3). This matrix exploration algorithm is chosen to reduce the usage of *if..then...else* constructs inside the programming code of ToPoliNano, to reduce their huge negative impact on performance when the code is executed in modern superscalar machines.

5.3 Magnetization Evaluation

During the matrix exploration the magnetization of every magnet is calculated according to the value of its neighbors, as shown in Fig. 20. The state of every magnets is calculated by the weighted sum of its neighbors.

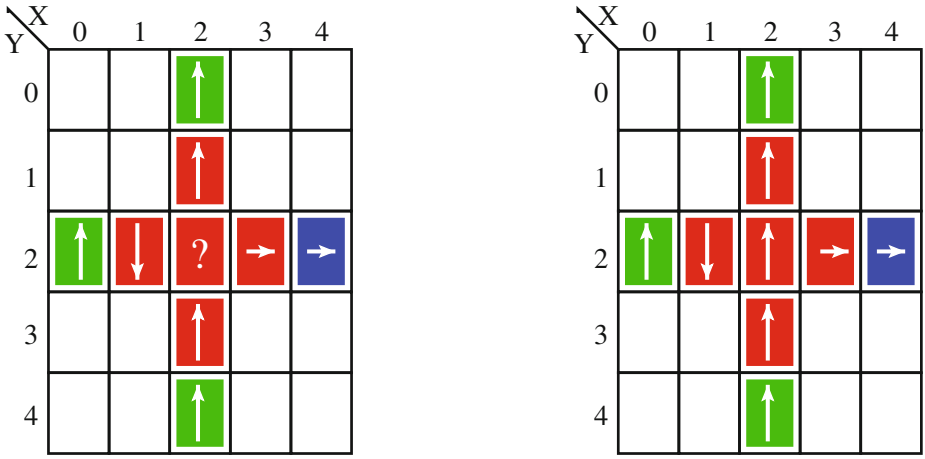


Fig. 20. Magnetization evaluation. The state of a magnet is calculated as a weighted sum of its neighbor elements.

6 Library Generator

The library generator is a tool that, together with the main application of ToPoliNano, allows the designer to describe new elementary components in order to be used within the internal library of the simulator. Its structure is shown in Fig. 21.

The library generator creates new components that can be used in ToPoliNano. However to create a component it needs three kind of information:

- *Ports* which contains information on the input/output pins. Every pin is identified by three parameters:
  - Direction: Input or Output.
  - Type: Standard Logic, Standard Logic Vector. This is required for compatibility with VHDL code, so that the created component can be identified by the parser.
  - Position within the circuit.
- *Layout* which contains information on the space occupied by the circuit.
- *Element Description* which contains the circuit layout, the gates which compose it and their position on the plane.

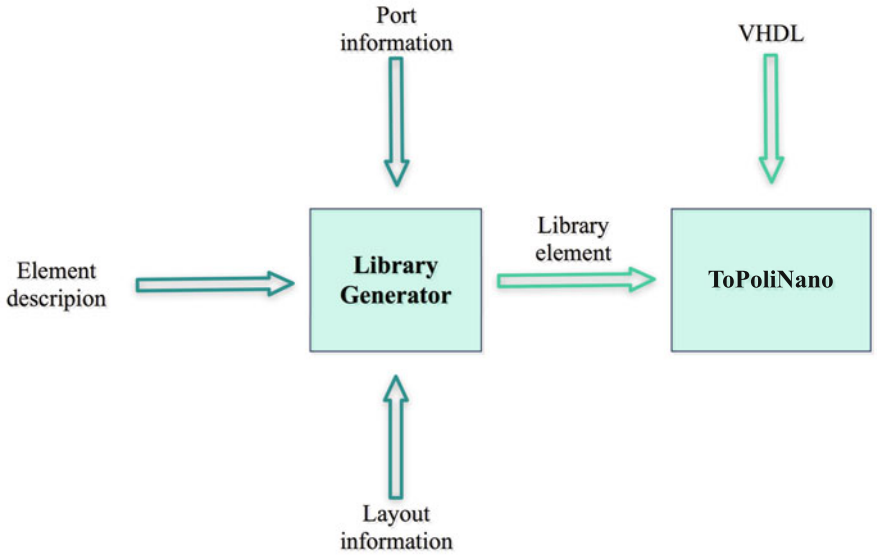


Fig. 21. Library Generation principle. To create a particular component the library generator requires three types of information. The component generated by the library generator can be used inside ToPoliNano.

The *Element Description* description of every component represents the circuit layout. It can be generated automatically by the *Place & Route* block or it can be created manually. There are two possibilities to manually describe a component. With the first possibility the user can draw circuit blocks using a vector graphic program called *Xfig*. The circuit drawn with *Xfig* is exported to an .svg file and loaded inside ToPoliNano. As a second option the circuit layout can be described manually writing the code which describes it. As a future work we should create an in-program editor of circuits. The layout is based on different logic gates defined as library components (e.g. majority voter, cross wire, wire, and, or).

7 Output Generation

At the end of the simulation, ToPoliNano offers two main ways to represent results.

- On the main program window displaying the output signal waveforms that can be saved on an Encapsulated PostScript file (EPS).
- A text file containing the timing samples that the depict the circuit behavior.

An example of output waveforms obtained after the simulation is reported in Fig. 22.

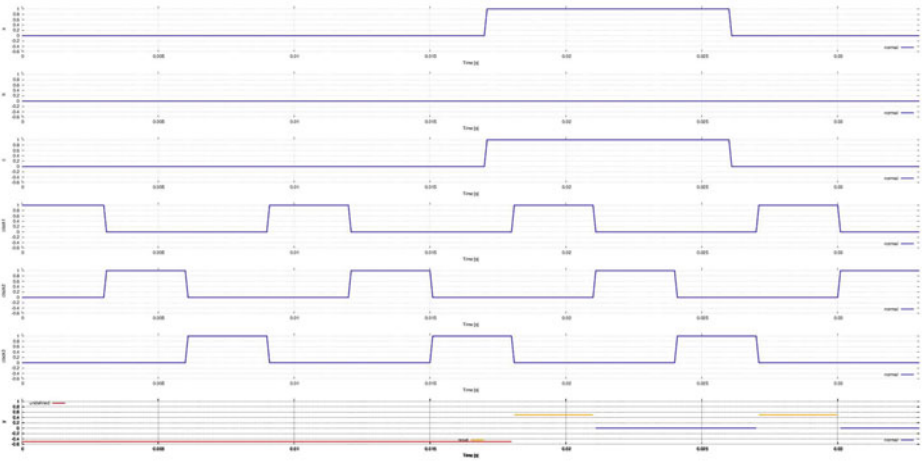


Fig. 22. Output waveforms example.

As well as the circuit timing behavior, additional information can be obtained, like circuit area, number of magnets, wasted area. Moreover a power estimation is also possible, up to now only considering the magnetic field clock.

8 Performance

One of our main targets during the development of ToPoliNano was to design a tool capable to handle high complexity circuits with reasonable execution times. While the software is still in development we were able to make some preliminary evaluations on performance. The tests are based on currently available machines, with *I-3*, *I-5* and *I-7* Intel processors, running both Linux and Mac-OS operating systems. As a benchmark we have used a simple Ripple Carry Adder, made by N full adders. Just for test purpose we have instantiated up to 10000 full adders. The placement of all magnets (around 2000000) took only 30 s. The in-memory occupation of such a circuit was just about 1.5 Gb. The simulation of one full adder for a time period of 80 µs with a simulation step of 1 ps, required just 0.3 s. To compare the performance of ToPoliNano with existing tools we have performed some simulations with two widely used micromagnetic simulators, NMAG [30] and OOMMF [29] on the same machine used for the testing of ToPoliNano. The simulated structure was a simple NML wire in three different cases, changing the length from 4 magnets, to 8 magnets and finally to 12 magnets. Results in terms of simulation time and memory usage are reported in Table 1. The time indicated is the machine time required to advance the state of the circuit of 1 ns. For example, in case of the 4 magnets wire simulated through NMAG, 43 s of machine-time are required to advance the state of the circuit of 1 ns. The memory usage and the simulation time increases with the circuit complexity. Starting from these values it is possible to get a rough estimation

Table 1. Simulation performance of two of the most widely used free micromagnetic tools, NMAG and OOMMF.

	Simulator	Memory [Mb]	Time [s/ns]
4 magnets wire	NMAG	49	43
	OOMMF	24	89
8 magnets wire	NMAG	79	75
	OOMMF	32	173
12 magnets wire	NMAG	104	127
	OOMMF	36	200

of the memory and the simulation time required to simulate a circuit made by around 2 milions of magnets. For NMAG a total of 12 TB of RAM memory and 70000 years of simulation time will be required. OOMMF instead requires only 4 TB of memory and just 35000 years of simulation times. These are just rough estimations but they clearly show the necessity of using a simulator like ToPoliNano to handle complex NML circuits. Comparing the performance of ToPoliNano with Modelsim [34], one of the most used VHDL simulators is more complex. However a simple comparison can be obtained looking at the microprocessor described in [26] and [27], or at the systolic array for biosequences analysis described in [55–57], which are also made by millions of magnets. In that case the simulation for a time of 100 µs required near 1–2 h of machine-time. Clearly ToPoliNano shows an advantage also over classical VHDL simulators.

Finally the Ripple Carry Adder was used also to test the performance of the Place & Route block, but only up to 32 bits. Using the most simple cross-wires minimization techniques, the *Barycenter*, the creation of the layout required only 20 ms. Optimizing the layout with the best technique available, the *Simulated Annealing*, the time required to synthesize a 32 bit adder was equal to few minutes.

9 General Code Structure

ToPoliNano has been developed in C++ language to be highly efficient in computation, thus providing ground to study multiple emerging technologies, demanding in terms of computational resources. With regard to the implementation, the C++ language was selected as programming language mainly for performance reasons, and the whole structure of the simulation is based on the use of classes grouped into macro-blocks:

- Controllers, which contains all the high-level logic of the application.
- GUI, which contains all the classes related to the management of graphical interface (configuration wizard, main window).
- HDL Graph Controller, about the logic of HDL parsing.
- Inputs And Clocks, for the generation of input and clock signals.

– NML, which contains all the classes for the implementation of the simulation algorithms and data structures of NML technology.
– NanoArray: which contains all the classes for the implementation of the simulation algorithms and data structures of NanoArray technology.

ToPoliNano has been designed as a single-source, cross-platform application. Currently, these are the supported platforms:

– Windows.
– Linux/X11.
– Mac OS X.

ToPoliNano is supported on a variety of 32-bit and 64-bit platforms.

– Ubuntu Linux 10.04 (32-bit).
– Ubuntu Linux 11.10 (64-bit).
– Microsoft Windows XP SP3 (32-bit).
– Microsoft Windows 7 (32-bit).
– Apple Mac OS X 10.6 "Snow Leopard" (64-bit).
– Apple Mac OS X 10.7 "Lion" (64-bit).
– Apple Mac OS X 10.8 "Mountain Lion" (64-bit).

ToPoliNano has been built on top of existing cross-platform frameworks. This enabled a development focused on core functionalities, rather then spending time re-creating commodity software. The framework used are well respected among developer communities. In particular, two are the cornerstones of ToPoliNano: the Qt framework and the Boost Library. Lots of resources are available on the web for both of them.

The Boost C++ Library is a collection of free libraries that extend the basic C++ functionality. It provides a wide range of platform agnostic functionality that the Standard Template Library (STL) missed. It can be regarded as a complement to STL rather than a replacement, according to its developers [58]. At the time of this writing, the current Boost release (1.54.0) contains over 80 individual libraries. The Boost C++ libraries are open-source and peer-reviewed.

The choice of the Boost library is due to the very high quality of the code and to its optimal performances. The use of such a library can speed-up the initial development: bugs are minimal, there is no need to "reinvent-the-wheel" and the maintenance cost is reduced to a bare minimum. The Boost libraries are mostly header-only, making them trivial to install, upgrade and configure. In most cases, installation and upgrading only requires the addition or the modification of the include path. Only few of them requires a full compilation. Moreover, Boost sub-libraries can be used independently of each other.

Nowadays computer architectures feature multicore designs, thus giving a calculator the opportunity to perform multiple tasks simultaneously. Not only operating systems can take advantage of the multiple cores in computers, but also applications. This is know as *multithreaded programming*. ToPoliNano supports multithreading to exploit the potential of modern microprocessor and to speed-up circuits simulation.

10 Other Supported Technologies

ToPoliNano was developed to support different technologies in one tool. Up to now it enables the study of NanoMagnet Logic and NanoArray, a promising technology that is briefly outlined in what follows. The NanoArray architecture is based on an intense research activity aimed at understanding the emerging nanotechnological devices and their constraints of realization. A NanoArray project is hierarchical, consisting of interconnected *tiles* that determine the architecture (Fig. 23) and it is optimized to meet the various constraints that come from the construction process, from materials and devices.

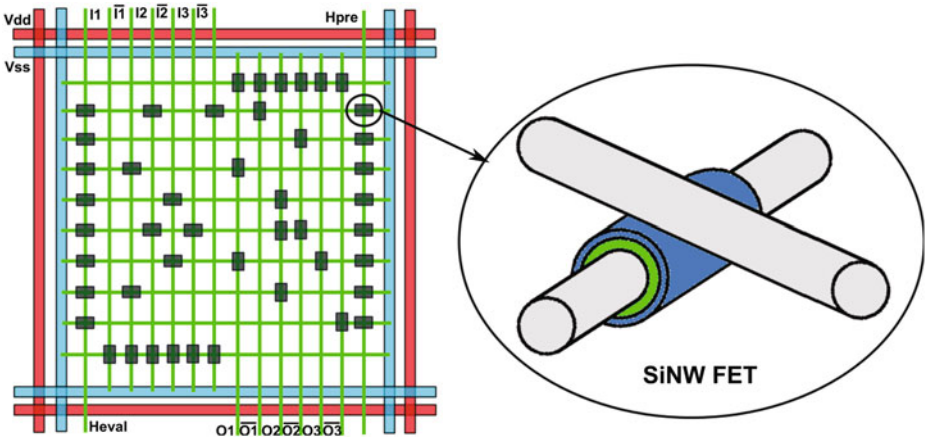


Fig. 23. Example of NanoArray made by Silicon Nanowires (SiNW) and SiNW FETs.

Among the devices that NanoArray architecture can use, there are FETs or diodes in 2-D structures based on crossed SiNWs to realize logical functions (Fig. 23); the various types of optimizations to overcome the limitations imposed by the layout, by the constructive process and by defects adopted in this approach differ from those enforced by other architectures.

The logic AND-OR (or equivalent according to DeMorgan) is carried out either statically, or dynamically. In a dynamic circuit style pipelining of circuits becomes possible. Signals like GND, Vdd, Vss and dynamic control signals necessary for operation are transported at micrometric scale by so-called microwires. The defects are masked directly into the circuit or architecture, without recurring to reconfigurability. The peculiarities of this approach can be summarized in four points:

– redundancy is added in each stage of AND-OR logic and outputs, in addition to the dual-rail redundancy; signals are intercalated and then combined as part of the AND & OR logic planes;

- modification of the logic of one stage to allow the masking of the defects in the current stage or, alternatively, in the following AND-OR stage;
- addition of wires with function of weak pull-up or pull-down in the device to maintain at a low logic level inputs potentially faulty (prior to an OR plane), or at a high logic level (prior to an AND plane);
- joint operation with majority voting circuits based on CMOS at key points of the architecture.

Nanotiles. The nanotiles represent the building blocks of NanoArray architecture. The crossing nanowires form a nanoarray, whose junctions (points of intersection) can be FETs or not connected. The nanoarrays are surrounded by microwires, that carry electrical power and, in this particular implementation, also signals for programming the interconnections. Each signal is present both in its original form and in complemented form.

Dynamic Pipelining. Due to design constraints (e.g. doping of the nanowire) and to topological constraints too, the latches are very difficult to implement within the nanoarray. Typically, latches or registers are used to obtain the functionalities of pipelining of data streams, for example in a datapath. This component is one of the most common within the microprocessor, so it is important to have an efficient way of pipelining. In NanoArray, when using circuits of the dynamic type, it is possible to obtain a temporary storage of information, without resorting to the explicit use of a latch. A pipelined NanoArray circuit can be realized by cascading dynamic nanoTiles, without requiring an explicit latching of the signals, which would entail a considerable reduction of density. A processor made in nanotechnology could have thousands of nanoTiles, so it is fundamental to have an efficient communication system, and this is one of the critical points in nanoarchitectures. In the NanoArray approach local communication occurs between nanowires, to maintain an optimum use of the area, while for the global communications microwires are used.

NanoArray Simulation. The NanoArray simulation engine belongs to the class of the event-driven simulators. It follows the information flow inside the structure under simulation and generates specific events, when necessary, to correctly handle the propagation of information. To better understand this process, we can imagine each sub-tile as a four-port device, with each port identified by a cardinal point.

A change in the information at a given port may need to be propagated inside the sub-tile, if there is an appropriate component to support propagation (e.g. a nanowire). We do not need to know anything about the electrical properties of the component to perform a logical analysis. As a function of the port at which the change in information happens, and the original direction of propagation of this piece of information, we can check whether there is support for further propagation and, if this is the case, to change the information on another port

of the sub-tile by means of the supporting element. This, in turn, will trigger an update event over the sub-tile, if any, connected to the first one by means of the output port. By following the very same process, the information is propagated inside the structure, only where it is needed.

There could be an active device inside the sub-tile, and the propagation of information could lead to a change in its status. Should this happen, another kind of event would be enqueued in the event queue, waiting to be processed to take into account a possible change of information in a direction of propagation that is orthogonal with respect to the one that originated the event. This approach is very flexible, indeed, because it allows for different kind of control of dynamic circuits (number of phases) since the phase sequence is not embedded into the simulator but is coded in the input control sequence and the same approach can thereby be used in many different scenarios.

References. The literature about NanoArray is quite wide, here some references are reported for interested readers. In [59,60] some of the work of University of Massachusetts Amherst is reported. Other solutions were proposed by Likharev [61], Dehon [62], a group of HP [63] and the Carniege Melon University [64].

11 Conclusions

In this chapter we have described ToPoliNano, a tool which aim is to design and simulate NML circuits and other emerging technologies following the same top-down methodology used in the CMOS case. This tool allows to easily describe and simulate complex NML circuits without loosing important details like the placement of each magnets, as it happens in case of VHDL modeling, and without the limitations in terms of speed and circuit sizes of low level simulators.

The tool is still in development so it does not have all the planned functionalities but we are still working hard on completing and improving it. Particularly our efforts are oriented in three directions: Custom circuits description, Place & Route algorithm and simulation engine.

While the automatic layout generation is a very important feature, we acknowledge that also the manual custom circuit description plays an important role. We are therefore trying to improve the means available to describe circuits manually, for example providing an interface that allows to import circuits designed with QCADesigner, opportunely converted substituting magnets to the QCA cells used by QCADesigner. This can represent an important step in the tool development, since many researchers in the QCA community use QCADesigner. We are also working to modify the program to be released to other people that work in this field, so that they can fully exploit the advantages given by ToPoliNano.

The Place & Route algorithm is the program core and also the most complex part. We are working in two directions, to improve the algorithm allowing it to handle also sequential circuits and implementing a floorplanning algorithm to create the layout of complex circuits with a hierarchical structure. Sequential

circuits require careful handling, because signals must be carefully synchronized. We would like to support in the future, multiple clock zones layout types and more NML types. Finally we are working with the aim of improving the simulation engine supporting other types of NML.

Overall it is possible to conclude that ToPoliNano has a great potential that must be still exploited, and we are doing our best toward this direction.

References

1. Porod, W.: Magnetic Logic Devices Based on Field-Coupled Nanomagnets. Nano and Giga, Tempe (2007)
2. Imre, A.: Experimental study of nanomagnets for quantum-dot cellular automata(MQCA)logic applications. Ph.D. thesis, University of Notre Dame, Notre Dame, Indiana (2005)
3. Lent, C., Tougaw, P., Porod, W., Bernstein, G.: Quantum cellular automata. Nanotechnology **4**, 49–57 (1993)
4. Csurgay, A., Porod, W., Lent, C.: Signal processing with near-neighborcoupled time-varying quantum-dot arrays. IEEE Trans. Circ. Syst. **47**(8), 1212–1223 (2000)
5. Awais, M., Vacca, M., Graziano, M., Masera, G.: FFT implementation using QCA. In: 2012 19th IEEE International Conference on Electronics, Circuits and Systems, ICECS, pp. 741–744 (2012)
6. Graziano, M., Chiolerio, A., Zamboni, M.: A technology aware magnetic QCA NCL-HDL architecture. In: 9th IEEE Conference on Nanotechnology. IEEE-NANO 2009, Genoa, Italy, July 2009, pp. 763–766. IEEE (2009)
7. Chiolerio, A., Allia, P., Graziano, M.: Magnetic dipolar coupling and collective effects for binary information codification in cost-effective logic devices. J. Magn. Magn. Mater. **324**(19), 3006–3012 (2012)
8. Pulimeno, A., Graziano, M., Piccinini, G.: Molecule interaction for QCA computation. In: IEEE International Conference on Nanotechnology, pp. 1–5 (2012)
9. Pulimeno, A., Graziano, M., Demarchi, D., Piccinini, G.: Towards a molecular QCA wire: simulation of write-in and read-out systems. Solid State Electron. **77**, 101–107 (2012)
10. Pulimeno, A., Graziano, M., Saginario, A., Cauda, V., Demarchi, D., Piccinini, G.: Bis-ferrocene molecular QCA wire: ab-initio simulations of fabrication driven fault tolerance. IEEE Trans. Nanotechnol. **12**(4), 498–507 (2013)
11. Das, J., Alam, S., Bhanja, S.: Ultra-low power hybrid CMOS-magnetic logic architecture. IEEE Trans. Circ. Syst. **59**(9), 2008–2016 (2012)
12. Karunaratne, D., Bhanja, S.: Study of single layer and multilayer nano-magnetic logic architectures. J. Appl. Phys. **111**, 07A928 (2012)
13. Niemier, M., et al.: Nanomagnet logic: progress toward system-level integration. J. Phys. Condens. Matter **23**, 493202 (2011)
14. Graziano, M., Vacca, M., Zamboni, M.: Magnetic QCA Design: Modeling, Simulation and Circuits. Cellular Automata - Innovative Modelling for Science and Engineering, InTech (2011). http://www.intechopen.com/articles/show/title/magnetic-qca-design-modeling-simulation-and-circuits
15. Vacca, M., Vighetti, D., Mascarino, M., Amaru, L., Graziano, M., Zamboni, M.: Magnetic QCA majority voter feasibility analysis. In: 2011 7th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME), pp. 229–232 (2011)

16. Vacca, M., Graziano, M., Zamboni, M.: Majority voter full characterization for nanomagnet logic circuits. IEEE Trans. Nanotechnol. **11**(5), 940–947 (2012)
17. Csaba, G., Porod, W.: Simulation of filed coupled computing architectures based on magnetic dot arrays. J. Comput. Electron. **1**, 87–91 (2002). (Kluwer)
18. Graziano, M., Vacca, M., Chiolerio, A., Zamboni, M.: A NCL-HDL snake-clock based magnetic QCA architecture. IEEE Trans. Nanotechnol. **10**(5), 1141–1149 (2011)
19. Niemier, M., Hu, X., Alam, M., Bernstein, G., Porod, W., Putney, M., DeAngelis, J.: Clocking structures and power analysis for nanomagnet-based logic devices. In: International Symposium on Low Power Electronics and Design, Portlan-Oregon, USA, pp. 26–31. IEEE (2007)
20. Ju, X., Niemier, M.T., Becherer, M., Porod, W., Lugli, P., Csaba, G.: Systolic pattern matching hardware with out-of-plane nanomagnet logic devices. IEEE Trans. Nanotechnol. **12**(3), 399–407 (2013)
21. Das, J., Alam, S., Bhanja, S.: Low power magnetic quantum cellular automata realization using magnetic multi-layer structures. J. Emerg. Sel. Top. Circ. Syst. **1**(3), 267–276 (2011)
22. Chung, T., Keller, S., Carman, G.: Electric-field-induced reversible magnetic single-domain evolution in a magnetoelectric thin film. Appl. Phys. Lett. **94**, 132501 (2009)
23. Csaba, G., Porod, W.: Behavior of nanomagnet logic in the presence of thermal noise. In: International Workshop on Computational Electronics, Pisa, Italy, pp. 1–4. IEEE (2010)
24. Ercan, I., Anderson, N.: Heat Dissipation Bounds for Nanocomputing: Theory and Application to QCA (2011)
25. Alam, M., Siddiq, M., Bernstein, G., Niemier, M., Porod, W., Hu, X.: On-chip clocking for nanomagnet logic devices. IEEE Trans. Nanotechnol. **9**(3), 348–351 (2010)
26. Graziano, M., Vacca, M., Blua, D., Zamboni, M.: Asynchrony in quantum-dot cellular automata nanocomputation: elixir or poison? IEEE Des. Test Comput. **28**(5), 72–83 (2011)
27. Vacca, M., Graziano, M., Zamboni, M.: Asynchronous solutions for nano-magnetic logic circuits. ACM J. Emerg. Tech. Comput. Syst. **7**(4), (2011)
28. Walus, K., Dysart, T., Jullien, G., Budiman, R.: Qcadesigner: a rapid design and simulation tool for quantum-dot cellular automata. IEEE Trans. Nanotechnol. **3**(1), 26–31 (2004)
29. Donahue, M., Porter, D.: OOMMF User's Guide, Version 1.0. Technical Report Interagency Report NISTIR 6376, National Institute of Standards and Technology, Gaithersburg, September 1999
30. Fischbacher, T., Franchin, M., Bordignon, G., Fangohr, H.: A systematic approach to multiphysics extensions of finite-element-based micromagnetic simulations: Nmag. IEEE Trans. Magn. **43**(6), 2887–2889 (2007)
31. Scholz, W., Fidler, J., Schrefl, T., Suess, D., Dittrich, R., Forster, H., Tsiantos, V.: Scalable parallel micromagnetic solvers for magnetic nanostructures. Comput. Mater. Sci. **28**, 366–383 (2003)
32. Vacca, M., Graziano, M., Zamboni, M.: Nanomagnetic logic microprocessor: hierarchical power model. IEEE Trans. VLSI Syst. **21**(8), 1410–1420 (2012)
33. Niemier, M., Varga, E., Bernstein, G., Porod, W., Alam, M., Dingler, A., Orlov, A., Hu, X.: Shape engineering for controlled switching with nanomagnet logic. IEEE Trans. Nanotechnol. **11**(2), 220–230 (2012)

34. Mentor graphics. http://www.modelsim.com
35. Awais, M., Vacca, M., Graziano, M., Masera, G.: Quantum dot cellular automata check node implementation for LDPC decoders. IEEE Trans. Nanotechnol. **12**(3), 368–377 (2013)
36. Graziano, M., Frache, S., Vacca, M., Turvani, G., Riente, F., RuoRoch, M., Zamboni, M.: Topolinano: Torino politecnico nanotechnology tool (2012). http://www.topolinano.polito.it
37. Vacca, M., Frache, S., Graziano, M., Zamboni, M.: ToPoliNano: a synthesis and simulation tool for NML circuits. In: IEEE International Conference on Nanotechnology, pp. 1–6, August 2012
38. Frache, S., Chiabrando, D., Graziano, M., Riente, F., Turvani, G., Zamboni, M.: ToPoliNano: nanoarchitectures design made real. In: IEEE International Symposium on Nanoscale Architectures NANOARCH, pp. 160–167 (2012)
39. Frache, S., Chiabrando, D., Graziano, M., Graziano, M., Boarino, L., Zamboni, M.: Enabling design and simulation of massive parallel nanoarchitectures. J. Parallel Distrib. Comput. **74**, 2530–2541 (2014)
40. Chiswell, I., Hodges, W.: Mathematical Logic. Oxford Texts in Logic, vol. 3. Oxford University Press, Oxford (2007)
41. Vacca, M., Frache, S., Graziano, M., Crescenzo, L., Cairo, F., Zamboni, M.: Automatic Place&Route of nano-magnetic logic circuits, pp. 58–63 (2013)
42. Ravichandran, R., et al.: Partitioning and placement for buildable QCA circuits. In: DAC, vol. 1 (2005)
43. Chung, W., et al.: Node duplication and routing algorithms for quantum-dot cellular automata circuits. IEEE Proc. Circ. Dev. Syst. **153**(5), 497–505 (2006)
44. Kernighan, B.W., Lin, S.: An efficient heuristic procedure for partitioning graphs. Bell Syst. Tech. J. **49**, 291–307 (1970)
45. Ravikumar, Si.Pi., Ravikumar, C.P.: Parallel Methods for VLSI Layout Design. Greenwood Publishing Group, Boston (1995)
46. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **4598**(220), 671–680 (1983)
47. Cerny, V.: Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. J. Optim. Theory Appl. **45**, 41–51 (1985)
48. Sechen, C., Sangiovanni-Vincentelli, A.: The Timberwolf placement and routing package. IEEE J. Solid-State Circ. **20**(2), 510–522 (1985)
49. Wang, D.: Novel routing schemes for IC layout part I: two-layer channel routing. In: Design Automation Conference (1991)
50. Sherwani, N.: Algorithms for VLSI Physical Design Automation. Springer, Heidelberg (2002)
51. Kahng, A., Lienig, J., Markov, I., Hu, J.: VLSI Physical Design Automation: From Graph Partitioning to Timing Closure. Springer, Heidelberg (2011)
52. Deutsch, D.: A dogleg channel router. In: Proceedings of 19th Design Automation Conference (1976)
53. Yoshimura, T.: An Efficient Channel Router (1984)
54. Sau, A., Pal, A., Mandal, T., Datta, A., Pal, R., Chaudhuri, A.: A Graph based Algorithm to Minimize Total Wire Length in VLSI Channel Routing (2011)
55. Urgese, G., Graziano, M., Vacca, M., Awais, M., Frache, S., Zamboni, M.: Protein Alignment HW/SW Optimizations. In: The IEEE International Conference on Electronics, Circuits, and Systems (ICECS), pp. 145–148 (2012)
56. Wang, J., Vacca, M., Graziano, M., Zamboni, M.: Biosequences analysis on NanoMagnet Logic. In: International Conference on IC Design and Technology, pp. 131–134, May 2013

57. Graziano, M., Frache, S., Zamboni, M.: A hardware viewpoint on biosequence analysis: what's next? ACM J. Emerg. Technol. Comput. Syst. **9**(4), 29 (2013)
58. Free peer-reviewed portable C++ source libraries
59. Wang, T., Narayanan, P., Moritz, C.A.: Heterogeneous 2-level logic and its density and fault tolerance implications in nanoscale fabrics. IEEE Trans. Nanotechnol. **8**(1), 22–30 (2009)
60. Comparison of Analog and Digital Nano-Systems: Issues for the Nano-Architect (2008)
61. CMOL: devices, circuits, and architectures (2004)
62. DeHon, A., Wilson, M.: Nanowire-based sublithographic programmable logic arrays (2004)
63. Snider, G., Kuekes, P., Williams, R.: CMOS-like logic in defective, nanoscale crossbars. Nanotechnology **15**, 881–891 (2004)
64. Goldstein, S., Budiu, M.: NanoFabrics: spatial computing using molecular, electronics. In: Proceedings of the 28th Annual International Symposium on Computer Architecture (2001)

Understanding a Bisferrocene Molecular QCA Wire

Azzurra Pulimeno, Mariagrazia Graziano(✉), Aleandro Antidormi, Ruiyu Wang, Ali Zahir, and Gianluca Piccinini

Dipartimento di Elettronica e delle Telecomunicazioni,
Politecnico di Torino, Turin, Italy
mariagrazia.graziano@polito.it

Abstract. Molecular QCA are considered among the most promising beyond CMOS devices. Frequency as well as self-assembly characteristics are the features that make them most attractive. Several challenges restrain them for being exploited from a practical point of view in the near future, not only for the difficulties at the technological level, but for the inappropriateness of the tools used when studying and predicting their behavior.

