Energy Reduction Coding

Paul P. Sotiriadis
Department of Electrical and Computer Engineering
The Johns Hopkins University
Baltimore, MD 21218, USA
e-mail: pps@jhu.edu

Abstract — **In classical communication systems, information transmission is always associated with energy consumption. Coding, can be used in certain channels to reduce the amount of energy required per transmitted information bit. The basic concepts of energy reduction coding and their relations are presented in the framework of noiseless finite state channels. The optimal relation between energy cost and communication rate is derived and demonstrated in circuit applications.**

I. Introduction

Error correction coding has a long and well established history. Its tremendous importance and applicability results from the need to build reliable communication, storage and computation systems. Reliability reflects to the completion of operations with very small probability of error.

During the past few years another type of coding has emerged in the field of VLSI circuits, the *energy reduction coding*. It is motivated by the need to build circuits, and therefore complete systems, that operate with minimal energy (power) requirements. The objective of *energy reduction coding* is to reduce power consumption in contrast to that of the error correction coding.

Data compression can be thought as an energy reduction mechanism since reduction of the size of data usually results to reduced transmission time and therefore reduced total energy consumption. The essential difference between data compression and energy reduction coding is that in the later case *redundancy is introduced in the data*. Although this may sound contradictory, the example of the next section is convincing.

The purpose of the paper is to present the idea, the basic concepts and the latest results in the area of *energy reduction coding* and provide a mathematical framework for a more general study of the problem. The motivation of this research comes from real applications in VLSI circuits.

II. Motivating Example

Energy reduction coding was introduced in the design of low power inter-chip communication [3, 16, 17, 15] as a technic to reduce the instantaneous and time-average power consumption of inter-chip communication circuitry (*buses*). It was demonstrated that by introducing temporal or spacial redundancy in the transmitted data in a certain way (or equivalently by expanding the communication device) it is possible to reduce the power (energy) consumption.

An inter-chip *bus* is a basic communication device that transfers data between chips. In many cases it can be regarded as being error-free and its basic structure looks as in figure 1. It consists of the drivers, the lines (wires laid out in parallel) and receivers. Here for simplicity we consider a bus with 4 lines.

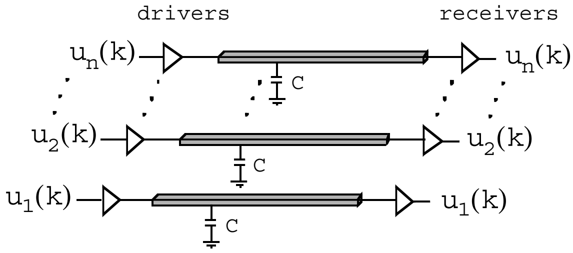

Figure 1: Simple inter-chip bus

A data vector $u(k) = [u_1(k), u_2(k), u_3(k), u_4(k)]$ of binary values is transmitted (from the left to the right) at every discrete time (clock cycle) $k = 0, 1, 2, \ldots$. The drivers on the left translate the binary values into voltages of the lines with the convention that 0 corresponds to 0 *Volts* and 1 corresponds to $V_{dd} > 0$ *Volts*. Every line has a (unavoidable) parasitic capacitance to ground equal to C *Farads*.

When there is a change of value, i.e. $u_i(k-1) = 0$ and $u_i(k) = 1$ or via versa, and only then, there is loss of energy E due to charging/discharging of the parasitic capacitor. It is: $E = CV_{dd}^2/2$ and so at time k there is total energy loss equal to $E(k) = \mathcal{W}(u(k-1) \oplus u(k))E$, where $\mathcal{W}$ is the Hamming weight function.

If $u_i(k)$, $i = 1, 2, 3, 4$, $k = 1, 2, \ldots$ are independent random variables uniformly distributed in $\{0, 1\}$, then the expected energy loss at time k is: $\mathcal{E}\{E(k)\} = 4E/2 = 2E$.

Suppose we modify our communication scheme (bus) by adding one more line, identical to the other ones, that carries a bit sequence $c(k)$. Also, instead of vector $u(k)$ we transmit another vector $x(k)$. The sequences x and c are defined as follows:

$$x(k) \doteq u(k) \ , \quad c(k) \doteq 0 \quad \text{if} \quad \mathcal{W}(\mathrm{u(k)} \oplus \mathrm{x(k-1)}) \leq 2$$

$$x(k) \doteq \overline{u(k)} \ , \quad c(k) \doteq 1 \quad \text{if} \quad \mathcal{W}(\mathrm{u(k)} \oplus \mathrm{x(k-1)}) > 2$$

