As we anticipated in our previous issue, the 2015 ITRS roadmap has been published. It makes an interesting reading for all those interested in understanding what is the point of view of the semiconductor industry. An industry that has seen its role progressively decrease by the reduction of the number of its members (from 19 in the 90’s to merely 4 today: Intel, TSMC, Samsung and Global Foundries) and, most importantly by the impact of novel devices and services brought in by the Internet. There are a number of interesting messages contained in this document but one seems to have got the attention of the media: “By 2040, computers will need more electricity than the world can generate”, as announces the title of the article by the science journalist Richard Chirgwin on The Register on July 25 2016. So the topic of “how much energy” is required to operate ICT devices and how to decrease it, is still the focus of many discussions. This is also the topic of the E.C. funded initiative ICT-Energy that recently published a new version of the Research Agenda (http://www.ict-energy.eu/sites/ict-energy.eu/files/ICTEnergySRA5.3.pdf). The picture in this page is from this document and presents a bird-eye view of the status of energy consumption in binary switches, made with different technologies.

On Aug. 16-19, in Aalborg, during the international conference “Energy efficiency and sustainability in ICT” two of the most revolutionary solutions in this field (TOLOP and LANDAUER in the figure) will be presented by the people who designed them. In this special number of ICT-Energy Letters, we publish the proceedings of this conference. (L.G.)
Energy efficiency and sustainability in ICT
Aalborg, 16-19 Aug, 2016

Conference proceedings

- **Fundamental Energy Dissipation Limits in Logic Circuits**, İlke Ercan, Pag. 3
- **Fundamental thermal machines for the thermodynamics of computation**, D. Chiuchiù, G. Gubbiotti, Pag. 5
- **Zero-Power computing**, M. López-Suárez, I. Neri and L. Gammanito, Pag. 9
- **Leveraging Non-Volatile Technology to Design Processor: Use Case of TLB CAM Cell Based on Nano-Electro-Mechanical Switch and CMOS**, A. Seyedi, V. Karakostas, S. Cosemans, A. Cristal, M. Nemirovsky, O. Unsal, Pag. 11
- **Underwater piezoelectric energy harvester**, M. Mattarelli, F. Orfei and F. Cottone, Pag. 13
- **Comparing CMOS and NEMS quasi-adiabatic logic**, Samer Houri, Pag. 17
- **xEnergy-Optimal Control of Multi-Core Applications**, W. Ahmad, M. Stoelinga, J. van de Pol, Pag. 19
- **Impact of Delay Propagation on NTV PCMOS Design**, L. Oudshoorn, G.A. Gillani, A.B.J. Kokkeler, Pag. 21
- **An Ultra-Low Power NVM-Based Multi-Core Architecture for Embedded Bio-Signal Processing**, R. Braojos, D. Atienza, Pag. 23
- **Kinetic energy harvesting at microscale: current progress and perspectives**, D. Galayko and E. Blokhina, Pag. 25
- **Energy harvesting based on Micro System technologies**, Cristina Rusu, Pag. 28
Abstract—In this study, we present FUELCOST methodology that allows us to obtain fundamental efficiency limits of complex computing structures. Heat dissipated by logically irreversible operations has a critical place amongst challenges imposed on the realization of nanocomputing paradigms. FUELCOST provides a common ground to study energetic cost of logic- and physical-state transformations in circuits, which allows us to analyze the correlation between dissipative cost of these operations and identify their contribution to unavoidable cost of information processing. Here, we provide an overview of the essential features of our approach and discuss applications to CMOS and post-CMOS paradigms to provide perspective on fundamental efficiency limitations of changing trends in computation. The dynamic and static logic circuits we have studied emphasize significance of circuit operation as well as physical structure on energy limits.

Keywords—FUELCOST, Fundamental bounds, Heat dissipation, Static Circuits, CMOS two-bit sort.

I. INTRODUCTION

FUNDAMENTAL Efficiency Limits for Complex Computing Structures (FUELCOST) is a methodology developed to obtain fundamental physical description of dynamics of information as it is processed by computing machines and calculate fundamental energy dissipation bounds based on this description for a given technology base [1]. These bounds provide insight into fundamental performance projections of post-CMOS paradigms and challenges ahead in emerging electronic technologies [2]. There are numerous sources of inefficiencies in computing strategies that interfere with improved performance of conventional circuits and realization of novel technology proposals. The practical sources of inefficiencies are inherently unpredictable and depend upon future technologies, however, fundamental limits are predictable as they are based on the generic physical requirements for implementation of a computational strategy. Therefore, calculating the fundamental lower bounds of a given technology base allows us to obtain best case scenarios of performance projections based on the underlying computing strategy by taking into account inefficiencies that are imposed by the nature of computing paradigms.

The foundations of FUELCOST methodology was laid out in our earlier work, where we obtained fundamental lower bounds on energy dissipation in circuit level for non-transistor-based [3][4] and transistor-based nanoelectronic technology proposals [4], as well as at the processor level [5] and in finite state machines [6]. Our analyses have revealed how the fundamental lower bounds on energy dissipation depend on the physical structure of the circuit, its control (e.g. clocking) scheme and the irreversibility of the logic operations performed during computation. Earlier applications of our methodology focus on dynamically clocked nanocomputing paradigms, where computation involves one-to-one mapping of logic- and corresponding physical-state-transformations in a circuit. In this work, we obtain fundamental lower bound on energy dissipation in a static CMOS two-bit sort circuit to (a) illustrate our methodology’s application to a conventional circuit technology, and (b) calculate fundamental energy dissipation bound in a system where the logic-state-transformation corresponds to a number of physical device state transformations. This provides a basis for comparison of the fundamental dissipation analyses in emerging and conventional electronic technologies. In addition, applying our methodology to a static circuit allows us to identify additional sources of dissipation in the fundamental lower bound resulting from correlation between logic- and physical-state-transformations.

II. THEORETICAL FRAMEWORK

FUELCOST is an evolving methodology for systematic determination of fundamental energy efficiency limits of computation in complex computing structures [1]. Its foundations are rooted in physical information theory, and it is tailored according to a given computing architecture, circuit design and control scheme. It depends explicitly on input statistics and computational task. The bounds obtained are truly fundamental as they are based solely on the fundamental physical description of information processing and independent of device model assumptions and parameterization, in which fundamental sources of inefficiency inherent in the computational strategy employed by the circuit are isolated; the approach allows us to calculate quantitative efficiency limits without explicit device models.

The methodology comprises of abstraction and analysis. First, we begin by constructing a functional abstraction of information flow in the computing structure, which captures the essential features of the circuit structure and control. Then we obtain limiting efficiencies via physical information theoretic analysis based on this abstraction. The efficiency analysis can be performed at the circuit level, architecture level or instruction level (for programmable processors). Here, we discuss an application in the circuit level. Abstraction of the computing structure is performed in two parts as physical and process abstractions. The physical abstraction of a computing structure is performed by constructing a globally closed and isolated universe that enables us to assume unitarily evolution via Schrödinger’s equation. The physical abstraction of a transistor-based circuit, treated as a composite quantum system, is shown in Fig. 1. The computational domain consisting of the information processing artifact, i.e. the circuit, C, as well as the computationally supporting subsystems such as the source, S, and drain, D. The input referent, R, is a system holding the physical instantiation of the initial input data. The environmental domain consists of a heat bath, B, in direct thermal contact with the computing artifact and nominally at temperature T, and an environment, B̄, including reservoirs that “rethermalize” B. During computation subsystems are drawn away from equilibrium but rethermalized as a part of
the process abstraction. These abstractions allow us to capture the paradigmatic operations that include all the essential functional features of underlying computational strategy. In order to perform a physical-information-theoretic analysis of the global system evolution, we construct a space-time decomposition of computation based on the abstractions. The details of the analysis step is discussed in relation to the illustrative example below.

III. APPLICATION OF THE METHODOLOGY

In transistor-based applications of FUELCOST, identifying the particle supply cost associated with processing of an input has proven to be crucial to accurately determine the lower bound on energy dissipated in computation. Our bounds are radically different than that of Landauer’s as our methodology is rooted in quantum dynamics, and thermodynamic cost of electron flow from S to D as a result of information processing logically follow from the formalism. We also define information with respect to a referent system, which allows us to identify the irreversible information loss with loss of correlation between the computing system and the referent holding memory of the initial input.

Here, we consider a combinational two-bit sort circuit implemented in CMOS proposed by Gershenfeld [7], as shown in Fig. 2. The design and operation of the circuit is different from earlier transistor-based implementations we studied as it is asynchronous—the computation of an input is composed of a single step rather than multiple computational steps operated by the clocking scheme—and each new input is loaded in the circuit without a prior reset—that there is no intermediate erase between computational cycles. The erasure of a given input is done via overwriting of the next input, which means that the energy associated with erasure of a given input is dissipated over the operation of two inputs. The lower bound on the energy dissipated into the environment during computation of an input is

\[
\Delta E_n \geq k N q V_{DD} + k T \ln \left( 2 \sum_p \rho_p S \left( \sum_i \rho_i^p \otimes \rho_i^p \right) - \sum_p \rho_p S \left( \rho_p^p \otimes \rho_p^p \right) \right).
\]

Here, k is Boltzmann constant, T is temperature, f represents a fraction of energy invested in SD. The second term represents a difference between average von Neumann entropies obtained after S and D are recharged and rethermalized following computation. Irreversible loss of information due to overwriting leads to this term as the cost of computing nth input depends on (n-1)th input. This term represents energetic cost of changing subsystem contribution as a result of partial erasure. The numerical value this bound can be calculated according to state transformations system undergoes during processing of a given input. The number, N, of electric charges owing in and out of the circuit depends on both nth and (n-1)th input. Four possible combinations of nth input followed by the four possible combination of (n-1)th input gives us sixteen scenarios, i.e. for equiprobable inputs, the above bound can be represented numerically based on values averaged over sixteen possible scenarios of charge transport and state transformations.

It is important to underline that this bound is independent of assumptions regarding material, device, or circuit dimensions or parameters. Analyzing CMOS circuits from the fundamental energy requirement point of view stands as mere academic effort due to the practical limitations imposed on the technology. The manufacturing techniques proposed for post-CMOS paradigms allow aggressive scaling that can mean higher performance, density, and power efficiency that can go far beyond the performance of CMOS technology.

Fig. 2. Logic (left) and physical circuit (right) diagram of CMOS implementation of Gershenfeld’s combinational two-bit sort circuit [7].

IV. CONCLUSIONS

FUELCOST can be applied to a wide range technologies operated under different control schemes. Dynamic circuits are usually faster than static counterparts, require less surface area, but are more difficult to design. However, from an information-theoretic perspective, this difficulty in design in dynamic logic circuits has proven to provide ease in analyzing limits of energetic cost of computation. Our studies have shown that sequential logic circuits contain certain inefficiencies from fundamental energy dissipation point of view that cannot be “engineered away” with improving technology. Our studies have also shown that with proper generalization of the theoretical framework, FUELCOST can be applied to unconventional emerging computing paradigms, such probabilistic computing and Brownian computers. Our research points out that this approach can also allow us to perform multi-level analysis of memory-logic-communication cost tradeoffs. The analyses performed with FUELCOST provide valuable insights into the long-term efficiency scaling capacities of emerging paradigms and can be used as “litmus tests” for checking consistency of performance projections in emerging technologies.

ACKNOWLEDGMENTS
The author would like to thank Neal G. Anderson for insightful comments on the manuscript. A part of this work has been completed in the IHTN at TU Darmstadt; the author gratefully acknowledges generous resources provided by the institute. Support for this research is provided by TÜBİTAK under 2219 Postdoctoral Research grant (No: 1059B191401520).

REFERENCES

Fundamental thermal machines for the thermodynamics of computation

D. Chiuchiù1, G. Gubbiotti2
1Dipartimento di Fisica e Geologia, NiPS Lab, Università degli Studi di Perugia,
2Dipartimento di Matematica e Fisica and Sezione INFN Roma Tre, Università degli Studi di Roma Tre

Abstract—Computing devices are thermal machines. This perspective is unusual, but the operations required to perform a computation are the same of thermal machine. Curiously enough, a strict thermodynamic description of computing devices is rare in the literature. One of the main reason is that such devices change their state once each nanosecond and, although two-hundred years old, thermodynamics is not yet capable of giving reliable analytic prescription of heat exchanges for finite time protocols, even for much simpler thermal devices. For example, there is no analytic formula describing the heat exchanged with the reservoir by a perfect gas compressed by a piston in a finite time. Here we address this latter system and develop analytic formula for the expected heat with the multiple scale expansion methods and regular perturbation techniques. Our aim is to introduce methodologies that could lead to a purely thermodynamic description of nowadays and future computing systems.