In this chapter we describe our methodology to simulate and model sequences of bisferrocene molecules aimed at understanding the behavior of a realistic MQCA wire. The simulations consider as variables distances between successive molecules, as well as different electric field applied (in terms of input and of clock). The method can be used to simulate and model also other more complex structures, and perspectives are given on the exploitation of the achieved results.

1 Introduction on Molecules for QCA

1.1 Molecular Implementation

Molecules, with their moving charges, present well-suited properties to perform information storage and processing in the ways expressed by the QCA paradigm [1]. Molecular charges can localize, under the influence of an external field, in determined locations of the molecule, resulting in different spatial charge configurations. Thus, molecular systems can be exploited as basic cells for binary computing: each configuration encodes a binary state and a switch from a state of the cell to another results from a charge flow through tunneling paths within the molecule itself. Molecular sites where charge localization occurs are called redox centers and represent the dots of the QCA device; they are effectively capable of attracting or releasing an electron thus becoming negatively or positively charged.

Since a molecule is neutral, a better performance is obtained with the oxidized and reduced forms of the same molecule; in the oxidized molecule, an electron is missing and its net charge is positive; in the reduced form, the molecule has gained an electron picking up a negative net charge.

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 307–338, 2014.
DOI: 10.1007/978-3-662-43722-3_13, © Springer-Verlag Berlin Heidelberg 2014

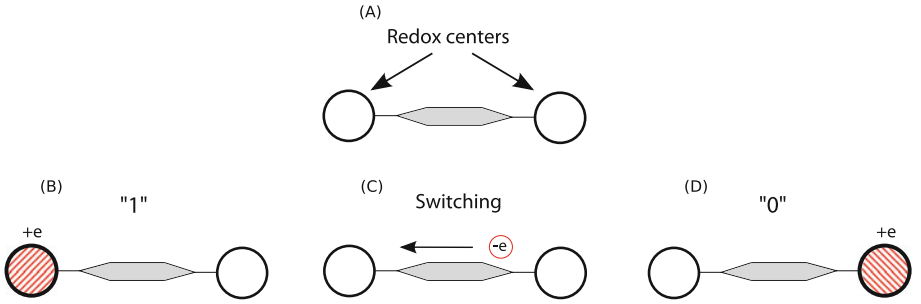


Fig. 1. Molecular implementation: (A) molecule sketch; (B)–(D) logic encoding and switching in case of oxidized molecule.

A simplified scheme of an oxidized molecule is shown in Fig. 1: an electron is free to move along the tunneling path and localize into one of the two redox sites available (the circles in this simplified representation). Encoding of binary information is related to the charge configuration of the cell: for example, when the positive charge is in the redox site on the left (the circle with internal pattern in this simplified representation, Fig. 1(B)), the cell is, for example, in a logic 1 state. When the positive charge is located on the right site (patterned circle on the right, Fig. 1(D)), then the molecule is in the 0 state. These charge configurations are represented in Fig. 1, along with the switching state corresponding to the free charge moving along the tunneling path from one dot to the other (Fig. 1(C)). The intermediate symbol (stretched hexagon) represents a part of the molecule slightly active from the electrostatic point of view, but acting like a channel to favor charge movement and as separator to favor charge localization in one of the two dots.

Fig. 2. State encoding of a 3-dot molecule.

The encoding scheme chosen in the previous example is ostensibly arbitrary, since other choices are possible. No choice can be made, instead, on the number of different possible configurations of charge within a given molecule, and correspondingly on the number of encoded states. The motivation for having more charge configurations in a single molecule resides in the mode of switching adopted in circuits with many QCA cells. A mode of switching consists in a procedure with several steps which allow the QCA cells to switch their state

thus transmitting or processing the information. To avoid metastability problems, the so-called adiabatic mode of switching has been devised and a single cell must dispose of six dots [2–4]. Consequently a molecule must have three redox centers, as sketched in Fig. 2. The third redox center is used to encode the logic NULL state which corresponds to the charge occupying the central dot. A molecule with three active dots reproduces a half- QCA cell and can be completed to form a whole cell aligning another molecule on its side (Fig. 3). The two half-cells interact through the electrostatic forces among the free charges which tend to localize within each molecule in order to minimize the relative potential energy. This configuration clearly recalls the theoretical QCA structures based on squared cells able to capture charges in the dots along the diagonal of the square [1].

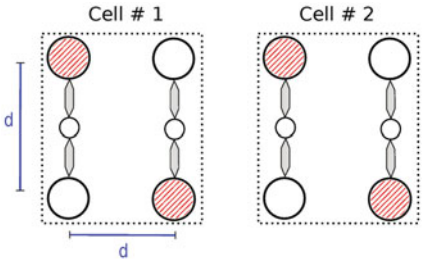


Fig. 3. Complete QCA cell and interaction between cells (top view).

1.2 Candidates Molecules

On the applicative point of view, the greatest effort in developing molecular QCA cells is in the quest for molecules with proper features as computing elements. The ideal molecule, indeed, encompasses many properties which reveal to be strictly necessary.

Many molecules present geometrical structures which limit the molecule itself to be used only in the gas phase or in solution; this is mainly due to the lack of a binding element for deposition on substrates. Consequently, practical QCA molecules should pose the functional groups that allow attachment and orientation on a surface.

As already mentioned before, a suitable molecule should have at least three redox centers in order to encode the NULL state, introduced for energy and stability reasons.

Finally, a simple and reliable mechanism is necessary to set and read molecular states at the edge of a molecular array. Devising solutions to this problem is a major issue, given the actual absence of instruments capable of revealing the motion of single charges on nanometer scale.

Many molecules have been presented and analyzed. Most of them are ideal [5–7,10,11,16,17,24] and their functionalities have been tested via computer simulation but never synthesized. Others, instead, are real molecules synthesized ad hoc for QCA purposes [8,9,18,19,25,26] and for some of them even early preliminary experiments have been carried out.

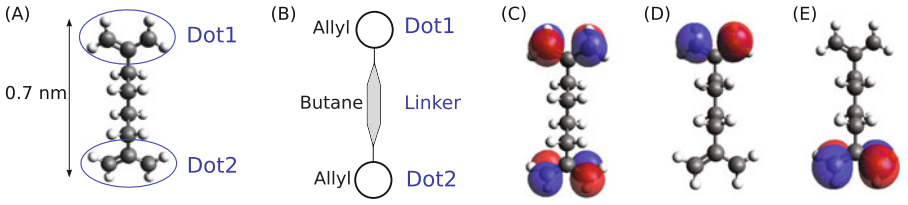


Fig. 4. The diallyl butane molecule: (A) structure, (B) scheme and (C)–(E) orbital localization.

The first molecule which has been proposed for QCA computing is the diallyl butane [7]. As shown in Fig. 4, it consists of two allyl groups connected by a butane bridge. The most suitable form for the molecule is the cationic one, where a positive charge is free to move inside the molecule. The corresponding unpaired electron can occupy one of the opposite allyl end-groups, which represent the dots (circled in Fig. 4(A) and sketched in Fig. 4(B)); the tunneling path between these redox centers is consequently given by the butane bridge. When the charge tunnels from one end to the other, a different charge configuration of the molecule is obtained and a different highest occupied molecular orbital (HOMO) is realized (Fig. 4(C)–(E)).

A molecular orbital is an eigenfunction of the Hamiltonian operator for a molecular system, corresponding to a determined value of energy (eigenvalue) of the molecule. According to quantum mechanics, this function is related to the spatial probability of finding an electron in a specific region of space with that energy. Thus, from a molecular orbital the most probable location of an electron can be evaluated. Moreover, each molecular orbital can be occupied by two electrons with opposite spin and, in general, all the electrons tend to arrange themselves in order to fill the orbitals starting from the one with the lowest energy. In this sense, the HOMO is the last energy level occupied by the available electrons and it has the highest energy.

In the diallyl butane cation an electron from the HOMO is removed, consequently the HOMO represents the localization of the unpaired electron. The possible HOMO conformations are shown in Fig. 4(C)–(E): in Fig. 4(C) the HOMO is symmetrically delocalized between the two allyl groups favoring the electron occupation neither of the top nor of the bottom group: it corresponds to an undefined state. In Fig. 4(D) and (E) instead, the HOMO (and thus the charge) is localized on one of the two dots, so these configurations could represent the logic states 1 and 0.

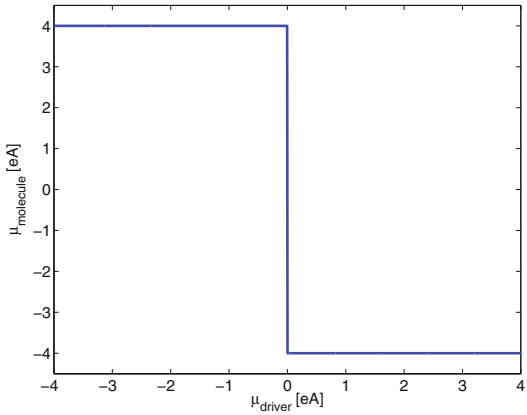


Fig. 5. The single-molecule response of the diallyl butane molecule to a point charge driver.

All these HOMO configurations differ for the dipole moment generated in the molecule from the charge localization: a positive moment in Fig. 4(D), a negative one in Fig. 4(E). Such a quantitative parameter, evaluated multiplying the electron positive charge and its distance from the center of the molecule, offers a way to analyze the molecular response when a change in the Coulomb field produced by neighboring molecules occurs. A QCA molecule is indeed expected to respond in non-linear way to the external perturbations, switching from one state to another as demonstrated in [7]. This nonlinear behavior is shown in Fig. 5 where the dipole moment of the molecule is shown as a function of the external dipole moment, considered as input of the system. The dipole of the molecule has opposite sign with respect to the driver dipole moment; thus the molecule interacts properly with the neighboring molecules assuming the opposite state.

The diallyl butane molecule constitutes only a half QCA cell; a whole QCA device can be obtained aligning two molecules along the axis thus forming a squared structure. Although very performing as a QCA device for its good charge confinement and non-linear molecular response, the diallyl butane is not suitable for real application. It lacks the possibility of encoding a NULL state for clock issues and has no binding element to be placed on a substrate following a specific layout. Other molecules should be devised which could overcome these limitations.

The decatriene [11] is one possible candidate to overcome the deficiencies of diallyl butane: its molecule presents three dots (a sketch of its structure is reported in Fig. 6). Each dot is an ethylene groups (the circle) and equivalently to the diallyl butane, no element is unfortunately able to allow attachment of the molecules on a surface.

The redox centers of the decatriene confine charge in the molecule according to three configurations (schematically reproduced at the bottom of Fig. 6): a simple correspondence can, then, be established among these charge

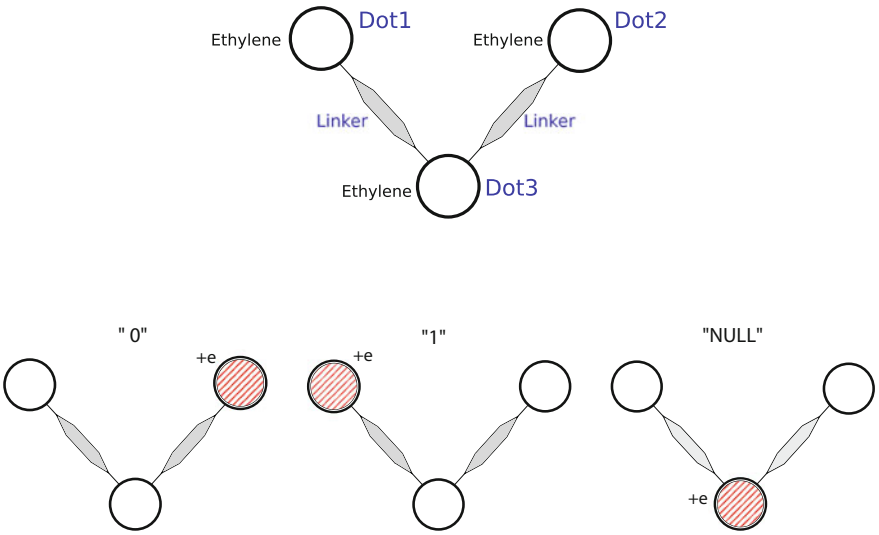


Fig. 6. The decatriene: structure (top) and state encoding (bottom).

conformations and logic states 0, 1 and NULL. In the ground-state, with no electric field applied, the molecule configuration is the one encoding the NULL state. When an electric field is applied in the vertical direction (called clock field) the molecule ends up in an excited state: a simultaneous application of another electric field parallel to the active dot axis (called switching field) implies the charge localization on one of the two dots, depending on the sign of the switching field, leading the molecule into the logic state 0 or 1. This organization recalls the behavior of the other QCA implementation currently considered as promising, the magnetic one [12,13]. As demonstrated in [14,15] for the case of nano magnetic Logic, respecting these constraints paves the way to the feasibility of complex circuits, once technology will be mature enough to manage the fabrication of molecular QCA structures.

One of the real molecules proposed as QCA device is a mixed-valence complex based on one iron atom (Fe) and one ruthenium atom (Ru) [8]. The molecule schematic structure is shown in Fig. 7(A): the two dots are vertically organized and are given by the metal atoms, while the ending nitrogen atom (N) plays the role of binding element allowing the molecule to be attached to a silicon substrate. Its functionality properties have been verified placing the oxidized form of the molecule between a silicon surface and a mercury tip, thus realizing a plate capacitor. Applying different voltages (electric fields) the free charge in the molecules moves from one dot to the other generating a change in the differential capacitance. By measuring a change in the differential capacitance an experimental demonstration of the charge motion inside the molecule has been given as a consequence of the external field (Fig. 7(B)).

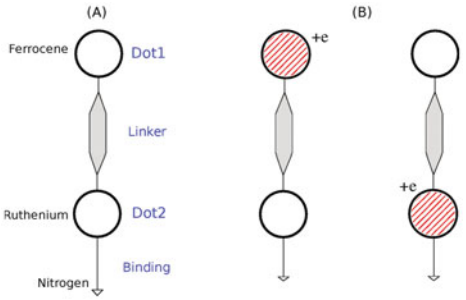


Fig. 7. The Fe-Ru mixed valence complex: (A) structure and (B) state encoding.

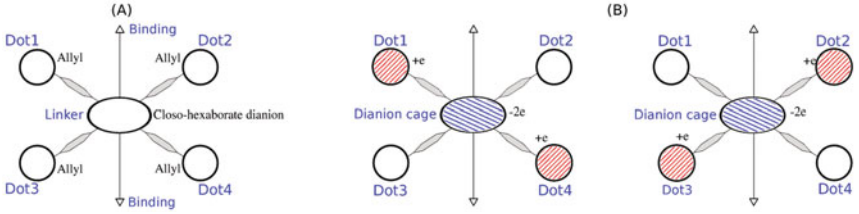


Fig. 8. The mixed-valence zwitterion: (A) structure and (B) state encoding.

Another real molecule proposed in literature is a mixed-valence complex based on iron atoms (Fe) [26]. This molecule is called meta-Fe2: the active dots are represented by two metal-ligand units, each composed of a cyclopentane (Cp) and in iron atom (Fe), while an ethynyl linker connects them to a central benzene ring. Scanning Tunneling Microscope (STM) imaging has revealed that the presence of perturbations in the molecular environment leads to a charge confinement inside the molecule in cationic form.

All the molecules already discussed need to be oxidized or reduced in order to employ their nonzero net charge to realize different charge configurations encoding binary information. As a consequence, particular attention has to be payed to control redox processes. The zwitterion, a molecule which incorporates a donor or acceptor site, permits to overlook such difficulties. This special site releases or attracts an electron, generating a free mobile charge in the molecule while preserving the overall neutrality.

A practical example of mixed-valence zwitterion based on boron clusters has been proposed as QCA device [16]. The octahedral cage of a closo-hexaborate dianion (sketched in Fig. 8(A)) is used as central linker group for four redox centers, while other two axial linker could be used to bind the molecule on a substrate. The boron cage attracts two electrons from two antipodal active dots, thus becoming doubly negatively charged; two choices of opposite dots are possible, subsequently two binary states can be encoded, as shown in Fig. 8(B).

Differently from the previous molecules, this molecule implements a complete QCA cell.

1.3 The Bisferrocene Molecule

In the continuous research for molecules suitable to be employed in QCA computing, recently a new candidate molecule has been proposed and studied: the bis-ferrocene molecule [18–23]. This molecule, shown in Fig. 9 is constituted by two ferrocenes, functioning as redox centers, and a carbazole bridge which acts as third dot for the NULL state. An alkyl chain is also present, which guarantees the attachment to a gold surface. The bis-ferrocene molecule is capable of realizing a QCA half-cell, thus two molecules aligned are necessary for an entire cell.

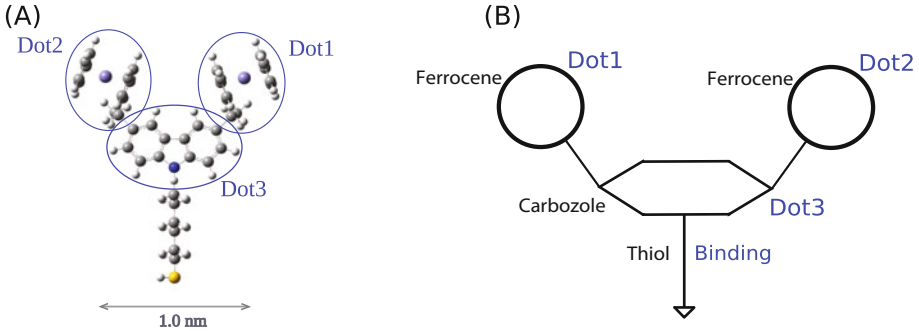


Fig. 9. The bis-ferrocene molecule: (A) structure and (B) scheme.

Its features have been demonstrated through theoretical and experimental analysis [18,19]. Three isomers of the molecule have been synthesized: two diastereoisomers (the (S,S) and (R,R), chiral compounds) and the meso isomer. Oxidation has also been performed chemically, with the introduction of iodine atoms (I) as counterions in the solution, and electrically, through cyclic-voltammetry (CV).

The possibility of binding bisferrocene molecules on a gold surface has also been demonstrated, as proved by the results of STM imaging shown in [19]. STM characterization has revealed the presence of structures with one and two lobes with almost the same density; further analysis has allowed to associate them with the (R,R and S,S) stereoisomers and (R,S) meso isomer, respectively.

The capacity of the bis-ferrocene molecule to correctly respond to a change in local electric field has been verified on the (R,S) *syn* isomer in its oxidized form [19]. In particular, after computing the electron density for an isolated molecule (the driver) in its cationic state, the electron density of a second molecule (the driven molecule) was deduced in presence of the point-charge electrostatic field of the first molecule. As expected, the external field seems to move the internal positive charge on the ferrocenes generating different charge configurations [19].

2 Methodology for the Analysis of Molecule and Ensamble of Molecules

This section describes a possible methodology to evaluate and simulate all the important parameters necessary for molecular QCA devices. It aims to represent both the bis-ferrocene features as QCA devices and the feasibility of computing with molecular QCA. Additionally, the methodology for analyzing defects concerning QCA wire fabrication is also provided.

The study is organized in two stages, as shown in the scheme of Fig. 10: **Stage I**, where ab-initio simulations are performed under different conditions as detailed in Sect. 2.1; and **Stage II**, where electrostatic equations and models are set up in order to define new figures of merit to describe the QCA molecular system (see Sect. 2.2) and to understand its potentials from an electronic point of view.

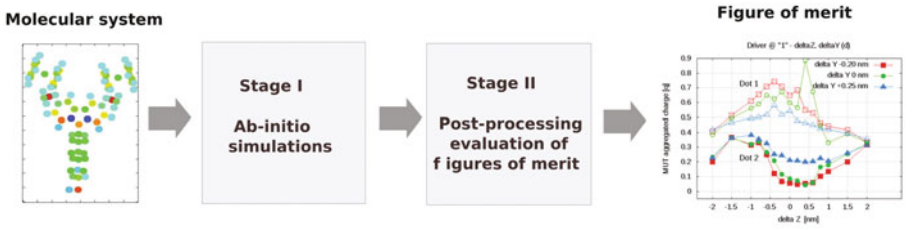


Fig. 10. A two stages analysis from molecular systems to device-level figures of merit.

2.1 Stage I – Ab-initio Simulations

The computations and evaluations of the electronic structure and properties we performed via ab-initio simulations (see Fig. 11) are mainly based on the laws of quantum mechanics and a set of physical constants (the speed of light, masses and charges of electrons and nuclei, Plank's constant).

They are based on the theory from the first principle and are highly accurate and, as a consequence, computationally intensive. In order to characterize a molecule as QCA device, a set of ab-initio simulations could be performed on different conditions of molecular systems, by defining proper methods and basis sets.

For all the simulations of the bis-ferrocene molecule the DFT-B3LYP theory was used with the LANL2DZ basis-set.

S I-A – Single Oxidized and Neutral Molecule Simulation. The binary information of the molecular QCA system could be evaluated by using the figure of merit of the highest occupied molecular orbital (HOMO), as mentioned in [7,10,11]. While based on this, in order to take clearly the electronic point of

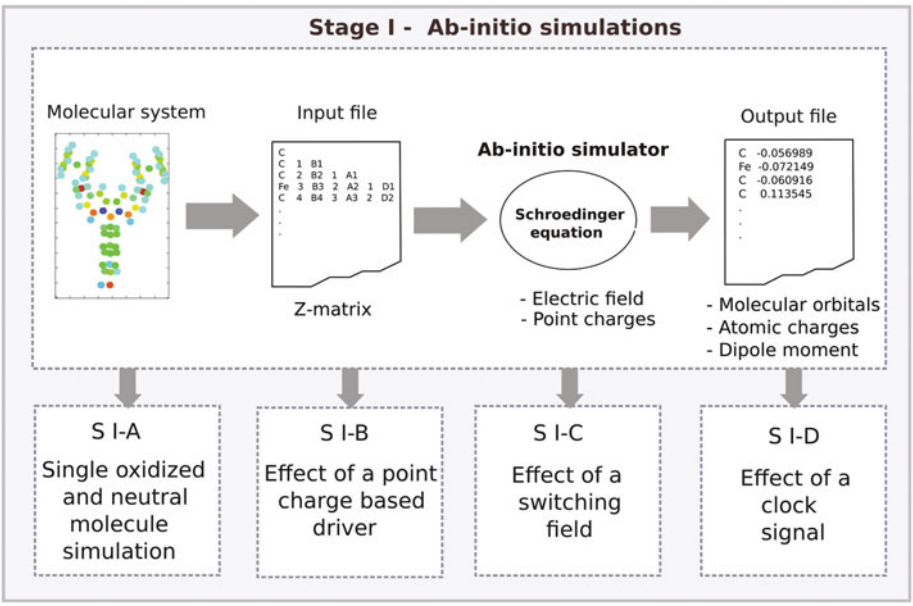


Fig. 11. Stage I – Ab-initio simulations: from molecular system chemical inputs to molecular system energetic and electrostatic quantities.

view about the molecular QCA, new figures of merit could be used to analyze the QCA molecule, both in its neutral and oxidized or reduced form. In particular, the charge distribution of the molecule has to be evaluated when the molecule is subject to a write-in system (like an electric field as described in [21]).

S I-B – Effect of a Point Charge Based Driver. Regarding the interaction between molecules, the methodology discussed here involves the simulation of a system considered as a complete QCA cell: it contains an ideal driver represented by point charges located at the distance d from the molecule as shown in Fig. 12(A). In this way, the ideal driver could emulate the other molecule of the complete QCA cell. In the case of neutral molecule, the point charges of the driver molecule are two and opposite (which are $+1e$ and $-1e$), placed respectively at the position of the two working dots, imitating an internal charge splitting that preserves the neutrality of the driver molecule. Thus, putting the polarized ideal driver next to the molecule and changing the state of the driver (swapping the position of the two point charges), it is possible to check whether or not the target molecule changes its logic state as well.

On the other hand, in the case of oxidized molecule, only one positive charge ($+1e$) is considered to simulate the ideal driver. In particular, the point charge is located on one of the two dots of the driver, according to its logic state. In this way, considering the case of driver at logic 1 (as represented in Fig. 12(B)), the charge distribution on the target molecule could be evaluated by means

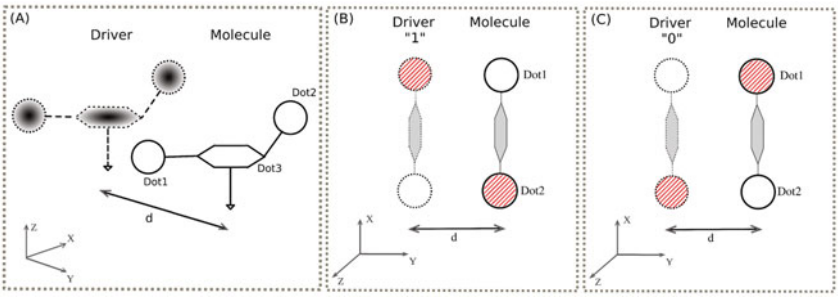


Fig. 12. Model of driver-molecule interaction.

of ab-initio simulations. Then, switching the driver to a logic 0 (as represented in Fig. 12(C)) the target molecule is expected to switch, as well, to the opposite state.

The amount of charge mentioned here is always referred to the elementary charge (e) as unit of measure and hereinafter the e will be omitted for sake of brevity.

S I-C – Effect of Switching Field. As already demonstrated in [21], the electric field generated by the write-in system of the molecular QCA cells is called *switching field* and the direction of the switching field is considered to be parallel to the working dot axis. The electric field should force the charge to localize in one of the two active dots, thus writing a logic state in a QCA molecule. By changing the sign of the switching field, the charge localization could be led in the opposite dot and, as a consequence, the molecule should change its logic state. The write-in system could be modeled and simulated applying a uniform electric field to the molecule as shown in Fig. 13 and the effectiveness of this method could be evaluated computing the charge distribution of the molecule: considering for example the oxidized bis-ferrocene and applying a positive switching field (Fig. 13(A)) the free positive charge should localize on Dot2; while in case of negative switching field (Fig. 13(B)) the charge should be on Dot1.

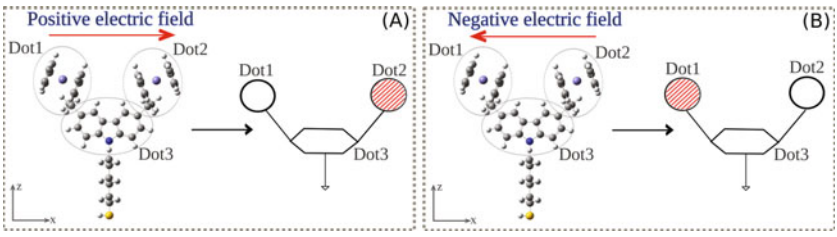


Fig. 13. Model of write-in system.

S I-D – Effect of Clock Signal. As mentioned before, in order to realize the adiabatic switching, the clock signal should be taken into account in the simulation. It is important to analyze the possibility of enhancing or hindering the communication between nearby molecules. In case of three-dot molecules, the clock signal could be implemented by applying an electric field along the vertical axis of the molecule (like the Z axis as illustrated in Fig. 14(A) for the bis-ferrocene molecule). The analysis could be performed also in presence of a polarized driver for both neutral and oxidized form of the target molecule. This means simulating the molecule under the simultaneous effects of a clock signal and a polarized driver, in order to evaluate its capability to interact with a nearby molecule.

