Extended Abstracts of the 2004 International Conference on Solid State Devices and Materials, Tokyo, 2004, pp. 402-403

P1-8L

Ultra Low-energy Computing via Probabilistic Algorithms and Devices: CMOS Device Primitives and the *Energy-Probability* Relationship

Suresh Cheemavalagu[1], Pinar Korkmaz[2] and Krishna V. Palem[3, 4]

[1,2,3]Georgia Institute of Technology, School of Electrical and Computer Engineering
[4]Georgia Institute of Technology, College of Computing
85 Fifth Street, NW, Atlanta, GA 30308, USA
Phone: +1-404-8943148 E-mail: kpalem@ece.gatech.edu

1. Introduction

Physical *energy consumption* and heat dissipation issues have become seriously limiting factors in the sustained impact of 'Moore's law', as an unprecedented propellant in the growth of the semiconductor industry. Apparently unrelated to this issue, *noise* has been viewed as an impediment to the successful design and deployment of physical devices supporting computing, starting with the early work of Stein [1]. Thus, significant efforts have been invested in innovative ideas meant to cope with and overcome the impact of noise, while sustaining Moore's law [2, 3, 4]. In [5, 6, 7], Palem showed for the first time that *rather than being viewed as a barrier or hurdle, harnessed as a (re)source of randomness, noise can in fact be used beneficially for energy savings.* Such energy savings were demonstrated through *probabilistic algorithms* [7], whose individual steps are realized through *probabilistic devices* [5]*;* in all of this work, the energy characteristics are modeled by extending the well-established framework of thermodynamics [8].

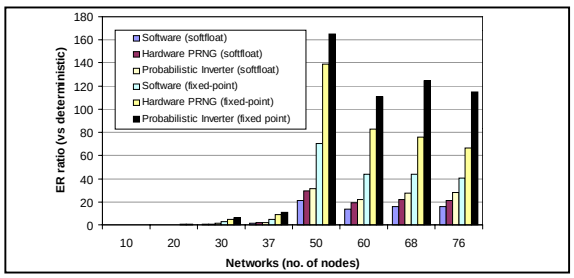

Fig.1 Improvements to energy X run-time (ER) metric through using probabilistic inverters.

In this paper, we extend the scope of the [5, 6, 7] by characterizing the behavior of probabilistic devices realized out of CMOS technology. To do this, the device of choice in this paper is the *inverter,* ubiquitous to the design of digital circuits [9] as well as being a key primitive in the widely used family of probabilistic algorithms referred to as Bayesian networks [10]. For example, it has been shown in [11] (Fig. 1) through simulation studies that the probabilistic inverter that constitutes the main topic of this paper, used in a probabilistic Bayesian network (of 76 nodes), results in a improvement by a factor of over 120 in the *energy* X *run-time* of the whole application, over a competing solution that is based on a conventional deterministic algorithm running on a SrongARM processor. (As shown in Fig. 1, this improvement is significantly higher when compared to the gains gleaned using a custom-designed pseudo-random number generator based on [12], or using the same algorithm implemented in software).

Intuitively, by allowing a higher probability of error in a CMOS device, computation can be realized with lower energy, since such a device can operate at a lower supply voltage (V_{dd}). As a result, such a device will be more susceptible to noise, which induces its probabilistic behavior. Thus, if the potential value of probabilistic devices as a basis for low-energy computing has to be fully realized, the relationship between the amount of *available noise*, the *probability of correctness p* ($0.5 \leq p \leq 1$) of the device, and the concomitant energy expended to realize the probability *p*, is a fundamental first step. To achieve this goal, the primary contributions of this '*Late News Paper*' are:

1. A characterization of the relationship between the energy consumed *E* and the probability, wherein the amount of available noise is a parameter.
2. A figure of merit that we refer to as the *noise to signal ratio* (NSR) that is uniquely associated with the amount of energy needed, independent of the amount of available noise (and other technology parameters) to realize a particular probability *p.*
3. Characterizing the variations of the relationship between the energy *E* and the probability *p* as technology and other parameters vary.

2. The Energy-Probability (*E-p*) Relationship

Given a particular CMOS device feature size, the *E-p* relationship characterizes the minimum energy consumed per-bit of information[1] produced (by an inverter), to achieve a particular probability of correctness with the amount of available noise, characterized through its rms value (RMS), as a parameter.

In Fig. 2 below, we show the *E-p* relationship for a CMOS inverter realized using a 0.5μ technology, estimated using HSpice simulations, and EKV transistor models [13]. As noted in Fig. 2, given a fixed amount of available noise, the energy needed to invert a single bit increases with the *p. As* we shall see in the sequel, through an analytical model reconciled against these HSpice estimates, this increase is exponential in *p.* Furthermore, with increasing RMS for a

[1] While probability plays a natural role in thermodynamics with its foundations in statistical mechanics, as well as in Shannons' theory of information [14], an explicit characterization of this relationship within the context of computational models, and hence its value to energy aware computing is lacking.

fixed probability value *p*, the energy consumed to invert a bit increases. Finally, as verified from the analytical modeling below, this increase is quadratic in RMS. In the following sections, we develop an analytical model of the inverter and validate it using our HSpice simulations. The noise in HSpice simulations is generated using Matlab®.

