Energy Aware Algorithm Design via Probabilistic Computing: From Algorithms and Models to Moore's Law and Novel (Semiconductor)Devices*

[Extended Abstract]

Krishna V. Palem
Center for Research in Embedded Systems and Technology
School of Electrical and Computer Engineering
Georgia Institute of Technology
777 Atlantic Drive
Atlanta, Georgia, USA.
palem@ece.gatech.edu
http://www.crest.gatech.edu

"Do not worry about your difficulties in Mathematics;
I can assure you that mine are still greater"
–Albert Einstein

1. SUMMARY

With their ever increasing proliferation, concerns of *power* (or *energy*) consumption have become significant in the context of the design and as well as the use of computing systems. While devices, computer architecture and the layers of software that reside and execute at higher levels of abstraction (such as operating systems, run-time, compilers and programming languages) all afford opportunities for being *energy-aware*, the most fundamental limits are truly rooted in the physics of energy consumption – specifically in *thermodynamics*. Based on this premise, this paper embodies the innovation of models of computing for energy-aware algorithm design and analysis, culminating in establishing, for the first time, the following central thesis of this work: *the computational technique referred to as randomization yielding probabilistic algorithms, now ubiquitous to the mathematical theory of algorithm design and analysis, when interpreted as a physical phenomenon through classical statistical thermodynamics with such pioneers as Maxwell, Boltzmann and Gibbs at the helm, leads to energy savings that are proportional to the probability* $\mathbf{p}$ *with which each primitive computational step is guaranteed to be correct (or equivalently to the probability of error of* $(1 - \mathbf{p})$*).*

*This work is supported in part by a DARPA Seedling Contract #F30602-02-2-0124.

CASES'03, Oct. 30–Nov. 2, 2003, San Jose, California, USA.

Historically, probabilistic algorithms were viewed as a mathematically very promising approach to algorithmic design and analysis elegantly stated by Schwartz [18]: "The startling success of the Rabin-Strassen-Solovay (see Rabin [17]) algorithm, together with the intriguing foundational possibility that axioms of randomness may constitute a useful fundamental source of mathematical truth independent of, but supplementary to, the standard axiomatic structure of mathematics (see Chaitin and Schwartz [1]), suggests that probabilistic algorithms ought to be sought vigorously." (For completeness we note that a fast probabilistic test for polynomial identities was also independently reported by DeMillo and Lipton [3].) Since this prediction, probabilistic algorithms have proliferated in a range of areas concerning the theoretical foundations of computer science. The present work departs from this tradition in a fundamental way by interpreting probabilistic computing as a physical, rather than as a mathematical process. Through this interpretation of probabilistic computing, energy savings can be achieved much as the purely mathematical formulations of probabilistic computing were shown historically to yield algorithmic improvements using conventional measures of computational complexity such as the running time.

Several novel constructs, techniques and insights are at the heart of establishing the central thesis of this work.*First*, idealized models for computing are introduced, and while on the one hand they capture the standard notions of describing algorithms, they also provide a basis for modeling energy consumption on the other, and are limited solely by thermodynamics. Thus, concerns based on the engineering realities of currently available microelectronics are abstracted away so as to be unfettered by legacy concerns. Specifically, the two novel models developed are the *ra*ndomized *b*it-level *r*andom *a*ccess *m*achine, or $RaBRAM$ developed by this author in [11] and [12], and the *probabilistic switch* as well as a *network* of switches, described by this author in [13]. These models are idealizations of practical notions of switching in so far as the process of computing is realized by dissipationless (or in classical thermodynamic terms, quasistatic) thermodynamics processes.

Second, based on these models, bounds on the energy advantages of probabilistic computing are proven quantitatively. Specifically, a single *deterministic* step–in a RaBRAM or in a switch–is shown to consume *at least* $(k.T. \ln(2))$ Joules no matter what techniques are used. Here, k is the well-known Boltzmann's constant, T is the *temperature* of the thermodynamic system, and ln is the natural logarithm. Furthermore, using probabilistic computing, each step (or switching) can be realized with as little as $(k.T. \ln(2\mathbf{p}))$ Joules of energy where $\mathbf{p}$ is no less than $\frac{1}{2}$. Thus, with the celebrated laws of thermodynamics guiding these determinations, we show that randomized computing offers the potential for energy savings of $k.T. \ln\left(\frac{1}{\mathbf{p}}\right)$ Joules.