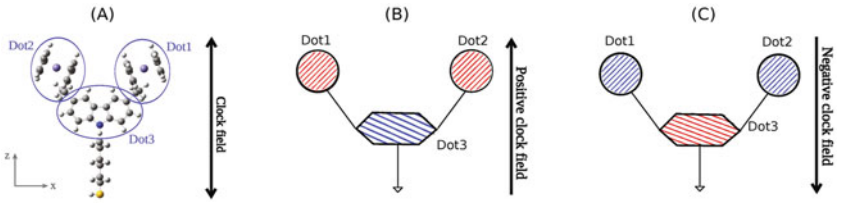


Fig. 14. Application of a clock signal to the bis-ferrocene molecule.

2.2 Stage II – Post-processing

Figure 15 gives an overview of the second stage of our analysis. Starting form the ab-initio simulation results obtained in the various conditions, we elaborate methods to reckon more usable quantities defined in the following.

S II-A – Charge Analysis. In order to model the molecular QCA system from an electronic point of view, it is still necessary to evaluate the results of the ab-initio simulations by defining new figures of merit. Thus, instead of focusing only on the HOMO of the molecule (as mainly done in the chemical approach proposed in literature), a new figure of merit, the *atomic charge* of the molecule, could be considered in different operating conditions (ground state and biasing). The atomic charge could be computed by means of ab-initio simulations using the Merz-Singh-Kollman (MK) approximation scheme (also called ESP) [27].

Furthermore, a new quantity defined as *aggregated charge* could be computed, based on the given atomic charges of the molecule. Particularly, the charge of the entire working dots (the two ferrocenes and the carbazole) and of the thiol group are computed simply summing the atomic charge of all the atoms that form each redox center. Figure 16 illustrates the methodology chosen to calculate the aggregated charges: the left picture shows the bis-ferrocene atoms (filled circles) colored in different ways according to the atomic charge of each atoms; summing all the atom charges included in each part of the molecule (the dots circled in the figure and the thiol), it is possible to model the molecule with a

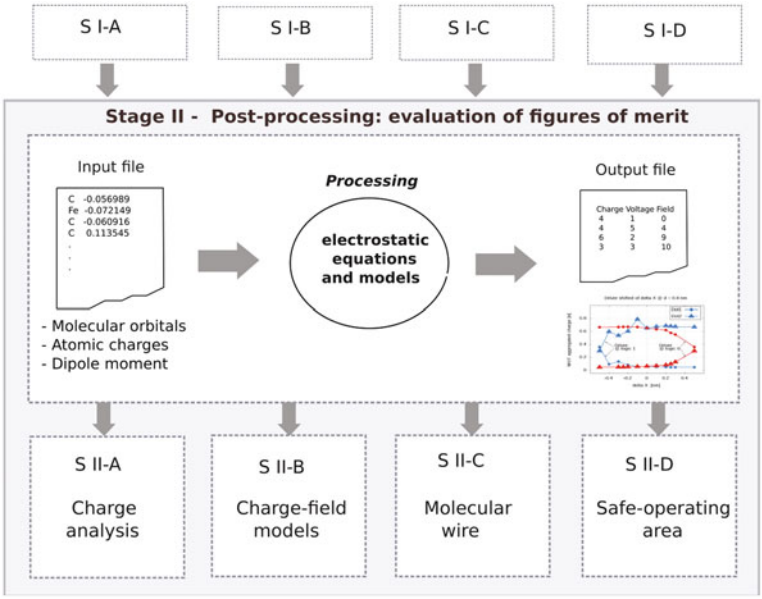


Fig. 15. Stage II – post-processing: from ab-initio to device level figures of merit.

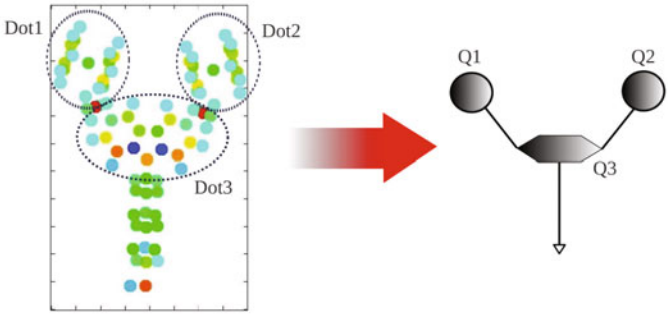


Fig. 16. Definition of the aggregated charge for each dot (Color figure online).

system of point charges. In Fig. 16, only the three dots are considered to model the charge distribution of the molecule and so the bis-ferrocene is modeled with three charges (Q1, Q2 and Q3) corresponding to the three dots (Dot1, Dot2 and Dot3, respectively). This is because the atomic charge itself is a theoretical approximation used to easily demonstrate the charge distributed on the molecule instead of an exact physical quantity. Although since the definition of *aggregated charge* is generated by means of summing the atomic charge in a particular way, it could demonstrate the charge distribution inside a molecule from a macroscopic point of view and at the same time, from the point of view of an application,

it could also be a readable quantity (as discussed in [21]). Hereinafter, when we refer to dot charge we will implicitly refer to the aggregated charges of the corresponding dots.

S II-B – Charge-Field Models. As long as the aggregated charges of the molecule under different bias conditions (point charge based driver, switching field or clock signal) are obtained by means of simulations, it is possible to compute the electric field generated by the charge distribution of the molecule at any specific working point through mathematical equations developed in MatLab/Octave. Particularly, a single charge $+q_1$ is considered to be placed at (x_0, y_0, z_0), and the electric field $E(x, y, z)$ can be evaluated by means of a positive test charge $+q_t$ positioned in (x, y, z). The distance $\overrightarrow{r_1}$ between q_1 and q_t and its modulus r_1 are defined as

$$r_1 = \sqrt[2]{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2} \tag{1}$$

$$\overrightarrow{r_1} = r_1 \cdot \widehat{r} \tag{2}$$

where $\widehat{r}$ is the unit vector of the axis that includes the space point of q_1 and q_t. Following the Gauss law, the Coulomb force $\overrightarrow{F_1}$ experimented by the test charge can be evaluated as

$$\overrightarrow{F_1} = \frac{1}{4\pi\epsilon_0} \frac{q_1 \cdot q_t}{r_1^2} \cdot \widehat{r}. \tag{3}$$

Then, the electric field generated by q_1 and measured by the test charge is computed as

$$\overrightarrow{E_1} = \frac{\overrightarrow{F_1}}{q_t} = \frac{1}{4\pi\epsilon_0} \frac{q_1}{r_1^2} \cdot \widehat{r}. \tag{4}$$

Considering a system of N point charges, the different forces related to the multiple charge distribution of the whole system affect the test charge. This effect could be summarized by defining a total force calculated by

$$\overrightarrow{F_{tot}} = \sum_{i=1}^{N} \frac{1}{4\pi\epsilon_0} \frac{q_i \cdot q_t}{r_i^2} \cdot \widehat{r_i} \tag{5}$$

and so the equation for the total electric field becomes

$$E_{tot}(x, y, z) = \frac{\overrightarrow{F_{tot}}}{q_t} = \sum_{i=1}^{N} \frac{1}{4\pi\epsilon_0} \frac{q_i}{r_i^2} \cdot \widehat{r_i}. \tag{6}$$

Finally, considering a generic test charge placed at a generic point of the space (x, y, z), the three components (E_x, E_y and E_z) of the electric field generated by a system of point charges are given by

$$E_x(x,y,z) = \sum_{i=1}^{N} E_{xi}(x,y,z) \tag{7}$$

$$E_y(x,y,z) = \sum_{i=1}^{N} E_{yi}(x,y,z) \tag{8}$$

$$E_z(x,y,z) = \sum_{i=1}^{N} E_{zi}(x,y,z). \tag{9}$$

Following this procedure, once the aggregated charges of the molecule in a specific condition is obtained, it is possible to compute the electric field generated by this system of charges (Fig. 17(A)) at any point in the space surrounding the molecule. Just as depicted in Fig. 17(B), the electric field is "measured" putting an ideal receiver nearby the molecule at the ideal distance from it to form a squared cell. In order to get the measurement of the electric field, it is necessary to define a new quantity called *equivalent voltage at the receiver* (see Fig. 17(C)). In particular, the equivalent voltage at the receiver could be computed by considering the component of the electric field parallel to the molecule and integrating it along the width of the molecule.

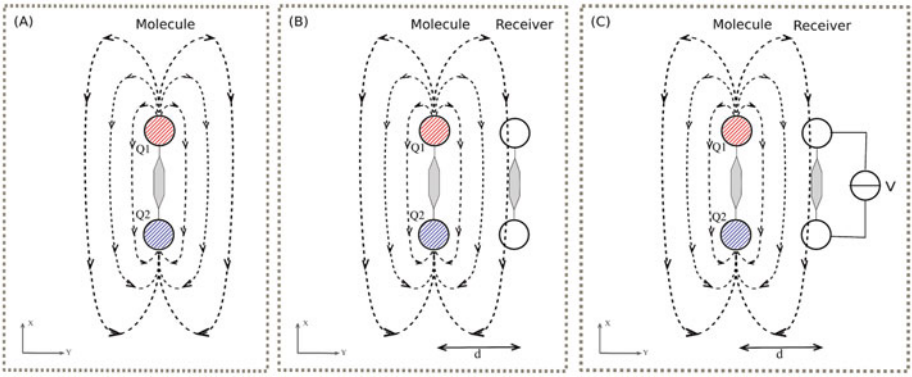


Fig. 17. Definition of the generated electric field and the equivalent voltage at the receiver.

S II-C – How to Simulate a Molecular Wire. After characterizing the single bis-ferrocene molecule as QCA device, it is important to simulate a molecular QCA wire made of oxidized bis-ferrocene molecules. To do this, an iterative method could be adopted as sketched in Fig. 18. In particular, the write-in system is applied simulating the input molecule (Molecule 1 in Fig. 18(A)) under the effect of an electric field (also named switching field). Thus, the logic state of the first molecule is given by the charge distribution and the dot charges are

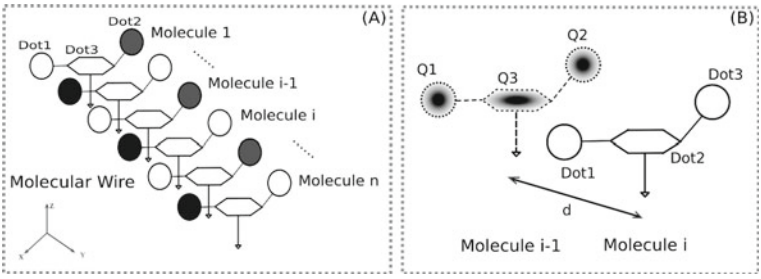


Fig. 18. Simulation of a molecular QCA wire made of bis-ferrocene molecules.

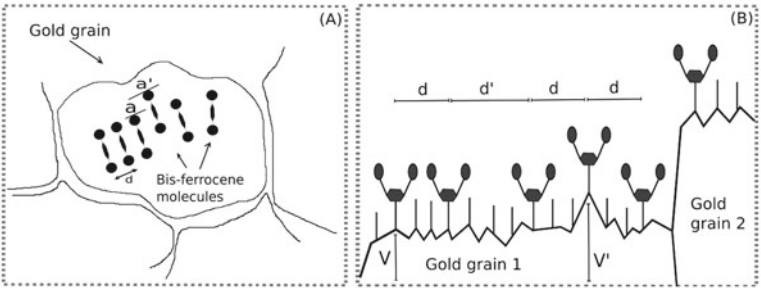


Fig. 19. Fabrication defects in the case of a QCA wire made of bis-ferrocene molecules. (A) Top view of a gold grain after molecule deposition: misalignment and tilt defect are sketched. (B) Section of a bis-ferrocene SAM on gold: vertical shifts may occur due to the roughness inside a gold grain or at the interface of two grains; higher molecule-molecule distance is caused by the number of hexane-dithiol elements.

computed. Then, they could be used as the driver for the second molecule (Molecule 2) of the wire. So, the second bis-ferrocene molecule is simulated in presence of a driver, whose charge distribution emulates the logic state of the Molecule 1. In this way, the charge distribution of Molecule 2 is computed and used as driver for the following cell (Molecule 3) and so on. As depicted in Fig. 18(B), at a generic point of the wire the charge configuration of *Molecule i* is computed as the response to *Molecule i − 1*, assuming that the *Molecule i* is in the neutral state (charge delocalized) at the beginning and ready to switch its logic state under the effect of *Molecule i − 1*. As a result, the aggregated charges (D1, D2 and D3) of the (*Molecule i*) become the driver system of the adjiacent molecule (*Molecule i + 1*) in the following step of information propagation. By iterating this method for all the molecules in the wire, it is possible to simulate the information propagation through the wire.

S II-D – How to Create SOA for Molecules. In order to experimentally demonstrate the QCA functionalities, a bis-ferrocene molecular wire like the one in Fig. 18(A) should be fabricated, using a gold nanowire upon which the

molecules can be bonded through the thiol element [21]. During the technological processes, possible fabrication defects may happen [23]. For this reason, it is important to classify the possible defects and analyze their effects on the molecular wire, generating thus a Safe Operating Area (SOA) that highlights the fault tolerance of the molecular QCA wire.

Defects Classification. As mentioned above, some defects may occur when fabricating the gold substrate and after depositing the molecules on the gold nanowire. These defects are schematically classified in Fig. 19: misalignment of two nearby molecules, due to the gold grain irregularity (Fig. 19(A)); a variation in the vertical distance between the active dot axis of two nearby molecule, as a consequence of the gold layer roughness (Fig. 19(B)); different distances between two nearby molecules, due to the different number of spacers placed in between them (Fig. 19(B)). These defects might cause faults and misbehaviors of the cells in the molecular wire.

Methodology. In order to perform a fault tolerance analysis for a molecular QCA wire, a methodology that could be adopted focuses on a part of the molecular wire made of the following elements: an ideal driver emulated with the point charge system as described before, a bis-ferrocene molecule (defined as molecule under test, MUT) and an ideal receiver that acts as third molecule. In particular, the bis-ferrocene is considered in its oxidized form (a positive net charge) and so the ideal driver is modeled with a single positive point charge. When no defects occur (as in the ideal case), the driver is located at an ideal distance d equal to the distance between Dot1 and Dot2 of MUT (equal to 1.0 nm), so that a squared QCA cell is formed along with the receiver and the MUT. Furthermore, the logic state of the driver is given by the position of the positive point charge as shown in Fig. 20: when it is localized on Dot1, the driver is in the logic state 1 (Fig. 20(A)); while, the charge localization on Dot2 encodes the logic state 0 (Fig. 20(D)). So, the MUT is affected by the presence of a polarized driver and, as consequence, it should encode a logic state opposite to the one of the driver (Fig. 20(B) and (E)). The dot charges of the MUT ($Q1$ and $Q2$) in a specific logic state generate an electric field as shown in Fig. 20(B) and (E), that at the same time affects the receiver through the above mentioned equivalent voltage. Therefore, the effect of the equivalent voltage should be such that it leads the receiver in a logic state opposite to the one of MUT, thus making the information propagate along the wire.

Regarding the real fabrication defects, they could be modeled varying the position of the driver with respect to the MUT as sketched in Fig. 21: the misalignment between the driver and the MUT that occurs along the dot axis is defined as $X\text{-}displ$, since in the ab-initio simulation the two working dots are located along the X axis; the variation of the driver-MUT distance with respect to the ideal distance d is called ($Y\text{-}displ$); the vertical displacement of the driver from the ideal position is defined as ($Z\text{-}displ$).

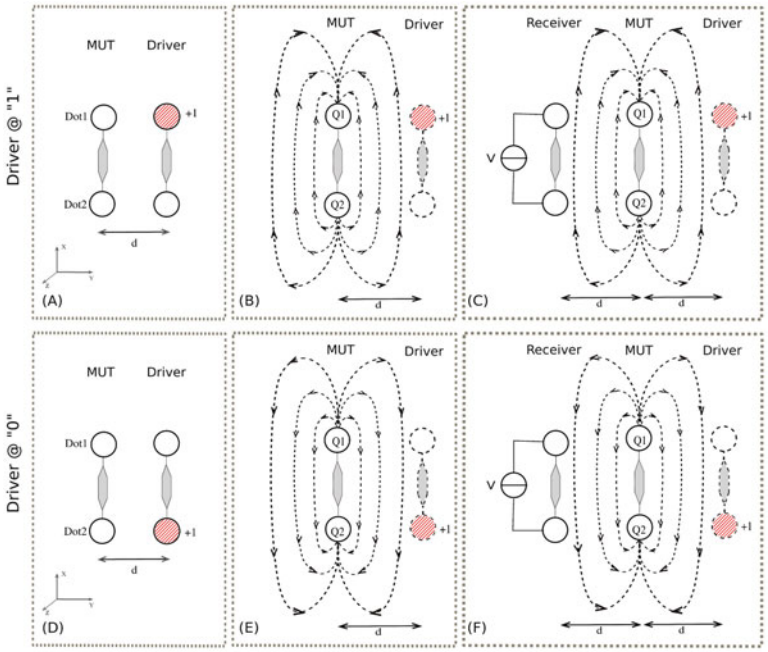


Fig. 20. Methodology scheme for the fault tolerance evaluation in case of molecular QCA wire.

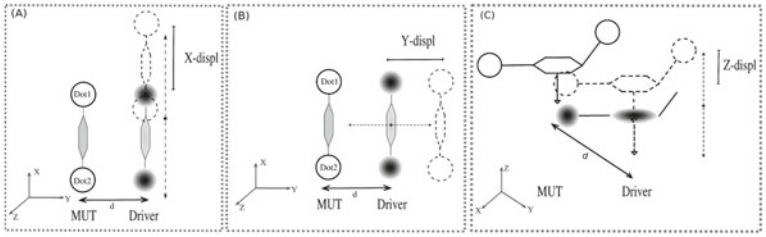


Fig. 21. Models of three possible defects in a QCA wire.

Once the charge distribution of the MUT is obtained for each case of driver displacement, the electric field generated by these charge distributions could be computed and measured at an ideal distance d of 1.0 nm from the MUT molecule, so that an ideal receiver is emulated. Furthermore, the equivalent voltage at the receiver could be also computed to check whether or not the information propagates even in case of fabrication defects. Finally, a *Safe Operating Area (SOA)* could be defined in order to evaluate the fault tolerance of a molecular QCA wire: in the SOA, it is possible to highlight the working points of the receiver that represent all the receiver positions in which a driver displacement

could be tolerated and the communication through the wire is not compromised. These working points could be computed measuring the equivalent voltage at the receiver and checking whether the receiver could still switch its logic state following the changing logic state of the MUT. In particular, a different SOA could be drawn for each driver displacement on a given area of interest centered on the MUT. The electric field generated by the MUT in the two cases of driver logic state is used to compute the equivalent voltage at the receiver, defined as V_{rx1} and V_{rx0}, for the driver logic 1 and 0 respectively. Defining a set of threshold voltages (V_{th}) ranging from 0.1 to 1.0 V, a generic point (x, y) of the area could be considered *safe* for any threshold voltage (V_{th}) that satisfies simultaneously the two following conditions:

$$V_{rx1}(x, y) > |V_{th}| \tag{10}$$

$$V_{rx0}(x, y) < |V_{th}|. \tag{11}$$

3 Simulation Results and Models

3.1 Single Molecule Characterization

The neutral bis-ferrocene molecule in the ground state has the HOMO that is de-localized along the molecule and mainly centered on the carbazole as shown in Fig. 22(A). The application of an electric field along the X axis (switching field) leads to a localization of the HOMO around one of the two ferrocenes, as shown in Fig. 22(B) and (C), depending on the sign of the switching field. Following the methodology adopted in literature [5–7], the HOMO localization could represent the logic state of the molecule and for the bis-ferrocene molecule a write-in system based on electric field seems to work properly [20].

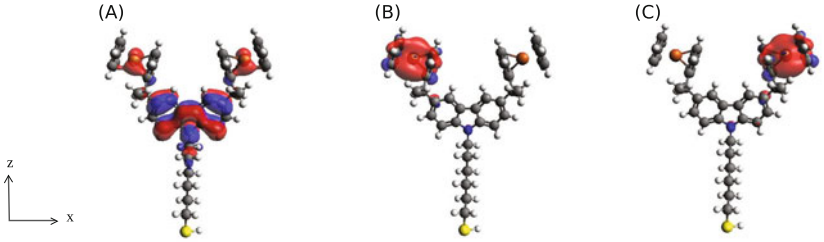


Fig. 22. Bis-ferrocene molecule: HOMO localization at (A) the equilibrium and (B, C) in presence of switching field.

As shown in Fig. 23, applying the switching field the charge distribution inside the molecule varies and the molecule exhibits a charge localization in favor of one of the two dots, depending on the sign of the switching field. In particular, with a positive switching field Dot2 (green curve) becomes more positively charged

than Dot1 (red curve), while the carbazole charge (Dot3, blue curve) does not significantly vary. Changing the sign of the switching field the situation is dual. For both positive and negative applied fields, the displacement of charge between Dot1 and Dot2 is small and this represents a problem in the encoding and identification of the two logic states. Thus, for QCA application it could be necessary to oxidize or reduce the molecule.

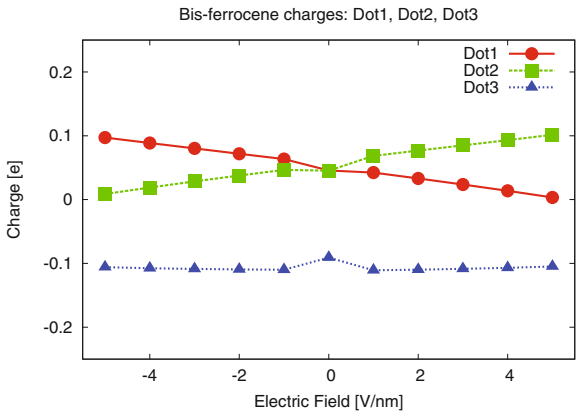


Fig. 23. Bis-ferrocene molecule: dot charges as function of the switching field (Color figure online).

Considering the oxidized version of the bis-ferrocene, at the equilibrium the free positive charge is mainly delocalized between the working dots, while the third dot is almost neutral. The application of the switching field moves the positive charge entirely on one of the two dots, depending on the sign of the electric field, and the third central dot remains almost neutral. The trend of the dot charges of the oxidized bis-ferrocene as a function of the switching field is reported in Fig. 24: the oxidized bis-ferrocene has a non linear behavior, since for a given value of the switching field the charge distribution "saturates", that means that the whole positive charge is confined in one of the two active dots and even increasing the magnitude of the electric field the charge distribution does not significantly vary. The results shown in Fig. 24 are important because they highlight the QCA properties of the bis-ferrocene molecule, as well as the suitability of a write-in system based on electric field.

Given the charge distributions of the oxidized bis-ferrocene in different bias conditions, the generated electric field was computed as second part of the second stage of the analysis flow (post-processing stage described in Sect. 2.2). In particular, the X component (parallel to the dot axis) of the electric field (E_x) generated by the molecule at the equilibrium was computed considering an ideal receiver placed at the distance of 1.0 nm from the molecule along the Y axis, so that a squared QCA cell could be formed. The values of the electric field magnitude are reported in Fig. 25 for all the points of the XZ plain.

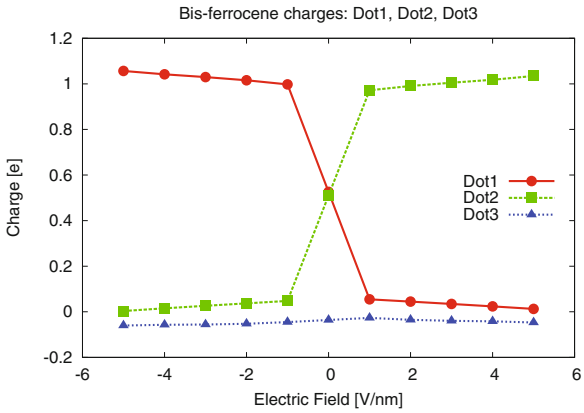


Fig. 24. Oxidized bis-ferrocene molecule: dot charges as function of the switching field.

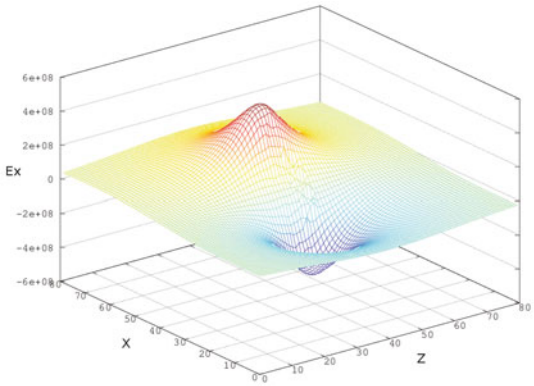


Fig. 25. X component of the electric field generated by the bis-ferrocene at the equilibrium and measured at a distance of 1.0 nm from the molecule.

In the same way, the electric field generated by the molecule under the effect of a switching field was computed, and in order to evaluate the actual effect on an ideal receiver, it is necessary to cut the curves along the axis where the active dots of the receiver stand, that means fixing the position of the receiver on the Z axis since the active dots of both the molecule and the receiver lie along the X direction. Figure 26 shows the results for the molecule at the equilibrium (SW-Field = 0, red line) and in two cases of switching field (SW-Field = −0.5 nm and SW-Field = +0.5 nm, respectively blue and green line). The two vertical violet lines indicate the width of the receiver, that is equal to 1.0 nm (the width of a bis-ferrocene molecule). In order to evaluate the effects of the electric field on the receiver, the curves have to be considered only along the receiver width. At the equilibrium (red curve) the two dots of the receiver (placed at the edge

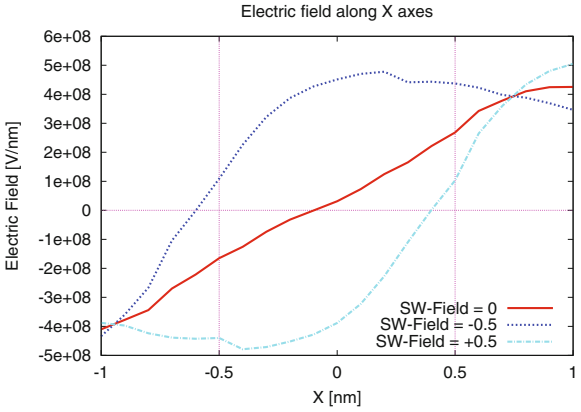


Fig. 26. X component of the electric field generated by the bis-ferrocene in different bias conditions and measured at position of an ideal receiver (Color figure online).

of the receiver and so where the violet lines stand) are subject to an electric field, whose intensity is almost the same, as well as the direction, but the sign is opposite and so the result is almost null. This means that the molecule at the equilibrium does not influence an ideal receiver. For negative applied switching field (blue curve) the positive peak of the generated electric field moves backward and it is centered on the receiver. For this reason, the sign of the electric field that influences the receiver (positive) is opposite with respect to the sign of the switching field applied to the molecule (negative), thus the receiver should encode a logic state opposite to the molecule state. On the contrary, in case of positive switching field (green curve) the situation is dual, since the negative peak of the generated electric field is centered on the receiver.

3.2 Interaction Between Two Molecules

A complete QCA cell could be implemented aligning two bis-ferrocene molecules with an ideal inter-molecule distance equal to the width of the molecule. For this reason, the following step of this analysis concerns the interaction between molecules: the charge distribution of two nearby molecules was evaluated emulating one of the two molecules with a polarized driver, both in case of neutral and oxidized bis-ferrocene.

In case of two neutral molecules placed at a distance of 1.0 nm, the dot charges at the equilibrium are almost zero as reported in Fig. 27. To model the effects of a polarized driver the single bis-ferrocene molecule was simulated in presence of a system of point charges placed at the reference distance from the molecule. Figure 28 reports the results obtained: considering a neutral bis-ferrocene molecule (Fig. 28(A)), the driver was emulated with a positive and a negative charge to maintain the driver neutrality (Fig. 28(B)) and the effect on the target molecule is a slight displacement of charge between the working dots

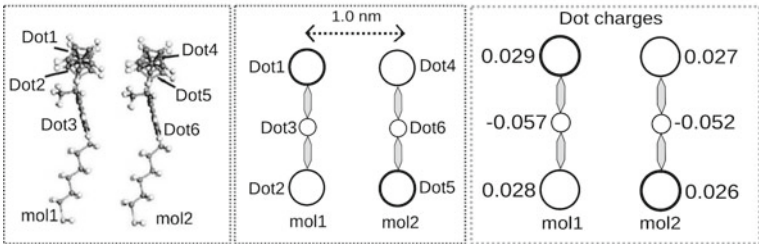


Fig. 27. Charge distribution of two nearby bis-ferrocene molecules (complete QCA cell).

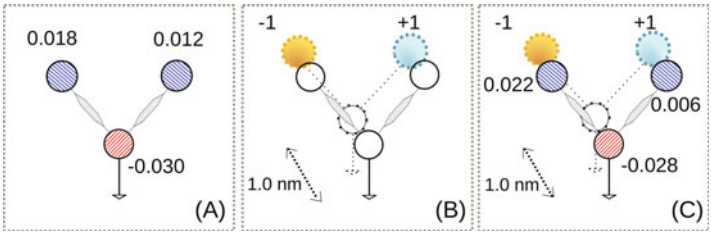


Fig. 28. Charge distribution of a bis-ferrocene molecules in presence of a polarized driver.