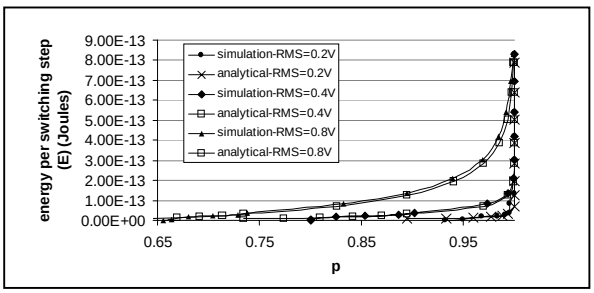

Fig. 2 The *E-p* relationship for an inverter realized in 0.5µ technology.

Modeling Noise and the Values 0 and 1

Our characterization of the amount of available noise in CMOS technologies will follow the canonical approach of using the standard deviation (σ) (or rms value) of the associated probability distribution [15]. Following Stein [1], the digital interpretations of the values 0 and 1 are shown in Figure 4 below. As shown there, noise is normally distributed. Because of symmetries, $p = (1- e_0) = (1-e_1)$.

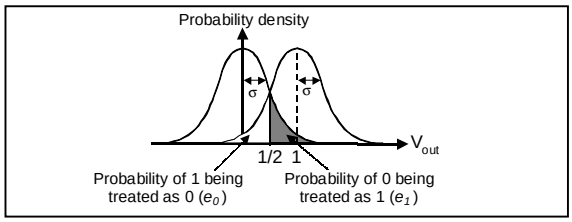

Fig. 3 Modeling noise and the 'digital' values 0 and 1 where a 0 is modeled using the voltage interval (0, ½] and a 1 using the interval [1/2, 0) in normalized voltage range in the interval (0, 1).

The Analytical Model

Our analytical approach to the *E-p* relationship will be based on the canonical approximation to an inverter used in Fig. 4 where V_n^* is the noise source.

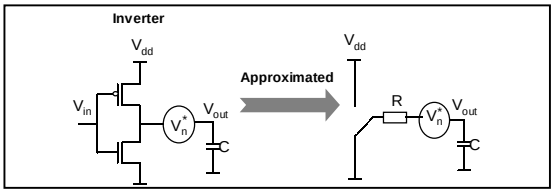

Fig. 4 The approximation for a CMOS inverter.

Analytical Model for the *E-p* Relationship

$$E = 4C\sigma^2\left[erfinv(2p-1)\right]^2 \qquad (1)$$

where *erfinv* is the inverse of the error function [16]

Using a capacitance value of 64 femtoFarads for the output capacitance value of a 0.5micron inverter that were used by the Spice simulation results, our analytical model deviates by 7 % on the average (Fig. 2).

3. The Noise-to-Signal Ratio and the Probability of Correctness *p*

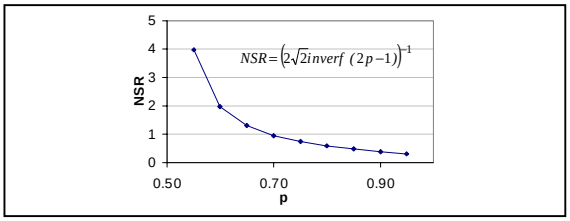

Fig. 5 NSR vs. probability of being correct *p*.

As defined in Fig. 6 where the relationship between NSR the probability value *p* is shown, an increasing probability of being correct---equivalently a decreasing probability of error---will require a lower NSR value. *The primary advantage of* NSR *is its ability as a figure-of-merit that combines the attributes of noise,* RMS, *as well as the impact of* V_{dd} *into a single numerical quantity that is technology independent.*

4. Remarks

Due to space constraints, additional results have not been described in this submission. Notably, we have extended our simulation and analytical estimates from the EKV model discussed above, to include 0.5micron AMI as well as 0.25micron TSMC processes for realizing a probabilistic inverter. Test chips with both of the above devices have been sent for fabrication through MOSIS. If accepted, these additional results as well as those characterizing variations of the *E-p* relationship and NSR across technology generations, as well as across temperature variations and V_{dd} fluctuations, will be presented at the conference. Additional details about this work can be found at [17].

References

[1] K.-U. Stein, IEEE J. Solid-State Circuits **SC-12** (1977) 527.
[2] B. Kish, Phys. Letters A **305** (2002) 144.
[3] Natori, and N. Sano, J. Applied Phys. **83** (1998) 5019.
[4] K. L. Shepard, ***Design Automation Conf.*** (1998).
[5] K. V. Palem, Georgia Inst. of Technology Tech. Rep. GT-CC-03-16 (2003).
[6] K. V. Palem, *Intl. Conf. on Compilers, Architecture and Synthesis for Embedded Systems* (2003).
[7] K. V. Palem, *Intl. Symp. on Verification (Theory and Practice)* (2003).
[8] H.S. Leff and A.F. Rex, Institute of Physics (2003).
[9] J. D. Meindl, Proceedings of the IEEE **83** (1995) 619.
[10] J. Pearl, Morgan Kaufmann (1988).
[11] L. N. Chakrapani, K. Palem, and B. Seshasayee, Unpublished Manuscript (2004).
[12] S.K. Park and K.W. Miller, Comm. of the ACM (1988) 1192.
[13] EKV Compact MOSFET Model, http://legwww.epfl.ch/ekv.
[14] C. E. Shannon, Bell System Technical J. **27** (1948) 379.
[15] A. Van der Ziel, Prentice-Hall (1970).
[16] A. J. Strecok, Mathematics of Computation **22** (1968) 144.
[17] www.crest.gatech.edu/palempbitscurrent.