The crux of this work is based on a fundamentally novel definition of *value*, as detailed by this author in [12] and [13]. The notion of a *value* denotes the mechanism through which a single "bit" of information (0 or 1) is represented in the physical instantiation of a computing device, and it is detected through an instantaneous "measurement" of the existence of a *microstate* from classical thermodynamics. An example of an instantaneous measurement is the detection of the position of a gas molecule in a cylinder containing it, as described by Feynman [4]; more generally, a measurement is the result of an experimental outcome, as defined in thermodynamics in the context of properties such as pressure. This novel definition of value constitutes a *third* contribution of this work.

Using these foundations (from [11], [12] , [13] and [14]), the author has, in collaboration with Cheemalavagu and Korkmaz [10], compared the energy savings attained using probabilistic methods with the traditional definition of value–the traditional definition of value uses the measurement of an average voltage at the output of an elementary switch, which is an inverter as defined by Stein [19] and by Meindl [8]. Surprisingly, we show that the conventional definition of value can support probabilistic computing, but however, and as shown in Figure 1 below, yields less gains in energy when compared to the novel definition of value (from [12] and [13]). To digress, an impressive contribution of Meindl's work, which continues the philosophical tradition set by Szilard [20], von-Neumann [22] and Landauer [7], among others, is the ability to model a switch in an idealized manner—without dissipation for example, much as we do—while, at the same time, reconciling the delicate and pragmatic balance needed to model the realities of modern semiconductor devices.

Our formulations of *energy complexity* characterizing the energy consumed in the context of the RaBRAM and a network of switches, respectively, are dependent only on thermodynamics and thus independent of specific technology parameters of a particular device implementation. These formulations constitute the *fourth* contribution of the work, and they parallel the conventional notions of time and space complexities, built on the foundational work of Rabin [16] and culminating in the work of Hartmanis and Stearns [5], that have become invaluable in the context of designing and analyzing algorithms; see Knuth's introduction for example [6].

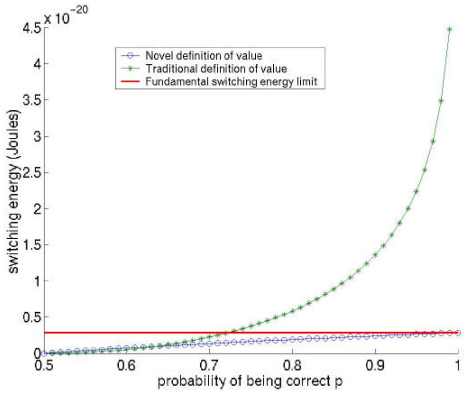

Figure 1: Comparing energy savings due to randomized switching based on conventional and novel definitions of value

A *fifth* contribution of this work is the demonstration of asymptotic energy savings in the RaBRAM model, in the context of the basic question of detecting whether a given vector of n elements which are either 0 or 1, contains any "0 elements"; this is the *distinct vector problem* (or DVP) [14] for which the following results are established. Using combinational techniques for bounding the energy complexities in the deterministic and probabilistic cases respectively from below and above, it is proved that the probabilistic approach to solving the DVP has asymptotically lower energy consumption than any deterministic variant in the RaBRAM model. Further it is proved that the resulting energy savings grow as $\Theta\left(n \log\left(\frac{n}{n-\epsilon \log(n)}\right)\right)$ for $0 < \epsilon < 1$. An interesting aspect of this result is that (as far as can be determined), it is the first asymptotic demonstration of energy savings derived from a probabilistic algorithm over any deterministic counterpart, wherein the complexity of the running time is identical $(\Theta(n))$ in both cases. Thus we clearly demonstrate that the energy savings are due to probabilistic "switching" as opposed to the (trivial) case of the savings being a by-product of an improvement to the running time since intuitively, a lower running time may imply lower energy consumption.