(Fig. 28(C)). Swapping the two point charges (that means switching the driver logic state) the situation is dual. This means that the molecule interacts with the driver, but the intrinsic properties of the neutral molecule already discussed in the previous section affect also the driver-molecule interaction.

3.3 Effect of Clock on Molecule Switching

The next step of the analysis was the simulation of a clock signal in order to evaluate the possibility to enhance or hinder the communication between nearby molecules. In the case of bis-ferrocene molecule, the clock signal could be implemented with an electric field applied along the vertical axis of the molecule. The analysis was performed both on the neutral and the oxidized bis-ferrocene and also in presence of a polarized driver. The molecule was simulated with all its elements including the thiol, whose charge is considered as component of the third dot charge.

In case of neutral molecule, the application of a negative clock field generates a displacement of charges inside the molecule and in particular the working dots become negatively charged, while the carbazole charge becomes positive. The charge distribution of the clocked molecule is depicted in Fig. 29(B). On the other hand, the application of a positive clock signal leads to a dual configuration in which the working dots have a positive charge, almost equal, while the carbazole becomes negatively charged (Fig. 29(C)). In this case, the charge distribution

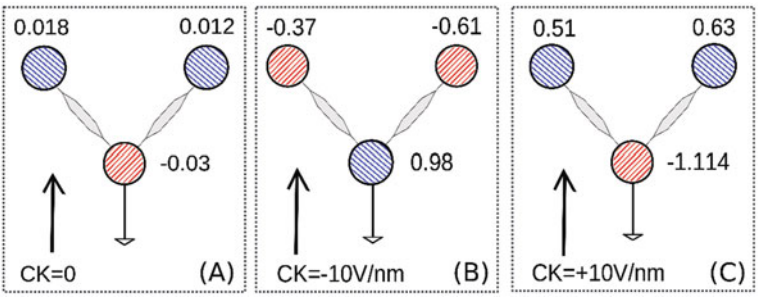


Fig. 29. Charge distribution of a clocked bis-ferrocene molecule.

is similar to the charge distribution of an oxidized molecule at the equilibrium, since the working dots have a similar amount of charge, while the carbazole of the clocked neutral molecule compensates the neutrality of the cell.

To completely characterize the effects of the clock, a polarized driver was introduced during the simulations, in both the logic state 1 and 0. The results are reported in Fig. 30 and show that the clocked molecule reacts better than the non clocked one, since in the former case the displacement of charge between the working dots is more pronounced then in the latter. For this reason, the clock signal could be used to enhance the charge localization inside a neutral molecule and, so, to help the interaction between neighboring cells.

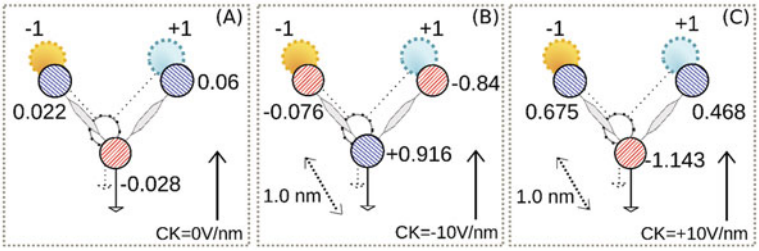


Fig. 30. Charge distribution of a clocked bis-ferrocene molecule in presence of a polarized driver.

Finally, thanks to this analysis it was possible to evaluate the properties of the bis-ferrocene molecule as QCA device. As shown by the results discussed here, the molecule is a suitable candidate for QCA technology, especially in its oxidized form. In fact, the oxidized bis-ferrocene exhibits a good charge confinement when a write-in signal is applied and it reacts properly to a polarized driver. Regarding the neutral bis-ferrocene, its performance are lower than the oxidized version, but the application of a clock signal could enhance the charge confinement inside the molecule and then the interaction with a nearby molecule or a polarized driver.

3.4 Interaction Among Molecules in a Wire

Following the methodology described in Sect. 2.2 and concerning the post-processing stage of the analysis flow, the interaction among molecules in a QCA wire was evaluated, considering two inter-molecule distances: the ideal one (1.0 nm) equal to the width of the bis-ferrocene, so that a squared QCA cell is formed, and a lower distance equal to 0.8 nm. In Fig. 31(A) the charges of the two main dots (Dot1 and Dot2) are reported in case of standard inter-molecule distance: for sake of brevity, only the first part of the wire was reported, considering only five molecules. In this case, the logic state of the molecules alternates along the wire, but the charge displacement between the two dots is smaller and smaller while increasing the number of molecules. In particular the fifth molecule is in an undefined state since the dot charges are almost the same. For this reason, the logic signal could be considered valid only for the first three molecules, while from the fourth molecule on the state is not defined.

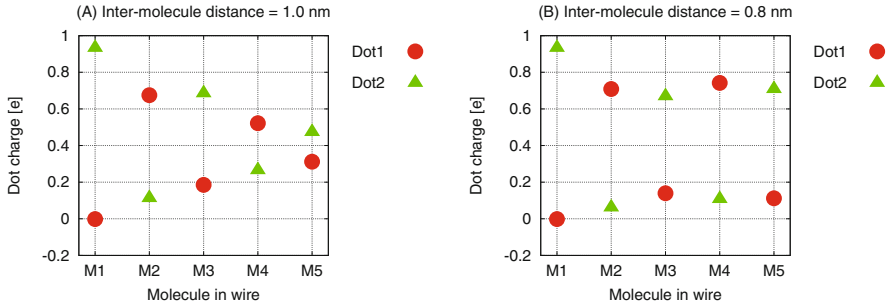


Fig. 31. Molecular QCA wire: dot charges of the molecules along the wire with a molecule-molecule distance equal to 1.0 nm.

On the other hand, when the molecules are placed at 0.8 nm far from each other, the logic state of the molecules along the wire alternates as well, as shown in Fig. 31(B). Moreover, the difference of charge between the two dots is still huge enough to consider all the molecules in a defined logic state.

As additional figure of merit for this analysis, the electric field generated by the molecules along the wire was computed in both the cases of distance. In particular, each molecule along the wire, depending on its logic state, generates an electric field whose absolute value is maximum near the occupied dot. Figure 32 shows a top view of the electric field generated by the first five molecules of the wire, computed at the position of an ideal receiver. The picture shows how the picks of the electric field move following the position of the free charge inside the molecule, but the intensity of the electric field decreases with increasing number of molecules.

On the contrary, in case of distance equal to 0.8 nm, the positive and negative peaks of the electric fields alternate as well (Fig. 33), but the values of the

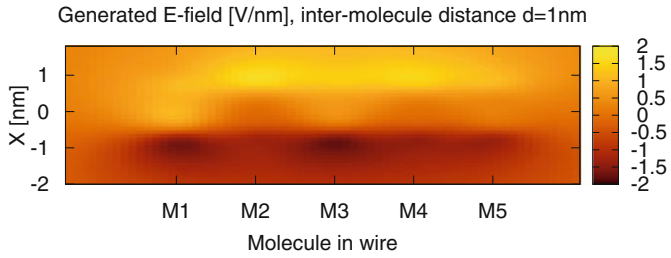


Fig. 32. Molecular QCA wire: electric field generated by the molecules along the wire with a molecule-molecule distance equal to 1.0 nm.

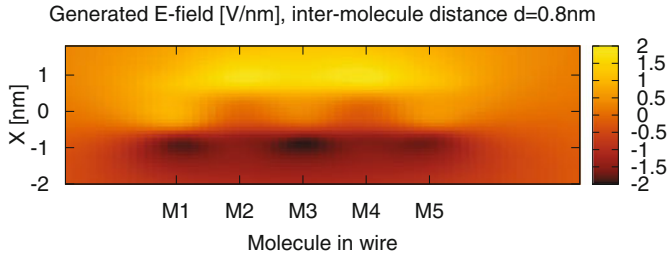


Fig. 33. Molecular QCA wire: electric field generated by the molecules along the wire with a molecule-molecule distance equal to 0.8 nm.

electric field are kept constant for all the molecules, as consequence of the charge distribution inside the molecules.

These results reveal the strength of the logic interaction in two molecular QCA wires. In particular, the wire with the ideal distance for the bis-ferrocene molecule ($d = 1.0$ nm) shows a degradation already at the fourth molecule, while in case of $d = 0.8$ nm the signal seems to be preserved for a larger number of molecules. This means that the molecular QCA wire could be built with an inter-molecule distance lower than the ideal one, in order to ensure the information propagation through the wire.

3.5 Fault Tolerance Analysis Based on Real Fabrication Defects

Considering the technological steps of the fabrication of a molecular QCA wire made of bis-ferrocenes [21,23] and the classification of the possible defects that may occur reported in Sect. 2, a fault tolerance analysis of the bis-ferrocene wire was carried out. As described in Sect. 2, the evaluation was performed computing

the charge distribution inside a bis-ferrocene molecule (MUT) in presence of a driver, for different positions of the driver (ideal and subject to a defect).

Figure 34 displays the results for the driver shifting along the vertical axes, defined as Z-displ. The graph includes the results for three different distances from the MUT (equivalent to three values of Y-displ), considering the driver in both the logic state. The charge difference between the two dots of the MUT decreases when the value of Z-displ is increased, which is enhanced for longer distances from the molecule (higher Y-displ). In particular, in the range related to the roughness of each grain of the gold substrate ($\pm 0.2 \div 0.4$ nm) the molecule still works properly, that means that the free positive charge is mainly localized on one of the two dots, encoding a valid logic state. On the contrary, when the driver-MUT interaction is at the interface between two gold grains (equivalent to $Z\text{-}displ = \pm 2.0$ nm) the molecule is in an undefined state, because the charge of the two dots is almost the same.

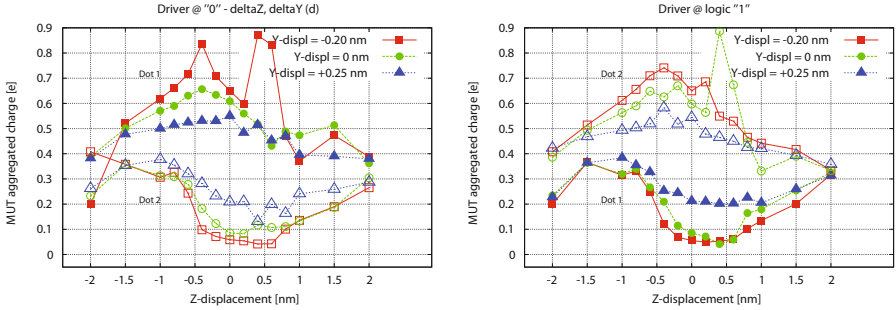


Fig. 34. MUT dot charges as function of the driver vertical shifting (Z-displ), in three case of driver-molecule distance (Y-displ).

For what concerns the driver misalignment with the respect to the two dots of the MUT (X-displ), the dot charges are reported in Fig. 35 for the ideal distance between the MUT and the driver (Y-displ = 0) and two other cases of driver-molecule distance (Y-displ = -0.20 nm and Y-displ = $+0.25$ nm). In these graphs, both the driver logic states at 1 and at 0 are reported simultaneously, in order to check immediately the MUT capability to encode the binary information for a specific driver position. In particular, for a given X-displ value it is possible to check if the positive charge inside the MUT moves between the two dots when the driver changes its state from 1 to 0 and the difference between them is large enough to consider the MUT in a valid logic state. Figure 35 reveals that in some cases of driver misalignment the interaction with the MUT is still preserved.

Finally, the simultaneous concurrency of driver and receiver defects were analyzed in order to draw a Safe Operating Area (SOA), that shows all the positions of the receiver that do not prevent the communication among the molecules. The results are reported in Fig. 36, where different corner cases of

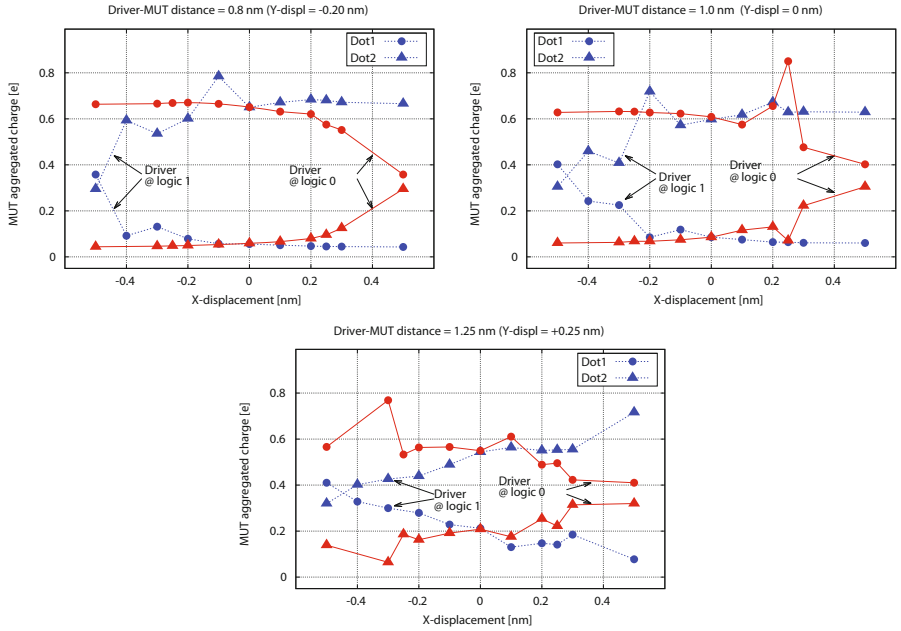


Fig. 35. MUT dot charges as function of the driver misalignments (X-displ), in three case of driver-molecule distance (Y-displ).

driver defects are considered: ideal position (no defects), lateral misalignment and vertical shift at the ideal driver-molecule distance, vertical shift at a lower driver-molecule distance. The results in case of ideal position of the driver are shown in Fig. 36(A): the positions (the barycenter of the molecule) of the receiver, in which it could still encode the two logic states, are highlighted. Two red circles indicate the working dot of the MUT, that was kept fixed. In case of driver misalignment equal to $X\text{-}displ = 0.5\,\text{nm}$, the region where the receiver could be safely placed (Fig. 36(B)) is a little bit wider than the previous case, especially in the Y direction, but the equivalent voltage at the receiver is quite lower. A vertical shifting of the driver due to the roughness of the substrate ($Z\text{-}displ = 0.2\,\text{nm}$) is reflected at the receiver level with a SOA that includes a wide range of both misalignments and higher distances from the molecule (Fig. 36(C)). In this case, the values of the equivalent voltage at the receiver are intermediate, confirming again that the gold roughness ($0.20\,\text{nm} \pm 0.1\,\text{nm}$) obtained at the experimental level does not affect the information propagation. In case of a bigger vertical shifting of the driver ($Z\text{-}displ = 0.8\,\text{nm}$), the charge distribution of the molecule is such that the SOA for the receiver is quite limited, as shown in Fig. 36(D). Finally, the simultaneous concurrency of a vertical shift and a variation in the driver-molecule distance ($Z\text{-}displ = 0.4\,\text{nm}$ and $Y\text{-}displ = -0.2\,\text{nm}$) is quantified at the receiver level in a very small SOA, depicted in

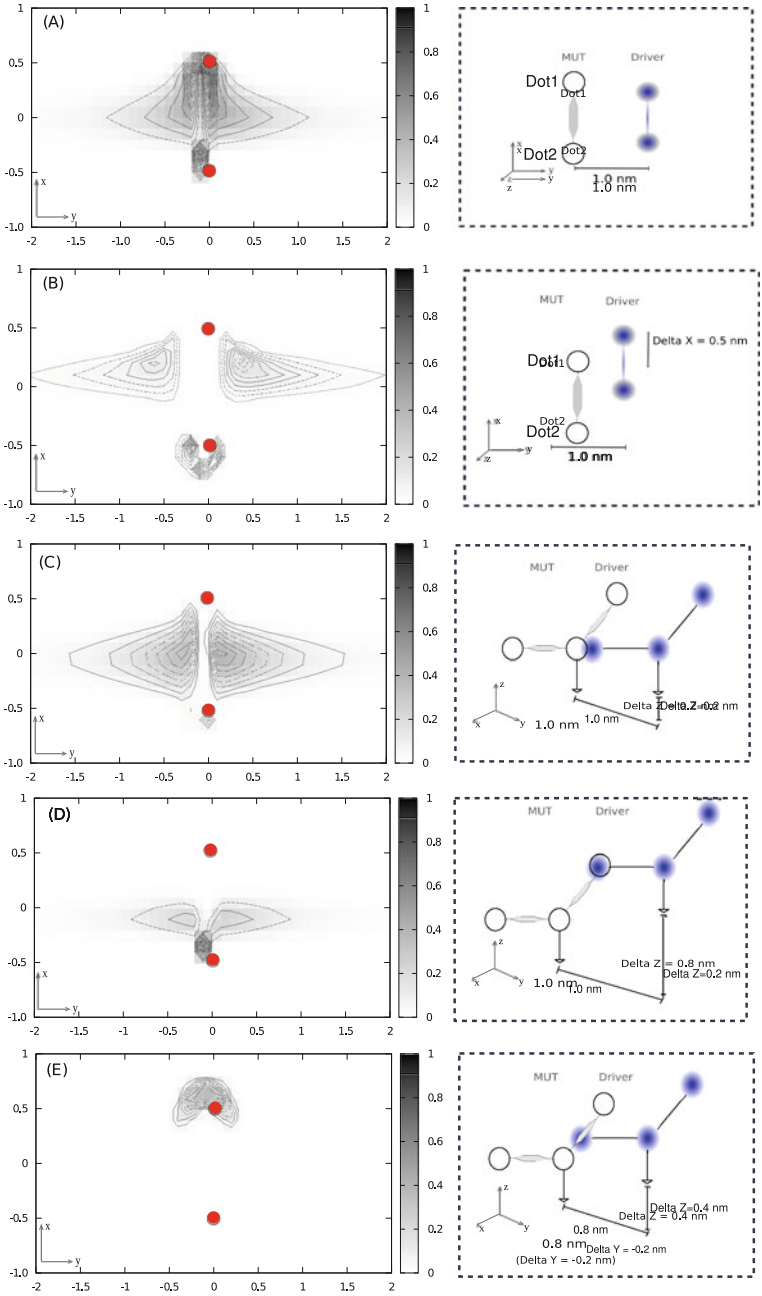


Fig. 36. Safe Operating Area (SOA) for different cases of driver defects (Color figure online).

Fig. 36(E), mainly localized near the molecule, with low values of equivalent voltage.

In summary, this analysis reveals that the tolerance of the QCA wire to some possible defects in the fabrication of the QCA wire is quite good. In addition, the results obtained and the data highlighted in Fig. 36 give an important feedback to the technologist about which are the critical points and which could be the improvement to assure a correct information propagation.

4 Conclusions

We summarized the most important characteristics of molecules currently under study as potential candidates as QCA devices. In particular we focused and discussed our method aimed at solving one of the gap that is preventing those molecules from being exhaustively studied from an electronic point of view and used as elements in a circuit. We indeed presented our two stages methodology, which starts from ab-initio simulations in several conditions followed by a second stage where post-processing is executed and electrostatic inspections are performed. The outcome is twofold. First, we improved the understanding of a MQCA wire based on a bis-ferrocene molecule and assess the conditions under which it can be used as well as the constraints the technological process should be subjected to. Second, we proposed a method to systematically and thoroughly study this and other molecules to be used as perspective MQCA devices. Though several steps are still necessary, our contribution enables the study of MQCA with an electronic perspective, allowing then to move from the single device level to a circuit level, still maintaining a strong link with the technological aspects.

References

1. Lent, C.S., Tougaw, P.D., Porod, W., Bernstein, G.H.: Quantum cellular automata. Nanotechnology **4**, 49–57 (1993)
2. Vankamamidi, V., Ottavi, M., Lombardi, F.: Clocking and cell placement for QCA. In: IEEE-NANO 2006, vol. 1, pp. 343–346 (2006)
3. Graziano, M., Vacca, M., Chiolerio, A., Zamboni, M.: A NCL-HDL Snake-Clock based magnetic QCA architecture. IEEE Trans. Nanatechnol. **10**(5), 1141–1149 (2011)
4. Graziano, M., Chiolerio, A., Zamboni, M.: A technology aware magnetic QCA NCL-HDL architecture. In: 9th IEEE Conference on Nanotechnology, 2009, IEEE-NANO 2009, Genova, Italy, pp. 763–766. IEEE (2009)
5. Hennessy, K., Lent, C.S.: Clocking of molecular quantum-dot cellular automata. J. Vac. Sci. Technol. B **19**, 1752–1755 (2001)
6. Lent, C.S., Isaksen, B.: Clocked molecular quantum-dot cellular automata. IEEE Trans. Electron Devices **50**, 1890–1896 (2003)
7. Lent, C.S., Isaksen, B., Lieberman, M.: Molecular quantum-dot cellular automata. J. Am. Chem. Soc. **125**, 1056–1063 (2003)

8. Qi, H., Sharma, S., Li, Z., Snider, G.L., Orlov, A.O., Lent, C.S., Fehlner, T.P.: Molecular quantum cellular automata cells. Electric field driven switching of a silicon surface bound array of vertically oriented two-dot molecular quantum cellular automata. J. Am. Chem. Soc. **125**, 15250–15259 (2003)
9. Jiao, J., Long, G.J., Rebbouh, L., Grandjean, F., Beatty, A.M., Fehlner, T.P.: Properties of a mixed-valence (Fe-II)(2)(Fe-III)(2) square cell for utilization in the quantum cellular automata paradigm for molecular electronics. J. Am. Chem. Soc. **127**, 17819–17831 (2005)
10. Lu, Y., Lent, C.S.: Theoretical study of molecular quantum-dot cellular automata. J. Comput. Electron. **4**, 115–118 (2005)
11. Lu, Y., Liu, M., Lent, C.S.: Molecular quantum-dot cellular automata: from molecular structure to circuit dynamics. J. Appl. Phys. **102**, 034311–034317 (2007)
12. Chiolerio, A., Allia, P., Graziano, M.: Magnetic dipolar coupling and collective effects for binary information codification in cost-effective logic devices. J. Magn. Magn. Mater. **324**(19), 3006–3012 (2012)
13. Vacca, M., Graziano, M., Zamboni, M.: Majority voter full characterization for nanomagnet logic circuits. IEEE Trans. Nanotechnol. **11**(5), 940–947 (2012)
14. Graziano, M., Vacca, M., Blua, D., Zamboni, M.: Asynchrony in quantum-dot cellular automata nanocomputation: Elixir or Poison? IEEE Des. Test Comput. **28**(5), 72–83 (2011)
15. Vacca, M., Graziano, M., Zamboni, M.: Asynchronous solutions for nano-magnetic logic circuits. ACM J. Emerging Tech. Comp. Syst. **7**(4), 15:1–15:18 (2011)
16. Lu, Y., Lent, C.S.: Self-doping of molecular quantum-dot cellular automata: mixed valence zwitterions. Phys. Chem. Chem. Phys. **13**, 14928–14936 (2011)
17. Wang, X., Ma, J.: Electron switch in the double-cage fluorinated fullerene anions: new candidates for molecular quantum-dot cellular automata. Phys. Chem. Chem. Phys. **2011**(13), 16134–16137 (2011)
18. Zoli, L.: Active bis-ferrocene molecules as unit for molecular computation. Ph.D. dissertation (2010)
19. Arima, V., Iurlo, M., Zoli, L., Kumar, S., Piacenza, M., Della Sala, F., Matino, F., Maruccio, G., Rinaldi, R., Paolucci, F., Marcaccio, M., Cozzi, P.G., Bramanti, A.P.: Toward quantum-dot cellular automata units. Nanoscale **4**, 813–823 (2012)
20. Pulimeno, A., Graziano, M., Abrardi, C., Demarchi, D., Piccinini, G.: Molecular QCA: a write-in system based on electric fields. In: IEEE Nanoelectronics Conference (INEC), June 2011 (2011)
21. Pulimeno, A., Graziano, M., Demarchi, D., Piccinini, G.: Towards a molecular QCA wire: simulation of write-in and read-out systems. Solid State Electron. **77**, 101–107 (2012). (Elsevier)
22. Pulimeno, A., Graziano, M., Piccinini, G.: Molecule interaction for QCA computation. In: IEEE NANO2012 12th International Conference on Nanotechnology, Birmingham (UK), 20–23 August 2012 (2012)
23. Pulimeno, A., Graziano, M., Sanginario, A., Cauda, V., Demarchi, D., Piccinini, G.: Bis-ferrocene molecular QCA wire: ab-initio simulations of fabrication driven fault tolerance. IEEE Trans. Nanotechnol. **12**, 498–507 (2013)
24. Wang, X., Chen, S., Wen, J., Ma, J.: Exploring the possibility of noncovalently surface bound molecular quantum-dot cellular automata: theoretical simulations of deposition of double-cage fluorinated fullerenes on Ag(100) surface. J. Phys. Chem. C **117**, 1308–1314 (2012)
25. Lu, Y., Quardokus, R., Lent, C.S., Justaud, F., Lapinte, C., Kandel, A.: Charge localization in isolated mixed-valence complexes: an STM and theoretical study. J. Am. Chem. Soc. **132**, 13519–13524 (2010)

26. Quardokus, R., Lu, Y., Wasio, N.A., Lent, C.S., Justaud, F., Lapinte, C., Kandel, S.A.: Through-bond versus through-space coupling in mixed-valence molecules: observation of electron localization at the single-molecule scale. J. Am. Chem. Soc. **134**, 1710–1714 (2012)
27. Singh, U.C., Kollman, P.A.: An approach to computing electrostatic charges for molecules. J. Comput. Chem. **5**, 129–145 (1984)

Irreversibility and Dissipation

Reversible and Adiabatic Computing: Energy-Efficiency Maximized

Ismo Hänninen[2(✉)], Hao Lu[1], Enrique P. Blair[1], Craig S. Lent[1], and Gregory L. Snider[1]

[1] Center for Nano Science and Technology, University of Notre Dame, Notre Dame, IN 46556, USA
{hlul, eblair, lent, snider.7}@nd.edu
[2] Department of Pervasive Computing, Tampere University of Technology, P.O. Box 527, 33101 Tampere, Finland
ismo.hanninen@nd.edu

Abstract. Emerging devices promise energy-efficient computing on a massively parallel scale, but due to the extremely high integration density the previously insignificant dissipation due to information erasure (destruction) becomes a prominent circuit design factor. The amount of heat generated by erasure depends on the degree of logical reversibility of the circuits and successful adiabatic charging. In this paper, we design an adiabatic arithmetic-logic unit to prototype the locally-connected Bennett-clocked circuit design approach. The results indicate one or two orders-of-magnitude energy savings in this physical circuit implementations vs. standard static CMOS. Previous work on computer arithmetic suggests that common hardware implementations erase much more information than would be required by a theoretical minimal mapping of the addition operation. A Bennett-clocked approach can reach the theoretical minimum number of bit erasures in the binary addition, though simulations show that a transistor technology has energy loss due to parasitic components that can exceed the information loss heat. In this paper, we describe the relationship between adiabatic and logically reversible circuits, and predict the potential of the arithmetic implementations based on quantum-dot cellular automata, which enable the full benefits of reversible, locally connected circuits to be realized.

Keywords: Reversibility · Addition · Arithmetic · Adiabatic circuit design

1 Introduction

The computing performance of integrated circuits has been tightly connected to the energy-efficiency of the underlying device technology, and this connection also exists for the emerging circuits [1]. In fact, due to their inherent efficiency in conserving the signal energy, technologies such as quantum-dot cellular automata (QCA) and nanomagnetic logic (NML) will potentially have the dissipation due to physical state compression and irreversibility as the limiting factors for their overall dissipation, which in turn influences the maximum operating frequency or battery life. The combination of molecular circuits and switching frequencies reaching the terahertz regime makes the dissipation associated with the logical and physical irreversibility a

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 341–356, 2014.
DOI: 10.1007/978-3-662-43722-3_14, © Springer-Verlag Berlin Heidelberg 2014

significant circuit design factor [2, 3]. With QCA, the irreversibility-induced heat dissipation may present surprisingly tight operating frequency limits for computer arithmetic based on high-density nanoscale components. Recent work suggests that binary adders and multipliers might have maximum operating frequency in tens of gigahertz instead of hundreds expected of the pipelined designs, with a typical power density constraint of 100 W/cm^2 [4, 5]. While NML circuits do not reach molecular device densities or gigahertz operating frequencies, their inherent signal energy conservation suggest use in battery-life limited applications, where the energy dissipated due to information loss could make the difference between a battery life of months vs. years.

Computer arithmetic circuits represent highly optimized logic designs usually laid out with extreme care, but the optimization goals have traditionally been the result latency, throughput, circuit area, or CMOS power. Logical reversibility has not been one of the goals, and this has the consequence that the existing or proposed arithmetic circuits are highly sub-optimal from the perspective of information loss. Our recent study suggests that typical QCA adders generate multiple times the number of bit erasures than the theoretical minimum for the addition operation [6], and that QCA multipliers erase a square-law number of bits vs. operand word length, compared to the potential sub-linear loss of the theoretical operation [7]. For an introduction to irreversibility induced density and frequency limits in QCA, the reader is referred to [8].

We believe that logical reversibility is connected to the physical reversibility of the system, that is, the physical, thermodynamically described state of the system has to mirror to some degree the computation that is performed. Fifty years ago, Rolf Landauer proposed this connection in [9], and finally in 2012, the Landauer's Principle was confirmed with a generic one-bit memory experiment [10]. A bit erasure at the room temperature has an inevitable energy cost of at least 3 zJ, which must be dissipated as heat into the environment. To achieve a lower dissipation circuits have to utilize adiabatic charging in forming the logic or clock signals, and achieve a degree of logical reversibility. An erasure-aware, partially reversible circuit involves tradeoffs between performance, timing, circuit area, and power, and must balance the effects of the erasures and adiabatic operation.