Since sequence x is transmitted through the bus, the energy loss at time k, is:

$$E'(k) = min\{\mathcal{W}(u(k) \oplus x(k-1)), 4 - \mathcal{W}(u(k) \oplus x(k-1))\} + \\ +\mathcal{W}(c(k) \oplus c(k-1))E$$

If the random vectors $u(k), k = 1, 2, \ldots$ are independent and uniformly distributed then so are $u(k) \oplus x(k-1), k = 2, 3, \ldots$. Moreover, $c(k), k = 2, 3, \ldots$ are independent and uniformly distributed as well. Therefore:

$$\mathcal{E}\{E'(k)\} = \frac{1}{16}\left[\left(\begin{array}{c}4\\0\end{array}\right)0 + \left(\begin{array}{c}4\\1\end{array}\right)1 + \left(\begin{array}{c}4\\2\end{array}\right)2 +\right.$$

$$+\left(\begin{array}{c}4\\3\end{array}\right)\cdot(4-3+1)+\left(\begin{array}{c}4\\4\end{array}\right)\cdot(4-4+1)\Bigg] E$$

which gives $\mathcal{E}\{E'(k)\} = 25E/16$ that is less than $2E$, i.e. the expected energy loss when no coding was used! So, by expanding the bus and encoding the data we were able to reduce the energy consumption!

III. The Dual View of Redundancy

In the previous example we started with a 4-line bus, expanded it by adding one more line and then applied a coding scheme that allowed for the transmission of 4 bits every time k, in other words, the capacity of the expanded-and-encoded bus was equal to that of the original bus.

We can pose the energy reduction coding problem differently. Suppose that we are given the 5-line bus but we are *not* allowed to modify it. When the sequence of transmitted vectors is formed out of independent and uniformly distributed random vectors, the information rate through the bus is 5 (information) bits per clock cycle $k = 1, 2\ldots$ and the expected energy cost (also per clock cycle) is: $\mathcal{E}\{E(k)\} = 5E/2 = 2.5E$. This of course translates to *energy per bit* equal to $\mathcal{E}\{E(k)\}/5 = E/2$.

Now we can ask the question: *Is it possible to transmit information at a lower energy cost (per information bit), assuming we can accept a lower transmission rate?*

The example of the previous section provides a positive answer. If we transmit only 4 information bits each time and use the 5^{th} line of the bus to transmit a "coding bit" it is possible to achieve an energy cost per bit equal to $\mathcal{E}\{E'(k)\}/4 = 25E/16/4 = 25E/64$ which is less that the $E/2$ of the fully-used 4-line bus. The penalty of course is the rate reduction, the 5-line bus is *utilized* only by a factor of 4/5.

IV. General Mathematical Framework for Energy Reduction Coding

The simple example of Section II motivates us to seek for more general classes of communication channels where energy reduction coding is meaningful. Moreover, is it interesting to know what is the best possible trade off between energy reduction and utilization (percentage of the useful capacity) of the channel.

Definition IV.1 *The Noise-Less Finite State channel, (NLFS), is a device capable of transmitting arbitrary sequences of symbols in a finite set $\mathcal{S}$. The cost of transmission of a symbol $y \in \mathcal{S}$ is a real nonnegative number, $E(x, y)$, that depends on y and the symbol x that was transmitted exactly before y. The state of the channel is identical to the symbol that was transmitted last. The channel starts from an initial state x_0. (For technical reason we assume that there is some $x \in \mathcal{S}$ such that $E(x, x) \leq E(y, z)$ for every $y, z \in \mathcal{S}$. Without loss of generality we can also assume $E(x, x) = 0$[1].)*

Note that any symbol can follow any other symbol, therefore the channel is memoryless as far as state sequencing concerns. The channel has memory regarding the transmission cost of symbols. We refer to the quantity $E(x, y)$ as the *transition* cost, exactly because it is associated to the transition from state x to state y.

[1] If $E(x, x)$ is not zero, we can substract it from function E and added it to any estimate of the operating cost of the channel.

Example IV.1 *For example, the set of symbols (and states) of the following NLFS channel is $\mathcal{S} = \{1, 2, 3, 4\}$ and the transition costs are shown in figure 2. Assuming the channel starts from state 3, the cost of transmitting sequence 1, 4, 2, 1 is:* $1 + 4.8 + 2.5 + 1.5 = 9.8$.