I. INTRODUCTION

We consider the simplest thermodynamic machine: a perfect gas enclosed by a cylindrical canister with a movable piston and in contact with a heat reservoir (Fig.1). The simplest way to describe this system is through a small set of macroscopic variables: the piston position \(x\), the gas internal temperature \(T\), the reservoir temperature \(T_b\) and the external force \(F\) applied on the piston. Among those variables, \(F\) and \(T_b\) can be changed according to some external time dependent protocol while \(x\) and \(T\) evolve as a consequence. Cerino et al. proposed in [1] a satisfactory set of equations to describe \(x\) and \(T\) dynamics. Without going too much into details, such macroscopic equations are obtained by averaging microscopic properties with the aid of heavy assumptions. The first one is that the piston and each gas particle undergoes elastic collisions. The second assumption is that the velocity of a gas particle is randomly changed according to the Maxwell-Boltzmann distribution of the reservoir when reservoir-gas particle collisions occurs [2]. The third assumption is that the gas distribution is always Maxwellian although gas-reservoir and gas-piston collisions change the gas temperature \(T\) over time, ruling out shock-waves propagation in the gas. These macroscopic equations can be linearized and cast in the dimensionless form [3]

\[
\begin{align*}
\ddot{x} + \frac{e^2}{T_b} \left( x - \frac{T_b}{F} \right) + 2 \frac{eF}{\sqrt{\pi} T_b} \dot{x} + \frac{1}{\sqrt{\pi} T_b} \left( T - T_b \right) &= 0 \quad (1.a) \\
\ddot{\hat{T}} + 2\hat{F} \dot{\hat{T}} + \frac{e}{\sqrt{\pi} T_b} \left( \hat{T} - T_b \right) &= 0, \quad (1.b)
\end{align*}
\]

where the ratio \(e\) between the total mass of the gas and the mass of the piston is the only relevant physical parameter [4] required to describe the system. If the time evolution of the solution to eq.1 for a given \(F\) and \(T_b\) protocol is known, then the heat exchanged by the system with the reservoir between from \(t=0\) to \(t=\tau\) is given by [5]

\[
Q(\tau) = -\int_0^\tau \left[ F \dot{x} + \dot{x} \ddot{x} + \frac{T}{2} \right] \, dt. \quad (2)
\]

In this extended abstract we report that we can find an approximated analytic solution to eq.1 with the multiple scales method if we assume that that \(e\) is small and that the \(F\) is slowly varying over time, i.e. \(F = F(\varepsilon t)\), we proceed to an asymptotic expansion using the multiple scales method. Such solution contains all the relevant physical behaviors of the system and allows to derive analytic formulas for the heat exchanged when finite time transformations are performed on the systems.

II. MULTIPLE SCALE EXPANSION METHOD

The multiple scale method is an old technique in perturbation theory introduced to avoid the problem of secular perturbations [6] appearing when differential equations are solved with the standard perturbation techniques. Leaving out all the technicalities (interested readers can refer to [3] and references therein), this method supposes that the solution of an arbitrary differential equation can be decomposed in various terms, each one related to a single specific physical phenomenon. For example, the multiple scales method applied to the damped harmonic oscillator decomposes the motion of the oscillator into a first contribution that yields the oscillations and a second one that fully characterizes the damping. This is possible thanks to the introduction of a set of functions, called scales. These and generated by the structure of the differential equation, and each one of them describes a single and specific physical phenomenon. What we did in [3] is to assume that \(\varepsilon\) is a small parameter and that the transformations we are interested in are the ones for which (i) \(T_b\) is constant and equal to one, and (ii) \(F\) is slowly-varying over
time, i.e. \( F=F(\varepsilon t) \). These hypotheses allowed us to apply the multiple scales method to eq.1 giving us an approximated expression for the \( x \) and \( T \) for a whole class of external forces. The space here available does not allow reporting the exact formulas and passages but we can comment two interesting results we found. The first one is that three scales are necessary to fully characterize the description of \( x \) and \( T \) time evolution and they are 

\[
(t_0, t_1, t_2) = \left( \int_0^1 F(\varepsilon t) \, dt, \int_0^1 F(\varepsilon t) \, dt, \int_0^1 F(\varepsilon t) \, dt, \varepsilon t \right).
\] (3)

Among those scales, \( t_0 \) is the fastest one and characterizes an exponential-relaxation of the system toward the equilibrium. It is also characteristic of dissipative mechanisms emerging from the indirect coupling of the piston with the reservoir; probably what remains of shock waves propagation in the medium after that the working hypothesis have crudely ruled them out. The \( t_1 \) scale is instead the one at which the oscillations of the system are established. The last scale, \( t_2 \), describes the suppression mechanism related to the mechanical damping the gas acts on the piston and is the one where the exponential suppression of the oscillations described by \( t_1 \) appears. These three scales are therefore able to quantitatively emerging in the system, but that are not easy to extract from the direct analysis of the underlying Hamiltonian dynamics of its constituents.

### III. THERMODYNAMICS OF FINITE-TIME COMPRESSIONS

In this section we show the second interesting result. As before, the passages are not included, but the interested reader can refer to \cite{3}.

Let us consider the case in which the gas, initially at equilibrium for \( T_0 = 1 \) and \( F = 1 \), undergoes a linear increase of the external force over a finite time and then relaxes to the new equilibrium condition. This encompasses all the isothermal compressions occurring in a finite time. Without the loss of generality, this is modeled by taking

\[
F(\varepsilon t) = \begin{cases} 
1 & \text{for } t < 0 \\
1 + f & \text{for } 0 \leq t < (a\varepsilon)^{-1} \\
1 + f & \text{for } t \geq (a\varepsilon)^{-1}
\end{cases}
\] (4)

with \( x(0) = 1, \dot{x}(0) = 0, T(0) = 1 \) as initial conditions. The additional parameters appearing here are the amount of force \( f \) by which \( F \) is increased and the constant \( a \) which fixes the time-span \( (a\varepsilon)^{-1} \) over which the the compression occurs. Qualitatively speaking, \( a\varepsilon \) can be interpreted as the speed of the compression. Putting eq.4 and eq.5 together with the approximated solutions of eq.1 obtained with the multiple scales expansion method in eq.2 one gets that the heat exchanged with the reservoir during the whole process is

\[
Q(\infty) = \ln(1 + f) - (f\varepsilon)^2 G_1(f) + (f\varepsilon) G_2(f),
\] (6)

where \( G_1 \) and \( G_2 \) are complicated functions of \( f \) (see \cite{3} for their structure) and which is valid only if

\[
f > 0, \quad a \leq f^{-1}, \quad f - \ln(1 + f) \gg \varepsilon^2
\] (7)

holds. Eq.6 is our main thermodynamic result: it estimates the heat produced by a perfect gas under the action of a finite-time compression and, to our knowledge, known analytic formulas capable to do so are rare, at best. In addition eq.6 is thermodynamically consistent: taking the limit of quasi-static transformations, i.e. \( a \to 0 \), yields

\[
\lim_{a \to 0} Q(\infty) = \ln(1 + f).
\] (8)

This is exactly the minimum value prescribed by Clausius theorem, so the remaining terms of eq.6 are contributions coming to the fact that the system is driven with a non-negligible speed. Clearly, such contributions must be positive if the second law of thermodynamics must be satisfied, but this requirement is always true within the constraints given by eq.7. This check, together with a numerical analysis of the exchanged heat, confirm us that eq.7 is a reliable result.

### IV. CONCLUSIONS

The multiple scales method allowed us to find a physically good approximated analytic description of a gas enclosed by a piston and in contact with a thermal reservoir when the temperature of the reservoir is fixed and the force applied on the piston varies slowly. Emergent phenomena like the piston-gas friction and the indirect energy transfer between the gas and the piston are naturally described with this approach. In addition, we showed that the heat produced when the gas-piston system undergoes an isothermal compression in a finite time can be computed in a closed form. The results obtained are consistent with thermodynamic constraints, which tells us that the multiple scales method is a promising methodology to access the non-equilibrium thermodynamics of a system. Additional results and considerations can be found in \cite{3} and \cite{7}.

### ACKNOWLEDGMENTS

We would like to thank L. Gammaitoni, A. Vulpiani and L. Cerino for the useful discussion on the topic. DC is supported by the European union (FPVII(2007-2013) under G.A. n.318287 LANDAUER). GG is supported by INFN IS-CSN4 Mathematical Methods of Nonlinear Physics.

### REFERENCES


Optimized Photovoltaic Solar Cells combined with Improved Thermal Efficiency

L. Ferre Llin1, S. Thoms1, A. Mellor2, D. Alonso Alvarez2, D. J. Paul1, N. Ekins-Daukes2
1School of Engineering, University of Glasgow, United Kingdom, 2Department of Physics, Imperial College London, United Kingdom

Abstract—The electrical performance of photovoltaic (PV) cells can be degraded by the thermal radiation that these devices are exposed to. For example, the efficiency of a crystalline Si PV cell can be reduced by 0.4% per 1°C. Therefore, there is the necessity to engineer these solar cells to maintain their high efficiency while also take advantage of the extra heat generated by the system. Hybrid photovoltaic-thermal generators normally use standard PV cells, which are only optimized for electrical photovoltaic performance. We propose a developed PV cell with a nano-textured back surface to reduce heat losses. The nano-textured rear reflector has been designed to increase the absorption near the infrared region of the solar spectrum in order to generate more heat.

I. INTRODUCTION

Photovoltaic solar cells are able to absorb 80% of the solar irradiation, however, not more than 20% can be converted into useful electricity. The excess of this energy is converted into heat, which in sunny days can increase the temperature of the cell to 80°C. Depending on the PV cell technology used, this increase in temperature as a function of time can quickly decrease the electrical efficiency of the cell [1].

The use of thermal collectors, either air or fluid collectors, attached to the PV cells has become an interesting technology to harvest the heat generated by the PV cell and at the same time deliver some cooling to the cell to remain its electrical properties. The most common PV-thermal system at present consists of a PV cell with a fluid collector at the back to conduct heat away from the cell that can be used for heating in buildings or hot water systems. The best of these systems has an overall efficiency approaching to 70%, where 20% corresponds to electrical efficiency and over 50% is due to thermal efficiency. Therefore it has become interesting not only to optimize PV cells for electrical conversion but also for thermal generation.

Different mechanisms contribute to the conversion of solar irradiance into thermal generation in the cells. In mono-junction cells, only photons with an energy equal or greater than the bandgap energy will potentially be absorbed into the cell, generating an electro-hole pair per photon. In addition, photons with energy greater than the bandgap will generate excess in energy, which will contribute to the generation of heat in the cell [2]. In order to reduce heat losses in the cell, spectrally-selective low-emissivity coatings can be deposited at the top of the cell to suppress thermal emission at long wavelengths (>3000 nm).

On the followings sections it is described the design and the fabrication process to produce the metallic-nano-textured back reflector.

II. METALLIC-NANO-TEXTURED BACK REFLECTOR TO OPTIMIZE ABSORPTION IN THE NEAR-INFRARED SOLAR SPECTRUM

As mentioned earlier, depending on the PV cell technology used, the electrical efficiency of a PV cell can drop drastically as a function of temperature. Thus the need of using PV cells with a low temperature coefficient in order to create efficient PV-thermal systems, such as heterojunction solar cells [3]. Heterojunction solar cells, however, present a low-doped emitter and therefore they are expected to have a low absorbance in the solar spectrum at wavelengths between 1100 and 2500 nm. This absorbance could be increased by the generation of surface plasmons at the rear reflector [4]. Next, we present the design and fabrication process to fabricate an array of nano-holes and nano-pillars at the back of a cell.

The plasmons structures were fabricated by using electron beam lithography (EBL), dry etch processes and electron beam metal evaporation. First the Si surface was spin coated with electron beam resist (ZEP520A) and both, holes and pillars were patterned by (EBL). The arrays were covering areas of 12 by 12 mm. After patterning the desired etch mask, the samples were loaded into an STS-ICP (Inductively Coupled Plasma) tool from Surface Technology Systems. 200 nm of Si were etched using a mixed process of C4F8 and SF6 gases, creating an anisotropic etch with straight sidewalls to define the holes and pillars. After the etch, the resist was removed by ashing the surface with oxygen plasma, and 200 nm of aluminum (Al) were electron-beam evaporated to coat the plasmon structures. While the metal evaporation process the sample holder was tilted by 12º to assure a complete coating of the structures. A scanning electron microscope (SEM) image of the etch nano-holes and nano-pillars in Si is shown in Fig. 1.