This paper explores the tradeoffs in a prototype arithmetic-logic (ALU) unit and is organized as follows: Sect. 2 gives an introduction into adiabatic circuit operation, logical reversibility, and the related heat generation. Section 3 describes our adiabatic transistor circuit, highlighting the challenges and gains of locally connected adiabatic operation. Section 4 describes the simulation setup and Sect. 5 the corresponding results on the power consumption. Section 6 presents predictions for the future technologies, while Sect. 7 the concludes the discussion.

2 Signal Energy Recovery

The energy-efficiency of any integrated circuit technology is closely related to the method of signal representation and the associated signal energy, which must be larger than the thermal noise floor by a significant margin [1]. In standard static CMOS, every switching event leads to the dissipation of all the signal energy stored on the

circuit node. Most of this loss can be avoided by utilizing adiabatic charging principles, which can be fully implemented only by logically reversible circuits. Logically reversible circuits save energy by avoiding the bit erasures and the related heat, but in addition, these circuits have to avoid the other types of loss, like static leakage and dynamic signal energy loss related to the switching mechanism. The operating principle must not lead to the loss of all signal energy during every switching event, like in traditional CMOS. One key property in QCA and NML is the signal energy conservation: the cells settle to the ground state while the signal energy is transmitted from cell to cell. In these technologies, the loss is low and the signal level high [2, 3].

2.1 Adiabatic Charging

Adiabatic charging is one of the pre-requisites of practical reversible circuits. For QCA, this implies that the clock field potentials must be switched at a lower rate than the highest possible rate of the cellular automata. The predicted terahertz devices would have adiabatic switching speeds of tens to hundreds of gigahertz. This is a speed vs. power tradeoff.

The energy dissipation in standard circuits occurs when electrical currents are driven through transistors, with a finite on-resistance, and resistive signal lines. The resistive losses are proportional to the voltage drop, for example between the terminals of a transistor device, which points to an approach of limiting this voltage difference and avoiding abrupt currents as a means to limit dissipation. For example, a static CMOS inverter gate in Fig. 1(a) represents information by the output node voltage, and dissipates all of the signal energy at each switching operation. During a switching event, either the pull-up or pull-down network loses

$$E_{\mathrm{CMOS}} = {}^{1}\!/_{2} C V_{\mathrm{DD}}{}^{2}, \tag{1}$$

where C is the output node capacitance including the wiring and next gate input, and V_{DD} is the operating voltage. This energy is practically all the signal energy.

This circuit can be modified to recover signal energy by utilizing ramped power-clock signals instead of the static operating voltage and ground. An example of such energy-recovering 1n1p-logic, or Split-Level Charge Recovery Logic [11, 12] inverter

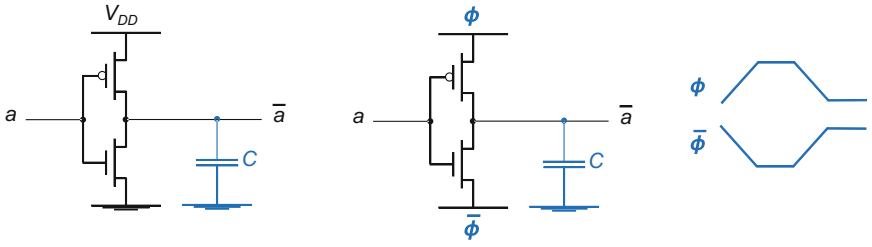


Fig. 1. CMOS inverter. (a) Standard static CMOS implementation, (b) adiabatic 1n1p CMOS implementation, and (c) dual-rail power-clock.

is shown if Fig. 1(b) with the corresponding dual-clock waveforms in Fig. 1(c), which can input energy into the circuit and recover it back. The adiabatic energy loss is

$$E_{\text{adiabatic}} = RC^2{V_{\text{DD}}}^2/t_{\text{ramp}}, \tag{2}$$

where t_{ramp} is the time duration of the ramp. It should be noted that the split-rail signals ensure that the circuit does not lose the typical minimum energy $E_t = {}^1\!/_2 CV_t^2$ associated with the transistor threshold drop V_t, which a single-rail approach would lose. The pipelining of this type of asymptotically adiabatic logic, where there is no lower bound on the dissipation, is challenging due to the need to utilize reversible gates or include garbage signals. For an introduction to adiabatic circuit families and their classification the reader is referred to [13], which describes also quasi-adiabatic approaches enabling simple pipelining but losing some part of the energy. Our design differs from the previous circuits [12, 14–16], since it has no logic overhead for the reversibility, utilizing the Bennett-clocking scheme described in the next section.

2.2 Reversible Logic and Operations

While the circuits presented here are based on adiabatic operation, we believe that maintaining a high degree of energy conservation requires either reversible logic gates or an operating principle that implements logical reversibility. Fully reversible circuits can be constructed using reversible logic gates, like the Toffoli gate [17], but this results in large circuit area and many "garbage" signals needed to retain the state throughout the computation. Another alternative is to utilize Bennett-clocking [18], where the logic structure of the circuit remains unmodified, but the timing is altered to include a compute step and a de-compute step, at the cost of reduced pipelining and throughput. Recovering the signal energy to any desired extent is possible using asymptotically adiabatic logic, where the energy is transferred inside the circuit avoiding any abrupt charging or discharging of the circuit nodes. However, designing such circuits implies that the utilized logic operations have to be reversible in nature.

Logical reversibility is connected to the physical reversibility of the system, that is, the physical, thermodynamically described state of the system has to mirror to some degree the computation that is performed. Fifty years ago, Rolf Landauer proposed this connection in [9] and the Landauer's Principle was confirmed with a generic one-bit memory experiment recently reported in [10]. A bit erasure at the room temperature has an inevitable energy cost of at least 3 zJ, which usually must be dissipated as heat into the environment. This part of the signal energy cannot be adiabatically recovered, unless we incorporate logical reversibility into the circuit.

In traditional circuits the bit erasure energy is insignificant compared to the other losses. For example, the end-of-the-roadmap CMOS will dissipate about three orders of magnitude higher energy per switching event since it loses the full signal energy at each switching event [1]. Losses in the emerging technologies like quantum-dot cellular automata (QCA) [2] vary, but not counting the information loss, the dissipation in all of them is like friction in nature: it can be made as small as desired by switching more slowly, while the energy-per-bit-erasure is unaffected by the speed.

Therefore, it is necessary to design the system to utilize adiabatic charging for the logic or clock signal to recover the energy, in addition to achieving some degree of logical reversibility. From the circuit design perspective, reversible logic and adiabatic operation are desirable, but they incur various costs. An erasure-aware circuit involves tradeoffs between performance, timing, circuit area, and power, balancing the effects of the erasures and adiabatic operation. There are two approaches for logical reversibility.

The first approach is based on using logically reversible gates like the Toffoli or Fredkin gates gate [17]. The truth table of this type of gate contains only one-to-one mappings between the input and output spaces, and therefore the physical trajectory of the evolving computing system can be logically tracked and reversed. In contrast, the truth table of an irreversible operation can be embedded into a larger logically reversible operation by adding "garbage" outputs. This has significant costs in the circuit area and complexity.

The second approach is based on designing the timing of the circuit in such a way that logical information is retained and energy recovery enabled, following the ideas of Bennett [19]. The circuit first computes from the input side to the output side, the result is obtained, and then the circuit de-computes from the output to the input in reverse order. This can be efficiently implemented by the Bennett-clocking technique [18], which is feasible for both adiabatic CMOS circuits and many emerging technologies. Reversibility is achieved by holding the predecessor parts of the circuits steady while successor stages compute, then relaxing after the whole computation has been finished. This is illustrated in the following example utilizing 1n1p asymptotically adiabatic logic style and dual-rail power-clock signals [11].

A simple circuit of two inverters in series and a 2-level concatenated clock is used to illustrate how Bennett-clocking works. Figure 2(a) shows the circuit schematic and the Bennett clocks at each end-terminal, and Fig. 2(b) the timing diagram of the clocks and voltages V_1 and V_{OUT}. For the purpose of illustration, the input is kept high all the time, and the logic swing of $\text{Clock}_{1..2}$ is $0–(+1/2\text{V}_{\text{DD}})$ and the swing of $\text{Clock}_{1..2\_n}$ is $0–(-1/2\text{V}_{\text{DD}})$. The circuit nodes have three stable states: logic "0" $(-1/2\ \text{V}_{\text{DD}})$, logic "1" $(+1/2\ \text{V}_{\text{DD}})$, and relaxed "R" (0 V). Before transitioning, the clocks are held at 0 V, which turns on M_{N1}, but not M_{P1}. Therefore capacitor C_1, the parasitic capacitance at middle node V_1, is discharged until voltage at V_1 becomes 0 V.

1. **Compute V_1**: Input V_{IN} is at stable logic value "1" $(+1/2\ \text{V}_{\text{DD}})$. When the first level clocks start to ramp, the transistor M_{N1} turns on and transistor M_{P1} turns off. Capacitor C_1 gets discharged through M_{N1} gradually. The voltage level at V_1 drops gradually until the clocks become constant.
2. **Compute V_{OUT}**: When the first level clocks stop ramping and become constant, V_1 has stable logic value "0" $(-1/2\ \text{V}_{\text{DD}})$, and the second level clocks start to ramp. Transistor M_{P2} turns on and transistor M_{N2} turns off. Capacitor C_2 gets charged through M_{P2} gradually. The voltage level at V_{OUT} increases gradually until the clocks become constant. Stable logic value "1" is read $(V_{OUT} = +1/2\text{V}_{\text{DD}})$.

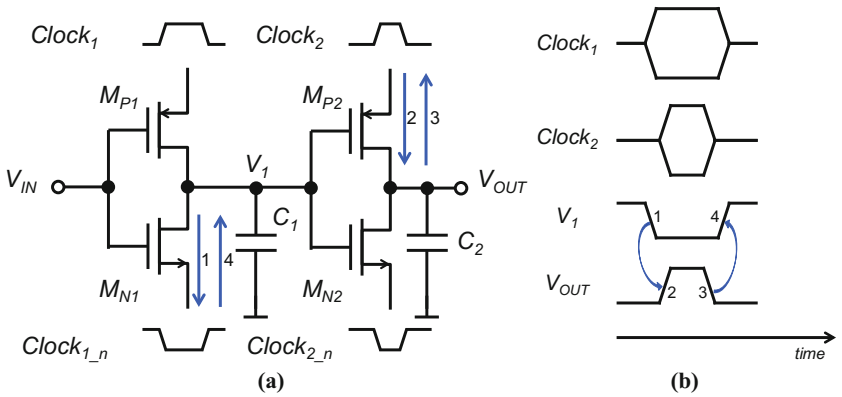


Fig. 2. (a) Schematic of a simple circuit consisting of two inverters connected in series, utilizing 2-level Bennett clock. The blue arrows indicate the current flow during the computation (1 and 2) and de-computation (3 and 4). (b) Timing diagram of the Bennett clocks and the voltages V_1 and V_{OUT}, transitions during the computation (1 and 2) and the de-computation (3 and 4) (Color figure online).

3. **De-compute V_{OUT}**: The second level clocks starts to ramp back to 0 V while the first level clocks are still held constant. Capacitor C_2 gradually gets discharged through M_{P2}, and the voltage level at V_{OUT} gradually drops back to 0 V.
4. **De-compute V_1**: When the second level clocks are relaxed, the first level clocks start to relax back to 0 V. Capacitor C_1 is gradually charged, and the voltage level at V_1 gradually rises back to 0 V. After this, the input V_{IN} is relaxed.

3 Arithmetic-Logic Unit Design

A Bennett-clocked arithmetic logic unit (ALU) was designed based on the commercially available SN74S381N [20], which can perform three arithmetic and three logic operations on two active high unsigned 4-bit words A = (A_3, A_2, A_1, A_0) and B = (B_3, B_2, B_1, B_0), producing the 5-bit result word F = (F_4, F_3, F_2, F_1, F_0). Negative result numbers are represented in the two's complement format. The function mode is set with three additional active high select signals S_2–S_0, to implement an operation from the set {clear/reset, B minus A, A minus B, A plus B, A xor B, A or B, A and B, preset} according to Table 1. The implemented structure is capable of operating either in irreversible standard CMOS mode or the reversible Bennett clocked mode.

3.1 Logical Structure and Implementation

The logic schematic of the ALU using standard gate set {not, nand, nor, xor} is shown in Fig. 3, having the longest combinatorial path of 11 logic levels in the final implementation. In this schematic, no clocks are shown, and only the data signals

Table 1. Functions of the ALU

Control signals			Function mode
S_2	S_1	S_0	
0	0	0	Clear
0	0	1	B minus A
0	1	0	A minus B
0	1	1	A plus B
1	0	0	A xor B
1	0	1	A or B
1	1	0	A and B
1	1	1	Preset

A, B, and F and the select signals S_2–S_0 are included since the structure is inherently combinatorial without any sequential elements.

The design was laid out in 2 μm CMOS to produce the layout in Fig. 4, which was fabricated at the University of Notre Dame. The n-type fabricated transistors have a width/length ratio of 6 μm/2 μm with a threshold voltage of 0.4 V, while the p-type transistors have a ratio of 12 μm/2 μm with a threshold voltage of −1.0 V. The gate oxide thickness is 20 nm, MOSIS Scalable CMOS (Revision 8.00) lambda = 1 μm design rule. To enable Bennett clocking, power to each level of logic is supplied by a separate clock signal. These signals Clk1–Clk11 and Clk1_n–Clk11_n can be configured to run the circuit in either an irreversible or a reversible mode.

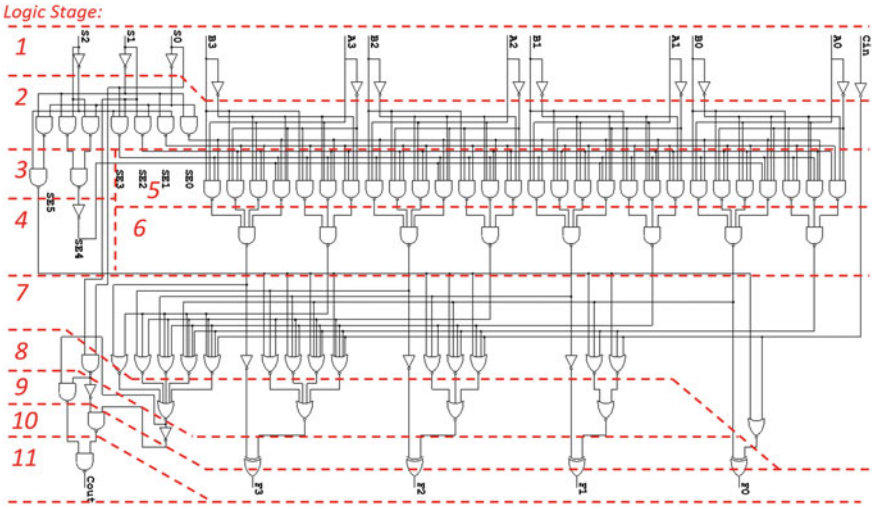


Fig. 3. The logical structure of the ALU, and the 11 stages of Bennett-clocking.

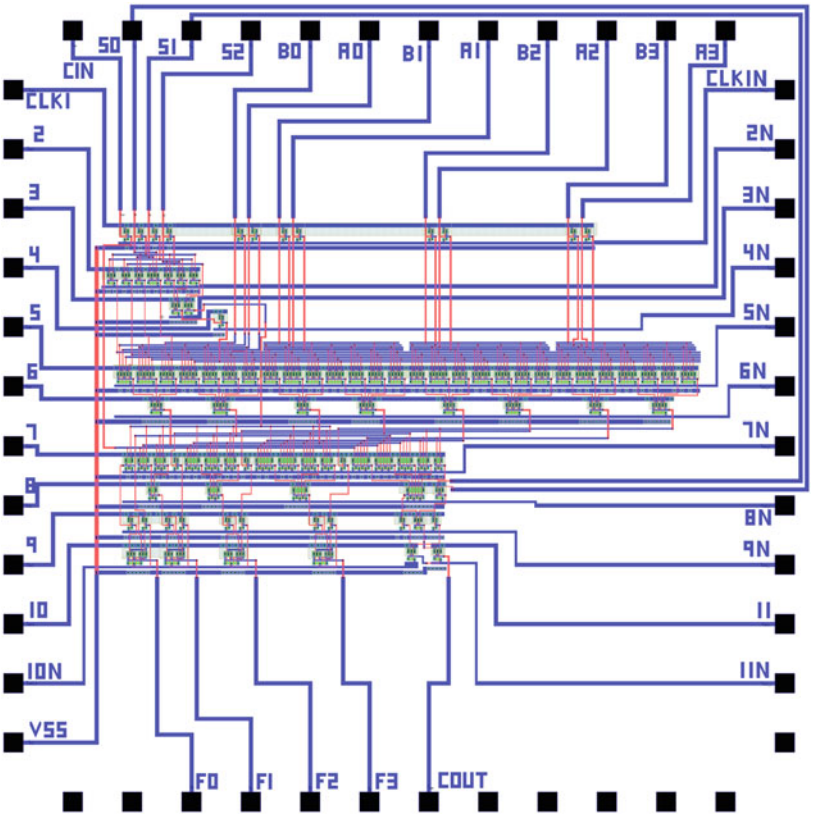


Fig. 4. The layout of the ALU in 2 μm CMOS.

3.2 Irreversible Standard CMOS Operation

The combinatorial ALU unit can be run in irreversible mode like a standard CMOS design by setting the power-clock lines to static values, which are held constant throughout the operation. The positive power-clocks Clk1…Clk11 are tied to the operating voltage V_{DD}, while the negative power-clocks Clk1N…Clk11N are connected to ground GND. In this configuration, the unit implements a standard combinatorial CMOS ALU, without pipelining or any sequential components. This mode of operation erases information and uses energy exactly like traditional irreversible CMOS logic.

3.3 Reversible Bennett-Clocked Adiabatic CMOS Operation

The ALU unit can be configured into fully reversible mode by utilizing the dual-rail power-clock signals with the ramp-up and ramp-down timing defined by the requirements of Bennett-clocking, setting the signals Clk1–Clk11 and Clk1_n–Clk11_n to ramp in concatenation as illustrated in Fig. 5. The design effectively forms

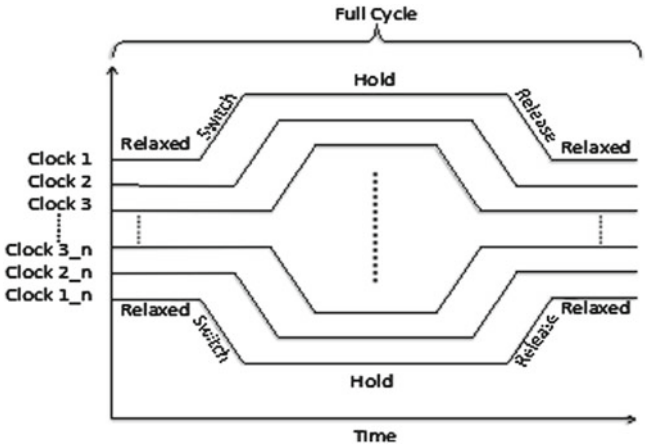


Fig. 5. Reversible Bennett-clocking, the dual-rail power-clock waveforms. The design has 11 positive clocks Clk1–Clk11 and 11 negative Clk1_n–Clk11_n, vertically offset for clarity.

an 11-stage 1n1p-type asymptotically adiabatic logic circuit depicted on logic level in Fig. 3, where the computing part takes 11 steps and de-computing part 11 steps for each arithmetic/logic operation. The design in Figs. 3 and 4 contains all the logic needed for computing and de-computing using the Bennett-clocking scheme, with the only additional overhead the clock generator not shown. By slowing down the frequency, all of the signal energy can be asymptotically recovered, with no additional cost in CMOS logic complexity and area. However, the unit is not capable of pipelining, and the generation of the complicated power-clock signals is challenging.

4 Simulation Setup

The irreversible and reversible ALU were simulated in Synopsys HSPICE 2012, in both the 2 μm and the 20 nm technology. In the **2 μm technology**, the irreversible operation had a static Clk1…Clk11 = V_{DD} = 5 V and Clk1N…Clk11N = GND = 0 V, while the reversible operation had the swing of the Bennett clocks Clk1…Clk11 between 0–2.5 V and the swing of Clk1N…Clk11N between 0–(−2.5) V. A level 3 semi-empirical MOSFET model [21] with parameters extracted from the devices made in the fabrication facility at the University of Notre Dame was used, and a 2 μm gate length was chosen to match the circuits being fabricated. The simulated n-type transistors had a width/length ratio of 6 μm/2 μm and a threshold voltage of 0.5 V, while the p-type transistors had a ratio of 12 μm/2 μm and a threshold voltage of −0.7 V. The gate oxide thickness was 20 nm, and a series resistance of 120 Ω at both the drain and the source was included in the model. The intrinsic transconductance parameter, which is the product of mobility and gate capacitance was KP_{NMOS} = 100 μA/V^2 for NMOS, and KP_{PMOS} = 80 μA/V^2 for PMOS. The velocity saturation was incorporated into the model by setting the parameter VMAX = 210 k for both NMOS and PMOS, and the channel length modulation was excluded by setting the

field correlation factor KAPPA = 0. The effective fast surface state density parameter was NFS = 0.01, in order to turn on the flow of the subthreshold current. The following directives were used for NMOS and PMOS:

```
.model nmos nmos LEVEL=3 KP=100u Vt0=0.5 TOX=20n RS=120 RD=120
+VMAX=210k kappa=0 NFS=0.01
.model pmos pmos LEVEL=3 KP=80u Vt0=-0.7 TOX=20n RS=120 RD=120
+VMAX=210k kappa=0 NFS=0.01
```

In the **20 nm tri-gate technology** simulation, the irreversible operation had a static Clk1…Clk11 = V_{DD} = 0.9 V and Clk1N…Clk11N = GND = 0 V, while the reversible operation had the swing of the Bennett clocks Clk1…Clk11 between 0–0.45 V and the swing of Clk1N…Clk11N between 0–(−0.45) V. A 20 nm LSTP transistor model from Arizona State University Predictive Technology Model library was used [22]. This is a level 72 model based on the BSIM_CMG model for multi-gate devices, and we used the default parameter values in the simulation.

5 Design Analysis

The implemented adiabatic CMOS design can be viewed as a prototype of a reversible, locally connected cellular automata based ALU. The Bennett-clocking approach enables the conservation of signal energy and information in the Landauer/Bennett meaning. This prototype design was simulated in both modes of operation with parameters extracted from the devices made at the University of Notre Dame. The actual 2 μm node chips have been fabricated, but the measurements are ongoing, and therefore we report only the simulation results here.

5.1 Importance of Adiabatic Operation

The adiabatic charging approach can be very beneficial, since it will minimize the resistive heat generated by the circuit. The adiabatic CMOS circuit model simulated with Bennett-type clocking demonstrates that while the logic signal level is very high, the losses in the circuit can be kept very low. Comparison of irreversible standard clocking and reversible Bennett-clocking in a 4-bit ALU at the 2 μm node is shown in Fig. 6. The reversible mode generally offers two orders-of-magnitude improvement in average power consumption in the low frequency end up to 3 MHz, and about an order-of-magnitude improvement up to 50 MHz clock frequency, while the modes are equal around 200 MHz frequency. Including the parasitic capacitances affects both modes of operation. The leakage power can be identified in Fig. 6 as the constant value of power at low frequencies: the irreversible mode has high voltages continuously applied over the transistor drain-source, producing a static level up to 1 mW. The reversible mode avoids most of the leakage, since the voltages are ramped and kept at for much of the clock period, producing a static power around 0.01 mW.

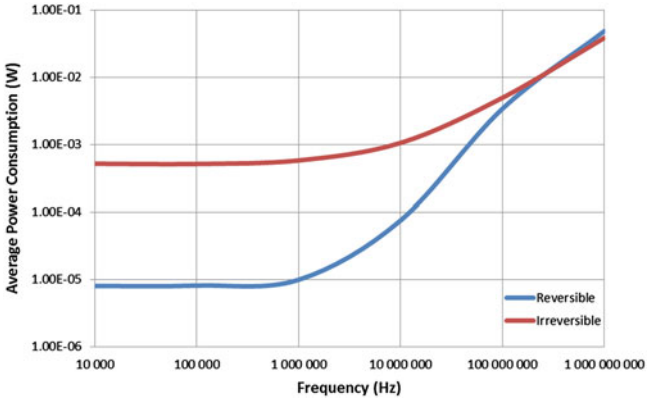


Fig. 6. Simulated average power consumption of the ALU on 2 μm standard CMOS, irreversible vs. reversible mode, as a function of operating frequency.

The comparison of irreversible standard clocking and reversible Bennett-clocking in a 4-bit ALU at the 20 nm node with tri-gate CMOS model is shown in Fig. 7. Although the transistor circuit with reversible/adiabatic operation consumes power approximately with a square-law dependency on the frequency, the approach is still

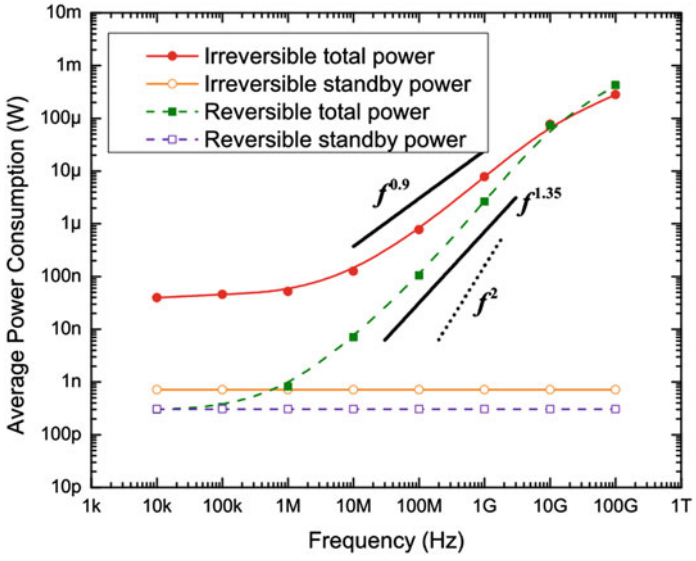


Fig. 7. Simulated average power consumption of the ALU on 20 nm tri-gate CMOS, irreversible vs. reversible mode, as a function of operating frequency

about an order-of-magnitude better at 100 MHz. The leakage/standby power of the irreversible mode reaches 1 nW, and the reversible mode around 0.2 nW.

The cost of the adiabatic approach lies in timing and the complicated clocks, since this design requires eleven clock phases. Adiabatic clocking can recover energy from the logic circuit, but to achieve system-wide energy efficiency the clock generator must be able to recycle the energy. Perhaps the most promising approach for energy recycling is MEMs based resonant clock generators [23]. This is a challenge common to all reversible circuits, including adiabatic CMOS and the QCA.

5.2 Logical Reversibility and Heat Generation

The heat cost of irreversible bit erasures can be approximated and directly related to the logic gates forming the circuit and the timing of the logic operation. We summarize this for the two operating modes of the designed ALU and predict the limits of emerging cellular automata circuits.

Irreversible mode of operation. The irreversible mode has the same information loss as a standard combinatorial static CMOS circuit; for the worst case, an estimate can be based on the worst-case bit erasures of the underlying gates. Assuming the truth tables and per-gate erasures in Table 2, the designed ALU has an upper-bound of logical information loss of

$$N_{\text{not}} \times R_{\text{avg,not}} + N_{\text{nand}} \times R_{\text{avg,nand}} + N_{\text{nor}} \times R_{\text{avg,nor}} + N_{\text{xor}} \times R_{\text{avg,xor}} \approx 85 \text{ bits} \quad (3)$$

per arithmetic/logic operation performed, with N_i the number of each specific gate in the design and $R_{\text{avg},j}$ the weighted average of gate erasures for that gate, with the parameters for each gate type defined in Table 2. The per-gate weighted bound is somewhat pessimistic, since the exact information loss in the logical space depends on the particular operands, as in the various adders modeled in [24]. However, the corresponding worst-case information loss heat of $E_{\text{erasures}} = 85$ bits $\times$ 0.003 aJ/bit $\approx$ 0.250 aJ per arithmetic/logic operation is insignificant, compared to the other losses in the CMOS circuit, the dissipation of the signal energy and static loss. Operating at 1 GHz, the information loss heat power is $P_{\text{erasures}} = (10^9 \text{ Hz} \times 1 \text{ s}) \times 0.250 \text{ aJ} = 250 \text{ pW}$ in the irreversible mode, which is comparable to a quarter of the static leakage power using 20 nm technology.

Reversible mode of operation. The Bennett-clocked mode of operation avoids all the internal bit erasures of the combinatorial ALU structure, leaving only the input operand words A and B to be erased after the de-computing sequence. With the ALU-design, this corresponds to the loss of 8 bits, with $E_{\text{erasures}} = 8$ bits $\times$ 0.003 aJ/bit $\approx$ 0.024 aJ per arithmetic/logic operation, which is an order-of-magnitude better than the irreversible worst-case bound. However, since the Bennett-clocked structure has 11 stages, the heat power at 1 GHz clock frequency is $P_{\text{erasures}} = (10^9 \text{ Hz} \times 1 \text{ s}) \times 0.024 \text{ aJ}/11 = 2.2 \text{ pW}$ in the reversible mode, about 1 % of the static leakage using 20 nm technology. It should be noted, that also the computing throughput is only 1/11 of the throughput of the standard implementation at the same clock frequency.