The talk will present a comprehensive overview of all of the above concepts and foundations, culminating in an analysis of their collective impact on the technology road map for semiconductors, as well as novel directions for microelectronics research to help realize the semiconductor devices based on these thermodynamic foundations. For a complete list of related references, comparisons with other work, and general remarks, the reader is referred to [12] [13]. Time permitting, recent efforts to validate the expected energy improvements derived through probabilistic computing and its relationship to noise in the context of a CMOS inverter gate will be described [2].

2. MOORE'S LAW, PROBABILISTIC COMPUTING AND RELATED REMARKS

Gordon Moore's prophetic prediction [9] about the exponential rate at which transistor feature sizes are supposed to shrink, popularly referred to as Moore's law, embodies a crucial factor at the heart of the unprecedented growth of the computer industry. However, as feature sizes continue to shrink at the current rate, noise, instability and other hurdles pose a threat which leads to the ongoing quest for "sustaining Moore's law", supported by significant (multi-billion dollar) investments. Probabilistic semiconductor devices—as envisioned above—offer several interesting alternatives to these hurdles and challenges. First, as shown in Figure 2, the probability parameter **p** is truly a "third" parameter or dimension essential to characterizing the space of behaviors represented by Moore's law.

First, as shown there (Figure 2), the same energy consumption can be achieved at an earlier point in time, albeit with a lower value of **p**. For example, according to this projection which in turn is a refinement of the well-known ITRS road map—this road map corresponds to the case where **p** = 1— with **p** = 0.6, the energy achieved using a device from the year 2001 will be achieved by a deterministic device or switch (with probability **p** = 1) in 2004. Thus, the implications of Moore's law to energy consumption can effectively be "accelerated" through probabilistic devices without having to incur the substantial investment needed to realizing the projections of the ITRS road map of Figure 2. This alternative of course is of interest to scenarios where the applications being considered are amenable to probabilistic computing approaches and algorithms.

Another interesting issue and one relevant to computing in general has to do with the value of the probabilistic switching methods and algorithms summarized here, to realizing general purpose processors and systems. This issue is best understood against the backdrop that it is infeasible to sustain Moore's law and preserve deterministic switching beyond a certain limit—thus, in this scenario, the only type of switching that is feasible below a certain feature size is probabilistic. Faced with this phenomenon, a challenge is to be able to use such probabilistic or error-prone devices and switches, and yet realize deterministic processors. This line of enquiry deviates from the issues of energy-aware computing into the very feasibility of computing deterministically in the future, at the rate of growth projected by Moore's law. To this end, a particularly attractive approach and one that this author is investigating collaboratively involves using error correcting schemes in conjunction with the error-prone probabilistic computing frameworks and devices described above—see von-Neumann [21] (also Pippenger [15]) for a pioneering study of this issue—as a potential path to sustaining Moore's law beyond all anticipated limits!

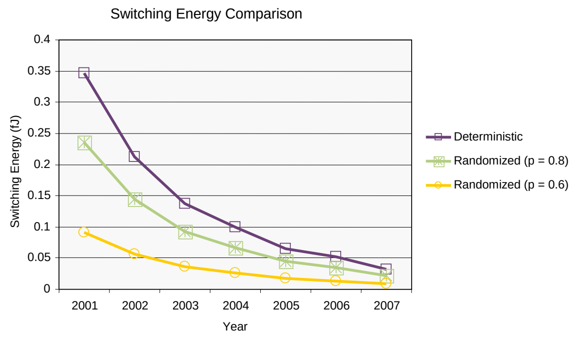

Figure 2: Accelerating and sustaining Moore's law through probabilistic computing