Fig. 1. SEM images showing the nano-hole and nano-pillar arrays after dry etch, with a diameter of 400 nm and spacing of 100 nm.

Fig 2. (left) shows an atomic force microscopy (AFM) image of the nano-holes after performing the ICP etching and Al deposition,
while Fig. 2 (right) shows a cross section SEM image of the nano-holes with Al.

In addition, a texturing at the front surface of the cell was developed to decrease the reflection of Si and also to promote light trapping inside the cell. The texturing design was based on an array of inverted pyramids covering areas of 12 by 12 mm. In this case, the Si surface was coated by 40 nm of thermal oxide followed by 60 nm of low-plasma-chemical-vapor-deposition (LPCVD) silicon nitride (Si$_3$N$_4$). Electron beam resist (PMMA) was spin coated on top of the surface and holes with 20, 5 and 3 μm diameters were patterned by EBL. The Si$_3$N$_4$ holes were etched by a reactive ion etching (RIE) process using CHF$_3$ and O$_2$ as gases. Once the Si$_3$N$_4$ etch mask was defined, a 55°C heated solution of de-ionized water, potassium hydroxide (KOH) and isopropylalcohol (IPA) was used to perform a wet etch into the Si. This wet etch promotes a parallel etching to the (100) crystal orientation while it remains almost zero-etching to the parallel (111) planes. An angle of 70° is created when the (111) planes meet, and because of their slow etching rate, the process is then terminated creating an inverted pyramid. The Si$_3$N$_4$ wet etch mask was finally removed by an RIE process. Fig. 3 shows an SEM image of the 20 μm diameter base inverted pyramids.

In order to investigate the improvement of the nano-texture structure developed at the rear of the cell, different configurations [5] have been proposed as shown in Fig. 4. The optical properties for these samples will be characterized optically at wavelengths from 1100 to 2500 nm.

III. CONCLUSIONS

PV-thermal systems are currently used to produce electrical power through the photovoltaic effect and also to harvest part of the heat generated for the PV-cell. To increase the heat generated at the cell a nano-texture surface has been proposed and develop on c-Si samples to absorb phonons at wavelengths between 1100 and 2500 nm. Different samples configurations have been already assembled and optical results will be presented somewhere else, focusing this paper on the fabrication technology employed to create such structures.

ACKNOWLEDGMENTS

We wish to thank EPSRC Project number EP/M02512/1 for support in undertaking the research in this paper.

REFERENCES

Zero-Power computing
M. López-Suárez, I. Neri and L. Gammaitoni
1Department of Physics, Università degli Studi di Perugia, Italy

Abstract— After the publication on 1961 of Landauer most famous work, where he established a relation between information and physical entropy, the claimed linkage between logical and physical irreversibility raised long debates. It is clear that a net entropy reduction on a physical system implies the dissipation of a certain amount of energy. This statement has been wrongly extended also to the reduction of informational entropy on computing devices even if this is not accompanied with a physical entropy reduction. In this paper we present an experimental demonstration of the lack of linkage between informational and physical entropy in computing logic gates.

I. INTRODUCTION

In modern computers computation is performed by assembling together sets of logic gates made by transistors acting as logic switches. The scaling of transistor dimensions allowed a continuous increase of performance accompanied by an increase of heat production. The latter will represent a road-block for the development of future computing devices. Landauer theorized on the minimum heat production required to perform a reset operation of one bit [1], i.e. $Q_L = k_BT \log_2$, a result that has been recently experimentally shown by S. Ciliberto et al [2]. However, a question that is still somehow controversial [3-5] is the following: is dissipation intrinsic to all logically irreversible operations? In the following we will present the implementation of a memory bit and a logic gate with MEMS technology. The values for the dissipated energy during the Landauer reset are compatible with $Q > Q_L$. Moreover, we have also studied the effect of the error probability on the total dissipated energy. Finally, the lack of linkage between informational entropy and physical entropy is demonstrated by operating a logically irreversible logic gate well below the Landauer limit. It is also demonstrated by coupling different MEMS based logic gates the feasibility of a full-adder.

II. LANDAUER RESET

Fig. 1(a) shows the schematic of the device considered to perform the Landauer reset. Bistability is induced through the interaction of two magnets with opposite magnetization placed one at the tip of the cantilever and the other in front of it [6]. The state of the bit is encoded in the cantilever’s tip deflection, i.e. ‘0’ (x<0) and ‘1’ (x>0). Varying the position of the counter magnet allows tuning the separation of the two stable states, the barrier height separating them and also the asymmetry of its total potential energy. V_u and V_d are used to tilt the potential energy of the system by polarizing the cantilever and inducing an electrostatic force allowing to switch the state of the bit, similar to [7]. We compute the Stratonovich integral as in [8] in order to evaluate the heat produced, Q.

Fig. 2(a) shows the average heat produced during the reset operation as function of the lateral alignment $\Delta x$. For $\Delta x < 0$ the counter magnet is moved to the bottom and the 0 state (x < 0) is favored. Accordingly, when $\Delta x > 0$ the 1 state is more favored. Introducing an asymmetry on the potential, Q decreases, which is reflected in the probability of success, $P_s$, that tends to decrease (see color coded map). The success rate of the reset operation as function of lateral alignment is shown in Fig.2(b). Solid violet circles represent the overall success rate while black and red symbols account for the success rate resetting to 0 and 1 state, respectively. The maximum overall success rate happens when the system is almost symmetric, $\Delta x \approx 0$. Please note that the measurements have been performed at effective temperature, well above room temperature, in order to make static dissipation negligible respect $Q_L$.

III. ZERO-POWER COMPUTING

Now we move towards the realization of a MEMS logic gate. The logic gate (see Fig.1(b)) is designed to be operated in analogy to a transistor based logic gate and, therefore, the output of the device must be read while the input forces are applied to the system. Once the inputs are removed the output logic state recovers its original state, i.e. ‘0’. In this case both electrostatic probes are at one side
of the cantilever, thus if one or both inputs are set to ‘1’, a negative deflection (x<0) is caused. There are a total of four possible input configurations, i.e. <00>, <01>, <10> and <11>, and the estimation of the total heat dissipated during the operation of the logic gate accounts for all four input configurations.

The protocol duration, \( \tau_p \), is the control parameter that set the speed of the operation. In order to measure the energy dissipation, the logic gate is operated cyclically sweeping all the possible input combinations. Each combination is applied for different protocol duration \( \tau_p \). For each \( \tau_p \), Q is evaluated over ~1000 cycles. Fig. 3(a) shows the obtained values for Q as a function of the protocol duration for all the input configurations except <00> since it is always equal to zero (by definition). As expected the total heat dissipated decreases as the time protocol increases, indicating that the system is reaching the adiabatic regime where \( Q=T\Delta S \). As it is apparent, energy dissipation falls well below \( k_B T \) even for the shortest \( \tau_p \) values. Fig. 3(b) shows the distribution of Q for \( \tau_p=9\,\text{ms} \). Negative tails can be observed although the mean value for Q is positive.

More complicated operations as where a full-adder device can be realized with four micro-cantilevers.

**ACKNOWLEDGMENTS**

The authors gratefully acknowledge financial support from the European Commission (FPVII, Grant agreement no: 318287, LANDAUER and Grant agreement no: 611004, ICT- Energy), Fondazione Cassa di Risparmio di Perugia (Bando a tema Ricerca di Base 2013, Caratterizzazione e micro-caratterizzazione di circuiti MEMS per generazione di energia pulita) and ONRG grant N00014-11-1-0695.

**REFERENCES**


Leveraging Non-Volatile Technology to Design Processor: Use Case of TLB CAM Cell Based on Nano-Electro-Mechanical Switch and CMOS

Azam Seyedi1,2, Vasileios Karakostas1,2, Stefan Cosemans3, Adrian Cristal1,2,5, Mario Nemirovsky1,4, Osman Unsal1
1Barcelona Supercomputing Center, 2Universitat Politecnica de Catalunya, 3IMEC, 4ICREA, 5IIIA-CSIC

Abstract—In this paper we describe a new Content Addressable Memory (CAM) cell design, NEMsCAM, based on both Nano-Electro-Mechanical (NEM) switches and CMOS technologies. The memory component of the proposed CAM cell is designed with two complementary non-volatile NEM switches and located on top of the CMOS-based comparison component. As a use case for the NEMsCAM cell, we design first-level data and instruction Translation Lookaside Buffers (TLBs) with 16nm CMOS technology at 2GHz. The simulations show that the NEMsCAM TLB reduces the energy consumption per search operation (by 27%), write operation (by 41.9%) and standby mode (by 53.9%), and the area (by 40.5%) compared to a CMOS-only TLB with minimal performance overhead.

I. INTRODUCTION

Nano-Electro-Mechanical (NEM) switches have been suggested as a promising candidate for replacing the CMOS technology [1]. NEM switches provide some unique characteristics which are not available in conventional MOS, such as near-zero leakage current and infinite subthreshold slope. Such characteristics make them ideal for designing highly energy-efficient structures. However, NEMs have relatively long mechanical switching delay [1] compared to the intrinsic delay of CMOS devices, and to this date, they suffer from low endurance (1011 write cycles) [2]. Also, NEMs do not offer high turn on current like CMOS transistors.

We describe a new CAM cell design, NEMsCAM [3], based on both NEMs and CMOS technologies, to employ in processor structures where writes are relatively infrequent, for example TLBs. As a use case, we leverage the design of the proposed NEMsCAM cell to build fully associative NEMsCAM TLBs and our analysis shows that the NEMsCAM TLB exhibits significant benefits over the CMOS-only TLB in terms of energy consumption and area. The main contributions of this paper are: 1- We explain the NEMsCAM cell design based on complementary non-volatile NEM switches and CMOS transistors. 2- We explain the design of highly efficient first-level TLBs for data and instruction accesses based on the NEMsCAM cell. 3- We evaluate the proposed designs at both circuit and system level, and compare to CMOS-only TLBs.

II. DESIGN OF NEMSCAM CELL

In this section, we present the circuit details of our proposed NEMsCAM cell. We use the memory structure proposed in [4] to implement the storage part of NEMsCAM. That memory structure provides full-select behavior which are essential to design a CAM cell; it also uses electrostatic pull-out and pull-in and does not require a cell selector device in the write path. The non-volatile memory design is based on the NEM switch described in [5] which has the ability to eliminate net-disturbing electrostatic force. Figure 1 shows the schematic of the NEMsCAM cell. The outputs, Out and OutB, are connected to the transistors of the comparison circuit. The memory component of NEMsCAM cell is designed with two complementary non-volatile NEM switches while the comparison circuits are designed with CMOS transistors to allow fast search operation and avoid the long delay of the NEMs that occurs due to the mechanical movement of the beams, and that would slow down the search operation.

Figure 2 presents the three-dimensional view of two adjacent NEMsCAM cells located in the same column index of the CAM array. Since NEMs have the potential to be fully integrated with CMOS devices [6], we place them on top of the CMOS layer and substantially reduce the layout area. The SL wires run parallel to the BL wires, while the matchlines (ML) and wordlines (WL) are orthogonal to the BLs. By employing vertical NEM switches [7], the requirement of a long beam has little impact on the layout area, as it is out-of-plane. Two Gate1s are aligned and connected to their corresponding WL while the two Gate2s are connected to 0. The drains are connected from the opposite sides and form a cross shape. Due to this organization, our proposed NEMsCAM cell reduces the wire lengths which considerably reduces the energy consumption along with the near-zero leakage characteristic of NEMs.

![Fig. 1. Schematic of the proposed NEMsCAM cell.](image1)

![Fig. 2. 3D view of two adjacent NEMsCAM cells in a CAM array.](image2)

III. A USE CASE FOR NEMSCAM: TLB

As mentioned before, we leverage the design of the proposed NEMsCAM cell to build a fully associative NEMsCAM TLB. The TLB hierarchy has been reported to account for an important percentage of the energy spent in the chip. Intel recently reported that 13% of the total core power comes from the TLBs for memory-intensive workloads [8]. Based on our evaluation...
infrastructure [3], we find that the TLB energy is overwhelmingly dominated by the first-level TLBs in terms of accesses across the TLB hierarchy. Moreover, by breaking down the energy consumption in the first-level TLBs, we find that the CAM part contributes by 94%. To reduce this source of energy consumption without affecting the performance, we leverage the NEMsCAM cell to design a highly energy-efficient first level TLB. We design the NEMsCAM TLB with our proposed CAM cell and with typical SRAM memory circuits. The CAM part (Figure 3) consists of the NEMsCAM cells and the necessary peripheral circuitry optimized for both search and write operations. Similarly, the SRAM cells and the associated circuits are designed with CMOS technology.