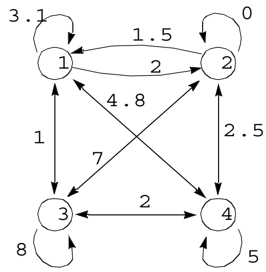

Figure 2: NLFS Channel

The capacity of an NLFS channel is $\mathcal{H} = log_2(|\mathcal{S}|)$ bits of information per use the channel. This information rate is achieved only when the channel transmits the outcomes of a source emitting independent symbols uniformly distributed in $\mathcal{S}$. In this case the expected cost *per channel use* is:

$$\mathcal{E} = \frac{1}{|\mathcal{S}|^2} \sum_{x,y\in\mathcal{S}} E(x, y) \tag{1}$$

The corresponding *cost per bit*, of information transmitted, is $\mathcal{E}_b^H = \mathcal{E}/\mathcal{H}$. In the previous example it was: $\mathcal{H} = 2$, $\mathcal{E} = 3.39$ and $\mathcal{E}_b^H = 1.69$.

V. Entropy and Cost; Definitions

A code $\mathcal{C}_P^L$ of length L and associated probability distribution P, is a subset of $\mathcal{S}^L$ with $\sum_{c\in\mathcal{C}_P^L} P(c) = 1$. For convenience we consider P as a function on $\mathcal{S}^L$ with $P(c) = 0$ for every $c \in \mathcal{S}^L - \mathcal{C}_P^L$.

Definition V.1 *The entropy per channel use of the code is:*

$$\mathcal{H}(\mathcal{C}_P^L) = -\frac{1}{L} \sum_{c\in\mathcal{C}_P^L} P(c) \log_2 P(c) \tag{2}$$

The cost per channel use of the code is:

$$\mathcal{E}(\mathcal{C}_P^L) = \frac{1}{L}\left[\sum_{c\in\mathcal{C}_P^L} P(c) \sum_{i=1}^{L-1} E(\pi_i(c), \pi_{i+1}(c)) + \right.$$
$$\left. + \sum_{c,d\in\mathcal{C}_P^L} P(c)P(d)E(\pi_n(c), \pi_1(d))\right] \tag{3}$$

where $\pi_i(c)$ is the i^{th} entry of vector c [2]*. The expected cost, per bit of information transmitted, is:*

$$\mathcal{E}_b(\mathcal{C}_P^L) = \mathcal{E}(\mathcal{C}_P^L)/\mathcal{H}(\mathcal{C}_P^L). \tag{4}$$

The utilization factor of the code is the ratio of the entropy per channel use of the code over the capacity $\mathcal{H}$ of the channel.

$$\alpha(\mathcal{C}_P^L) = \mathcal{H}(\mathcal{C}_P^L)/\mathcal{H}. \tag{5}$$

[2] Note that the first sum expresses the cost of the codewords whereas the second sum expresses the cost of transition between consecutive codewords.

Example V.1 *Consider the NLFS channel of figure 2 and the code* $\mathcal{C}_U^2$ *with codewords* $\{(2,2),(1,2),(2,1),(1,3)\}$ *and uniform associated probability distribution* U*. Then,* $\mathcal{H}(\mathcal{C}_U^2) = 2$*,* $\mathcal{E}(\mathcal{C}_U^2) = 3.14$*,* $\mathcal{E}_b(\mathcal{C}_U^2) = 1.57$ *and* $\alpha(\mathcal{C}_U^2) = 0.5$*. Note that the energy per bit of the code is less than that of the uncoded channel (*1.57 *vs.* 1.69*). The utilization of the code though is only* $1/2$*.*

In Section VIII, analytical expressions are provided for the theoretically minimum possible energy per bit, using coding, as a function of the entropy per channel use.

VI. Properties of Energy Reduction Codes; Concatenation and Probability Mixing

Definition VI.1 *Let* $\mathcal{C}_P^L$ *and* $\mathcal{C}_Q^M$ *be two codes of lengths* L, M *and with associated probability distributions* P, Q *respectively. We define their concatenation,* $\mathcal{C}_P^L \odot \mathcal{C}_Q^M$*, to be the code of length* $L+M$ *that has the codewords* $(z,w) : z \in \mathcal{C}_P^L, w \in \mathcal{C}_Q^M$ *and associated probability distribution,* $P \odot Q$ *such that:* $P \odot Q(x_1,\ldots,x_{L+M}) = P(x_1,\ldots,x_L)Q(x_{L+1},\ldots,x_M)$ *for every* $x_1,x_2,\ldots,x_{L+M} \in \mathcal{S}$*. We also use the "power" notation,* $\left(\mathcal{C}_P^L\right)^{\odot k}$ *for the code,*

$$\underbrace{\mathcal{C}_P^L \odot \mathcal{C}_P^L \odot \cdots \odot \mathcal{C}_P^L}_{k-times}$$

Same basic yet useful properties of concatenated codes are summarized in the next lemma.