Table 2. Bit erasures of the standard gate set {not, nand, nor, xor} for each operand combination, and the number N_i of each gate in the design and the weighted average gate erasures $R_{avg,j}$.

Inputs	Output	Information loss
a	*not(a)*	
0	1	$log_2(1) = 0$ *bits*
1	0	
$N_{not} = 18$, $R_{avg,not} = 0$ *bits/gate*		

a	*b*	*nand(a,b)*	
0	0	1	$log_2(3) \approx 1.585$ *bits*
0	1	1	
1	0	1	
1	1	0	$log_2(1) = 0$ *bits*
$N_{nand} = 49$, $R_{avg,nand} = (3 \times 1.585 + 1 \times 0)/4 \approx 1.19$ *bits/gate*			

a	*b*	*nor(a,b)*	
0	0	1	$log_2(1) = 0$ *bits*
0	1	0	$log_2(3) \approx 1.585$ *bits*
1	0	0	
1	1	0	
$N_{nor} = 19$, $R_{avg,nor} = (1 \times 0 + 3 \times 1.585)/4 \approx 1.19$ *bits/gate*			

a	*b*	*xor(a,b)*	
0	0	0	$log_2(2) = 1$ *bit*
0	1	1	
1	0	1	
1	1	0	
$N_{xor} = 4$, $R_{avg,xor} = 1$ *bit/gate*			

6 Predictions for Future Technologies

Reversible logic has great potential in the *Beyond-CMOS* technologies, while quasi-adiabatic CMOS can operate reasonably well with less concern for the information loss [13]. For asymptotically adiabatic circuits in either CMOS like our ALU, or in one of the emerging technologies including QCA [2], logical reversibility is a prerequisite for reaching the full potential of recovering all energy. Our previous work on QCA arithmetic units can be used to predict the operating frequency limits of the studied ALU, if implemented in the emerging extreme low-power technologies [8].

We determined the average logic density of adders and multipliers based on cellular automata, assuming a constant-width square cell as the basic device conceptualized in Fig. 8(a). The cell width w was used as a measurement unit similar to λ

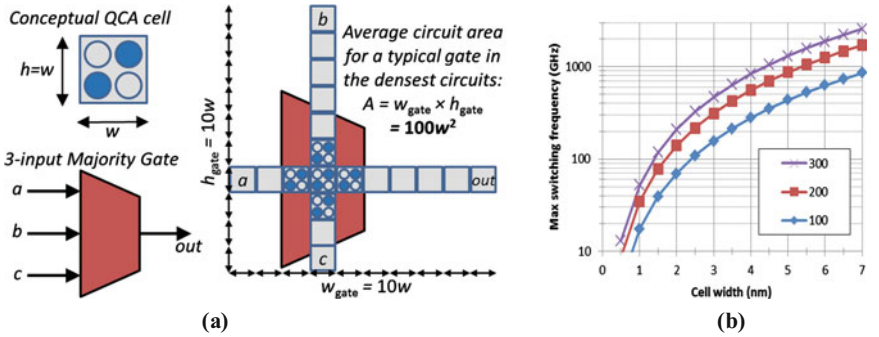


Fig. 8. (a) Definition of logic density in QCA, the average gate area expressed as a multiple of molecular cell footprints. (b) Maximum operating frequency of QCA circuits as a function of cell width, for $A = 300w^2$, $200w^2$, and $100w^2$ dissipation area per majority gate. High-density computer arithmetic designs found in the literature typically use about 200 cell footprints of circuit area per each logic gate (the middle curve) [8].

in standard CMOS layout rules, enabling us to express scalable QCA layout lengths/distances as a multiple of w. Based on the selection of arithmetic units found in literature, we calculated the average area A available for each standard gate, which in QCA is a 3-input Majority Voter Gate (MG) residing in an imaginary layout rectangle of area A, as illustrated in Fig. 8(a). We found that the average area per gate inside the densest core logic (e.g. a full adder) was around $A = 100w^2$, while the wiring overhead of an optimized multi-bit arithmetic unit increased the area per gate typically to $A = 200w^2$. In random logic, the area per gate was typically to $A = 300w^2$ [8].

Figure 8(b) shows the resulting maximum operating frequency vs. QCA cell width w, when the heat generation is limited to 100 W/cm^2. In this comparison, the nominal area per logic gate was considered to be $A = n_{fp}w^2$, where the gate span n_{fp} was 300, 200, or 100 cell footprints, corresponding to the studied relative gate densities.

Table 3. Estimates for the worst-case bit erasures in QCA arithmetic units, vs. operand word length n. The logical operations must perform only a linear number of erasures, but the adders discard 2–6 times that much and the multipliers a square-law amount of information [8].

Binary addition with unsigned n-bit operands *(typically linear)*	
Theoretical addition operation	n
Ripple carry adder, lower bound	$2n$
Ripple carry adder, upper bound	$6n$
Binary multiplication with unsigned n-bit operands *(typically square-law)*	
Theoretical complete multiplication	$n + 1$
Theoretical non-trivial multiplication	$123 \times \log(n + 241) - 673$
Array multiplier	$8n^2$
Serial-parallel multiplier	$16n^2$
Serial-parallel optimized multiplier	$16n^2 - 12n$
Radix-4 recoded multiplier	$26n^2 + 86n - 2$

The typical gate span was $n_{fp} = 200$ footprints and the area $A = 200w^2$ in the studied optimized arithmetic layouts. The results indicate that irreversibility heat is limiting the operating frequency of nanometer-scale molecular QCA cells, while possible sparser implementations will suffer less.

The existing computer arithmetic designs have not been optimized for logical reversibility, which is apparent from the conclusions of our previous study summarized in Table 3. The fundamental baseline information loss for binary addition is n bits, where n is the operand word length of the arithmetic unit. The 4-bit CMOS ALU unit in this paper discards about 85 bits per operation in the irreversible mode, while the known basic adder structures for QCA would have between 8–24 bit erasures per operation. However, the Bennett-clocked reversible CMOS ALU discards only 8 bits, which appears to be the ultimate lower bound even for the future technologies.

7 Conclusion

The adiabatic charging and reversible computing approaches are related to each other, and both will eventually be necessary for the efficient design of future digital circuits in the emerging technologies. In this paper, we considered the relationship between adiabaticity and information loss and designed a configurable CMOS ALU with irreversible and reversible operation modes. The results indicate that even using standard CMOS devices, adiabatic charging and reversible Bennett-clocking together would potentially yield, on average, one or two orders of magnitude improvement in power consumption, compared to the standard static CMOS approach.

The presented adiabatic CMOS design relies on local interaction between the consequent states and we consider it a prototype for future reversible circuits especially based on quantum-dot cellular automata. The existing computer arithmetic structures appear sub-optimal from the perspective of information loss, even though the Bennett-clocked circuit can reach the theoretical lowest bound of loss in the addition operation. The cost in throughput and clocking complexity suggests that a block-reversible scheme should be utilized in larger designs, to seek a compromise between information conservation and the design complexity.

Acknowledgments. The authors wish to thank the organizers and attendees of the *Workshop on Field-Coupled Nanocomputing, February 7–8, 2013, Tampa, Florida, USA*, for their suggestions and constructive critique on the design of emerging reversible circuits. This work was supported by the Academy of Finland under research grant 132869 and the Finnish Foundation for Technology Promotion.

References

1. International Technology Roadmap for Semiconductors. ITRS report [Online] (2012). http://www.itrs.net/Links/2012ITRS/Home2012.htm
2. Lent, C., Tougaw, P.: A device architecture for computing with quantum dots. Proc. IEEE **85**(4), 541–557 (1997)

3. Timler, J., Lent, C.: Maxwell's demon and quantum-dot cellular automata. J. Appl. Phys. **94**, 1050–1060 (2003)
4. Hänninen, I., Takala, J.: Binary adders on quantum-dot cellular automata. J. Sig. Proc. Syst. **58**(1), 87–103 (2010)
5. Hänninen, I., Takala, J.: Binary multipliers on quantum-dot cellular automata. Facta Universitatis **20**(3), 541–560 (2007)
6. Hänninen, I., Takala, J.: Irreversible bit erasures in binary adders. In: Proceedings of the 10th IEEE Conference on Nanotechnology, Seoul, Republic of Korea, 17–20 August 2010, pp. 223–226 (2010)
7. Hänninen, I., Takala, J., Lent, C.: Irreversible bit erasures in binary multipliers. In: Proceedings of the 2011 IEEE International Symposium on Circuits and Systems, Rio de Janeiro, Brazil, 15–18 May 2011, pp. 2369–2372 (2011)
8. Hänninen, I., Takala, J.: Irreversibility induced density limits and logical reversibility in nanocircuits. In: Proceedings of IEEE/ACM International Symposium on Nanoscale Architectures, Amsterdam, The Netherlands, 4–6 July 2012
9. Landauer, R.: Irreversibility and heat generation in the computing process. IBM J. Res. Dev. **5**, 183–191 (1961)
10. Berut, A., Arakelyan, A., Petrosyan, A., Ciliberto, S., Dillenschneider, R., Lutz, E.: Experimental verification of Landauer's principle linking information and thermodynamics. Nature **483**, 187–189 (2012). http://dx.doi.org/10.1038/nature10872
11. Valiev, K.A., Starosel'skii, V.I.: A model and properties of a thermodynamically reversible logic gate. Russian Microelectron. **29**(2), 83–98 (2000)
12. Younis, S.G.: Asymptotically zero energy computing using split-level charge recovery logic. Ph.D. thesis (1994). http://dspace.mit.edu/handle/1721.1/7058
13. Starosel'skii, V.I.: Adiabatic logic circuits: a review. Russian Microelectron. **31**(1), 37–58 (2002)
14. Lim, J., Kim, D.-G., Chae, S.-I.: Reversible energy recovery logic circuit and its 8-phase clocked power generator for ultra-low-power applications. IEICE Trans. Electron. **E82-C**(4), 646–653 (1999)
15. Lim, J., Kim, D.-G., Chae, S.-I.: nMOS reversible energy recovery logic for ultra-low-energy applications. IEEE J. Solid-State Circuits **35**(6), 865–875 (2000)
16. Yibin, Y., Roy, K.: Energy recovery circuits using reversible and partially reversible logic. IEEE Trans. Circuits Syst. I: Fundam. Theory Appl. **43**(9), 769–778 (1996). doi:10.1109/81.536746
17. Dueck, G.W.: Synthesis of Toffoli Networks: status and challenges. In: International Symposium on Electronic System Design (ISED), 19–22 December 2012, pp. 11–16 (2012). doi:10.1109/ISED.2012.26
18. Lent, C.S., Liu, M., Lu, Y.: Bennett clocking of quantum-dot cellular automata and the limits to binary logic scaling. Nanotechnology **17**(16), 4240–4251 (2006)
19. Bennett, C.: Logical reversibility of computation. IBM J. Res. Dev. **17**, 525–532 (1973)
20. Texas Instruments Inc.: The TTL Data Book for Design Engineers, 2nd edn, pp. 7-484–7-486. Texas Instruments Inc., Dallas (1976)
21. Vladimirescu, A., Liu, S.: The simulation of MOS integrated circuits using SPICE2. Technical report no. UCB/ERL M80/7, University of California, Berkeley (1980)
22. Arizona State University Predictive Technology Model (PTM) Website [Online]. http://ptm.asu.edu
23. Weinstein, D., Bhave, S.A.: The resonant body transistor. Nano Lett. **10**(4), 1234–1237 (2010). doi:10.1021/nl9037517
24. Hänninen, I., Lent, C.S., Snider, G.L.: Models of irreversibility for binary adders. In: Proceedings of IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), Invited paper, Columbus, OH, USA, 4–7 August 2013, pp. 1071–1074 (2013)

Modular Dissipation Analysis for QCA

İlke Ercan and Neal G. Anderson(✉)

Department of Electrical and Computer Engineering,
University of Massachusetts Amherst, Amherst, MA, USA
{iercan,anderson}@ecs.umass.edu

Abstract. A modular approach for determination of lower bounds on heat dissipation in clocked quantum-cellular automata (QCA) circuits is proposed, and its application is illustrated. This approach, which is based on a methodology developed previously for determining dissipation bounds in nanocomputing technologies, simplifies analysis of clocked QCA circuits that are designed according to specified design rules. Fundamental lower bounds on the dissipative costs of irreversible information loss for a (generally large and complex) QCA circuit are obtained in the modular approach by (i) decomposing the circuit into smaller zones, (ii) obtaining dissipation bounds for the individual circuit zones, and (iii) combining results from the individual-zone analyses into a single bound for the full circuit. The decomposition strategy is specifically designed to enable this analytical simplification while ensuring that the consequences of intercellular interactions across zone boundaries - interactions that determine the reversibility of local information loss in individual zone - are fully preserved and properly captured in the modular analysis. Application of this approach to dissipation analysis of a QCA half adder is illustrated, and prospects of using the modular approach for automation of QCA dissipation analyses is briefly discussed.

Keywords: Nanocomputing · QCA · Heat dissipation

1 Introduction

Field-coupled nanocomputing (FCN) paradigms - both electronic and magnetic - offer the possibility of computation at energy efficiencies far superior to what CMOS will ultimately provide. Logical operations are driven by interactions between primitive computing elements that irreversibly discard energy and information without exchanging mass. This obviates the energetically costly requirement for constant "opening" of the elements to one another and to particle reservoirs, as is required in CMOS to transfer information and maintain the computational "working substance."

However, any computing paradigm that requires irreversible loss of information – e.g. during implementation of logical transformations and erasing and/or overwriting information remaining from previous computations – is necessarily dissipative [1,2]. FCN circuits are no exception in this regard. Heat dissipation

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 357–375, 2014.
DOI: 10.1007/978-3-662-43722-3_15, © Springer-Verlag Berlin Heidelberg 2014

from irreversible information loss may be substantial in FCN circuits with densities and computational throughputs far exceeding "end-of-the-roadmap" CMOS (e.g. [3,4]). This motivates investigation of fundamental lower bounds on dissipation in FCN circuits, and of ultimate performance limits that follow from these bounds under various assumptions regarding heat removal capability.

In previous work [3,5], we introduced a very general methodology that enables determination of fundamental heat dissipation bounds for concrete and nontrivial nanocomputing scenarios. We showed how irreversible information loss can be isolated in any clocked quantum-cellular automata (QCA) circuit, and how lower bounds on the resulting energy dissipation can be obtained. The bounds resulting from this approach are truly fundamental, in that they depend only on the circuit structure, clocking scheme, and temperature of the circuit's environment; implementation-specific quantities such as the kink energy do not appear in these energy bounds. Combining results from such analyses with assumptions on circuit scale and clock rate for specified QCA circuits, lower bounds on areal heat dissipation can be obtained at any desired computational throughput for arbitrary QCA circuits.

In this work, we "modularize" this methodology to facilitate – and possibly automate – the determination of fundamental dissipation bounds for large, complex QCA circuits designed according to specified rules. We begin in Sect. 2 with a review of our general methodology, which enables fundamental lower bounds on heat dissipation to be determined for nanoelectronic circuits designed within various paradigms through a manual analysis (hereafter the "general approach"). We then show how the general approach, applied to QCA, can be "modularized" by decomposing a circuit into smaller zones – such that dissipative contributions can be evaluated separately for each zone and summed – and how this can simplify dissipation analysis of QCA circuits (hereafter the "modular approach"). We emphasize the enabling feature of this decomposition; that it preserves the effects of field interactions across tile boundaries that influence the reversibility of information loss while avoiding spurious contributions that can be introduced if boundaries are improperly placed. Comparative application of the general and modular analyses to a QCA adder circuit is presented in Sect. 3. In Sect. 4, we briefly discuss prospect for automation of QCA dissipation analysis using the modular approach. Automated analysis could, for example, enable evaluation of dissipation bounds for full QCA-based processor architectures [6] via the approach of [7] by simplifying circuit-level dissipation analyses of the constituent circuit blocks. We conclude this work in Sect. 5.

2 Dissipation Analysis

2.1 General Approach

We begin by sketching our general approach for determination of fundamental lower bounds on the dissipative costs of digital computation. The approach allows fundamental bounds to be obtained for specified circuits realized in a

concrete nanocomputing paradigm by bringing physical law directly to bear on the underlying computational strategy that defines the paradigm.

For any given nanocomputing circuit operated under a certain clocking scheme, a fundamental lower bound on the dissipative cost of executing the circuit's computational function is determined via a two-stage process: *abstraction* and *analysis*. In the abstraction phase, an idealized physical abstraction of the circuit and its operation is constructed. This abstract description is constructed so it captures the essential functional features of the underlying computational strategy, implemented precisely as envisioned in the computational paradigm. (We call this "paradigmatic operation.") In the analysis stage, a physical-information-theoretic analyses of the abstract circuit and its surroundings is performed for the time intervals relevant to each computational step in the circuit's computational cycle. These step-specific analyses yield a lower bounds on the amount of energy that is unavoidably dissipated into the circuit's local environment under paradigmatic operation, including both the dissipation required to execute logically irreversible operations and other unavoidable paradigm-dependent "overhead" costs e.g. particle supply costs required to maintain the computational "working substance" in transistor-based paradigms). The dissipation bounds obtained for each computational step are finally summed over all steps in the computational cycle to lower bound the input-averaged energy cost of each computation performed by the circuit. These two phases are described in further detail below.

Abstraction. The abstraction is composed of two main stages. First, we create a physical abstraction of the circuit and its surroundings in a globally closed and isolated universe. Second, in the process abstraction, we present assignments to identify local physical operations each of which is decomposed into a control and a restoration operations. The circuit's interaction with the other information-bearing subsystems and the bath is described by the control operations. The restoration operations represent the coupling between the remote environment and the bath. The control and restoration operations provide the circuit evolution required to implement computation. We now discuss the abstraction procedure in detail.

Physical Abstraction: The physical abstraction of the circuit and its surroundings is depicted schematically in Fig. 1. The upper half of the figure represents the computationally relevant domain. This domain includes an information processing artifact $\mathcal{A}$ which is the computing circuit of interest and computationally supporting subsystems such as external registers and adjacent circuit stages, as well as an input referent $\mathcal{R}$ that holds a physical instantiation of the input data that will be processed by the artifact. The lower half represents the environmental domain consisting of a heat bath $\mathcal{B}$, which is the part of the environment that is in direct thermal contact with the artifact and nominally at temperature T. The greater environment $\bar{\mathcal{B}}$ includes heat reservoirs that "rethermalize" the bath and anything else that is required to ensure that the universe is globally closed. Constructing a globally closed universe as presented above enables us to assume it evolves unitarily via Schrödinger's equation. The global unitarity, together with identification

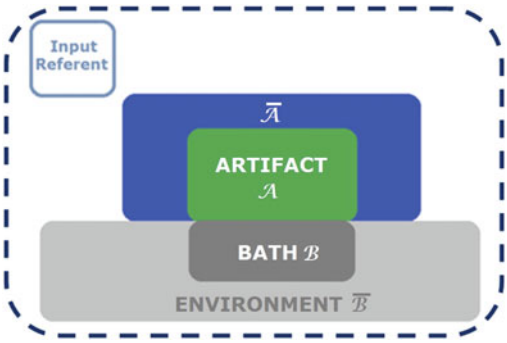


Fig. 1. Physical abstraction of an information processing artifact, such as a QCA circuit, and its surroundings in a globally closed universe.

of the subsystems that are coupled in each step, allows determination of the fundamental lower bounds that we are after in this work.

Process Abstraction: During computation the subsystems are driven away from equilibrium but then rethermalized as a part of the process abstraction. We identify a set of local physical operations $\phi_t \in \{\phi_t\}$, each of which is decomposed into a control process and a restoration process. Control operations are the local operations that act during specified time intervals to change the states of representational elements in the artifact either unconditionally or conditioned on the states of other representational elements. Typically, they involve interaction between the artifact, other information-bearing subsystems, and the bath. Restoration processes are the local operations that couple the remote environment $\bar{\mathcal{B}}$ to the bath $\mathcal{B}$ and local particle reservoirs in $\mathcal{A}$. These operations rethermalize the bath and recharge the reservoirs after they have been driven from their nominal states by control operations. Together, the control and restoration phases make up the sequence of global system evolutions required for implementation of computation in the circuit.

Analysis. The second step in our approach involves spacetime decomposition of the circuit function (operational decomposition) and physical-information-theoretic analyses of local dissipation into the bath throughout the computational cycle (cost analysis). Any local information about the initial state of $\mathcal{A}$ that is irreversibly lost during a computational step induces dissipation in $\mathcal{B}$ before being swept completely into $\bar{\mathcal{B}}$ during the restoration process. Note that, loss of initial-state information from part of $\mathcal{A}$ is locally irreversible if it is erased in the absence of interaction with other parts of $\mathcal{A}$ or $\bar{\mathcal{A}}$ that hold or receive copies of the initial state during the clock step. This locally irreversible information loss affects the state of $\mathcal{B}$ during an operation's control process, which is precisely the point at which the dissipation costs are "cashed out" in our approach. The manual calculation of our bounds involves calculation of dissipation for each step of and summing the contribution to obtain the total cost of the complete computation cycle. The details of this procedure is outlined below.

Operational Decomposition: In this analysis step, we perform a spacetime decomposition of the circuit. We first define clock zones and subzones. A clock zone is a set of representational elements that are simultaneously subjected to the same control operation in any given time step. Each clock zone may consist of physically disjoint subsets of representational elements - called clock subzones - that do not interact with one another directly as they change state. We denote the u^{th} clock zone as $C(u)$ and the l^{th} clock subzone of $C(u)$ as $C_l(u)$. We, then, define a clock step which represents a time step during which a specified set of control operations are applied to various clock zones. We denote the assignment of control operation ϕ_t to clock zone $C(u)$ as $(C(u); \phi_t)$, the v^{th} clock step ϕ_v is specifically defined as an assignment $\phi_v : \{(C(u); \phi_t)\}_v$ of control operations to all clock zones. The restoration processes that rethermalize the bath and recharge the artifact's local particle reservoirs after the associated control operation drives these subsystems from their nominal states is also included in a clock step. Hence, we can define a clock cycle as one period of the periodic sequence $\phi = \phi_1\phi_2\phi_3\ldots$ of clock steps applied to the artifact to enable its operation.

This allows us to define the computational steps and cycle. The computational step c_k, defined for the η^{th} input $x(\eta)$ in an input sequence $\ldots x(\eta - 1)x(\eta)x(\eta{+}1)\ldots$, is the k^{th} of K clock steps required for evaluation of the output. The computational cycle is then represented as $\mathit{\Gamma}^{(\eta)} = c_1\ldots c_k\ldots c_K$ of the K clock steps required to fully implement the logic operation for the η^{th} input $x(\eta)$ including the phase that loads $x(\eta)$ into the artifact, the phases that evaluate the output $(x(\eta))$, and the phases that transfer the output to the outside world and erase all information about the input from the artifact. We denote c_1 as the LOAD phase and c_K as the phase in which all correlation between the computational state of the artifact and the i^{th} referent is lost. The computational cycle $\mathit{\Gamma}^{(\eta)}$ may include clock steps from multiple clock cycles, and, in artifacts that pipeline input data, $\mathit{\Gamma}^{(\eta)}$ may exclude clock steps that implement operations belonging only to other computational cycles (e.g. $\mathit{\Gamma}^{(\eta-1)}$ and $\mathit{\Gamma}^{(\eta)}$), i.e. clock steps that do not affect representational elements whose states depend on the η^{th} input. Thus, the η^{th} computational cycle includes only clock phases that contribute directly to evaluation of output in the information processing artifact.

Cost Analysis: We move on to the calculation of total dissipative cost associated with one computational cycle based on information dynamics which involves data zones and subzones. For the η^{th} computational cycle, the k^{th} data zone is the set of representational elements that, at the completion of the k^{th} computational step c_k, hold information about the input data $x(\eta)$. A data zone may contain clock subzones belonging to multiple clock zones, and need not include all subzones belonging to any given clock zone. It may consist of physically disjoint subsets of representational elements - called data subzones - which do not interact with one another directly during some or all of the computational steps. We denote the data zone associated with computational step c_k as $D(c_k)$, and the w^{th} data subzone of $D(c_k)$ as $D_w(c_k)$. Note that, regardless of the circuit implementation, there is one data zone defined at the end of computational step of the computational cycle from c_1 to c_{K-1}. By definition, there are no

data zones at the completion of step c_K. As a computation progress through steps in the computational cycle, the data zone will change its size and topology while propagating from input to output. Data subzones will generally split and merge throughout a computation, generally changing in number from step to step.

The total information loss is the sum contributions from information lost to the bath in each computational step, with "information loss" for the k^{th} step defined as the amount of information about the state of each subzone in the $(k-1)^{th}$ data zone that is not in the state of the corresponding subzone in the k^{th} data zone, summed over all data subzones.

We assume that data subzones that do not interact with one another essentially "act alone" for the purposes of dissipation calculations, as the interactions that erase information about the prior state will necessarily be local. This highlights the importance of correctly identifying and classifying the relevant physical interactions that occur throughout the computational cycle, as identification of the data subzones - and thus the appropriate level of analysis for obtaining dissipation bounds - depends on the nature of these interactions.

The dissipation bounds are then obtained as follows. In any given computational step, any information lost from $\mathcal{A}$ that is not completely transferred to $\bar{\mathcal{A}}$ results in local energy dissipation into the bath. Information is "lost" from data subzone $D_w(c_{k-1})$ during computational step c_k if, at the conclusion of c_k, the initial states of erased clock subzones of $D_w(c_{k-1})$ cannot uniquely inferred from the final states of clock subzones $C_l(u) \in D(c_k)$ that interacted directly with $D_w(c_{k-1})$ during c_k. The total dissipative cost of one computational step is taken to be the sum of contributions from all informationally lossy data-subzone-to-data-subzone transitions, and the cost of processing one input is then the sum of contributions from each computational step. Based on these assumptions, only one of the clock subzones contributes to the dissipative cost associated with each data subzone in the k-th step. For the w-th data subzone and k-th step, this clock zone is denoted $\mathcal{C}_w^{(k)}$. The total dissipative cost of one computational step is obtained by summing the contributions from all informationally lossy data-subzone-to-data-subzone transitions, and he cost of processing one input is obtained by summing of contributions from each computational step. This is to say that the total energy cost is additive at the subzone level, i.e. the energy costs associated with these individual data-subzone information losses can be "cashed out" individually and summed over a full cycle to get

$$\Delta\langle E\rangle = \sum_{k=1}^{K} \Delta\langle E\rangle_k = \sum_{k=1}^{K} \left[\sum_{w\in\{w\}_{k-1}} \Delta\langle E\rangle_{k-1}^{w} \right] \tag{1}$$

where $\langle E\rangle_k$ and $\Delta\langle E\rangle$ represent the average energy transferred to the bath in the k^{th} computational step and over the full cycle, respectively. This again assumes that data subzones that do not interact with one another for the purposes of dissipation calculations, as their environmental interactions will be local. With this assumption – that the $\mathcal{C}_w^{(k)}$ interact locally with the bath since by definition

they do not interact with one another – the dissipative cost of information loss from subzones in a given computational step is additive and the net dissipative cost of erasure for all subzones erased during the k^{th} computational step can be lower bounded as

$$\Delta \left\langle E^{\mathcal{B}} \right\rangle_k \geq \sum_{\mathcal{C}_w^{(k)}} -k_B T \ln(2) \Delta \mathcal{I}^{\mathcal{C}_w^{(k)} \mathcal{D}_w^{(k)}} \tag{2}$$

where $-\Delta \mathcal{I}^{\mathcal{C}_w^{(k)} \mathcal{D}_w^{(k)}}$ is the amount of information about the initial state of clock subzone $\mathcal{C}_w^{(k)}$ that is locally and irreversibly lost from $\mathcal{C}_w^{(k)} \cup \mathcal{D}_w^{(k)}$ during the k^{th} computational step [2].

2.2 Modular Approach

The general approach outlined above enables isolation and quantification of irreversible information loss in digital circuits implemented in a wide variety of nanocomputing paradigms. Full account is taken of interactions between neighboring circuit structures that, in some paradigms, can render information erasure locally reversible (e.g. the ERASE WITH COPY operation in QCA). Circuit regions associated with irreversible information loss – which may be related to logic gates or other functional circuit blocks in ways that are not obvious – need not be known *a priori* to apply the general approach, allowing fundamental dissipation bounds to be obtained for new circuit structures and/or in unfamiliar nanocomputing paradigms where the appropriate "rules of thumb" have not yet been identified. This generality comes at an analytical cost, however, as the identification of data zones and subzones is required on each computational step to allow tracking of information flow and proper isolation of irreversible information loss. This process is laborious for large circuits analyzed via the general approach.

The modular approach of the present work allows these same dissipation bounds, obtained "holistically" in the general approach, to be obtained under certain conditions through a simplified "reductionist" procedure of (*i*) decomposing a circuit into smaller *fixed* regions or zones, (*ii*) evaluating dissipation bounds for those zones that are necessarily dissipative (hereafter "dissipation zones"), and (*iii*) adding the bounds for each dissipation zone to obtain a dissipation bound for the full circuit. This simplification is possible, however, only after it has been established that a circuit decomposition that properly captures cross-boundary interactions can be performed and the appropriate decomposition rules have been identified.