IV. EXPERIMENTAL EVALUATIONS AND RESULTS

We design NEMsCAM-based TLBs for data (DTLB) and instruction (ITLB) accesses based on [9] using the TLB organization of a modern AMD server-oriented processor [10] (Table 1). For both NEMsCAM and CMOS-only TLB, we construct the transistor level netlists with all the necessary circuitries, equivalent capacitances and resistance of wires. We simulate and optimize both TLB structures with Cadence Spectre using 16nm Predictive Technology Model at T=25°C targeting 2GHz processor frequency. To estimate the impact of the NEMsCAM TLBs at system’s performance, we use the Sniper simulator [11] with the configuration of Table 1, and run the TLB-intensive workloads from Spec2006 with the reference input set and execute for one billion instructions.

First, we observe that the area reduces by 40.5% for the DTLB. The reason for this improvement is thanks to the novel structure of the NEMsCAM cell. Second, we observe that the energy per search and write operation and standby mode reduces by 27%, 41.9% and 53.7% respectively for the DTLB. This happens due to the lower dimensions of the circuit leading to lower parasitic wire capacitances and resistances on the searchlines and the matchlines which in turn require fewer driving buffers. Moreover, the energy consumption further reduces due to the near-zero leakage current that NEMs provide. Similar results hold for the ITLB as well. Furthermore, we show that the NEMs’ increased write latency introduces minimal performance overhead (0.27% on average).

<table>
<thead>
<tr>
<th>Table 1. TLB organization of a modern processor [10].</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
</tr>
<tr>
<td>Data (DTLB)</td>
</tr>
<tr>
<td>Instruction (ITLB)</td>
</tr>
<tr>
<td>Level 2</td>
</tr>
<tr>
<td>Data (L2-DTLB)</td>
</tr>
<tr>
<td>Instruction (L2-ITLB)</td>
</tr>
</tbody>
</table>

V. CONCLUSIONS

We describe the NEMsCAM cell design that combines both NEMs and CMOS to design low power and highly efficient processor structures such as TLBs. Our analysis shows that the NEMsCAM TLB exhibits significant benefits over the CMOS-only TLB in terms of energy consumption and area. However, the limited write endurance of current NEMs may delay their adoption until technology improves.

REFERENCES

Underwater piezoelectric energy harvester

Maurizio Mattarelli, Francesco Orfei, Giacomo Clementi and Francesco Cottone
Department of Physics, University of Perugia, Perugia, Italy

Abstract— In this works we present an energy harvester for underwater applications. It is based on a piezoelectric cantilever and it can be used to harvest energy in water pipelines. We present a finite elements analysis and the experimental results of the tests conducted in a reduced size model of a real pipeline.

I. INTRODUCTION

In this paper we describe how we are powering a sensor to monitor physical and chemical parameters in a water pipeline. In order to provide self-powering capability to wireless sensor devices, we decided to not use batteries but an energy harvester [1, 2]. Generally, a turbine (e.g. Pelton, Darrieus) for capturing the energy of the water flux is used, but we have chosen to exploit the vibrations induced by the Von Karman Vortex [3] generated by turbulence around a cylindrical or blunt body.

Fig. 1 shows a simulation of the vortex that a blunt body generates in a water flux. Self-sustaining drag and transversal lift force are generated by the vortex and after a transient the excitation becomes periodic.

This energy harvester works in linear mode and its resonance frequency, when in the water, is about 30 Hz. The excitation due to the water flux is a sinusoidal force at the frequency \( f = St \frac{v}{d} \), where \( St \) is the Strouhal number [4], \( v \) is the speed of the water and \( d \) is the cylinder diameter. At the speed of 0.7 m/s, this corresponds to a frequency of 6 Hz. Even if the two frequencies are different, we obtained a good conversion confirming that our harvesting is working on a wide band.

II. ENERGY HARVERSTER DESIGN

The first version of the water flux energy harvester is based on piezoelectric elements, thus we named it Piezoelectric Vortex Generator (PVG). The overall mechanical design of the generator is shown in Fig. 2.

The transduction element is a bimorph piezoelectric beam that is provided by Midé Volture company (www.mide.com, model V21BL). Two piezoelectric layers are deposited on the side of inner steel layer and have copper electrodes on both sides. Then the beam is packaged with FR4 and presents a thickness of 0.8 mm. The mechanical water-air interface is disk-cap printed in ABS. This element has the important function of feedthrough for electrical connections.

Fig. 2. Design of the piezoelectric linear energy harvester.

This energy harvester works in linear mode and its resonance frequency, when in the water, is about 30 Hz. The excitation due to the water flux is a sinusoidal force at the frequency \( f = St \frac{v}{d} \), where \( St \) is the Strouhal number [4], \( v \) is the speed of the water and \( d \) is the cylinder diameter. At the speed of 0.7 m/s, this corresponds to a frequency of 6 Hz. Even if the two frequencies are different, we obtained a good conversion confirming that our harvesting is working on a wide band.

III. TEST SETUP

In our lab at NiPS we have built a reduced scale (1:10) water circuit in order to test the energy harvester (Fig. 3). A pump pulls and pushes the water from a reservoir through a 2 inches wide pipeline where we placed the piezoelectric generator (PVG). Smaller flexible tubes (1 inch in diameter, in yellow) connect the test pipeline (in grey) to the pump and to the reservoir creating a closed loop circuit.

The pump is a common one for water. It is made by Lowara, model CEAM 80/5 A. It can provide a flux of 100 ltr/minute, requiring a power of 0.75 kW at 230 Vac. We also provided a power regulator to change the rpm of the pump and the flux of the water.

Our target is to make this system work with a flux of about 100 ltr/s: we can expect much more power in a real scenario. Thus this setup is useful to understand the capability of such kind of energy harvesters.
IV. PRELIMINARY RESULTS

We conducted several tests to evaluate our system. First, we investigated if the water flux was enough to make our generator to oscillate. After this we tried to optimize the depth of insertion of the oscillator in the pipeline. We performed this operation empirically trying to find the maximum output voltage with an oscilloscope.

Finally we evaluated the amount of power that can be converted in a typical working condition where the water flux was travelling at the speed of 0.7 m/s. We loaded the piezoelectric with a variable resistor and we have found that the optimal load value is 300 kOhm. In these conditions the RMS value of the output voltage is approximately 7 V and the corresponding power is about 160 µW.

Fig. 4 depicts the output voltage when the piezoelectric is loaded with a 300 kOhm resistor. As it can be seen, this signal is not sinusoidal. This is due to the impacts of the cylinder on the internal walls of the pipeline.

V. CONCLUSIONS

We have demonstrated an alternative technology to extract energy from a water flux with minimal impact on the pressure in the pipeline (0.01 bar drop).

We have also demonstrated that Von Karman Vortex can be used to make extract energy from a flux by using a piezoelectric and that this energy harvester has a wide frequency response.

The proposed system can be further optimized: e.g. experimenting different shapes of the submersed oscillating body.

We are also working on self-tuning of the mechanical resonance frequency of the energy harvester with excitation: water speed can change and consequently the frequency of the excitation.

ACKNOWLEDGMENTS

This work is part of the research activities funded by the European Union (EU) Horizon 2020 Programme for research, technological development and demonstration (Grant agreement n. 644852, PROTEUS) and by The Ministry for European Affairs-National Agency of Turkey under the Grant No: 2015-1-TR01-KA203-021342 – IESRES.

REFERENCES

Single Electron Devices and Circuits

M. F. Gonzalez-Zalba$^1$, S. Kaxiras$^2$, R.D. Levine$^2$, F. Remacle$^4$, S. Rogge$^5$, M. Sanquer$^6$

$^1$Hitachi Cambridge Laboratory, Cambridge, UK, $^2$Division of Computer Systems, Department of Information Technology, Uppsala University, Sweden, $^3$The Fritz Haber Research Center for Molecular Dynamics, Hebrew University of Jerusalem, Israel, $^4$Département de Chimie, Université de Liège, Belgium, $^5$Centre for Quantum Computation and Communication Technology, School of Physics, University of New South Wales, Australia, $^6$SPSMS, UMR-E CEA/UJF-Grenoble 1, France

Abstract— We introduce single-electron devices and circuits as an alternative route to transistor-based computing. The non-monotonic I-V characteristics of devices such as the single-electron transistor and single-atom transistor provide an opportunity to implement complex logic functions or even multi-valued logic which allow reducing the number of devices necessary to perform an operation. Additionally we combine single-electron devices with magnetic elements to perform reconfigurable logic operations.

I. INTRODUCTION

O

ver the past decades, transistors, the workhorses of the semiconductor industry, have been continuously scaled down in size to increase the performance and reduce the power consumption of microelectronic circuits. This trend, known as miniaturization, has achieved such a level of sophistication that currently more transistors can be fitted in a single 32GB SD card than the number of stars in the Milky Way [1]. However, the downscaling of conventional transistors is bound to reach its fundamental limit as the critical dimensions of the devices head to the atomic scale. In this after-Moore scenario, novel devices, with enhance switching characteristics not dominated by size, could provide an alternative way to extend the life of the digital era.

Moreover, if these novel devices were to be fabricated in legacy foundries using the same nanofabrication recipes as conventional transistors this would significantly reduce the financial cost of adoption.

In this article, we introduce single electron devices and circuits as an alternative route to complementary metal-oxide-semiconductor (CMOS) transistor-based computing. These novel devices, such as the Single-Electron Transistor (SET) [2], the Single-Atom Transistor (SAT) [3] and the Magnetic Single-Electron Transistor [4], base their functionality on the discreteness of charge as only a stream of single electrons flow through the structure. This phenomenon, known as Coulomb Blockade, gives single-electron devices non-monotonic I-V characteristic allowing them to perform complex logic operations. In this article, we present a few examples of how single-electron devices and circuits can perform more efficiently that their equivalent transistor designs.

II. SINGLE-ELECTRON FUNCTIONALITY: COULOMB BLOCKADE

The simplest implementation of a single-electron device consists of an isolated electrode, known as the island, which contains a small number of electrons. The island is weakly coupled to two electron reservoirs, source and drain, via tunnelling junctions.

Finally, a third gate electrode is capacitively coupled to the island and allows continuous control of the electron density. Fig. 1(a). The system must fulfil two conditions to observe the discreteness of the electron charge in the electrical transport properties. First, the energy cost to add an electron to the island with total capacitance $C_\Sigma$ should be larger than the thermal energy, $e^2/C_\Sigma > k_BT$, where $e$ is the electron charge and $k_B$ the Boltzmann constant. Secondy, the source ($R_s$) and the drain ($R_d$) resistance should be larger than the resistance quantum, $R_s, R_d > e^2/h$, where $h$ is Planck’s constant. If both conditions are met, the Coulomb energy will be dominant and the system is said to be in the Coulomb Blockade regime. In this situation, the system exhibits periodic conductance peaks known as Coulomb Blockade oscillations (Fig. 1(b)) with a gate voltage spacing given by $\Delta V = e/C_g$. This non-monotonic behaviour of the conductance with respect to gate voltage gives single-electron devices enhanced possibilities beyond simple binary arithmetic.

III. SINGLE ELECTRON DEVICES: SET and SAT.

We explored two types of single-electron devices for logic applications: SETs and SATs. For SETs the island typically contains a small number of electrons which are confined by an externally applied electrostatic potential. On the other hand, for
SATs, electrons are confined electrostatically around a single atom, typically a single-impurity atom in a semiconductor host. We studied undoped silicon tri-gate nanowire transistors fabricated in fully depleted SOI substrate at LETI facilities. First the active regions were patterned by etching the SOI layer above the 145 nm buried oxide (BOX) forming Si channels of thickness $t=12$ nm. After short oxidation of the channel ($0.8$ nm) the gate stack is formed ($1.9$ nm HfSiON, $5$ nm TiN and $50$ nm polycrystalline silicon) and etched (see Fig. 2(a)). These devices are designed to operate as conventional transistors at room temperature. However at low temperature, the electron transport becomes quantised as can be seen in Fig. 2(b). Here Coulomb Diamonds [5], characteristics of Coulomb blockade regime in SETs, become apparent. Additionally, we studied doped silicon transistors in which the electron transport was mediated by single impurity energy levels in the channel of the transistors, i.e. SAT. The compatibility of both single-electron devices with conventional CMOS fabrication lines makes them promising candidates for large scale production.