Lemma VI.1 *For every two codes* $\mathcal{C}_P^L$ *and* $\mathcal{C}_Q^M$ *of lengths* L, M *and distributions* P, Q *respectively, we have that:*

$$\mathcal{H}\left(\left(\mathcal{C}_P^L\right)^{\odot k}\right) = \mathcal{H}(\mathcal{C}_P^L) \quad (6)$$

$$\mathcal{E}\left(\left(\mathcal{C}_P^L\right)^{\odot k}\right) = \mathcal{E}(\mathcal{C}_P^L) \quad (7)$$

$$\mathcal{E}_b\left(\left(\mathcal{C}_P^L\right)^{\odot k}\right) = \mathcal{E}_b(\mathcal{C}_P^L) \quad (8)$$

$$\mathcal{H}(\mathcal{C}_P^L \odot \mathcal{C}_Q^M) = \frac{L}{L+M}\mathcal{H}(\mathcal{C}_P^L) + \frac{M}{L+M}\mathcal{H}(\mathcal{C}_Q^M) \quad (9)$$

$$\mathcal{E}(\mathcal{C}_P^L \odot \mathcal{C}_Q^M) = \frac{L}{L+M}\mathcal{E}(\mathcal{C}_P^L) + \frac{M}{L+M}\mathcal{E}(\mathcal{C}_Q^M) + \frac{2e}{L+M} \quad (10)$$

$$\mathcal{H}\left(\left(\mathcal{C}_P^L\right)^{\odot k} \odot \left(\mathcal{C}_Q^M\right)^{\odot r}\right) = \frac{kL}{kL+rM}\mathcal{H}(\mathcal{C}_P^L) + \frac{rM}{kL+rM}\mathcal{H}(\mathcal{C}_Q^M) \quad (11)$$

$$\mathcal{E}\left(\left(\mathcal{C}_P^L\right)^{\odot k} \odot \left(\mathcal{C}_Q^M\right)^{\odot r}\right) = \frac{kL}{kL+rM}\mathcal{E}(\mathcal{C}_P^L) + \frac{rM}{kL+rM}\mathcal{E}(\mathcal{C}_Q^M) + \frac{2e}{kL+rM} \quad (12)$$

where $|e| \le E_{max}$ *and* $E_{max} = \max_{x,y\in\mathcal{S}} E(x,y)$*.*

Another operation between two codes of the same length is *probability mixing.*

Definition VI.2 *Consider two codes* $\mathcal{C}_P^L$*,* $\mathcal{C}_Q^L$ *of the same length* L *and a constant* $\alpha \in (0,1)$*. We define the new code*

$$\alpha\mathcal{C}_P^L \oplus (1-\alpha)\mathcal{C}_Q^L = \mathcal{C}_{\alpha P+(1-\alpha)Q}^L.$$

The following lemma states some useful facts of the probability mixing operation on codes. The proof follows directly from the concavity of entropy function and the linearity of the cost per channel use.

Lemma VI.2 *For every two codes* $\mathcal{C}_P^L$ *and* $\mathcal{C}_Q^L$ *of the same length we have:*

$$\mathcal{H}\left(\alpha\mathcal{C}_P^L \oplus (1-\alpha)\mathcal{C}_Q^L\right) \ge \alpha\mathcal{H}(\mathcal{C}_P^L) + (1-\alpha)\mathcal{H}(\mathcal{C}_Q^L) \quad (13)$$

and

$$\mathcal{E}\left(\alpha\mathcal{C}_P^L \oplus (1-\alpha)\mathcal{C}_Q^L\right) = \alpha\mathcal{E}(\mathcal{C}_P^L) + (1-\alpha)\mathcal{E}(\mathcal{C}_Q^L). \quad (14)$$

For every given length L there are two codes that are useful in achieving technical results. The first one, $\mathcal{C}_0^L$, contains only one element, a vector $(x,x,\ldots,x) \in \mathcal{S}^L$ for which $E(x,x) = min_{y,z}E(y,z)$[3]. It is of course,

$$\mathcal{H}(\mathcal{C}_0^L) = 0,$$

and *if we follow the assumption of definition IV.1*, i.e. $E(x,x) = 0$, we also have that,

$$\mathcal{E}(\mathcal{C}_0^L) = 0.$$

The properties above can be combined with (13) and (14) of lemma VI.2 to give the following result.