Suppose that such a decomposition *is* possible for a given circuit, that N_{diss} dissipation zones d_n are identified, and that each dissipation zone is "used" once for processing of each input to the full circuit. Where this is the case, the bound on the total energy dissipated in processing one input over a computational cycle simplifies to

$$\Delta E_{diss} = \sum_{n=1}^{N_{diss}} \Delta \langle E^{\mathcal{B}} \rangle_{d_n} \tag{3}$$

where $\Delta\langle E^{\mathcal{B}}\rangle_{d_n}$ is the amount of energy dissipated in the n-th dissipation zone on each use. Lower bounds on the $\Delta\langle E^{\mathcal{B}}\rangle_{d_n}$, which are fixed by the amount $\Delta\mathcal{I}_{d_n}$ of information irreversibly lost from the dissipation zones on each use, would be obtained from dissipation analyses performed individually for each dissipation zone.

The advantages of this approach would, of course, be questionable if new decomposition rules had to be sought and codified anew for each dissipation zone in every circuit that is to be analyzed. A more productive strategy would be to tie circuit decomposition rules to *design rules* rather than to individual circuits, so a single set of decomposition rules applies to *all* circuits realized within a given nanocomputing paradigm that are designed according the specified design rules. If this is possible, it would enable the analytical simplification promised by the modular approach while imposing only mild constraints on the allowed space of circuit structures: Adherence to specified design rules is a modest and familiar constraint from a circuit design perspective. Below we show that decomposition rules can indeed be formulated for Landauer-clocked QCA circuits that are constructed according to example design rules that reflect common design practice, and we demonstrate that simplified modular analysis based on these rules yields bounds identical to those of the general approach for a concrete QCA circuit.

3 Illustrative Example: A QCA Half Adder

As a vehicle for exploring and demonstrating the modular approach, we consider the Landauer-clocked QCA half adder circuit depicted in Fig. 2. The cell layout, timing diagram, and logic diagram are all shown, with color coding corresponding to clocking zones. This circuit, which is designed so it is free of wire crossings, is composed of 135 cells that make up five majority gates (each with one fixed input), two inverters, and required interconnects. The cells in each clock zone cycle periodically through switch, hold, release and then relax phases to propagate input information through the circuit via Landauer clocking [8]. New input data is loaded into the circuit every time the adder cells adjacent to the input cells A and B go through the switch phase of the clock cycle. The corresponding outputs are available in the adder cells adjacent to the output cells S and C two full clocking cycles later, and are loaded into the output cells when these clock zones are in the subsequent hold phase.

We sketch dissipation analysis of this circuit via the general approach in Sect. 3.1. In Sect. 3.2, we present the set of QCA circuit design rules obeyed by this adder design[1], establish corresponding decomposition rules that enable modular dissipation analysis of any circuit that adheres to these design rules, and employ these decomposition rules in a simple modular dissipation analysis

[1] Studies of a similar QCA adder circuit, including detailed dissipation analysis via the general approach, have been presented elsewhere [3,5]. Differences in the adder of these previous studies and the adder of this work (Fig. 2) stem from reconciliation of the previous design with the QCA design rules presented in Sect. 3.2. Statement of these design rules is deferred to Sect. 3.2 since the procedures used for dissipation analysis via the general approach are independent of them.

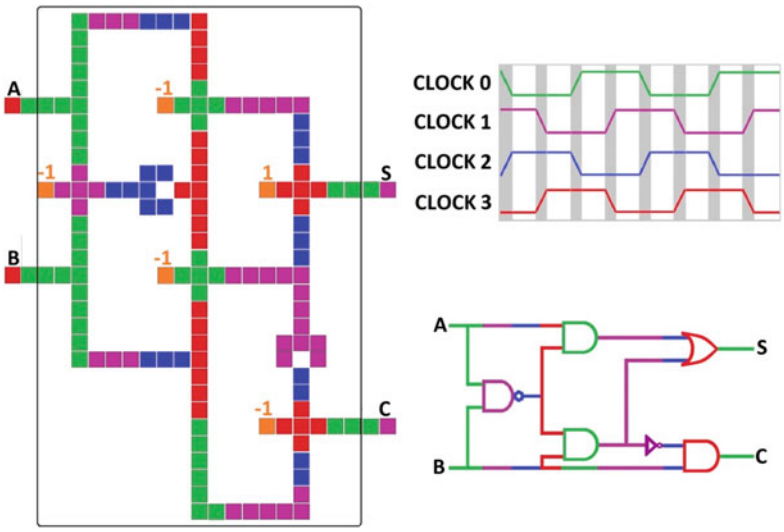


Fig. 2. Cell layout, timing, and logic diagram for the Landauer-clocked QCA half adder circuit used in this work to illustrate and compare general and modular dissipation analyses.

of the adder. The two possible input bit values are assumed to occur with equal probabilities for inputs A and B in these dissipation analyses, although uniform input probabilities are not required by the formalism.

3.1 Analysis via General Approach

The initial step in applying the general approach is to construct the physical abstraction of the circuit and its surroundings. Here the QCA adder is the information processing artifact $\mathcal{A}$ of Fig. 1. The input referent system $\mathcal{R}$, which holds a physical instantiation of the string of two-bit inputs to be processed by the adder, is included in $\bar{\mathcal{A}}$, as are the input and output cells A, B, S, and C. The environmental domain includes the heat bath $\mathcal{B}$ at temperature T, which is in direct thermal contact with the artifact, and the remote environmental subsystems $\bar{\mathcal{B}}$ that "rethermalize" the bath when it is driven from a thermal state by implementation of adder operations. The next step is to construct the process abstraction. The four clock phases defined for Landauer clocking of the QCA circuit – switch, hold, release and relax – are mapped onto the four physical operations $\phi_1 \ldots \phi_4$ that act locally to implement computation over the $K = 11$ steps of the computational cycle.

The final step in analysis via the general approach is the cost analysis. This begins with identification and assignment of data zones and data subzones, as described in detail in our previous work [3,5]. This can be quite involved even for relatively simple circuits. For the adder circuit of the present work, the data zones and subzones are

$$
\begin{aligned}
D(c_1) &= D_1(c_1) \cup D_2(c_1)\\
D(c_2) &= D(c_2)\\
D(c_3) &= D_1(c_3) \cup D_2(c_3) \cup D_3(c_3)\\
D(c_4) &= D_1(c_4) \cup D_2(c_4) \cup D_3(c_4)\\
D(c_5) &= D(c_5)\\
D(c_6) &= D_1(c_6) \cup D_2(c_6) \cup D_3(c_6)\\
D(c_7) &= D_1(c_7) \cup D_2(c_7) \cup D_3(c_7) \cup D_4(c_7)\\
D(c_8) &= D_1(c_8) \cup D_2(c_8)\\
D(c_9) &= D_1(c_9) \cup D_2(c_9)\\
D(c_{10}) &= D_1(c_{10}) \cup D_2(c_{10})
\end{aligned}
$$

where each data subzone corresponds to clock subzones (numbered from top to bottom and then from left to right) as

$$
\begin{aligned}
D_1(c_1) &= C_1(1)\\
D_2(c_1) &= C_2(1)\\
D_1(c_2) &= C_1(1) \cup C_1(2) \cup C_2(2) \cup C_2(1) \cup C_3(2)\\
D_1(c_3) &= C_1(2) \cup C_1(3)\\
D_2(c_3) &= C_2(2) \cup C_2(3)\\
D_3(c_3) &= C_3(2) \cup C_3(3)\\
D_1(c_4) &= C_1(3) \cup C_1(4)\\
D_2(c_4) &= C_2(3) \cup C_2(4)\\
D_3(c_4) &= C_3(3) \cup C_3(4)\\
D_1(c_5) &= C_1(4) \cup C_5(1) \cup C_2(4) \cup C_2(5) \cup C_3(4) \cup C_3(5)\\
D_1(c_6) &= C_1(5) \cup C_1(6)\\
D_2(c_6) &= C_2(6) \cup C_2(5)\\
D_3(c_6) &= C_3(5) \cup C_4(6)\\
D_1(c_7) &= C_1(6) \cup C_1(7)\\
D_2(c_7) &= C_2(7) \cup C_2(6) \cup C_3(7)\\
D_3(c_7) &= C_3(6) \cup C_4(7)\\
D_1(c_8) &= C_1(7) \cup C_1(8) \cup C_2(7)\\
D_2(c_8) &= C_3(7) \cup C_2(8) \cup C_4(7)\\
D_1(c_9) &= C_1(8) \cup C_1(9)\\
D_2(c_9) &= C_2(8) \cup C_2(9)\\
D_1(c_{10}) &= C_1(10)\\
D_2(c_{10}) &= C_2(10).
\end{aligned}
$$

The data subzones for the data zones $D(c_3)$ and $D(c_4)$ are, for example, shown in Fig. 3. With this, the cost analysis can be completed. The energy dissipated

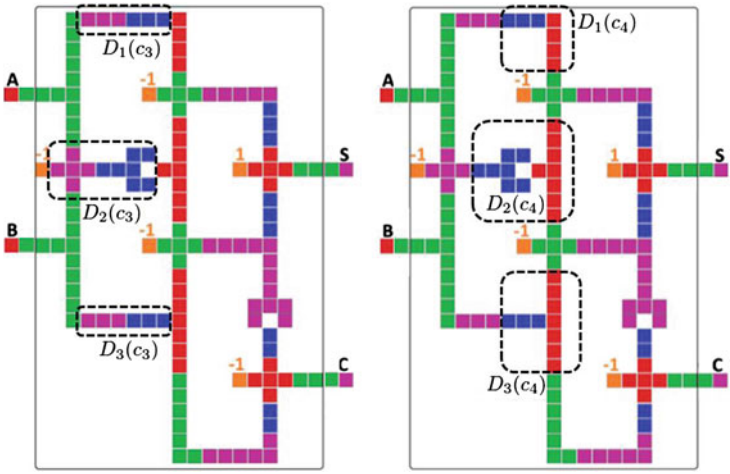


Fig. 3. Data subzones comprising data zones $D(c_3)$ (left) and $D(c_4)$ (right), which are the circuit regions holding input information at the conclusion of the third and fourth computational steps, respectively. Computational step c_4 is one of the steps that unavoidably suffers irreversible information loss resulting energy dissipation.

in one computational cycle of the QCA half adder is the sum of contributions from each step

$$\Delta E_{diss} = \sum_{k=1}^{11} \Delta E_{diss}(c_k) = \sum_{k=1}^{11} \Delta \langle E^{\mathcal{B}} \rangle_k \tag{4}$$

where $\Delta \langle E^{\mathcal{B}} \rangle_k$ is the change in the expected energy of the bath $\mathcal{B}$ during the control phase of the k^{th} computational step c_k. Under paradigmatic operation, our analysis reveals that only computational steps c_4, c_6 and c_8 are dissipative. This dissipation is a result of irreversible loss of input information from associated clock subzones that occur as the data zones evolve during these steps. Physical-information-theoretic evaluation of the corresponding single-step lower bounds on $\Delta \langle E^{\mathcal{B}} \rangle_k$ for the dissipative steps via Eq. (2) yields

$$\begin{aligned} \Delta \langle E^{\mathcal{B}} \rangle_4 &\geq 1.19 k_B T \ln(2) \\ \Delta \langle E^{\mathcal{B}} \rangle_6 &\geq 1.38 k_B T \ln(2) \\ \Delta \langle E^{\mathcal{B}} \rangle_8 &\geq 1.19 k_B T \ln(2). \end{aligned}$$

The resulting lower bound on the dissipative cost of one computational cycle in the Landauer-clocked QCA half adder is then

$$\Delta E_{diss} \geq (3.76) k_B T \ln(2). \tag{5}$$

This full-cycle bound is about 7.5× the $0.5 k_B T \ln(2)$ lower bound for any half adder that erases its input, with the excess irreversibility resulting from locality

of information loss associated with the specific circuit structure and clocking scheme.

3.2 Analysis via Modular Approach

Design Rules. We begin discussion of our modular approach by articulating an example set of QCA design rules, since circuit decomposition rules defined for a set of design rules can be applied to modular dissipation analysis of any circuit designed according to these rules. Our example design rules, which are specific to Landauer-clocked combinational QCA circuits with no wire crossings, are as follows:

1. *Wires:* All wires are linear arrays of "90-degree" wires, i.e. with adjacent cells oriented as in Fig. 4[2], and with right-angle corners. Wire segments corresponding to individual clock zones are of length

$$2 \leq N \leq \exp[E_k/k_B T]$$

   (in units of cells), where E_k is the kink energy and k_B is Boltzmann's constant. The minimum allowable wire pitch is three cells.
2. *Inverters:* Inverters are of the "complex" form shown in Fig. 5, with identically clocked, two-cell input leg and an identically-clocked two-cell output leg as shown.
3. *Majority Gates:* Majority gates are of the standard three-input configuration in Fig. 6. The four input and output legs adjacent to the central cell – hereafter the "device cell" – are identically clocked, and the four identically clocked input and output legs are of equal length.

Fig. 4. Three example sections of QCA wires with "90-degree" cell orientation, two of which include right angle corners.

Fig. 5. A complex inverter with two-cell input and output legs and specified clock zones.

[2] We do not consider "45-degree" wires since the design rules we propose are not intended for wire crossings in a plane.

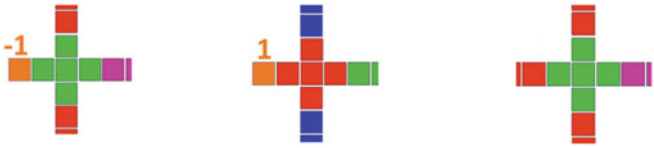


Fig. 6. QCA majority gates, with and without one fixed input, and associated clocking. The majority gates with cell polarizations fixed as −1 and +1 function as two-input AND or OR gates, respectively.

The lower and upper bounds on the number of identically clocked cells in the wire segments, which are based on considerations discussed in Refs. [9] and [10] respectively, are selected to help ensure reliable information transfer. The minimum pitch is selected to minimize crosstalk from adjacent wires. The requirement of equal-length input and output legs in majority gates helps to ensure simultaneous arrival of new inputs at the device cell, and thus fair voting [9].

We emphasize this simple set of design rules is presented for the purposes of illustrating our modular dissipation analysis. Having said that, these simple rules allow for construction of QCA circuits that implement any desired Boolean function[3] and are generally consistent with common QCA design practice. It is easily verified that the adder design of Fig. 2 adheres to these simple design rules.

Decomposition Rules. Decomposition of a QCA circuit for modular analysis requires that the circuit first be segmented into zones according to a design-rule-specific set of decomposition rules. These rules stipulate how boundaries between the zones are to be placed. For the set of example design rules presented above, the decomposition rules are simply as follows:

1. Every cell in the circuit must belong to one and only one zone.
2. All zone boundaries must be placed *between* adjacent, identically clocked cells, perpendicular to the direction of information flow. The same applies to the boundary enclosing the full circuit.
3. Zone boundaries are to be placed between the two cells of the input legs and between the two cells of the output legs of inverters, with no boundaries in between.
4. Majority gates must be enclosed within a single zone as in Fig. 7, i.e. with zone boundaries placed so they enclose (a) the device cell and the identically clocked, equal-length input and output legs, (b) one cell adjacent to each input leg in the neighboring clock zone[4], and (c) one cell adjacent to the output leg in the neighboring clock zone.

[3] AND, OR, and NOT form a universal set of primitives, and three-input majority gates can implement two-input AND and OR functions if one if the inputs is appropriately biased as shown in Fig. 6.

[4] Alternatively, this cell could be fixed if the corresponding input is to be biased.

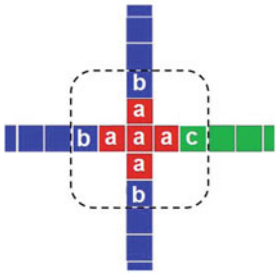


Fig. 7. A dissipation zone, including placement of boundaries, for circuits designed according to the design rules of this work. For such circuits, all dissipation zones resulting from modular dissipation analyses performed according to the decomposition rules of this work will have this form. The design rules preclude irreversibility in all circuit structures other than majority gates.

Such a decomposition is always possible for circuits designed according to the design rules presented above. Use of the decomposition rules presented here greatly simplifies dissipation analysis, as shown below.

Dissipation Analysis. Recall that the modular approach aims to simplify evaluation of the fundamental dissipation bounds obtained via the general approach by partitioning the circuit into smaller zones – once and for all at the beginning of the analysis – and applying the general approach piecemeal to determine the dissipative contributions from each zone. The partitioning process can, however, introduce an artifact that will cause the modular approach to overestimate the dissipative contributions from the individual zones, and thus from the circuit as a whole when the individual contributions are summed. We now describe the origin of this artifact, and show that it is avoided in circuits that are designed according to the above design rules and partitioned according to the above decomposition rules.

Information propagation in Landauer-clocked QCA is not dissipative under paradigmatic operation. Information lost from a block of adjacent, identically clocked cells in a QCA wire during the "relax" phase of the clocking cycle are always erased in the presence of (and in interaction with) an identical copy that has already been transferred to – and is locked in – an adjacent block of cells that is immediately "downstream." This is the reversible ERASE WITH COPY operation. If the block of cells being erased belongs to a particular circuit zone, but the downstream copy does not, then the erasure will appear *irreversible* – and thus dissipative – in a dissipation analysis that treats the zone including the erased cells as independent and isolated. Neglect of the cross-boundary interactions that renders the erasure reversible are lost, causing the simplified modular analysis to fail.

If the decomposition rules stated above are followed, however, any group of identically clocked cells is necessarily a wire segment that belongs to two circuit

zones. Information in the furthest downstream cell(s) of the "upstream" circuit zone is always also held in the furthest upstream cell(s) of the "downstream" circuit zone, since these two groups of adjacent "boundary cells" are identically clocked. Dissipation analysis of the upstream zone can thus neglect any apparent contributions from the clock phase where information is erased from the furthest downstream cells, since this information also belongs to the furthest upstream cells of the downstream circuit zone, and any dissipation that would result from irreversible erasure of this information in a subsequent clocking phase will be captured in analysis of the downstream zone.

This simple constraint on circuit decomposition results in major simplification of the dissipation analysis. Dissipative contributions from each circuit zone can be calculated independently and added, and the effects of cross boundary interactions captured by the general analysis are properly reflected. Furthermore, the only circuit zones that are necessarily dissipative – those designated as "dissipation zones" – are those enclosing majority gates; there is no irreversible information loss in zones that correspond to wire segments and inverters. Dissipation analysis thus requires only that the dissipation zones be identified and their contributions calculated.

The analysis is simplified even further by the fact that, on each "use," dissipation zones defined as above irreversibly lose information during one and only one clock transition: the clock transition in which the information-bearing state of the "core" of the dissipation zone – the device cell and surrounding identically clocked cells belonging to the input and output legs ((a) in Fig. 7) – is relaxed.

With this, we proceed to modular dissipation analysis of the adder circuit of Fig. 2. One can immediately identify $N_{diss} = 5$ dissipation zones, which are delineated and labeled in Fig. 8.

These dissipation zones can be analyzed separately, with each regarded as an independent "information processing artifact" (*cf.* Fig. 1). As per Eq. (3), the amount of energy dissipated in the processing of each input by the circuit is

$$\Delta E_{diss} = \sum_{n=1}^{5} \Delta \langle E^{\mathcal{B}} \rangle_{d_n} \tag{6}$$

where $\Delta \langle E^{\mathcal{B}} \rangle_{d_n}$ is the amount of energy dissipated during the critical clock phase in dissipation zone d_n. $\Delta \langle E^{\mathcal{B}} \rangle_{d_n}$ is lower bounded as [2]

$$\Delta \langle E^{\mathcal{B}} \rangle_{d_n} \geq k_B T \ln(2) \Delta \mathcal{I}_{d_n} \tag{7}$$

where $\Delta \mathcal{I}_{d_n}$ is the amount of information irreversibly lost from zone d_n during the dissipative clock phase. Using the same assumptions of pure, orthogonal QCA data states that were made in general analysis of Sect. 3.1, $\Delta \mathcal{I}_{d_n} = H_n(X|Y)$ where $H_n(X|Y)$ is the conditional Shannon entropy for the zone (gate) input and output random variables X and Y. Obtaining the probability mass functions (pmfs) for the various gate inputs that result from a uniform adder input pmf,

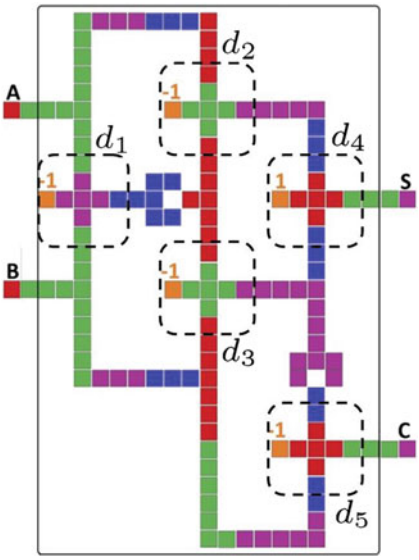


Fig. 8. Dissipation zones identified by application of the circuit decomposition rules to the QCA half adder of this work.

and evaluating the five required conditional entropies, we have

$$\begin{aligned}
\Delta\langle E^{\mathcal{B}}\rangle_{d_1} &\geq 1.1887 k_B T \ln(2) \\
\Delta\langle E^{\mathcal{B}}\rangle_{d_2} &\geq 0.6887 k_B T \ln(2) \\
\Delta\langle E^{\mathcal{B}}\rangle_{d_3} &\geq 0.6887 k_B T \ln(2) \\
\Delta\langle E^{\mathcal{B}}\rangle_{d_4} &\geq 0.5 k_B T \ln(2) \\
\Delta\langle E^{\mathcal{B}}\rangle_{d_5} &\geq 0.6887 k_B T \ln(2)
\end{aligned}$$

for the five dissipation zones. Summing these results, we obtain the bound

$$\Delta E_{diss} \geq (3.76) k_B T \ln(2) \tag{8}$$

on the dissipative cost of processing one adder input, which is indeed identical to that obtained from the general approach.

Evaluation of this dissipation bound, which was shown in Sect. 3.1 to be somewhat involved even for this simple circuit in general approach (and is laborious in more complex circuits like the $>10^5$-cell QCA ALU studied in [11]), is straightforward and simple in the modular approach. The vast analytical simplification was enabled by the consistency of the circuit structure with stated design rules, and the identification and formulation of an appropriate set of circuit decomposition rules specific to these design rules. With decomposition rules in hand for our design rules, modular dissipation analyses could be performed in exactly the same manner for any Landauer-clocked QCA circuit constructed according to the same design rules.

4 Prospects for Automation

The modular dissipation analysis presented here is much better suited to automation than is the general approach. For easy comparison, the flow of the general and modular dissipation analysis procedures is shown schematically in Fig. 9. An algorithm could certainly be devised that would enable automation of the general approach, but it would be complex and difficult to formulate and implement. The general approach requires that data zones and subzones be identified at each step, which requires that the flow of input information be tracked in space and time throughout the computational cycle. Irreversible information loss and associated energy dissipation depend upon changes in the amount of correlation between the states of these zones during each step.

It would be comparatively straightforward to formulate an algorithm for modular dissipation analysis of QCA circuits designed according to the design rules of this work and driven by a random input string with specified input pmf. The first step is simply to identify the device cells in the circuit, which could easily be performed by searching a simple matrix representation of the circuit layout. The second step is to determine the required joint pmfs and marginal (input and output) pmfs for the gates corresponding to the dissipation zones associated with each device cell. This could be achieved, perhaps in a simulator embedding QCADesigner simulation of the circuit, by building appropriately weighted input-output histograms for all gates in simulations that step through all adder inputs. The third step is to evaluate the conditional entropies, and thus the corresponding lower bounds on the dissipative contributions, for each dissipation zone, which is easily done once the joint input-output pmfs have

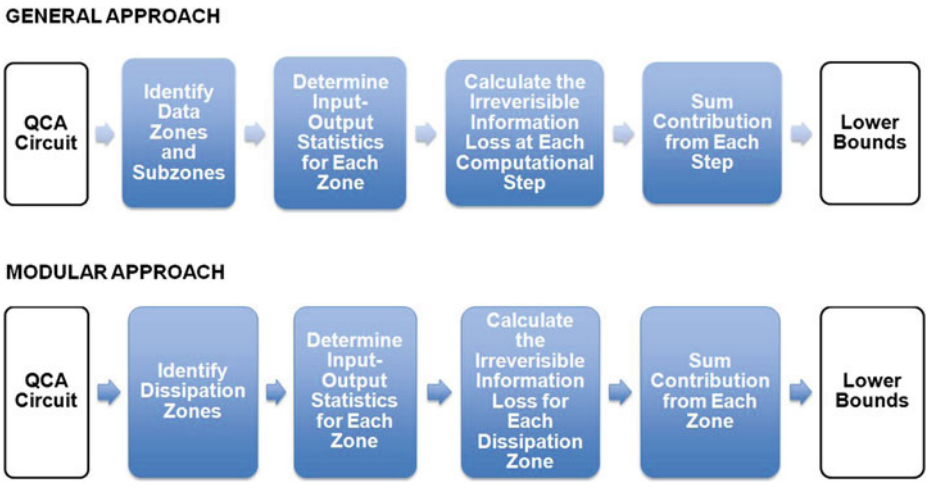


Fig. 9. Schematic representation of the general (top) and modular (bottom) dissipation analysis procedures discussed in this work.

been determined. The fourth and final step, is simply to sum the zone contributions to obtain a dissipation bound for the full circuit. None of these steps pose formulation or implementation difficulties. We leave implementation of modular dissipation analysis in a QCA simulator for future work.

5 Conclusion

In this work, we have described a modular approach for determination of fundamental lower bounds on heat dissipation in Landauer-clocked QCA circuits. This approach can provide dramatic analytical simplification over a much more general approach to dissipation analysis introduced previously, provided that circuits are designed according to specified design rules and are "modularized" through proper application of corresponding circuit decomposition rules. We described both the general and modular approaches to dissipation analysis, illustrated their application to a QCA adder circuit designed according to a specified set of example design rules, and verified that they yield identical results. The circuit decomposition rules used in the illustrative modular analysis were presented, and their necessity and physical justification was discussed in detail. We finally argued that the modular dissipation analysis is well suited to automation, which could enable determination of fundamental lower bounds on dissipation for large and complex QCA circuits such as full processors.

Acknowledgments. This work was supported in part by NSF under Grant CCF-0916156.

References

1. Landauer, R.: Irreversibility and heat generation in the computing process. IBM J. Res. Dev. **5**, 183–191 (1961)
2. Anderson, N.G.: On the physical implementation of logical transformations: generalized L-machines. Theoret. Comput. Sci. **411**, 4179–4199 (2010)
3. Ercan, İ., Anderson, N.G.: Heat dissipation bounds for nanocomputing: theory and application to QCA. In: Proceedings of the 11th IEEE International Conference on Nanotechnology, pp. 1289–1294 (2011)
4. Hanninen, I., Takala, J.: Irreversibility induced density limits and logical reversibility in nanocircuits. In: Proceedings of the 2012 IEEE/ACM International Symposium on Nanoscale Architectures, pp. 50–54 (2012)
5. Ercan, İ., Anderson, N.G.: Heat dissipation in nanocomputing: lower bounds from physical information theory. IEEE Trans. Nanotechnol. **12**, 1047–1060 (2013)
6. Walus, K., Mazur, M., Schulhof, G., Jullien, G.A.: Simple 4-bit processor based on quantum-dot cellular automata. In: Proceedings of the 16th IEEE International Conference on Application-Specific Systems, Architectures, and Processors, pp. 288–293 (2005)
7. Anderson, N.G., Ercan, İ., Ganesh, N.: Toward nanoprocessor thermodynamics. IEEE Trans. Nanotechnol. **12**, 902–909 (2013)

8. Lent, C.S., Liu, M., Lu, Y.: Bennett clocking of quantum-dot cellular automata and the limits to binary switch scaling. Nanotechnology **17**, 4240–4251 (2006)
9. Liu, W.L., Lu, L., O'Neill, M., Swartzlander Jr., E.E.: Design rules for quantum-dot cellular automata. In: IEEE International Symposium on Circuits and Systems (ISCAS), pp. 2361–2364 (2011)
10. Niemier, M.T., Kogge, P.M.: Origins and motivations for design rules in QCA. In: Shukla, S.K., Iris Bahar, R. (eds.) Nano, Quantum and Molecular Computing, pp. 267–293. Kluwer Academic Publishers, Dordrecht (2004)
11. Stearns, K.J., Anderson, N.G.: Throughput-dissipation tradeoff in partially reversible nanocomputing: a case study. In: Proceedings of the 2013 IEEE/ACM International Symposium on Nanoscale Architectures, pp. 101–105 (2013)

The Road Ahead: Opportunities and Challenges

Opportunities, Challenges and the Road Ahead for Field-Coupled Nanocomputing: A Panel Discussion

Neal G. Anderson(✉) and İlke Ercan

Department of Electrical and Computer Engineering,
University of Massachusetts Amherst, Amherst, MA, USA
{anderson,iercan}@ecs.umass.edu

Abstract. The FCN'13 Workshop, held at the University of South Florida in February 2013, concluded with a panel discussion on opportunities and critical challenges facing research in field-coupled nanocomputing. The panelists were Craig Lent (Notre Dame), Wolfgang Porod (Notre Dame), and Robert Wolkow (Alberta). Questions were posed to the panel and to all workshop participants by moderator Neal Anderson (UMass Amherst), and by the panelists and participants themselves. What follows is an edited transcript of that discussion. All participants in this discussion had the opportunity to review the edited transcript and offer corrections and clarifications. Editor's notations are enclosed in square brackets.

Anderson (Neal): Let's start out with the most urgent and critical challenges facing the research community. What are the things we should be focused on to really enable further progress in field-coupled nanocomputing?