IV. RECONFIGURABLE SINGLE ELECTRON CIRCUITS: Magnetic SETs

Additionally, we explored magnetic single-electron transistors for reconfigurable logic applications [6, 7]. The device consists of an aluminium single-electron transistor with a GaMnAs magnetic back-gate (Fig. 3(a,b)). Changing between different logic gate functions is realized by reorienting the magnetic moments of the magnetic layer, which induces a voltage shift on the Coulomb blockade oscillations of the MSET. We show that we can arbitrarily reprogram the function of the device from an n-type SET for in-plane magnetization of the GaMnAs layer to p-type SET for out-of-plane magnetization orientation. Moreover, we demonstrate a set of reprogrammable Boolean gates and its logical complement at the single device level (Fig. 3(c)). Finally, we propose two sets of reconfigurable binary gates using combinations of two MSETs in a pull-down network.

ACKNOWLEDGMENTS

We thank D.A. Williams, C. Ciccarelli, M. Klymenko, M. Själander, M. Klein, J. Van der Heijden, S. Barraud and A.C. Betz for their contribution to the project. The research leading to these results has been supported by the European Community’s seventh Framework under the Grant Agreement No. 318397 and EU European Research Council (ERC) advance grant no. 268066.

REFERENCES

Comparing CMOS and NEMS quasi-adiabatic logic

Samer Houri

1Department of Quantum Nanoscience, Delft University of Technology, Lorentzweg, 1, Delft, The Netherlands

Abstract—In this work, an analysis and comparison of nanoelectromechanical systems (NEMS) and CMOS technologies for low power quasi-adiabatic logic implementation is presented. Two NEMS-based solutions, a switch-based and a resonator-based solution, are investigated. The contribution of the power-clock or energy recovery generator is estimated in order to compare CMOS and NEMS-based quasi-adiabatic architectures at the system level.

I. INTRODUCTION

Adiabatic computing architecture was proposed as a means to circumvent energy dissipation in conventional CMOS logic circuits [1]. However, the implementation of fully adiabatic circuits also required the use of reversible logic, which led to significant circuitry overhead compared to more conventional logic. Quasi-adiabatic logic was therefore proposed as a half-way solution to reduce energy dissipation with minimal circuit overhead [2]. Unfortunately both reversible and quasi-adiabatic solutions tend to perform poorly when implemented with CMOS devices, as leakage current suffered in transistors degrades their performance significantly [3]. The use of nanoelectromechanical (NEM) devices as elements to construct logic circuits was proposed as an alternative to transistors as these devices suffer from no leakage current, and therefore no static dissipation [4]. Different types of NEM-based devices have been proposed, where in the simplest implementation electrostatically actuated NEM Ohmic-type switches are proposed as a drop-in replacement for transistors as logical ON-OFF switching elements [5], this switch approach tend to have inherent limitations due to physical contact and electrostatic pull-in, and therefore dissipation cannot be scaled to arbitrarily small values. Other NEM-based devices show promise in their ability to scale energy dissipation to vanishingly small values, these include capacitively coupled electromechanical systems [6], and non-linear resonator-based systems [7]-[8]. In this work, after introducing briefly the adiabatic principle and the concept behind a quasi-adiabatic logic circuit, we shall investigate the energy dissipation in a NEM-resonator based logic solution and compare its performance with previously published results concerning the NEM-switch and CMOS-based solutions.

II. ADIABATIC PRINCIPLE AND QUASI-ADIABATIC CIRCUITS

At its simplest circuit-level abstraction, CMOS logic operations consist of charging and discharging a load capacitance \( C_L \), representing the capacitance of the logic gates, through a series resistance \( R \) mainly dominated by the resistance of the logic switches. In conventional logic circuits, where charging is done under constant bias voltage \( V_{dd} \) and discharging is done directly to ground, the total energy dissipated is equivalent to \( C_L V_{dd}^2 \) regardless of the value of series resistance \( R \). In order to improve on this value of dissipated energy two things are necessary, first reduce energy dissipated during the charging process, second efficiently recuperate the charge on the load capacitance during the discharging process. The adiabatic charging principle indicates that to reduce energy dissipation during the charging of a capacitor, a voltage ramp like the one shown in Fig 1a should be employed [9]. With the use of a linear voltage ramp, the energy dissipated upon charging or discharging of a capacitor is now given by:

\[
E_{ad} = 2 \frac{R_{CL}}{T} C_L V_{dd}^2
\]

where \( E_{ad} \) is the adiabatically dissipated energy, and \( T \) is the ramp-up or ramp-down duration. Note that the energy saving term is now given by \( R_{CL}/T \).

By employing a waveform generator, called power-clock, the logic gates are slowly charged and discharged, where in addition to applying a voltage ramp the charge now is recuperated into the power-clock at the end of each cycle, as shown schematically in Fig 1b.

Note that the above dissipation term only speak to the adiabatically dissipated energy by the charge transfer, other terms should be added to account for non-electrical and non-adiabatic terms, such as mechanical dissipation in the case of NEM devices and non-adiabatic residual terms (both electrical and mechanical) in the case of quasi-adiabatic circuits.

Finally it is important to note that the mechanical switching time constant is considered significantly longer (at least an order of magnitude) than the \( RC \) time constant of the circuit or the ramp up period \( T \) of the voltage ramp.

![Fig. 1. Schematic representation of the adiabatic charging principle (a), where a voltage ramp is applied to charge a load capacitor \( C_L \) thorough a series resistance \( R \). And (b) shows a schematic representation of an adiabatic logic circuit where the energy is transferred in a charge-discharge cycle between the power-clock generator and the logic blocks (taken from [3]).](image)

III. NEM RESONATORS BASED SOLUTION

Nonlinear NEM resonators offer an interesting route for implementing quasi-adiabatic logic functions, and although some bitwise operations have been demonstrated [7]-[8] the full extents and limitations of their potential as computing elements has yet to be explored.

The NEM-resonator solution considered here is based on a duffing resonator that can be switched from a high-amplitude vibration state to a low-amplitude vibration state [8] upon the application of a suitable pulse.
As a first order approximation of the energy dissipation calculation, the resonator is considered to spend half its time in the high state and half in the low state. With every up-down transition costing \( E_{\text{Mech}} \approx K(X_{\text{High}}^2 - X_{\text{Low}}^2) \), where \( E_{\text{Mech}} \) is the mechanically dissipated energy, \( K \) is the linear stiffness of the NEM structure, and \( X_{\text{High}} \) and \( X_{\text{Low}} \) are the high and low vibration amplitudes respectively. And having a switching rate, i.e. the rate of logical \( R = f_{\text{Mech}}/Q \), where \( f_{\text{Mech}} \) is the mechanical resonance frequency of the resonator and \( Q \) is its quality factor.

IV. POWER-CLOCK GENERATOR

A fair estimate of energy dissipation in an (quasi-)adiabatic circuit, should also include the energy cost of the waveform generator known as the power-clock generator. Despite the fact that a linear voltage ramp is near-optimal it is easier to implement a sinusoidal voltage waveform generator by using an LC tank circuit [3]. Such circuit uses the load capacitor of the logic circuit itself in series with an inductor \( L_i \), thus forming a resonant LC circuit as shown schematically in Fig. 2. This configuration also means that all the logic circuit’s capacitances are charged and discharged during each cycle, thus it would not be possible to implement power gating.

By applying a sinusoidal ramp instead of a linear voltage ramp the adiabatic energy saving factor in (1) is slightly reduced to \( 1.23 \frac{R_{CL}}{T} \). Since the generator is a resonant circuit, the clock frequency of the circuit is given as \( f_{LC} = \left( \frac{4\pi n_G c L_i}{L} \right)^{1/2} = f_{\text{Mech}} \), where \( n_G \) is the total number of logic gates, and \( L_i \) is the value of the inductance.

Fig. 2. Schematic representation of the LC power-clock generator, with \( R_i \) representing the internal resistance of the inductor, the drive transistor provides periodic current pulses to make for circuit losses (taken from [3]).

Now it is possible to account for electrical losses in such a circuit by using a modified expression from the one available in [3], as follows:

\[
\frac{1}{2} E_{\text{Mech}} = \frac{\pi^2}{2} n_G C_i V_{dd}^2 \left( \frac{R_i n_G C_i}{T} \right)
\]

Thus giving a total dissipation as \( E_T = E_{\text{Elec}} + E_{\text{Mech}} \).

Table 1 gives a comparison between energy dissipation and operating frequency for the different solutions assuming \( n_G = 10^6 \) and \( R_i = 10^4 \, \Omega \), and considering the following values, for CMOS: \( V_{dd} = 1 \, \text{V}, \, C_i = 1 \, \text{fF}, \, \) and for NEM-switches (values taken from [3]): \( V_{dd} = 1 \, \text{V}, \, C_i = 4 \, \text{fF}, \, R_{\text{Contact}} = 2 \, \text{K} \). For NEM-Resonators (values taken from various literature data): \( V_{dd} = 1 \, \text{V}, \, C_i = 1 \, \text{fF}, \, K = 9 \, \text{N/m}, \, X_{\text{High}} = 10 \, \text{nm}, \, X_{\text{Low}} = 1 \, \text{nm}, \, Q = 100 \).

<table>
<thead>
<tr>
<th>Device type</th>
<th>Energy /operation (fJ)</th>
<th>Operating frequency (MHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMOS</td>
<td>0.1</td>
<td>100</td>
</tr>
<tr>
<td>NEM-switch</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>NEM-Resonator</td>
<td>1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

V. CONCLUSIONS

The values obtained in Table 1 indicate that a NEM-resonator based system could be a tempting means to move forward in the implementation of NEM-based logic. It is important to underline that while energy dissipation for CMOS and NEM-switch solutions are relatively well established, those for NEM-resonator solution could be well improved by considering smaller resonators (the one considered here had dimensions of \( 5 \mu \text{m} \times 0.3 \mu \text{m} \times 0.16 \mu \text{m} \)), since NEM resonator are easier to scale than NEM switches. However, even when scaled operating frequencies are likely to remain in the low MHz range.

In conclusion, NEM-resonator based logic have excellent potential for the implementation of quasi-adiabatic circuits based on well-known and easy to scale device technology. Yet this topic is significantly understudied in the scientific literature, and currently no resonator-based gate-level design or circuit level design exist. Critical questions regarding the cascadibility and performance of such circuits remain, thus requiring serious investigation before solid conclusions maybe drawn on their viability.

ACKNOWLEDGMENTS

The research leading to these results has received partial funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 649953 (Graphene Flagship).

REFERENCES

xEnergy-Optimal Control of Multi-Core Applications

Waheed Ahmad, Mariëlle Stoelinga, Jaco van de Pol
Formal Methods and Tools, University of Twente, Enschede, The Netherlands

Abstract — Software applications for mobile platforms require a delicate balance between throughput and resource consumption. Programming them is a complex multi-level job: tasks must be assigned to heterogeneous processors, and must be scheduled on time, to meet throughput constraints. The processors must be controlled for minimal power consumption, using techniques like DPM (Dynamic Power Management) and DVFS (Dynamic Voltage and Frequency Scaling). How to guarantee that the system meets all throughput constraints, while minimizing the energy consumption?

This paper reviews our previous work, which proposed to model streaming applications by Synchronous Data Flow, and hardware characteristics by a Processor Platform Model. Both models are transformed into a network of Timed Automata. Subsequently, we use the tool-suite of UPPAAL, for analysis and synthesis of schedules with guaranteed throughput and optimal energy consumption. We will discuss several extensions to this basic framework, using Priced Timed Automata and Stochastic Hybrid Automata.

I. INTRODUCTION: POWER MANAGEMENT

Power consumption of computing systems is still increasing exponentially. To minimize energy usage, dynamic system-level power management gained significant value in recent years [7, 21]. Two popular techniques are DVFS (Dynamic Voltage and Frequency Scaling) and DPM (Dynamic Power Management). The rationale behind DVFS is that decreasing frequency leads to a linear decrease in voltage, leading to a cubic decrease in power consumption. DPM allows switching off idle processors, saving static power consumption.

