Energy Consumption Management in Design

J.Smit
University of Twente, Department of Electrical Engineering
Network Theory and VLSI-Design
PO Box 217, 7500 AE Enschede, The Netherlands
tel. +31 53 489 2838
J.Smit@el.utwente.nl

Abstract— **A survey of the basic issues in low power design is presented, including techniques for the analysis of energy consumption in the early design phase of analog and digital circuits. The concept of energy complexity will be introduced in conjunction with techniques for parameterized energy management. The technique of energy consumption management will be applied to some analog and digital designs, to shows how the designer can identify design bottlenecks and develop design alternatives, before starting any tedious design steps.**

I. Introduction

Speed and functionality were the topics of main interest for the VLSI designer in recent years, however it is now clear to a group of designers of increasing size, that design for low-power will be the most challenging issue of the sub-micron era. The low power issue has become important due to the fact that extreme speed levels can only be reached for given (RISC-) architectures with absurd consequences for the energy consumption or a by a complete rethinking of the overall concept [1,2]. The compactness of sub-micron circuits, offers new opportunities for designs of unprecedented complexity. The nature of the energy consumption on a chip is however such that the designer is faced with a rapidly decreasing freedom as far as placement is concerned.

A thorough understanding of the concept of energy management should not be restricted to the traditional aspects of VLSI design, like scheduling and allocation, logic synthesis and placement. Instead it may be used to find new alternatives for complete systems of a mixed analog and digital nature [3]. Some of the consequences of the concept of energy complexity have already been settled without being made explicit. The new digital TV standard which replaces the proposed HD-MAC standard is a painful example.

Other application area where energy consumption is important is mobile communication, i.e. GSM & DECT [3] and Digital Audio Broadcast [4]. Energy management techniques can be used to study the feasibility and cost in terms of energy consumption of fully integrated IF filters/amplifiers in such products which are partially analog and partially digital expressed in protocol parameters, like bandwidth, dynamic range and rejection of adjacent channels [3]. Such an analysis should however not only be used to optimize the chip at hand, it is important as well to give feedback to those who propose a (telecom) standard, such that adaptations to the protocols may be made in favor of the price of all the products which conform to the standard.

The early evaluation of a standard with respect to the implications on the cost of the VLSI chip, is nowadays not performed, due to the assumption that the calculation of the energy complexity is an ill posed problem given the complexity of the tools at hand. It should be noted however that it is sufficient to calculate a lower bound within a fair margin. This task is quite well feasible. The omission to evaluate the implications of a proposed standard, degrades the system design aspect, to a discipline of trial and error. Energy complexity management is a technique which can be used to derive design and specification alternatives, using an energy-based lower bound to make design decisions at high levels of abstraction.

II. Energy complexity

The idea of energy complexity is to define a realistic lower-bound for high level fundamental operations, like arithmetic, communication, data-storage, IO in

TABLE I
FUNDAMENTAL OPERATIONS

Digital design:	Describes:	Point of reference
□ Arithmetic	Logic	→ $\mathcal{E}_{fa}$ Of a best full-adder design
□ Communication	Wiring	→ Switching-energy / meter
□ Data-storage	Memories	→ $\mathcal{D}_{cell}$ Diameter RAM/ROM-cell
□ IO-energy	IO-PAD	→ IO standards
▲ Latch	Temporary variable	→ Best design
▲ D-flip-flop	State variables	→ Best design
Analog design:	**Describes:**	**Point of reference:**
♦ Resistor	Coefficients	dissipates energy
♦ Capacitor	State	conserves energy
♦ Communication	Wiring	conserves energy
□ Filtering	Selectivity	→ Signal to Noise ratio, Degree, Q
□ Mixer	Non-linear operation	→ Signal to Noise ratio, distortion
□ Quantization	Comparator	→ Random-Signal to Noise ratio
▲ Analog-Latch	Sample & Hold	→ Signal to Noise ratio
▲ Analog-D-ff	Analog-State	→ Signal to Noise ratio
▲ Amplifier	Signal-gain	→ Signal to Noise on the input
▲ Power-Amp.	Power-gain	→ Power to be delivered, Amplifier Class

the digital domain, and filtering, mixing, quantization, etc. in the analog domain, for a given CMOS process, based on calibrated energy consumption figures from primitive operations. Examples of this methodology can be found in [5,6,7].

A. Calibration

Note that the point of reference is not constant, for instance $\mathcal{E}_{fa}$ and $\mathcal{E}_{/m}$ depend on the supply voltage V_{DD} and $\mathcal{E}_{fa}, \mathcal{E}_{/m}$ and $\mathcal{D}_{cell}$ depend on the dimension λ of the process. The energy consumption of the primitive operations need a proper calibration procedure. For a given process one takes the best design which consumes the smallest average amount of energy taken by a full-adder $\mathcal{E}_{fa}$ as the reference value for bit-addition, from which the cost of high level operations like multiplication can be derived [5,6]. The situation is slightly different for IO-Pads [1], as the standardization on a given V_{DD} and the desired capability to drive a large capacitive load are the important issues, dictated by the IO-standard at hand
The energy consumption data for a digital latch and for a digital D-flip-flop are derived in a way similar to those for a full-adder.