Lemma VI.3 *Let* $\mathcal{C}_P^L$ *be a code of entropy per channel use* H *and cost per bit* E_b*. For every* H'*,* $0 < H' < H$ *there is a code* $\mathcal{C}$ *of length* L *such that,* $\mathcal{H}(\mathcal{C}) = H'$ *and* $\mathcal{E}_b(\mathcal{C}) \le E_b$*.*

Proof: The function $f : [0,1] \to \Re$ such that

$$f(\alpha) = \mathcal{H}\left(\alpha\mathcal{C}_P^L \oplus (1-\alpha)\mathcal{C}_0^L\right)$$

is continuous. Moreover $f(0) = 0$ and $f(1) = H$, and so there is some $\alpha' \in (0,1)$ for which $f(\alpha') = H'$. In addition,

$$\begin{aligned}
\mathcal{E}_b\left(\alpha'\mathcal{C}_P^L \oplus (1-\alpha')\mathcal{C}_0^L\right) &= \frac{\mathcal{E}\left(\alpha'\mathcal{C}_P^L \oplus (1-\alpha')\mathcal{C}_0^L\right)}{\mathcal{H}\left(\alpha'\mathcal{C}_P^L \oplus (1-\alpha')\mathcal{C}_0^L\right)} \\
&\le \frac{\alpha'\mathcal{E}(\mathcal{C}_P^L) + (1-\alpha')\mathcal{E}(\mathcal{C}_0^L)}{\alpha'\mathcal{H}(\mathcal{C}_P^L) + (1-\alpha')\mathcal{H}(\mathcal{C}_0^L)} \\
&= E_b.
\end{aligned}$$

So $\mathcal{C} = \alpha'\mathcal{C}_P^L \oplus (1-\alpha')\mathcal{C}_0^L$ is an appropriate code. $\square$

The second code, often involved in the proofs, $\mathcal{C}_U^L$, contains all vectors in $\mathcal{S}^L$ and is associated with the uniform probability distribution U. Directly from the definitions we have that,

$$\mathcal{H}(\mathcal{C}_U^L) = \mathcal{H}$$

and

$$\mathcal{E}(\mathcal{C}_U^L) = \mathcal{E}$$

where $\mathcal{H} = \log_2(|\mathcal{S}|)$ and $\mathcal{E}$ is given by (1). We have the next lemma whose proof is similar to that of lemma VI.3.

[3] Note that there is at least one such element x in $\mathcal{S}$.

Lemma VI.4 *Let $\mathcal{C}_P^L$ be a code of entropy per channel use H, with $H < \mathcal{H}$, and cost per bit E_b. Consider the family of codes $\mathcal{C}(\alpha) = \alpha\mathcal{C}_U^L \oplus (1-\alpha)\mathcal{C}_P^L$ where $\alpha \in [0,1]$. The entropy per channel use, $\mathcal{H}(\mathcal{C}(\alpha))$, is a continuous and strictly increasing function of α, ranging from $\mathcal{H}(\mathcal{C}(0)) = H$ to $\mathcal{H}(\mathcal{C}(1)) = \mathcal{H}$. The cost per bit is a continuous function of α with $\mathcal{E}_b(\mathcal{C}(0)) = E_b$ and $\mathcal{E}_b(\mathcal{C}(1)) = \mathcal{E}$.*

VII. Optimal Cost VS. Rate Relation; The Limiting Cost per Bit Function and Its Properties

The desirable in using a code is to transmit information at a cost per bit lower than that of the uncoded channel and do so at an acceptable rate (transmitted information bits per channel use). This is expressed by the following definition.

Definition VII.1 *A pair (H, E_b) of entropy per channel use, cost per bit, is called achievable (using coding) if there is a sequence of codes $\mathcal{C}(k)$, $k = 1, 2, \ldots$ (not necessarily distinct, nor of the same length) such that for some $H' \geq H$ and some $E_b' \leq E_b$ it is $\lim_{k\to\infty} \mathcal{H}(\mathcal{C}(k)) = H'$ and $\lim_{k\to\infty} \mathcal{E}_b(\mathcal{C}(k)) = E_b'$.*

It is important to know what is the "best" possible trade off between the information rate through the channel and the energy cost per bit.

Definition VII.2 *The limiting cost per bit is the function of the entropy per channel use: $\mathcal{E}_b^\ell : (0, \mathcal{H}] \to [0, \infty)$ such that, $\mathcal{E}_b^\ell(H) = \inf\{E_b : (H, E_b) \text{ is achievable}\}$.*

Note that the set $\{E_b : (H, E_b) \text{ is achievable}\}$ is nonempty for every $H \in (0, \mathcal{H}]$. So the limiting cost per bit function is well defined.

Lemma VII.1 *The limiting cost per bit function, $\mathcal{E}_b^\ell$, is increasing and continuous. For every $H \in (0, 1]$ there is sequence of codes $\mathcal{C}(k)$, $k = 1, 2, \ldots$ such that $\lim_{k\to\infty} \mathcal{H}(\mathcal{C}(k)) = H$ and $\lim_{k\to\infty} \mathcal{E}_b(\mathcal{C}(k)) = \mathcal{E}_b(H)$. The function $f(H) = H\mathcal{E}_b^\ell(H)$ is convex.*

Proof: If a pair (H, E_b) is achievable then, for every $H' < H$, the pair (H', E_b) is achievable as well. Therefore $\mathcal{E}_b^\ell$ is increasing.