DVFS reduces the dynamic power consumption of modern processors, by lowering the voltage and clock frequency, at the expense of raising the execution time of a task. DPM switches the processor to a low power state when it is not used, thus reducing static power consumption in idle mode. When predicting energy savings, however, one should also take into account the non-negligible costs of switching between power states. It has been shown [11, 12] that optimal energy savings require a combination of DPM and DVFS.

In principle, DVFS can be applied locally per processor, but implementing many clock domains requires complex logic. To balance energy efficiency with design complexity, the concept of Voltage and Frequency Islands (VFI) has been put forward [13, 14, 18]. One VFI consists of a clustered group of processors, running on a common clock frequency/voltage domain.

DPM is widely used in modern processors by Intel and AMD. For instance, Intel Core i7 and NVIDIA Tegra 2 employ global DVFS; the multi-core IBM Power 7 series adopts VFI.

Our aim is to analyze the throughput and energy consumption of some application on a given platform. Obviously, these depend on the mapping of tasks to processors, their scheduling over time and on the power modes used. More importantly, we are interested in the automated synthesis of valid and optimal schedules.

II. PROBLEM: APPLICATION AND PLATFORM MODELS

In a series of papers [1, 2, 3], we have developed a methodology to model applications and hardware. We proposed to use the well-established Synchronous Data Flow (SDF graphs) to model applications, and devised the Processor Platform Model (PPM) for the characteristics of the hardware platform.

An SDF graph [17, 19] consists of a set of tasks \( \{A_1, A_2, \ldots\} \) and a set of channels \( \{A \rightarrow B\} \), communicating data between tasks. A task can execute if it has sufficiently many data tokens on its incoming channels. On firing, it produces data tokens on its outgoing channels. The channels function as token buffers. Each channel in the model is decorated with three numbers:

- The consumption rate (nr. of tokens required by \( B \))
- The production rate (nr. of tokens produced by \( A \))
- The initial capacity (nr. of tokens initially in \( A \rightarrow B \))

Assuming that the SDF graph is consistent, we can find an iteration, i.e., a multiset of tasks whose multiplicity respects all production and consumption rates.

The PPM [1, 2] contains all DPM, DVFS, and VFI information. It consists of a set of processors, partitioned in VFI-islands. In particular, it contains for each processor:

- Its capability, as a subset of all tasks it can run, \( \{A_1, A_2, \ldots\} \)
- A set of discrete frequency levels, \( f_1 < f_2 < \ldots < f_n \)
- Its static and dynamic power consumption at each \( f_i \)
- Its power overhead for switching from each \( f_i \) to \( f_j \)
- The best- and worst-case execution times for each task \( A \) at each frequency level \( f_i \)

The synthesis goal is a schedule for the SDF iteration, which maps each task to a concrete processor, determines the required frequency level, and schedules it at a certain starting time. The mapping should respect the processor capabilities. Frequency levels must be switched per whole VFI only. A scheduled task should find the required number of tokens in its incoming buffer, taking into account the execution time of its predecessors. A valid schedule guarantees a predefined throughput constraint, while an optimal schedule minimizes the power consumption.

III. TIMED AUTOMATA AND VARIATIONS

To achieve automatic mapping and scheduling, we will reuse the synthesis algorithms implemented in the UPPAAL tool suite, designed for modeling and analyzing Timed Automata.

Timed Automata [5] define transitions between locations, governed by constraints on clock variables. Discrete state variables can be used to model large automata concisely, while networks of timed automata, synchronizing on handshake or broadcast channels, provide modularity to the specification.

This basic formalism admits several extensions. Hybrid Automata feature continuous variables, governed by differential equations in states and discrete jumps on transitions. Priced Timed Automata form a special case, where prices are hybrid variables that cannot be used in guards. In Stochastic Automata, all choices and time delays are governed by stochastic distributions, like the...
uniform and exponential distribution. Timed Games distinguish controllable actions (e.g., scheduling choices by the system) and uncontrollable actions (e.g., inputs or time delays determined by the environment). Finally, Stochastic Hybrid Games combine all mentioned features.

Uppaal Stratego [9] supports strategy synthesis for Stochastic Hybrid Games. It integrates the symbolic algorithms for model checking Priced Timed Automata (Uppaal Cora [6]) and for synthesizing optimal strategies of Timed Games (Uppaal Tiga [8]), with the statistical model checking algorithms for Stochastic Timed Automata (Uppaal SMC [10]). Moreover, it implements reinforcement learning to synthesize near-optimal strategies for Stochastic Hybrid Games. Strategies are treated as first class objects that may be synthesized, compared, further optimized and restricted, and analyzed for correctness and performance.

IV. AUTOMATIC STRATEGY SYNTHESIS

We summarize our results on modeling, analysis and synthesis. The basis of our results is a modular translation of the SDF and PPM models to the appropriate Timed Automata. Note that the original SDF algorithms [19] compute the fastest schedule, based on maximal concurrency, i.e., not taking into account any restriction on the available processors. We reuse this to obtain the SDF iteration.

In [1], we mapped SDF graphs on a restricted number of constrained processors. We provided a transformation to Timed Automata, to compute a feasible schedule on the minimal set of processors, which tends to reduce the energy consumption as well. In [2], we extended the PPM with DVFS, DPM and VFs, and extended the transformation to Priced Timed Automata, modeling energy consumption as prices. This enabled us to use Uppaal Cora to compute valid schedules with minimal energy consumption. The work in [1, 2] considered deterministic timing, which leads to over-dimensioning due to pessimistic worst-case assumptions. In [3], we extended the translation with stochastic timing, distinguishing controllable scheduling decision from uncontrollable delays between the best- and worst-case time. Hence, the target of that translation is Stochastic Hybrid Systems. We have used Uppaal Stratego to compute adaptive strategies, which are valid and near-optimal. These correspond to a scheduler that adapts its decisions at runtime to the actual computation time of previously scheduled tasks.

The method is supported by the tools STARS\(^1\) and COMET\(^2\) with meta-models of SDF, PPM and UPPAAL, a graphical editor, and the model-to-model transformations described above. We applied the method to benchmarks from the literature, and to industrial case studies in the EU FP7-project SENSATION\(^3\).

V. CONCLUSION

Our method can be further extended by adding a hybrid battery model, allowing the synthesis of energy-aware schedules [20, 4]. Stochastic Scenario-aware SADF enables the treatment of complex dependencies between data and timing [19, 15]. Future work could incorporate realistic timing and energy consumption of the communication fabric, and properly integrate battery scheduling. Multi-core synthesis algorithms (following [16]) could boost the scalability of the approach. Thus, the near future could bring a full-blind methodology, supporting the automated design of energy-autonomous systems.

REFERENCES


\(^1\) https://github.com/utwente-fmt/STARS
\(^2\) https://github.com/utwente-fmt/COMET
\(^3\) http://www.sensation-project.eu, EU FP7 grant number 318490
Impact of Delay Propagation on NTV PCMOS Design

L. Oudshoorn, G.A. Gillani, A.B.J. Kokkelar
Faculty of Electrical Engineering Mathematics and Computer Science, University of Twente, Netherlands

Abstract— While modern energy efficient low voltage designs focus on near-threshold voltage (NTV) operation for general purpose computing and exploiting an application’s intrinsic error resilience by deploying probabilistic-CMOS (PCMOS) circuits for application specific computing, our results emphasize the impact of delay on PCMOS bits at NTV and lower voltage operations. This delay is very important to be considered while modeling PCMOS systems, as it propagates and has crucial effects on the most significant bits of the computations.

I. INTRODUCTION

Future low voltage noise dominated designs render probabilistic behavior of CMOS. This is acceptable as far as applications’ intrinsic error resilience allows quantified inaccuracy in results to save energy consumption, such as in applications like audio/video processing and sky image formation in radio astronomy. This introduces the trade-off between energy consumption (E) and probability of correctness (p) that provides an opportunity for inexact computing to attain higher energy efficiency.

To understand why probabilistic behavior can lead to fundamentally lower energy usage, the process of switching can be analyzed from a thermodynamic perspective. Palem [1] has shown the energy gain of PCMOS as $kT\ln(1/p)$, where $p$ is the probability of correctness, $T$ is the temperature and $k$ is the Boltzmann constant. Analytical discussion in [2] suggests to model a probabilistic system to find the total error which is probabilistically calculated sum ($s'$) at any stage ($i$) equals its deterministic counterpart ($s$) as,

$$
p = P(X \leq \frac{V_{dd}}{2}) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{V_{dd}}{\sqrt{2}\sigma}\right) \tag{1}
$$

where $\sigma$ is the noise RMS. We will compare (1) with our simulation results in section II to discuss the relation in behavior of E-p curves at NTV and lower voltages.

Moreover, it is important to calculate the error propagation within the probabilistic system to find the total error which is bounded by the application’s intrinsic resilience. In case of a probabilistic ripple carry adder, the probability of correctness of the sum output of stage $i+1$ depends on the probability of correctness of the adder block itself and also on the probability that its input carry from stage $i$ is correct. Keeping in view this effect, M. Lau [3] calculated the propagation error within a 4-bit ripple carry adder for each sum and carry output. We further derive for the probability of correctness of sum outputs, i.e. the probability that the probabilistically calculated sum ($s'$) at any stage ($i$) equals its deterministic counterpart ($s$) as,

$$
P(s'_{i+1} = s_{i+1}) = \frac{1}{2} + \left(\frac{p^{c}_{i+1} - \frac{1}{2}}{2}\right) \times \prod_{j=1}^{i} \left(1 - \frac{1}{2} \right) + \sum_{k=1}^{i} \prod_{j=1}^{k} \left(1 - \frac{1}{2} \right) \tag{2}
$$

Where $p^{c}$ and $p^{s}$ are the probabilities of correct outputs for carry and sum respectively. In section III, we will show from our simulation results, the impact of delay propagation in addition to error propagation as modeled in (2) and compare them to emphasize the importance of considering delay propagation in the PCMOS system design.

Based on the fact that modern digital design targets NTV region for optimal supply voltage [4], we are specifically interested in the energy consumption vs probability of correctness relation, i.e. E-p curves, within the NTV circuit operation. We have investigated the impact of variations in frequency of operation and noise levels on the E-p curves and found a different behavior between the analytical models and our simulation results in the NTV and lower levels of supply voltage.

II. SIMULATION SETUP AND RESULTING E-p CURVES

We have used the umc65 library in Cadence IC for the simulations. Our approach is to simulate 65nm technology with increased intrinsic noise due to channel resistance to represent the much smaller future transistors. A parameter called ‘noise scale’ is used to amplify the noise levels. The points on the E-p curves are estimated by simulating many bit periods for a supply voltage setting and counting the number of correct and incorrect samples. A long transient simulation is performed using Cadence, which can simulate time domain noise. The results are then exported to Matlab where the sampling and processing of the data is done.

The simulation in Cadence is performed using a default set of parameters with variably increased noise amplification scales as shown in Table 1. In order to get all the points for the E-p curve, a parameter sweep is performed for the $V_{dd}$ parameter. To keep the simulation time acceptable, the number of points (the supply voltage step size) for the E-p curves is kept relatively low. Although this results in a less smooth curve, it gives a reasonable representation of facts.

Table 1: Simulation Parameters.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOS type</td>
<td>Umc65I N/P 12_lrvl</td>
</tr>
<tr>
<td>NMOS gate length</td>
<td>60nm</td>
</tr>
<tr>
<td>PMOS gate length</td>
<td>60nm</td>
</tr>
<tr>
<td>PMOS gate width</td>
<td>160nm</td>
</tr>
<tr>
<td>$V_{dd}$</td>
<td>Range: 0-2V, 10mV steps</td>
</tr>
<tr>
<td>$V_{in}$</td>
<td>Alternating between 0V and $V_{dd}$</td>
</tr>
<tr>
<td>$C_{in}$</td>
<td>10fF</td>
</tr>
<tr>
<td>Temperature</td>
<td>27ºC</td>
</tr>
<tr>
<td>Noise amplification</td>
<td>50x/100x/200x/400x</td>
</tr>
<tr>
<td>Bit duration</td>
<td>1ns</td>
</tr>
<tr>
<td># Bits simulated</td>
<td>800</td>
</tr>
</tbody>
</table>

The CMOS inverter circuit has been simulated to show the variations in E-p curves at low voltage levels. In our simulations, the noise can be scaled by a factor, which multiplies all generated noise by the chosen amount. Simulated E-p curves for noise scaled by a factor of 50 to 400 times are shown in Figure 1. Though noise scaling factors of over 100 are not realistic for the contemporary CMOS feature sizes, these numbers are chosen in order to better show the qualitative influence of noise on the E-p curves. The dotted line in Figure 1 represents the theoretical performance for a certain noise standard deviation (here we assume 100mV) according to (1). A large difference between the predicted shape and the simulation results is the sudden drop around the threshold voltage (+0.5V) while lowering the supply voltage. This suggests that the inverter makes errors that are caused by malfunctioning rather than output misinterpretation due to noise. This is because the channel conductance of the inverter quickly becomes lower at low voltages, causing the output capacitance to charge (or discharge) slower. At a certain point the supply voltage becomes...
too low such that the output capacitance is not charged before the sample moment, resulting in abrupt decrease in probability of correctness.

