Energy Reduction and Fundamental Energy Limits in Digital VLSI Circuits

Paul P. Sotiriadis, Vahid Tarokh, and Anantha Chandrakasan
Dept. Electrical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
pps@mit.edu, vahid@deas.harvard.edu, anantha@mtl.mit.edu

I. Introduction

Many, widely used building *units* in digital VLSI circuits take an input vector $x(k) \in \{0,1\}^n$ and produce an output vector $y(k) = F(x(k)) \in \{0,1\}^m$ at every clock cycle $k = 0, 1, \ldots$. Examples of such units are, adders, multipliers etc. Associated to the operation F, there is a nonnegative cost $\mathcal{E}(x(k-1), x(k))$, that depends only on $x(k-1)$ and $x(k)$. There is also a communication unit, called *bus*, that simply reproduces the input vector at its output, i.e. $y = x$. We will use the bus as an example since its cost function is known [1]. *We address the problem of deriving the minimum cost (energy), necessary for a unit to process a bit of information. We also address the use of coding in order to achieve this limit.* Let $X : x(k)$, $k = 0, 1, \ldots$ be, a stationary random process in $\{0,1\}^n$, and input of a given unit with cost function $\mathcal{E}$.

Definition I.1 *The utilization, $a(X) \in [0,1]$, of the unit, by the process X is: $a(X) = \mathcal{H}(X)/n$.*

Definition I.2 *The energy cost per information bit corresponding to X is: $\mathcal{E}_b(X) = \overline{\mathcal{E}(x(k-1), x(k))}/\mathcal{H}(X)$ where the over-line stands for expectation over $(x(k-1), x(k))$.*

The function $\mathcal{E}_b(X)$ is, an induced by the unit, measure, of how efficiently the information is carried by the process X. It is important to relate $\mathcal{E}_b(X)$ to the corresponding utilization $a(X)$ of the unit by the process X. As shown in Figure 1 there are cases that the energy per bit can become arbitrary low for processes with arbitrary low entropy rate.

Definition I.3 *The minimum energy required, per information bit processed by the unit, at utilization a' of the unit, is:*

$$\mathcal{E}_b^*(a') = \inf_{a(X)=a'} \mathcal{E}_b(X)$$

where the infimum is taken over stationary processes.

II. Results

Theorem II.1 *[2] With the above setup and certain assumptions on the function $\mathcal{E}$ [2], for every $a \in [0,1]$ there exists a unique nonnegative number γ such that:*

$$\mathcal{E}_b^*(a) = \frac{\ln(2)}{\gamma - \frac{1}{\frac{\partial}{\partial\gamma} \ln(\ln(\mu(\gamma)))}} \tag{1}$$

$$a = -\frac{1}{n \ln(2)} \gamma^2 \frac{\partial}{\partial \gamma}\left(\frac{\ln(\mu(\gamma))}{\gamma}\right) \tag{2}$$

where $\mu(\gamma)$ is the maximal eigenvalue of the cost matrix

$$W(\gamma) = \left[e^{-\gamma\,\mathcal{E}(x,y)}\right]_{x,y\in\{0,1\}^n}.$$

Furthermore, the minimum is attained by the stationary ergodic Markov process with transition probabilities:

$$\Pr(y|x) = \frac{1}{\mu(\gamma)} \frac{g_y}{g_x} e^{-\gamma\,\mathcal{E}(x,y)}$$

where $g = (g_x)_x$ is the right eigenvector of matrix $W(\gamma)$ corresponding to $\mu(\gamma)$.

Introducing and using the notion of *energy typical sequences* we can prove the following theorem.

Theorem II.2 *[2] The minimum energy per information bit, given by (1), is asymptotically achievable using coding.*

Example: For *buses* the energy cost is given by $\mathcal{E}(x,y) = E_0\,(y-x)^T \mathcal{C}(y-x)$ where $\mathcal{C}$ is the symmetric tri-diagonal Toeplitz $(-\lambda, 1+2\lambda, -\lambda)$ matrix [1]. Figure 1 presents the ratio $\mathcal{E}_b^*(a)/\mathcal{E}_u$ as a function of the utilization a for $n = 8$ and $\lambda = 5$. $\mathcal{E}_u$ is the expected energy per bit when $x(k)$, $k = 0, 1, \ldots$ are i.i.d. and uniformly distributed random vectors.

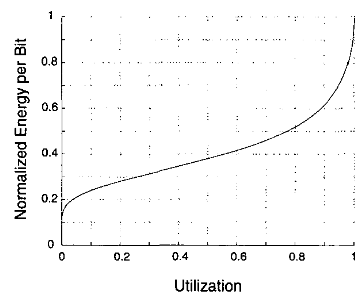

Figure 1: The function $\mathcal{E}_b^*(a)/\mathcal{E}_u$ for $n = 8$ and $\lambda = 5$.

Our results can be extended to units modelled as *Finite State Machines.* Other than energy costs objectives, like time for example, can be considered as well. There are some interesting relations between our results and Shannon's theory on discrete noiseless channels (DNC)s. The DNCs can be thought as the "asynchronous" versions of the our units.

References

[1] P. Sotiriadis, A. Chandrakasan, "A Bus Energy Model For Deep Sub-Micron Technology," *To appear in IEEE Trans. on VLSI.*

[2] P. Sotiriadis, V. Tarokh, A. Chandrakasan, "Energy Reduction in Deep Sub-micron Computation Modules An Information Theoretic Approach,", *Subm. to IEEE Trans. on Inf. Th.*

[1] This work was supported from the MARCO Focus Research Center on Interconnect funded at MIT through a subcontract from Gatech. The program is supported by MARCO and DARPA. Paul Sotiriadis is partially supported by the Alexander S. Onassis Public Benefit Foundation, the Greek Section of Scholarships and Research. Vahid Tarokh is supported by the National Science Foundation under the Alan T. Waterman Award, Grant No. CCR-0139398. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation.

[2] Like for example $\mathcal{E}(x,y) = 0 \Leftrightarrow x = y$

0-7803-7501-7/02/$17.00©2002 IEEE.