Suppose the function is discontinuous from the left at H. Then there is some $\epsilon > 0$ and a sequence H_k, $k = 1, 2, \ldots$ converging to H such that $H_k < H$, and $\mathcal{E}_b^\ell(H_k) < \mathcal{E}_b^\ell(H) - \epsilon$ for every k (recall that $\mathcal{E}_b^\ell$ is increasing). The sequence $\mathcal{E}_b^\ell(H_k)$ is bounded from below as well, therefore it must have a converging subsequence. Without loss of generality we assume that $\lim_{k\to\infty} \mathcal{E}_b^\ell(H_k) = \xi$, $\xi \leq \mathcal{E}_b^\ell(H) - \epsilon$. By the definition of the limiting cost per bit function, for every k we can find a code $\mathcal{C}(k)$ such that $\mathcal{H}(\mathcal{C}(k)) \geq H_k - \frac{1}{k}$ and $\mathcal{E}_b(\mathcal{C}(k)) \leq \mathcal{E}_b^\ell(H_k) + \frac{1}{k}$. Therefore, there is a subsequence of the codes that achieves $(H, E_b - \epsilon)$. This is a contradiction.

The continuity from the right is an immediate consequence of lemma VI.4 and the fact that $\mathcal{E}_b^\ell$ is an increasing function.

By the definitions of the achievable pairs and the limiting cost per bit as well as the fact that $0 \leq \mathcal{H}(\mathcal{C}) \leq \mathcal{H}$, for every code $\mathcal{C}$, we can conclude (using Heine-Borel theorem) that for every $H \in [0, 1]$, there is a sequence of (not necessarily distinct) codes $\mathcal{C}(k)$, $k = 1, 2, \ldots$ such that, $\lim_{k\to\infty} \mathcal{H}(\mathcal{C}(k)) = H'$ and $\lim_{k\to\infty} \mathcal{E}_b(\mathcal{C}(k)) = \mathcal{E}_b^\ell(H)$ for some $H' \geq H$. Using the result of lemma VI.3 we conclude the existence of another sequence of codes, $\widehat{\mathcal{C}}(k)$, $k = 1, 2, \ldots$ such that $\lim_{k\to\infty} \mathcal{H}(\widehat{\mathcal{C}}(k)) = H$ and $\lim_{k\to\infty} \mathcal{E}_b(\widehat{\mathcal{C}}(k)) = \mathcal{E}_b^\ell(H)$. This proves the third part of the lemma.

Now, consider two codes, $\mathcal{C}_P^L$ and $\mathcal{C}_Q^M$ and set $H_1 = \mathcal{H}(\mathcal{C}_P^L)$, $E_{b1} = \mathcal{E}_b(\mathcal{C}_P^L)$, $H_2 = \mathcal{H}(\mathcal{C}_Q^M)$ and $E_{b2} = \mathcal{E}_b(\mathcal{C}_Q^M)$. For any given $\alpha \in (0, 1)$ we can find a pair of strictly increasing sequences of positive integers, k_n, r_n, $n = 1, 2, \ldots$ such that, $k_n L/(k_n L + r_n M)$ approaches $1 - \alpha$ and $r_n M/(k_n L + r_n M)$ approaches α. For convenience we set $1 - \alpha = \widetilde{\alpha}$. Identities (11) and (12) give,

$$\lim_{n\to\infty} \mathcal{H}\left(\left(\mathcal{C}_P^L\right)^{\odot k_n} \odot \left(\mathcal{C}_Q^M\right)^{\odot r_n}\right) = \widetilde{\alpha}H_1 + \alpha H_2$$

and

$$\lim_{n\to\infty} \mathcal{E}\left(\left(\mathcal{C}_P^L\right)^{\odot k_n} \odot \left(\mathcal{C}_Q^M\right)^{\odot r_n}\right) = \widetilde{\alpha}H_1 E_{b1} + \alpha H_2 E_{b2}$$

Therefore, the pair

$$\left(\widetilde{\alpha}H_1 + \alpha H_2, \frac{\widetilde{\alpha}H_1 E_{b1} + \alpha H_2 E_{b2}}{\widetilde{\alpha}H_1 + \alpha H_2}\right)$$

is achievable and by the definition of the limiting cost per bit we have:

$$(\widetilde{\alpha}H_1 + \alpha H_2)\mathcal{E}_b^\ell(\widetilde{\alpha}H_1 + \alpha H_2) \leq \widetilde{\alpha}H_1 E_{b1} + \alpha H_2 E_{b2}.$$