Simulations with various operating frequencies also demonstrate the delay of the circuit. Interestingly, the model used in (1) shown as dotted line in Figure 2, proves to be the theoretical maximum for the E-p tradeoff, which is approached by the simulated curves with the decrease in frequency of operation. To plot the dotted line, the standard deviation (\(\sigma\)) of the samples has been calculated for each supply voltage setting, which is assumed to be caused completely by the noise in the system. The theoretical maximum performance of a system in the presence of the measured amount of noise is calculated by filling in the supply voltage and noise standard deviation in the model specified by (1).

III. DELAY PROPAGATION IN PCMOS SYSTEMS

We further investigated the influence that connected probabilistic building blocks have on each other. We simulated the 4-bit ripple carry adder comprised of 4 full adders in Cadence IC with the same assumptions as that of the inverter. Figures 3 and 4 present the simulated E-p curves for the carry and sum outputs respectively for the 4-bit ripple carry adder along with calculated ones according to (2). Theoretically, the outputs of stages 2, 3 and 4 are expected to be almost equal to that of stage 1, but in our simulations they are worse.

The theoretical curves are based on the assumption that the propagation of error is only due to probability of correctness metric. However, the delayed correct outputs of stage \(i\) can make the probability of correctness worse for stage \(i+1\) than calculated by (2). Therefore, a logical explanation for the theory and simulation not being equal is the delay propagation. At the start of a clock, the calculations are started using whatever value is present at the input from the previous calculation. There is a 50% chance that the next input will be different. However, the new value will not be available immediately. Therefore, the calculation may be underway when a new value settles on an input. The calculation of the output then starts again, but with less time left to complete it before the clock cycle ends. Unfortunately, this problem stacks for additional stages resulting in the most significant sum output to fail first.

IV. CONCLUSIONS

We have simulated an inverter and a 4-bit ripple carry adder in Cadence that showed the shortcomings of current analytical models for the probability of correctness metric at near-threshold voltage (NTV) and lower supplies. We further investigated the impact of delay propagation in a digital system composed of probabilistic building blocks, which provides a clear insight of timing delay affecting the higher significant computational bits more than its lower significant counterparts and hence contributing considerably to the total error.

ACKNOWLEDGMENTS

This work was conducted in the context of the ASTRON and IBM joint project, DOME, funded by the Netherlands Organization for Scientific Research (NWO), the Dutch Ministry of EL&I, and the Province of Drenthe. Furthermore, we are thankful to dr. ir. A.J. Annema, ICD group, University of Twente, for simulation support.

REFERENCES


An Ultra-Low Power NVM-Based Multi-Core Architecture for Embedded Bio-Signal Processing

Rubén Braojos, David Atienza
Embedded Systems Laboratory, École Polytechnique de Lausanne
Switzerland

Abstract—Healthcare delivery is evolving towards new Wireless Body Sensor Nodes (WBSN), which are miniaturized devices able to acquire, process and transmit subjects’ bio-signals in real time within a tiny energy budget. Recent efforts on AD converters and transmission schemes have enabled a major power consumption reduction of these components, thus leaving the embedded processing stage as the dominant power-hungry component. In this context, new multi-core architectures designed with smaller CMOS devices and aggressive voltage scaling greatly improve the energy efficiency of WBSNs, but originate reliability operation concerns. In this work we present a novel WBSN architecture equipped with a completely re-designed memory subsystem (including a low-voltage low-latency non-volatile partition), which operates in combination with an advanced code synchronization management to reduce the platform power consumption by up to 82%.

I. INTRODUCTION AND MOTIVATION

ONGOING lifestyle changes are increasing the prevalence of chronic disorders, which are now the major sources of death worldwide [1]. These ailments require extensive monitoring, which represent a major financial burden for healthcare providers. Wireless Body Sensor Nodes (WBSNs) can lower these costs by allowing to acquire and analyze the bio-signals of patients even outside of a hospital environment and with little intervention from the medical staff. These devices must autonomously sense, process and wirelessly transmit body signals (such as electrocardiograms) for extended periods of time, while relying on small batteries. Thus, energy-efficiency (from acquisition to transmission) is fundamental for their ubiquitous use.

With the reduction of the energy required by signal transmission, the efficient implementation of the digital signal processing (DSP) stage is key in order to minimize the power consumption of WBSNs. As shown in Fig. 1, most of the power dissipated by these devices is due to the processing of the acquired samples. In fact, the leakage power dominates the consumption of the overall system, which, based on our analysis, reaches up to 86% of the power devoted for DSP. To overcome this problem, voltage-frequency scaling has been proposed [2] [3], but aggressive reduction of supply voltage is unfeasible under certain levels and leads to undesired memory and logic reliability issues.

In this context, herein we proposed a novel WBSN architecture equipping a completely re-designed 2-level memory subsystem, which combines low-voltage, low-latency non-volatile memories (NVMs) with tiny volatile banks, to obtain superior energy-efficiency while meeting real-time constraints.

II. PROPOSED ARCHITECTURE

Typical bio-signal processing architectures based on volatile memories are designed to minimize the idle time by employing the lowest possible supply voltage and a clock frequency that allows to barely meet real time constraints [4] [2]. Conversely, our NVM-based architecture performs short computing bursts at a higher frequency in order to minimize active time. In this way, during long idle periods the full digital architecture can be power gated, while new samples are acquired in what we term “deep-sleep sensing”. This strategy is possible thanks to the availability of persistent memory provided by the NVM.

The proposed architecture, depicted in Fig. 2, is similar to the one introduced by in [3]. It features eight low-power RISC cores interfaced to 16 data memory banks and 8 instruction memory banks. The cores have access to the memory banks through a logarithmic interconnect [5], that provides single-cycle read/write operations and perform arbitration in case of conflict among several memory requests.

![Fig. 2. Proposed multi-core architecture featuring the 2-level memory subsystem consisting on a low-latency large NVM partition and a set of small instruction and data page buffers (I-PBs, D-PBs respectively).](image)

In the architecture of [3], the entire instructions and data contents (96 KB and 64 KB respectively) reside in volatile SRAM banks while in our proposed architecture those volatile memories are realized as tiny full-custom banks, termed “page buffers”, that collectively act as a cache for the unified non-volatile storage (160 KB). These buffers have been implemented as arrays of latches that incorporate a direct input line connected to each bit cell allowing a single-cycle massive page storage or readout. For the non-volatile partitions, low-voltage STTRAM [6] structures have been used as they significantly reduce the access energy with respect to standard solutions, such as FLASH-based NVMs.
In this paper, we have proposed a novel WBSN architecture featuring a completely re-designed 2-level NVM-based memory subsystem that allows for new power management strategies resulting in up to 82% power savings with respect to state-of-the-art alternatives. Moreover, the new memory design of this architecture enables further benefits by capitalizing on new nanoscale manufacturing technologies, but this is out of the scope of this paper. We refer to the interested reader to [9] for more details.

ACKNOWLEDGMENTS

The authors would like to thank G. Ansaloni (from USI Lugano) and T. Wu, M. Sabry and S. Mitra (from Stanford University) for their insights and support in the architecture and used NVM technology.

REFERENCES


IV. RESULTS

First, we explored the configuration of the NVM-based memory subsystem of our architecture (TARGET) to determine the optimal sizes of the page buffers, which result to be 8 words each. Even though small page buffers induce an increase in the amount of page transfers, the exchange timing overhead remains below 6% as shown in Table 1. The table also shows that the platform can meet the required real-time constraints while allowing for long deep-sleep sensing periods (>90%). Such amount of inactivity leads to a considerable decrease of the platform power consumption as depicted by Fig 3. As a result, the proposed TARGET architecture can obtain up to 82% reduction (3L-MF) with respect to the SOA architecture.

<table>
<thead>
<tr>
<th></th>
<th>8L-CS</th>
<th>3L-MF</th>
<th>3L-MMD</th>
<th>RP-CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active time (%)</td>
<td>5.5</td>
<td>4.7</td>
<td>8.2</td>
<td>7.0</td>
</tr>
<tr>
<td>Page exchange (%)</td>
<td>2.3</td>
<td>3.4</td>
<td>4.3</td>
<td>5.8</td>
</tr>
<tr>
<td>Processing (%)</td>
<td>97.7</td>
<td>94.6</td>
<td>95.7</td>
<td>94.2</td>
</tr>
<tr>
<td>Deep-sleep Sensing (%)</td>
<td>94.5</td>
<td>95.3</td>
<td>91.8</td>
<td>93.0</td>
</tr>
</tbody>
</table>

Fig. 3. Average power consumption (in µW) of the TARGET and SOA architectures for the studied benchmarks.

Finally, area-wise, the novel 2-level memory subsystem is more compact than a traditional SRAM-based structure. However, as depicted in Fig. 4, the routing of the latch-based page buffers incurs in a non-negligible area overhead, increasing the footprint of TARGET by 1.27x with respect to the SOA architecture.

Fig. 4. Area breakdown (in mm²) of the TARGET and SOA architectures.

V. CONCLUSIONS

In this paper we have proposed a novel WBSN architecture
Kinetic energy harvesting at microscale: current progress and perspectives
Dimitri Galayko\textsuperscript{1} and Elena Blokhina\textsuperscript{2}
\textsuperscript{1}Sorbonne Université, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France
\textsuperscript{2}University College Dublin, Belfield, Dublin 4, Ireland

I. INTRODUCTION

MINIATURIZATION and ultra-low power consumption are current trends in modern microelectronics and circuit design. They most certainly continue to be as we progress with the concepts of the Internet of Things (IoT) and wireless sensor networks (WSN). Smart buildings, cities and transport infrastructures will have a huge number of networks of wireless communicating devices. Making sensors and systems autonomous would ease the sensor installation and would, generally, offer a lighter, smarter and more reliable system.

To supply these devices with power, one requires independent, miniature and long lasting power sources. Currently, a battery is used in a majority of cases. A battery lasts from several days to several years depending on its size and application. However, there are applications where a battery is not suitable: a harsh environment, an inaccessible location prohibiting the battery replacement or incompatibility with ecological requirements.

The concepts of self-sustaining sensors and energy harvesting active networked \cite{1} propose to derive energy from the ambient environment of a sensor. Current research in the field of the energy harvesting is focused on the increasing the yield of miniature energy harvesting sources.

There are many ways to harvest ambient energy, which can take the form of the energy of sunlight, thermal gradient or the kinetic energy of motion. Kinetic energy, and more specifically vibrations, can provide great opportunities since they are present in the environment of many applications. The goal of the research on kinetic energy harvesting is to extract the largest possible part of energy out of the environment and to convert it into electricity.

The conversion of vibrations into electrical energy can be achieved by using electromagnetic, piezoelectric or electrostatic technologies \cite{2}. Each family has its advantages and drawbacks. Roughly, we can say that electromagnetic transduction is most effective at a macroscopic scale, but its performance drops as the device dimensions are scaled down. The piezoelectric transduction is efficient at all scales but requires constant stressing of an electroactive material, which raises reliability issues. Finally, the electrostatic transduction may be more complex to implement but is particularly suitable for miniaturized systems since they are fabricated through MEMS technologies \cite{3}. Electrostatic kinetic energy harvesters (EKEHs) employ capacitive transducers for electromechanical conversion.

This communication reviews the latest advances in electrostatic kinetic energy harvesting, presenting progress in the design of transducers and conditioning electronics. We first present the modern requirements on a kinetic energy harvesting system and discuss how they can be achieved. Then we present the principles of capacitive conversion and the state-of-the-art of capacitive transducers. We end with discussion on conditioning circuits and their most up-to-date configurations of EKEHs.

II. REQUIREMENTS ON ENERGY HARVESTERS

An energy harvester is a complex system whose design is a challenge. A schematic high-level block-diagram with a detailed description is shown in fig. 1. Below we discuss system requirements that are imposed on an EKEH.