Porod (Wolfgang): Well, as you know, we've been at this for a while, starting out with Craig Lent in the beginning. One of the things I've learned over the years is how tough it is to try to beat CMOS. Think back to the early days of the transistor. Even though the transistor was very reliable, it still had to compete with vacuum tubes, which were at a much more primitive level than we're at today. So I think you need some reliable primitives that you can engineer, aiming at experimental realizations of this discrete coupling phenomenon. Without that, I think it's very tough. And I think that, even if you have a reliable primitive, there is still the issue of what are you going to do with it.

But overall the real issue here is that, if you are targeting computing, you're up against a very high bar. The state of the art is just extremely complicated. Some of the lower hanging fruits might be in different kinds of applications, like sensing. I'm thinking, for example, about the beautiful work that we've heard about – single atoms and the like – which might make extremely sensitive chemical sensors. As a community, I think we might keep our eyes open for those kinds of applications and not just go after beating CMOS.

Wolkow (Robert): One of the problems we have is that we're sort of leading a group of one. You want to be prominent in a field and hopefully lead in some

N.G. Anderson and S. Bhanja (Eds.): Field-Coupled Nanocomputing, LNCS 8280, pp. 379–392, 2014.
DOI: 10.1007/978-3-662-43722-3_16, © Springer-Verlag Berlin Heidelberg 2014

areas. But we do things that are quite rare, and almost no one in the world has the particular tool set that we have. Well, that's great - we're the leaders! But there's no one else.

So I think we have the problem of needing to help enable some other groups to do what we do so that hopefully they compete with us and move the field forward ... but don't beat us too badly. It's a funny problem – there are not enough people able to do what we do – but maybe there will be after some of these new results get disseminated. Maybe it will convince a few more people that this is worth trying.

To me, QCA as an ambition is a little bit like quantum computing as an ambition. We may not achieve the desired ends exactly, but I'm quite sure that the road there is going to be worthwhile. There are just so many beautiful phenomena and capabilities emerging that will be good for something – many things – maybe even including QCA and quantum computing. And I think it's a worthwhile road now because we have these multiple building blocks emerging – this fabric that can provide passive components, active components, and sensors. I didn't show it in my talk, but we can make a little strip of silicon between two contacts that can exquisitely sense arriving molecules electrically. So, yes, it is tempting to think, in this tiny little patch of silicon, we could have a chemical nose, providing some activation so we could make a decision based on that and release the drug or whatever. Some kind of ambition like that. Maybe that is the kind of niche we should be seeking.

Lent (Craig): I guess I take a pretty long view. The initial motivation for QCA was to make something as small as possible. And at the time we did start as small as possible. People had just learned to make quantum dots out of a patterned, two-dimensional electron gasses, and that seemed like the smallest thing that we could pattern and make. We did not have CMOS in mind, but we were asking: What is the limit of small?

I start with the question: Can humans pattern matter at the smallest scale possible and do useful things? It's hard to see how you could pattern matter – engineer matter to do useful things – at a scale smaller than that of atoms, molecules, or artificial molecules. You have individual molecular orbitals and you're configuring them, you're laying them out in a particular way to do useful things. I don't know that there can be anything smaller than that. Just having that thought forces you to also think that it had better be really low power, because it will otherwise melt immediately. As we heard from Ismo [Hanninen], we're on the road to there with CMOS now, so it's a lesson we've already learned practically. So: what are the small size limits structures, and, intimately related to that, what are the low power limits for doing useful things?

The idea of using a field to couple two things that have binary states, translated into the magnetic domain, is a big bonus. As we've seen several times, and Wolfgang described beautifully, these are much easier to fabricate now and they work at room temperature. An awful lot of the thinking about the architectural implications in QCA maps one to one, and everything starts with the questions of how we can represent the information and how the information here be

coupled to nearby information in a way that is minimally dissipative. The answer is: just with a field. There's kind of two fields we can use, the electric and magnetic fields. We're unlikely to do it with the weak force, or the gravitational force. That would be engineering on too small a scale, or much too large a scale.

Wolkow: You should predict it now then be famous three years from now. [Laughter.]

Lent: We've really got those two choices, and we have the luxury of being able to pursue it several different ways at different length scales. The argument is that, in the long term, if matter is going to be patterned in a manner that is electronically useful, it very likely has to be something very much like this at the single atom or the molecular scale. Its clear that we're not going to get there quickly, but there will be a lot of technologies along the way and a lot of prototypes. I hope Wolfgang's exactly right that lower hanging fruit in the shorter term will keep us on the path.

Anderson: Are there questions, comments, suggestions or other impressions about the big challenges?

Karim (Faizal): Getting research support can be very difficult for something like QCA. I'm wondering what you think are the biggest challenges or hurdles that we have to overcome to continue doing research in QCA. What makes it so hard to get funding sometimes? Is it that people are still hanging on to CMOS, and they're just not receptive to new things? Is it that there are a bunch of bad publications out there, and that people read them and think that all of this is garbage? Or are there other things? What do you think?

Lent: Well, I think there are number of things, one of which is just that it is long term research. Up until fairly recently, industry's interest was not so great because there was an enormous amount of faith that the roadmap would just continue. My sense of things is that has changed – even in the last three years – and that there is now a lot more interest because there is a lot more panic about reaching the end of the roadmap (for financial reasons if not for technical reasons with CMOS scaling limits). That has changed late in the game, but it really is changing.

It will also be helpful to have more proof of concept deliverables, things that you can see working. Magnetic devices working. I frankly don't know why the metal dot devices do not have more credibility; fairly sophisticated things have been made that way. And the results Bob showed with these atomic scaled devices. Seeing those things working, I think, will help a lot in terms of sustaining the interest in the field.

Finally, it has to be said, there's kind of a human phenomenon. There was a time when all of this was brand new. No one had ever heard of it. That, you know, not the case anymore. The field is in it's awkward adolescence. [Laughter]

Wolkow: I would echo the point about the human phenomenon. Sometimes society goes down blind alleys, and everyone follows. You can see that through

trends in architecture or trends in science or anything. There are so many examples of where society goes in a direction that only very slowly is shown not to be worth following. As some areas that have enjoyed much support are shown to not pan out we will see redirection of funds to paths that lately haven't been mainstream. Greg Snider's work connecting CMOS to SETs and to field-controlled elements will gain new prominence and support, I expect.

Porod: I have a slightly different view. QCA has been around for quite some time; it's getting close to twenty years now. It's very hard to find examples of anything that has had sustained funding for twenty years – it's just a very long time. In terms of funding, I don't want to say there's been a tremendous amount of funding, but there certainly *has* been funding for this research. You always want to see more of it, obviously. And I think what Craig was saying is correct: The less you know about something, the easier it is to get funding because you live on promise. Then you start doing experiments and you realize how hard it is. Knowing a lot works against you. So I think we're seeing a little bit of that.

In parallel to knowing more, I think it would have helped to really come up with something that works reliably. We're still struggling with this.

Anderson: Any other comments along the lines of critical challenges?

Wang (Peng): Is QCA the only technique that could take over the common CMOS technology. How does the future of QCA look in comparison to other research, like research related to graphene and carbon nanotubes?

Wolkow: Looking at the prospects for those other approaches, I can't convince myself that they are more worth pursuing than what I'm pursuing. I'm doing what I'm doing because I think it has the best prospects. With nanotube transistors, for example, they really need to create tunnel transistors where you go from one discrete level to the other ultimately to try and beat the subthreshold slope problem. With years of effort in that area, there's really no significant progress. And there's the terrible fabrication problem, and not only of the nanotubes themselves and separation of the nanotubes, but also the placement of them. So I see (A) that they haven't shown great prospects, and (B) that even if they did, they wouldn't know how to make them. So that, for example, is one area that I wouldn't prefer to pursue. IBM has, to my knowledge, essentially dropped graphene as an electronic material, although they're still working on it for optical properties that they're more interested in. So I really think that this is the thing to pursue. Of course, I'm terribly biased.

Anderson: Not the gravitational QCA idea? [Laughter]

Walus: Planetary QCA!

Anderson: Planetary QCA - you heard it here first. So along these same lines, but maybe a little more focused, what's the one problem – for nanomagnet logic, for charge QCA (molecular or atomic) – what's the one problem that, if you could solve tomorrow, would really help you move down the road?

Lent: I think if either we had Bob's chain switching, or molecules switching each other, that would be a big step.

Another big step, next to that, is integrated detection. It's really very fortunate that there are scanning probes, or it would be hard to see how anything would be possible at this stage. Scanning probes are a big help; they are a sort of scannable electrometer. That's not exactly right, but they let one see very small structures and see changes in electronics. Chemists learn an empirical base of knowledge that allows them to synthesize structures, and they have a battery of tools to try to figure out what it is they made, spectroscopically, in solution. And the gold standard, without which we would have made hardly any progress, is X-ray crystallography, which allows you can see something of what you made in a crystalline structure. But it's much more difficult to tell the *charge state* of what you've made, or to detect something within the molecule that has changed. X-ray crystallography relies upon really indirect kinds of arguments, so scanning probes have been enormously helpful as a way to see. Without that, it would be very hard. We'd like that integratable, and we'd like to be able to see more.

But the short answer is probably device-device interactions.

Wolkow: So we have a whole list of things. One of the challenges – but advantages – of trying to start a company is that you must be organized and you must have a list of marching orders and things that you would deliver against money received. We're looking at the I/O problem. The "I" and the "O" are quite separate, the atom fabrication has to be better and better. We also have to address really dull-sounding things like laser cleaning processes to make more perfect surfaces because of defects, which we have for two reasons: (1) because we make things incorrectly, and (2) because there was a preexisting defect that prevents us from ever making things correctly. We need new algorithms. If, for example, this room was my substrate that I was going to build on, where the table tops are ideal regions and the regions in between were not ideal, I would have an algorithm that preselects the ideal regions and then knits those together. We are working on all of these things in parallel. So I wouldn't pick one technology, I would say there's a whole bunch.

But, echoing the earlier conversation, we have to be, hopefully, clever enough to deliver something impressive sooner than later. We have to stop the tendency to be researchers and wild explorers only, and really constrain ourselves to solve some of these kinds of dull problems so that we make tangible advances soon.

Anderson: Wolfgang, do you have a favorite problem that you'd love to solve tomorrow?

Porod: There are two sets of issues, in different directions. One issue is exactly along the lines that Bob mentioned: at some point, you have to demonstrate something. So I think it would be great if we were able to demonstrate an integrated system with electronic input and output and magnetic "something" inside. As far as I can see, there are no scientific barriers to doing that. It's just a hard thing to do. Industry creates large teams to work on things like that, and they make it happen. Like MRAM, for example. Something like that is very hard

to do in a university environment. So there are these kinds of practical issues. But if you can demonstrate things, then I think you can draw larger support so it would be very important to do.

On the other hand, there also are scientific issues related to these magnets, to making these magnets switch with less effort. The nice thing about magnets is they're very stable, but it's just hard to switch them. The two things obviously go hand in hand. We heard about some work at this workshop about different ways of switching magnets. Generating magnetic fields is costly, so having voltage control over magnetism would be a very important breakthrough. We heard some work from the Torino group on using strain to accomplish that, and I think it's very important work. We heard some work from Berkeley about using the spin Hall effect to do that, and I think that this is important work as well.

Getting somewhere has payoffs along the way that Bob mentioned before. If we, as a community here, were to come up with radical and different ways of voltage control, the first applications would not be in magnetic QCA. The first applications would be in MRAM, there's no question about it.

Walus (Konrad): I think in this community, there are two groups: people that really care about making individual devices and people who are thinking about circuits. There are issues that are being raised at the circuit level and issues that are being raised at the device level, and the circuit guys say, "Yeah, yeah, yeah, you guys will solve it" and the device guys will say, "Yeah, yeah, yeah, you guys will sort that out," right?

I think that one thing we really need to do is to find the killer app for QCA. We're talking about CPUs, and we're talking about adders, and the writing's kind of on the wall about deep pipelining, and we're talking about sensors. But we need to really define the killer app for this technology. Is it image processing, or something else? Then I think the sort of path will be more clear. Right now, we're sort of exploring everything and nothing is really definite.

Wolkow: Well, I won't go into great detail, but one of the lucky things that's happened to us is that we've managed to get people at TI and Lockheed to engage in a conversation with us. I'm impressed by what they've said so far. They've been able to think of things that would meet their needs, things I never would have thought of, some of which relate to things like A/D conversion. These would be things that I think we could make. With the capability to stamp out thousands of cells, which, given our current capabilities, would only take minutes, we might be able to make some circuits that would be useful in very rarified situations like satellite applications. Places where it wouldn't matter if the chip cost ten thousand dollars because that cost is insignificant compared to the cost of just getting the thing up there and making sure it works.

I always give this example of the CCD. When people were trying to perfect the charge coupled device – an optical detector – they were lucky to have those few customers who were willing to pay tens of thousands of dollars for one detector, and that allowed them to make a little money (not really profit, but at least to keep going). That led to reductions of cost and eventually widespread deployment. So, I'm trying to find those *first* customers, and I think maybe with

TI and Lockheed we'll find that. The other thing I'm doing is reading the history of how television developed, and what a small team did to make the most crucial developments. And I've read about how radar was developed and how the first computers were developed. I'm trying to take historical lessons, and see where a small team could make the critical steps. That's my approach. I'm trying to look for those little openings and I'm trying to learn from history a little bit.

Walus: So what have you learned? [Laughter.]

Wolkow: Well, I have learned that it *is* possible for a small team to take the most critical first steps. I'm motivated by that. I think it's still possible. You don't have to be a thousand-man team to get going.

Back to identifying the killer app: I'm hoping to make some ultra-low power circuitry that, no matter how expensive it is to make, will be just a little more efficient so it won't cause the clock in a satellite to be just a little bit inaccurate (which makes GPS positioning a little less accurate than it could otherwise be). That is one really narrow thing that could be a very good first objective. And everything just seems right, because when you talk about planning a new satellite, you're actually talking ten years in the future because everything takes a long time. So it doesn't matter how much it costs, we have a lot of time, they only need a few, and I've got a lot of people who can advise me. It seems like one example of just the right kind of thing for us.

So it's not the *killer* app, it's the first app.

Porod: I think you're absolutely right. And, as I mentioned at the beginning, you need something reliable, like a building block, that you can engineer with and that works. And even if you have that, you still need to know what you're going to build with it, right? This is the application you're asking for.

QCA is so different than CMOS. I think the main difference comes in the communication between the devices. In CMOS, you can have wires and can communicate over long distances at very high speeds. In QCA you pay a very heavy price for communication that carries a heavy penalty. So I think that the challenge to computer architects and for people who think about these applications is this: What kind of applications rely heavily on local computation, so you don't have to communicate very much but you compute at every step. This is one of the reasons why we're looking into systolic architectures, where you push data through and at each step you do some computation. There are different ways of thinking about applications, but I think the key is minimizing communication and maximizing computation. I think that is what QCA needs. And I think you're absolutely right, just taking the layout of a microprocessor and then just redoing everything the QCA way has very little chance in computing.

Vemuru (Srinivasa): Are systolic arrays the best solution?

Porod: I don't know, I don't have the answer. But I think that is a critical problem: What applications are there, what opportunities are there, what killer apps are there where you do heavy computation at every step, where you do mostly computation and very little communication. It seems to me that is key.

Wolkow: It's not a new idea, but if you could do analog computation and then judicious discretization of the data in a modular way, so that you never accumulate errors but get the extraordinary efficiency of analog computing, I still think that's the way to get extraordinarily improvements in computing power.

Anderson: Regarding killer apps, CMOS wouldn't be where it is without general purpose computing. There must be a hundred computers in this room, and that is because of general purpose computing. Taking account of considerations like those that Wolfgang just pointed out – the difficulty of communication in QCA and the possible need to shift architectures toward more local computation and less long-distance communication – should we be thinking of general purpose processors as a killer app in QCA? I am just curious to know what people's opinions are on this, because, unlike some other novel technologies, QCA has the capacity for general purpose computing. General-purpose computing may not be efficient in QCA *if* it's done the way it's done in CMOS, but should we still be thinking along these lines, or should we just be looking for special purpose niche applications? Or a mix of both?

Graziano (Mariagrazia): Maybe in the future we could think about general purpose computing. My view is that the real thing that might make QCA a winner is power consumption. If we look at the history of computers and of devices, the great steps from one technology to the next occurred when there was a big problem in terms of power. After the first computers, we moved to BJT logic and then to MOS. With respect to the other technologies emerging beyond CMOS technology, QCA – all forms of QCA – has the advantage of avoiding transport, which is a real problem. So, if this can be demonstrated, then other problems can be solved, so in the future one could think also about general purpose computation, maybe with thinking of the internal organization. Maybe in the future, our children – or the children of our children – could be thinking about general purpose computing.

Bhanja (Sanjukta): I think that we should reflect a lot. We have a tendency to map everything into a Boolean framework. I agree with Bob that we may have to think about something different, looking at the strength of our technology as opposed to mapping everything into things like full adders and half adders. A lot of people are thinking memristor-based logic and its potential major types of applications. There is a lot of work going on in RAM-based logic, resistance-based logic. I feel that there are opportunities there to explore different computing paradigms.

Wolkow: One thing that Greg (Snider) said to me just yesterday, and often points out, is that QCA isn't attractive only because of its inherent low power tendency. It has options to be used reversibly, so it could even escape the most fundamental energy costs. That's what Greg would say. Did I say that at all correctly, or do you want to say it better?

Snider (Greg): Well, no matter what you're doing, no matter what your device, no matter what your material, eventually it's going come down to the energy

required to do the computation. We can get into the discussion of reversibility, and what are the options there. But no matter what your material, if you're computing a bit of information, even if it's an analog computation, you have to put in energy to encode your information. And if you're doing a computation, even an analogue computation that has a certain amount of information in and a certain amount of information out, you're still faced with fundamental limits. So how you encode the information is less important than how you process the information (whether it's analog or digital). When you're doing computation, you have to worry about how you're doing the computation and how you're using the energy to do that computation.

Lent: I don't understand much about computer architecture, but it does seem to me there's a kind of rough, blurry convergence there. The most impressive QCA architectures that I have seen simulated [were by] Sarah Frost-Murphy, that had these little regions that stored some memory and did some computation. There were a lot of them. And she had this thing called bouncing threads, so threads of information move from one region to another, bouncing around, and they would do a computation which would alter the thread, and that would go elsewhere to an available little computational cluster. They were just integer operations.

Think about QCA as processing in memory, holding lots of state and processing while it's holding the state so it's breaking the von Neumann separation. That seems to me to be a promising way to think about it. It's not so different, except in scale, from multicore, so the problem you're running into if you have thousands of cores is a similar kind of problem: that there's a big hit for communication. (The scale is just bigger for what counts as local and what counts as remote.) But you can do lots of stuff locally. So how do you map important problems onto that architecture?

You have a lot of finite state machines and they're connected by more costly communication. But it's costly only in the sense of latency, so once things get there – I know that this doesn't sell well with lots of computer architects, the idea that you have to walk over there – but once there, you get a different bit every cycle. It does seem to me that if you think about that from the processing-in-memory point of view, then that's kind of promising.

Bhanja: There is a lot of work in that direction, with magnetic memories, a lot of work in non-volatile computing.

Wolkow: Another thing that comes to mind is that there's a lot of new effort in what you might call probabilistic computing, non-fully deterministic computing where people want to do, say, pattern recognition or image analysis but not seeking some absolutely precise numerical representation. They just want to sort of know "did that thing move?" in the simplest and just sufficient way and no more.

No matter how good a normal von Neumann binary computer you ever make, it's never going to be enough to handle all the data people want to handle. And so there need to be totally different approaches. So maybe we could, as a community, look at that more and think of radically different things we should

try to do. I was wondering if I could make a really useful and attractive random number generator that could seed other things.

Porod: I have a slightly different take perhaps. Maybe taking inspiration from biology might be a good thing. The secret behind the CPU is that it is so general; you just know how to crunch bits and you have the software to tell what this bits mean. You have hardware that can solve a large class of problems. So I think the issue is commercial; designing a chip is very hard, designing hardware is hard and expensive, and it has to justify your investment.

So I can imagine special purpose hardware for sufficiently large classes of problems that justifies this investment. Everything does not necessarily have to be general purpose. A class of applications that I think would map onto QCA is vision applications. If you think about how visual processing is done – Sanjukta [Bhanja] mentioned it this morning – it essentially involves near neighbor interactions or near neighbor computations. You extract features from that with very little movement of data and with very little communication overhead. I am familiar with some of the work that is going on in Berkeley at the Vision Research Lab suggesting that the early visual processing in the retina takes place in about then different layers in the retina. Each layer represents essentially a different connectivity between processing elements you can think about it as neurons. Each of these connections essentially extracts different kinds of features. So this maps onto cellular neural networks – or CNN – and some people are working on that. I could see that if one could come up with a QCA-like implementation, where some of these connections that you want for image processing were realized by physical interactions – improving power and speed and all that – that there might be a sufficiently large market. That might be a killer app. So I am keeping my eyes open for vision types of applications.

Bhanja: Professor Porod, do you have CNN implementations of processing low-level features for image processing?

Porod: Well, to some extent. We had a MURI Project associated with the vision research I mentioned at Berkeley. Part of this project involved a group in Budapest that is working on cellular nonlinear networks, trying to implement the kinds of templates they extract from the retina into hardware. We were a part of that project, looking to see whether or not this hardware could be naturally implemented in a QCA-like fashion. That was a few years back, and we were not far enough along with the magnetic implementations to have something reliable enough that you could actually do some engineering with. But if one had reliable building blocks, then [one could implement] something based entirely on near-neighbor computations where you do not have to think about threads, about moving data from memory into some processing element so that the processing element can crunch on it a bit. So I think vision applications might be a good target. But, again, I think we need some reliable building blocks first. We need some bricks to start constructing buildings.

Lent: Were they implemented in analog?

Porod: They were implemented in CMOS chips, but there were issues competing with general purpose processors. Some of these chips are actually fairly sophisticated, but they were made by university groups. Computers get faster and faster and faster, so even though they had parallel hardware, it was hard to compete with the new computer you buy at Christmas! Another issue was that they were working with a general purpose CNN processor, where they had to load in different templates encoding how each pixel looks to its nearest neighbors. There was a significant overhead in this template matching operation. But I think if you could identify – just like in the retina – say ten different connections, and you implement those in hardware in ten chips, the processing could be very fast and you just mix and match the information from these ten chips. Even though it would be special purpose in some sense, something like this would be relevant for a very large class of applications and, I think, worth some investment.

Wolkow: You know, another related thing, the European Union just announced a funding on a billion-Euro brain simulation project. They are trying to explore what kind of models begin to represent a brain. Maybe we should watch for what emerges there and see if we could offer, like you said, localized, special functions that might emerge as attractive and useful. Maybe we could deliver something that's just what they need.

Porod: So, just this one comment, since you mentioned the brain. I would like to make a distinction between the retina and the brain. The retina actually does a lot of computation. I do not think that this is generally appreciated. So the world around us presents this huge flux of information that we need to process. It's not processed in the brain. This huge flux of information is broken down into relevant pieces of information – which are the features – which are then handed on as a trickle of information back into the brain where some of the higher level cognitive functions occur. It seems to me that the higher cognitive parts could be done by a CPU-based computer, but that there are opportunities for specialized hardware that do what we do: We do the same thing over and over and over again, just breaking down this flow of information into lines or edges or whatever.

Wolkow: You know, some of the most advanced cochlear implants use that same knowledge. They pre-filter and they process information in almost mechanical ways, with waveguides and so on, so that minimal computing is needed afterwards. So you do not use more bits then you need and you do not general-purpose process anything. You try and cleverly handle information by pre-processing it.

Graziano: I have a comment here. Look at the work going on around the concept of dark silicon; that seems to be the direction. It is impossible to have bigger and bigger and bigger CPUs even with parallel organization. The trend seems to be that you have a big circuit with a lot of blocks then you switch them on only when you need them, and then you specialize some of them. So you have some general purpose part where some part of the computation is done, and, depending on the application, you have some blocks that are specialized. You only switch them on when you need them. This is clearly the case in CMOS.

So this trend could be also something to be followed for specific applications, where you have, say, something that needs very low power in a repetitive computation.

Anderson: So maybe it is natural to do more things outside of the processor.

Bhanja: Like co-processors and GPUs.

Wolkow: That reminds me of a talk at a recent Gordon Conference by Paul Solomon of IBM. Surprisingly, he mentioned that kind of future vision for continuing to make ever more powerful computers. It seems that people are recognizing that you are going to need these special functions, and even little chiplets that will be fabricated in different kinds of technology – seemingly very expensive and unattractive special assemblies to perform special functions. Maybe QCA can find special roles.

Wang: Is there a existing working full adder in QCA?

Lent: I don't think so.

Porod: It depends on what you mean by working. [Laughter.]

Wolkow: The magnetic one can show the essential features, but it is not very dynamic I guess.

Porod: With an external magnetic field, and looking at this assembly of magnets with MFM, we can show that an adder works. But it is not a prototype. This goes back to the point that I was trying to make before. If all of that can be integrated into a working prototype, that would be a major step. There is no science obstacle that I can see to making it happen; all of the technologies are there. From a scientific standpoint, it seems like a trivial problem. But it is still a tremendous problem to make it happen. So, yes in principle, but no in practice.

Wang: So is that too early to talk about the apps since we do not even have a working full adder.

Porod: Well, this might be a little bit too strong, but we do not really have a working anything. I think a fully integrated switch would go a long way.

Lent: The reason to think about the architecture is to know if you have a reason to put in the effort to get all of the things to work. Is there an architectural show stopper waiting for you once you get all of the devices working? You'd like to explore that well in advance. Is there an appropriate architecture, are there show stoppers, are the things that push back – like Konrad [Walus] was saying – on what you absolutely require. If you are just building half a dozen devices and tying them together, you will not need to worry about power. It is only because you have a picture of where this is going that you would say "Wow, this thing is dissipating watts per device! It is not worth going on."

Wang: What is the first priority? Is it more important to have a full adder or is it more important or to find a type of architecture for QCA?

Bhanja: Most groups don't have the infrastructure to monolithically grow something like an MRAM device where you start with CMOS, then you add metal 1, metal 2, metal 3, and then you anneal it, and then you put the magnetic logic down, and then you do metal 4. We don't have that. It is scientifically possible, but it is not economically possible. You also need the team.

Wang: So actually there is no technical problem to manufacture.

Bhanja: No, MRAMs are in market now. That's why as researchers we are focusing on different things.

Anderson: So we have only a few minutes left. Any closing comments or remarks?

Snider: Just a question. One of the things that made electronics so attractive for the last forty five years is Moore's law, which is this nice exponential thing. And an exponential function like that, as Gordon Moore realized, is a license to print money. So that has driven the industry for forty years. Is there any sort of exponential curve that you guys can see anywhere that would be an equivalent sort of thing, that would be an exponential increase in productivity of something even over a short term. (I don't think anybody is going to say forty years.)

Wolkow: It's such a funny question. It's essentially saying that even if you can make this fantastic new technology, it might not be that attractive from a business point of view. It might be delivered and mature, but if it did not have more and more stages to come it might not be that attractive. It is a funny question, but I guess a very practical one. I do not know the answer. I think if we made the first thing just work, and it was worth something to someone, then they would buy it, and that would be the first generation. And then there would be so many improvements, not in the way we follow Moore's law by making more or less the same thing better, but there would be more refinements. So I think it could display that trend for a good long while.

Lent: It seems to me that if you start at the limits of small, so we are talking about molecular or atomic size, then what you scale first is the number of devices. So you will be at the order of 10^{24} in terms of device density. Seems to me that is a reasonable scaling if you are trying to get more perfection, more control. At those kinds of densities you can go quite a distance by just having more control and perfection.

If the power dissipation is low enough, you could think about building them up. I have always had a hard time seeing [management of] the power dissipation in 3-D, but the answer is essentially the same answer in the orthogonal direction. If the units are small, then all you can really do is to add more of them and do it more perfectly and more reliably.

Wolkow: There are twenty seven atoms smaller than silicon so we can go through the periodic table. [Laughter.]

Snider: That goes to my followup question. Do we ask Bob [Wolkow] to write the paper "We are out of room at the bottom"?

Vemuru: I have a comment on the Moore's Law. Moore's Law is great not because you could integrate things, but price goes down exponentially. It is a good economic model for things to grow. If we can deliver QCA – say, processes to make a few cells – down the line if you could make a bigger area with the same cost then we will have Moore's Law in reverse order. I think that's the path we should choose.

Porod: I would like to echo the second part. There are obviously different versions of Moore's Law. Because we like technology we think about sort of the ones we relate to the technology feature size and number of transistors and so on. But what really drives things is economics. People don't spend money on a new computer because they are so proud of the technology that goes into it – because they have 22nm feature size transistors in there – no one cares about that. People care about performance. So as long as you can deliver performance, the way you do it does not matter and Moore's Law continues.

In recent years, as you know, the industry has actually backed off of raw transistor performance in favor of parallelism, and that has kept performance improving. So 2 cores, 4 cores, 8 cores and so on. In principle, you could envision going to million cores if you can place a billion transistors on a chip. One way of arranging a billion transistors is in an array of a thousand by thousand processing elements – an array of a million processing elements – each with thousand transistors (which is the order of complexity of one of the early microprocessors). So I think there is a long way to go for increased performance, if we just exploit the parallelism. I think this is a continuation of Moore's Law, if you will. It has moved from raw technology to architecture, but consumers do not care.

Anderson: Moore's Law by other means. Thank you all, and thanks to the panel.

Acknowledgments. We thank all of the FCN panelists and participants for their contributions, USF graduate students Srinath Rajaram, Ravi Panchumarthy and Dinuka Karunaratne for their work recording the panel discussion, and Katherine Anderson for transcription assistance.

Author Index