Given any H^1, H^2 in $[0, 1]$ there exist two sequences of codes $\mathcal{C}_1(k)$ and $\mathcal{C}_2(k)$, $k = 1, 2, \ldots$ such that $\lim_{k\to\infty} \mathcal{H}(\mathcal{C}_1(k)) = H^1$ and $\lim_{k\to\infty} \mathcal{E}_b(\mathcal{C}_1(k)) = \mathcal{E}_b^\ell(H^1)$ as well as $\lim_{k\to\infty} \mathcal{H}(\mathcal{C}_2(k)) = H^2$ and $\lim_{k\to\infty} \mathcal{E}_b(\mathcal{C}_2(k)) = \mathcal{E}_b^\ell(H^2)$. The existence of the sequences is implied by the part of the lemma that was just proved. Using the last inequality we have that for every k it is:

$$\begin{aligned}[\widetilde{\alpha}\mathcal{H}(\mathcal{C}_1(k)) + \alpha\mathcal{H}(\mathcal{C}_2(k))]\,\mathcal{E}_b^\ell(\widetilde{\alpha}\mathcal{H}(\mathcal{C}_1(k)) + \alpha\mathcal{H}(\mathcal{C}_2(k))) \leq \\ \leq \widetilde{\alpha}\mathcal{H}(\mathcal{C}_1(k))\mathcal{E}_b(\mathcal{C}_1(k)) + \alpha\mathcal{H}(\mathcal{C}_2(k))\mathcal{E}_b(\mathcal{C}_2(k)).\end{aligned}$$

Because $\mathcal{E}_b$ is continuous, taking the limit of the above expression we obtain

$$(\widetilde{\alpha}H^1 + \alpha H^2)\mathcal{E}_b^\ell(\widetilde{\alpha}H^1 + \alpha H^2) \leq \widetilde{\alpha}H^1\mathcal{E}_b^\ell(H^1) + \alpha H^2\mathcal{E}_b^\ell(H^2).$$

which concludes the proof. □

VIII. Limiting Cost per Bit Analytical Expressions

The limiting cost per bit function, $\mathcal{E}_b^\ell$, can be derived analytically and the result is summarized in the following theorem. The proof is long and tedious and can be found in two parts, in [8] or [9], and [10]. An alternative proof can be found in [11].

Theorem VIII.1 *The limiting cost per bit at entropy per channel use H is*[4],

$$\mathcal{E}_b(H) = \ln(2) \cdot \left(\gamma - \frac{1}{\frac{\partial}{\partial\gamma} \ln\left(\ln\left(\mu(\gamma)\right)\right)}\right)^{-1} \tag{15}$$

[4]There is one exception: if there are some ζ, θ_x, $x \in \mathcal{S}$, such that $E(x, y) = \zeta + \theta_x - \theta_y$, for every x, y, then it is $\mathcal{E}_b^\ell(H) = \zeta/H$.

where γ is the unique positive solution of the equation:

$$H = -\frac{1}{\ln(2)}\,\gamma^2\,\frac{\partial}{\partial\gamma}\left(\frac{\ln\left(\mu(\gamma)\right)}{\gamma}\right) \tag{16}$$

and $\mu(\gamma)$ is the maximal eigenvalue of the matrix:

$$W(\gamma) = \left[e^{-\gamma\, E(x,y)}\right]_{x,y=0}^{2^n-1} \tag{17}$$

Furthermore, the minimum is attained asymptotically by a sequence of codes $\mathcal{C}(k)$, $k = 1, 2, \ldots$ such that $\mathcal{C}(k)$ contains all vectors c in $\mathcal{S}^k$ and is associated with the probability distribution P_k such that $P_k(c)$ equals the probability of c being the outcome of the stationary, ergodic, Markov process having transition probabilities:

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

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

IX. Application: On-Chip Buses as NLFS channels

On-chip buses are communication channels within microprocessors that consist of parallel electrical lines, as in figure 1 but with more complex parasitic elements and more involved expressions of energy consumption [13]. In most applications they can be regarded error-free and can be treated as NLFS channels.

On-chip communication through buses consumes a significant part of the total operating power and many techniques have been proposed to reduce it, e.g. [1, 15, 12, 7, 6, 5, 4]. A question stated and answered in [8] was: *How much power (energy) reduction is possible using coding?* Using the energy model in [13] and theorem VIII.1 we get the graph of normalized $\mathcal{E}_b$ for a family of buses with $2, 4$ and 8 lines and coupling ratio $\lambda = 5$ (see [13]).