\textbf{Power:} The first fundamental question we are concerned with is what power is in principle available from the kinetic motion of the environment? Commonly accepted estimation of the energy available from sinusoidal vibrations energy is given in \cite{2} or \cite{3}. Taking typical values for a micro-scale KEH manufactured using MEMS technology (one cubic centimeter volume, a silicon material for the mobile mass, 100 Hz sinusoidal vibrations of 1g amplitude), the maximum available power is estimated to be 12mW. Although this seems a large quantity, the estimated limit can never be achieved by a realistic system.

\textbf{Wideband Response:} It is important that the resonator of a KEH is able to capture vibrations over a wide range of input frequencies \(\omega_{\text{in}}\) – this would make the KEH robust with respect to external conditions. Recently, ultra-wideband devices and devices employing frequency up-conversion have been suggested as an alternative to low-frequency MEMS design \cite{4}.

\textbf{Non-linearity:} Compared to the sensing/actuation application of capacitive transducers, where a high linearity is desirable, the energy conversion is nonlinear in majority of cases. A capacitive transducer is an intrinsically nonlinear device (cf. Sec. III). Moreover, a capacitive transducer controlled by a conditioning circuit brings additional nonlinearity into the system (cf. Sec. IV). Nonlinearity is a challenge for the analysis and design of EKEHs.
III. PRINCIPLES OF ELECTROSTATIC (CAPACITIVE) CONVERSION AND CAPACITIVE TRANSDUCERS

A capacitive transducer is a variable capacitor with one mobile electrode whose instantaneous position \( x(t) \) determines the value of the transducer capacitance \( C(x(t)) \). A capacitive transducer is a physical capacitor whose geometry can change in time so as to affect the value of the capacitance. Although a capacitor can be of any geometrical shape, in practice the most common is a parallel plate capacitor with parallel conductive planes (electrodes) spaced by some distance, called gap. The capacitance of such a device is

\[
C(t) = \varepsilon_0 \varepsilon_r A(t)/d(t)
\]

where \( d \) is the distance between the planes (the gap), \( A \) is the overlapping area of the planes, \( \varepsilon_0 \) \( \varepsilon_r \) is the permittivity of the medium between the plates. In order to achieve a variable capacitance, one can change with time the area \( A \) or the gap \( d \).

A number of attempts are applied to the design of capacitive transducers. The goal is to achieve as high ratio \( C_{\text{max}}/C_{\text{min}} \) as possible. While in the past this ratio was below two, recent studies present capacitive transducers that are able to achieve \( C_{\text{max}}/C_{\text{min}} > 2 \) [5]. With this ratio, there are very efficient conditioning circuits that can manage the operation of such a capacitor [6].

Energy conversion from the mechanical to electrical domain occurs when the mobile electrode of a charged transducer moves in such a way that the capacitance of the transducer decreases. In the case when one uses the kinetic energy of the environment to generate electricity, the motion of the mobile electrode is generated by a mechanical resonator driven by external vibrations. The work required to change electrical energy will be provided by external mechanical forces in the form of vibrations or any other kind of motion. A capacitive transducer is a passive device, and it requires electrical biasing. In steady-state mode, this biasing can be generated using the energy converted by the transducer from the mechanical domain. However, to start-up energy conversion one requires the presence of an external electrical energy source, which provides initial energy. The need of such a source is a serious drawback. For this reason, many studies address the solutions in which electrostatic biasing is physically embedded in the transducer. These solutions are based on the use of electret layers or on triboelectric effects. The electret is a dielectric material storing some amount of embedded charge, which serves as a source of an electric field and, hence, a voltage [7]. When an electret layer is deposited on one electrode of a variable capacitor, the capacitor behaves as if there is a DC source in series with it. Triboelectricity is a phenomenon of charge separation during friction between two materials. Its use for kinetic energy harvesting has been studied by several research groups [8].

IV. CONDITIONING CIRCUITS

The role of a conditioning circuit is to bias the transducer in such a way that electromechanical energy conversion is optimal. Conditioning of electromagnetic and piezoelectric transducers only requires impedance matching. However, electrostatic transducers need a dynamic flow of electrical charges, ideally according to the so-called “constant voltage” scenario described in [9]. The difficulty lies in the precise synchronization of this flow with the variation of the capacitance.

For this reason, the constant voltage scenario is difficult to implement in practice. More practical conditioning circuits are based on the charge pump shown in fig. 2, where the synchronisation of the charge flow to the transducer is automatically achieved by the use of diodes.

A basic charge pump conditioning circuit, shown in fig. 2, operates as follows. The variation of the transducer capacitance is used for pushing charges from a low voltage source \( V_{\text{res}} \) to a high voltage source \( V_{\text{st}} \). This way, the electrical energy received by \( V_{\text{st}} \) is greater than the electrical energy provided by \( V_{\text{res}} \). The difference \( \Delta W \) comes from the mechanical domain. It can be shown that converted energy is

\[
\Delta W = (V_{\text{st}} - V_{\text{res}})(V_{\text{res}}C_{\text{max}} - V_{\text{st}}C_{\text{min}})
\]

This circuit has two advantages: (i) the synchronisation of the charge flow with the transducer capacitance variation is obtained automatically by the use of diodes and (ii) the converted energy \( \Delta W \) is defined by the voltages \( V_{\text{st}} \) and \( V_{\text{res}} \). Usually, voltage sources are implemented as large, charged capacitors. The main question is how the voltages of these capacitances are practically generated and controlled. For this reason, several practical implementations of charge pump based circuits exist. In this communication we present two common architectures: one directly derived from the basic charge pump [10], and one inspired by the Bennet’s electricity doubler [6].

V. CONCLUSIONS

The communication has presented a short review of the state-of-the-art of capacitive energy harvesters. It is expected that in the next three to five years the capacitive transduction will be widely used for electromechanical conversion in kinetic energy harvesters providing tens µW of power.

REFERENCES


Energy harvesting based on Micro System technologies

Cristina Rusu
Sensor Systems, Acreo Swedish ICT, Gothenburg, Sweden

Abstract—Harvesting environmental energy as an alternative power source addresses the increasing demand for future energy-efficient autonomous sensor systems, especially for applications requiring miniaturisation and distributed sensing such as Wireless Sensors Network and Internet-of-Things. A functional energy harvesting system requires addressing simultaneously the harvester device, the energy storage and the powering management circuits. These components are described through examples of a miniaturized harvester system developed in the European project H2020 “Smart-Memphis”.

I. INTRODUCTION

The main drivers for technology development related to self-powering of smart devices are Internet of Things and Services (IoT/S) and Wireless Sensors Network (WSN) [1]. The vision of powering trillion of sensors [2] is a challenging task and a big percentage of IoT opportunities will not realise if batteries need to be changed often, are placed in inaccessible places or need to be in large quantities. Energy Harvester, device converting environmental energy into electric signal, offers a solution to this problem of device longevity.

Energy harvester can utilize Micro System Technology (MST) and thin-film as for microsystems and portable devices or bulk approach as for other electronic devices. MST harvesters generate energy mainly from mechanical vibrations and from heat transfer. Miniaturized and MST-based energy harvesters have advantages in applications were size and weight are critical, needs ultra-low power, have low data rate and low duty cycle.

Vibration energy harvesting based on piezoelectric transduction has received the most attention for MST devices due to high conversion efficiency (power density and voltage output), simple configuration (e.g. ease of application) and no required input voltage as comparing to electrostatic and electromagnetic harvesters [3-5]. However, the scaling remains an issue and the unfavourable scaling of power with miniaturization of MST harvester needs still to be solved. Utilization of MST allows for an integrated system containing sensor, electronics, communication and energy source.

A standard energy harvester system for a WSN consists of five main components: harvester, energy storage, microcontroller, sensor(s) and transceiver. A functional energy harvesting system requires addressing simultaneously all these components. In the following paragraphs, the main challenges for harvester device, energy storage and powering management circuits will be briefly presented with results from our project Smart-Memphis (Smart MEMS Piezo based energy Harvesting with Integrated Supercapacitor and packaging) http://www.smart-memphis.eu/.

II. PZT MEMS ENERGY HARVESTER

Low frequency (below 100Hz) and low amplitude (around 1g) vibrations are challenging requirements for a harvester as in Aeronautical Structural Health Monitoring application [6] where many various, autonomous and wireless sensors are needed to collect data on the changes in the material’s complex structures (e.g. micro-cracks in aircraft wings). Requirements become tougher when also dimension is a restriction (mm³ scale) as for smart implantable medical devices [7, 8] that require a solution to the problem of device longevity.

All aspects of the harvester need to be addressed and in case of Smart-Memphis project:
- PZT piezoelectric material [8, 9] and MEMS processing is developed by Silex Microsystem AB (Sweden) with figure of merit at industry-benchmark levels (Fig 1);
- Poling of piezoelectric material and its characterization is performed by AixACCT Systems Gmbh (Germany) that designed a special setup for non-destructive poling and reliability testing on wafer level as well as on device level (Fig 2) allowing for determination of e31 on wafer level, and for piezoelectric thin film test systems and implementing of in-situ e31 measurements during poling [8];
- Simulation [8, 10], design and characterization done by Acreo Swedish ICT (Sweden). There are various losses that have to be taken in consideration: transfer of mechanical excitation, mechanical to electrical conversion, and electrical conversion. Via simulation and design the influence of the losses may decrease. Characterization is done by combining Laser Doppler vibrometer with vibration shaker and vacuum chamber to obtain information on mechanical characteristics, piezo-mechanical coupling, damping and losses (air and vacuum, Fig. 3) and verification of harvester model.

Fig. 1. 1.2 μm <100> PZT with piezoelectric coefficient e31 of -15 C/m² with high wafer uniformity; Young’s modulus of 75GPa, relative permittivity tunable 700-1400; breakdown voltage 80-130 V/μm depending on PZT type, leakage current below 200 nA/cm².

Fig. 2. In-situ measurement of e31 during temperature poling.
III. ENERGY STORAGE: SUPERCAPACITOR

Novel materials are explored by Chalmers University of Technology (Sweden) for the supercapacitor’s components: electrodes, separators and electrolytes. Material requirements for electrodes are porosity, surface area, mechanical stability, electrical conductivity, electrochemical stability. Electrodes use 3D structured carbon nanofiber / MnO2 composite material (NCNF / MnO2) exhibiting the best performance with high specific capacitance (108.6 F/g @ 0.5A/g) and excellent power capability (84.3 F/g @ 15 A/g). The Separator is from glass fibre with high thermal stability up to 600°C, excellent mechanical property and high uptake of different electrolytes. The Electrolyte is a high temperature electrolyte EMIM Ac ionic liquid enabling high working voltage window up to 1.5 V and increases the energy density to 21.1 Wh/kg. Temperature durable PVA / H3PO4 gel electrolyte with reduced leakage risk and high package capability and can deliver 82mF after high temperature exposure.

IV. POWER MANAGEMENT UNIT

Ultra-low-power integrated power management unit (PMU) developed by Linköping University (Sweden) [8] contains the piezoelectric-harvester interface circuits (rectifier) for power-transfer, DC-DC power conversion, voltage regulation and control circuitry (Fig.5). The key challenges are the efficient extraction of very low power-levels (µW range) and also ultra-low-power / low-voltage IC design. The energy consumption bottlenecks are the Communication unit (RF standard, protocols), Memory unit (low-voltage RAM), Sensor interfaces (ADCs and drivers), Power management unit (efficiency) and Controller unit. The first generation test chip is manufactured in 0.18µm CMOS technology with high-voltage option, reconfigurable rectifier and DC/DC conversion.

V. CONCLUSIONS

Energy harvesting is a complex process involving interdisciplinary scientific and technological R & D for effective translation into validated systems. A functional harvester system requires simultaneous tackling of all components requiring knowledge from material science, modelling of components and their behaviour, up to processing and instrumentation development. Reaching high efficiency levels needs efficiency optimization in the entire chain, even if incompatible boundaries and their matching contradicts some of these.

ACKNOWLEDGMENTS

Smart-MEMPHIS project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 644378.

REFERENCES

SAVE THE DATE

<table>
<thead>
<tr>
<th>DATE</th>
<th>EVENT</th>
<th>WEBSITE</th>
</tr>
</thead>
</table>

ICT-ENERGY LETTERS is realized with the contribution of LANDAUER project and ICT-Energy Coordination Action, funded under the Future and Emerging Technologies (FET) programme within the ICT theme of the Seventh Framework Programme for Research of the European Commission.