Acknowledgments

This work is supported in part by a DARPA Seedling Contract #F30602-02-2-0124. Jack Schwartz and Zeke Zalcstein provided invaluable criticism and comments. Jack Schwartz also pointed out the importance of clarifying the deeper issues underlying the ergodicity of the gaseous system, as well as the concerns caused by dissipation in the domain of actual switching. For all of this, I am very grateful, especially for their willingness to take a look at this work within the constraints of non-disclosure that prevailed until recently. Suresh Cheemalavagu has been of immense help, as a "sounding board" in ensuring the engineering relevance of this work, and in verifying the correctness of several of the subtler relationships between the physical domain and the mathematical formalisms. My students Lakshmi Chakrapani and Pinar Korkmaz were very valuable sources of support including editorial and typesetting issues; my sincere thanks to them. Special thanks to Dick Karp for pointing out the connection to von-Neumann's work on the synthesis of reliable organisms from unreliable components.

3. REFERENCES

[1] G. J. Chaitin and J. T. Schwartz. A note on monte carlo primality tests and algorithmic information theory. *Communications on Pure and Applied Mathematics*, 31:521–527, 1978.

[2] S. Cheemalavagu, P. Korkmaz, and K. Palem. Design and validation of an energy-aware probabilistic inverter in CMOS. *Unpublished manuscript*, 2003.

[3] R. A. DeMillo and R. J. Lipton. A probabilistic remark on algebraic program testing. *Information Processing Letters*, 7(4):193–195, 1978.

[4] R. Feynman. *Feynman Lectures on Computation.* Addison-Wesley Publishing Company Inc, 1996.

[5] J. Hartmanis and R. E. Stearns. On the computational complexity of algorithms. *Transactions of the American Mathematical Society*, 117, 1965.

[6] E. Knuth, Donald. *The Art of Computer Programming.* Addison Wesley Publishing Company, 1997.

[7] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Research and Development*, 3:183–191, July 1961.

[8] J. D. Meindl. Low power microelectronics: Retrospect and prospect. *Proceedings of IEEE*, pages 619–635, Apr. 1995.

[9] G. E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8), 1965.

[10] K. Palem, S. Cheemalavagu, and P. Korkmaz. Probabilistic switching and the Meindl limit: Interpretation in an electrical domain. *Unpublished manuscript*, 2003.

[11] K. V. Palem. Thermodynamics of randomized computing, a discipline for energy aware algorithm

design and analysis. Technical Report GIT-CC-02-56, Georgia Institute of Technology, Nov. 2002.

[12] K. V. Palem. Energy aware computation: From algorithms and thermodynamics to randomized (semiconductor) devices. Technical Report GIT-CC-03-10, Georgia Institute of Technology, Feb. 2003.

[13] K. V. Palem. Energy aware computing through randomized switching. Technical Report GIT-CC-03-16, Georgia Institute of Technology, May 2003.

[14] K. V. Palem. Proof as experiment: Probabilistic algorithms from a thermodynamic perspective. *Proceedings of The Intl. Symposium on Verification (Theory and Practice), Taormina, Sicily, Italy*, 2003.

[15] N. Pippenger. Developments in 'the synthesis of reliable organisms from unreliable components'. In *Proc. of Symposia in Pure Mathematics*, pages 311–324, 1990.

[16] M. O. Rabin. Degree of difficulty of computing a function and a partial ordering of recursive sets. Technical Report 2, Hebrew University, Israel, 1960.

[17] M. O. Rabin. Probabilistic algorithms. *Algorithms and Complexity, New Directions and Recent Trends (ed. J. F. Traub)*, pages 29–39, 1976.

[18] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J ACM*, 27:701–717, 1980.

[19] K.-U. Stein. Noise-induced error rate as limiting factor for energy per operation in digital ICs. *IEEE Journal of Solid-State Circuits*, SC-31(5), 1977.

[20] L. Szilard. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *in Maxwell's Demon: Why warmth disperses and time passes by Leff. H, Rex, E.*, 1998.

[21] J. von Neumann. *Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components*. Princeton Univ.Press, 1956.

[22] J. von Neumann. *in Fourth University of Illinois lecture in Theory of Self-Reproducing Automata (A. W. Burks editor)*. University of Illinois Press, 1966.