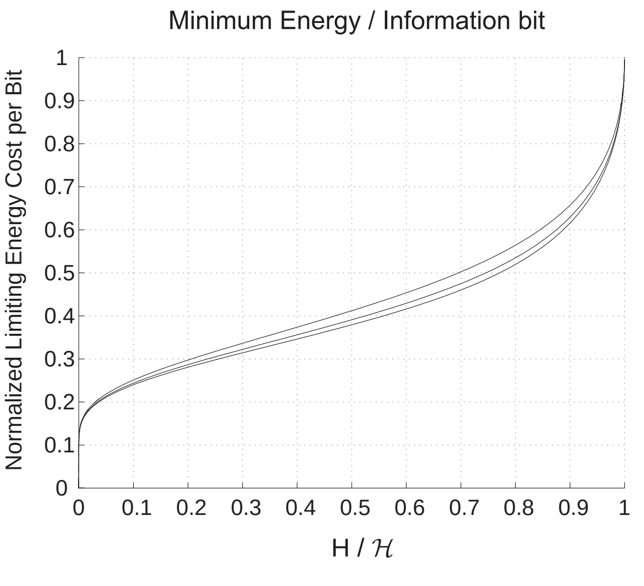

Figure 3: Best achievable energy vs. rate performance in a class of on-chip buses

X. Conclusions

The concept of energy reduction coding has been discussed and a general mathematical framework has been introduced to study it. The analysis has been based on the concepts of noiseless finite state channel, entropy per channel use, cost per channel use, cost per channel bit and channel utilization. The performance limits of energy coding have been derived.

References

[1] L. Benini, G. De Micheli, E. Macii, D. Sciuto, and C. Silvano, “Asymptotic zero-transition activity encoding for address busses in low-power microprocessor-based systems,” *Proc. of Great Lakes VLSI Symp.*, pp. 77-82, 1997.

[2] T.M. Cover and J.A. Thomas, “*Elements of Information Theory,*” John Wiley and Sons, 1991.

[3] R.J. Fletcher, “*Integrated circuit having outputs configured for reduced state changes,* U.S. Patent 4,667,337, May 1987.

[4] K.W. Kim, K.H. Baek, N. Shanbhag, C.L. Liu, S.M. Kang, “Coupling-Driven Signal Encoding Scheme for Low-Power Interface Design,” *Proceedings of ICCAD*, Nov. 2000.

[5] Y. Shin, T. Sakurai, “Coupling-Driven Bus Design for Low-Power Application-Specific Systems”, *Design Automation Conference*, 2001.

[6] P. Sotiriadis and A. Chandrakasan, “Low Power Bus Coding Techniques Considering Inter-wire Capacitances,” *Proceedings of Custom Integrated Circuits Conf.*, May 2000.

[7] P. Sotiriadis, A. Wang, A. Chandrakasan, “Transition Pattern Coding: An approach to reduce Energy in Interconnect”, *Proceedings of ESSCIRC,* Stockholm, Sweden, Sept. 2000.

[8] P. Sotiriadis, *Interconnect Modeling and Optimization in Deep Sub-Micron Technologies*, Ph.D. Thesis, Massachusetts Institute of Technology, May 2002.

[9] P. Sotiriadis, V. Tarokh, A. Chandrakasan, “Energy Reduction in VLSI Computation Modules: An Information Theoretic Approach,” *To appear in the IEEE Trans. on Information Theory* on April 2003.

[10] P. Sotiriadis, A. Chandrakasan, “Power Estimation and Power Optimal Communication in Deep Sub-micron Busses: Analytical Models and Statistical Measures,” *To appear in the Journal of Circuits, Systems, and Computers* on Feb. 2003.

[11] P. Sotiriadis, “Energy Reduction Coding and its Applications,” *submitted for publication*.

[12] P. Sotiriadis, A. Chandrakasan, “Bus Energy Minimization by Transition Pattern Coding (TPC) in Deep Sub-Micron Technologies,” *Proc. of IEEE/ACM Int. Conf. on CAD*, pp. 322-327, San Jose, Nov. 2000.

[13] P. Sotiriadis and A. Chandrakasan, “A bus energy model for deep submicron technology,” *IEEE Trans. VLSI Systems*, pp. 341-350, Vol. 10, No. 3, June 2002.

[14] S.K. Srinivasan and K.M. Mehta, “*Stochastic Processes*,” Tata McGraw-Hill Publishing Company Limited, New Delhi 1988.

[15] M.R. Stan and W.P. Burleson, “Bus-invert coding for low-power I/0,” *IEEE Trans. on VLSI*, pp. 49-58, Vol. 3, No. 5, March 1995.

[16] J. Tabor, “Noise reduction using low weight and constant weight coding techniques,” Master’s thesis, EECS dept. MIT, May 1990.

[17] R. Wilson, “Low power and paradox,” *Electronic Engineering Times*, pp. 38, Nov. 1, 1993.