B. Noise

In analog designs the signal to noise ratio at the input is, as can be seen from the table, the dominant figure of merit. It is calibrated for a given process, using matching data and noise measurements. Several noise sources, spot-noise, $1/f$ noise, burst noise and thermal noise, can be distinguished in a semiconductor [8]. All contribute to the equivalent noise at the input of the amplifier at hand. Thermal noise over an equivalent resistor r with an energy of $v^2/(r\Delta f) = 4KT = 1.6610^{-20}$ cannot be influenced within bipolar transistor circuits. Variations in Oxide thickness cause matching errors within CMOS transistors, with a magnitude in the $100KT$ range [12]. The presence of $1/f$ noise need not be a fundamental limitation as systems can be designed in such a way that frequencies approaching zero carry no information.

C. Latches

A sample and hold circuit is an analog equivalent of a digital latch. A cascode circuit can be used in an analog latch to improve its capability to keep the signal value, in the same way that a feedback circuit is used to retain the value in a digital latch. A pair of latches forms an edge triggered D-flip-flop.

D. Filters

An analog continuous time filter, should be properly scaled to reach its complexity limit. The proportionality with the degree is due to the fact that multiple state variables are present. Note also the proportionality to the Q factor for RC-active filters.

E. Signal representation

It is good practice to use data-words with random bit-values in the definition of the energy complexity of a digital design. This is quite useful when the signal

statistic is 'unknown', an additional advantage is that it makes no difference whether the random data comes from one source or the other, i.e. the cost of on-chip communication is independent of the way in which the signals are transported, i.e. in which busses are connected.
For analog filters and mixers it is good practice to assume an harmonic signal, and to relate its effective value to the effective value of the noise. In-band distortion is considered as an additional noise source in analog designs.

F. Comparators

Comparators are fundamental building blocks in A/D converters. They are used to build a circuit which forms a digital quotient together with an analog remainder, which is usually a signal with a random signal distribution. So the statistics of the remainder are better described with a uniform signal distribution over the analog output range.

G. Signal to Noise ratio

Any ratio of a signal to the noise level should take effective signal values (RMS) into account. The signal to noise ratio of an analog system should at least fit in the dynamic range at the critical location(s), i.e. it is good practice to consider the noise and the signal level at the input of an amplifier. An increased area of an input transistor has as effect that the noise voltage declines, as well as the signal voltage, provided that the signal to noise ratio is kept constant, as the signal and the noise voltage are present on the same capacitor. A desired signal to noise ratio is typically reached through sizing of the critical (input) transistor(s) and adaptation of the signal level.

H. Passive components

Resistors and capacitors are used together with amplifiers within an analog filter to produce the desired transferfunction, which actually solves a linear difference equation. Note that a *passive* analog filter does not introduce any noise, nor does it consume energy. This indicates that it is a possible option to realize an analog filter function at low energy consumption levels using a combination of non-tuned passive filter functions, in combination with a tuned analog counterpart. This technique may be needed to extend the use of analog filters to applications in which extremely high frequencies are used in combination with a high dynamic range.
In most applications it will be desired to design an RC-*active* analog filter, i.e. without passive inductors. The R component is frequently realized using a two port with a given transconductance.

I. Beyond components

The energy consumption of complex digital building blocks like for instance adders and multipliers can be calculated from the energy consumption of the building blocks: 'full-adders' and 'and' circuits. Extra energy is dissipated, when compared to the energy-sum due to the individual building blocks. This is due to the fact that computations are started in a partially arbitrary input state which can be for instance the initial value of C_{in} signal, in an adder. A combinatorial digital circuit will 'iterate' to the steady state without any need for a synchronization or clock signal. This effect is known as 'rippling'
The overall energy consumption of an m-bits wide carry-ripple adder can be described in terms of the average energy consumption of a full-adder, multiplied with the number of bits m, per word, multiplied with the ripple factor: $\rho(m)$.
One may argue that the ripple factor of an arithmetic building block should always be equal to one. This situation can be reached for individual full-adders, however a substantial overhead should be taken into account when a ripple factor of 1 should be reached for a complete circuit as well, like for instance an m-bits adder or an m $\times$ n bits multiplier.

1. The overhead is either:

$$\mathcal{E}_{ff} + \mathcal{E}_{fa})/\mathcal{E}_{fa} = \\ = (1.8pJ + 2.5pJ)/2.5pJ = 1.72,$$

   due to the insertion of extra D-flip-flops, or:
2. A similar factor due to the use of asynchronous logic.

J. The effect of wiring

Wiring causes an increased energy consumption within a digital circuit, whereas most analog designs can be arranged such that the wiring capacitors are placed in parallel with the capacitors which are present anyhow. It is in this latter case possible to take the wiring fully into account in the nominal design without affecting the energy consumption. This is why capacitors and wiring (communication) is classified as energy conserving. The easiest way to take

wiring into account in a digital design is to extend all wires of a cell until the amount of energy consumed due to data transport equals the amount of energy due to arithmetic. This value is by definition the power radius $\mathcal{R}_{pd}$ of the cell at hand.
The power radius of a full-adder is 0.7 mm in an 1μm 5V process. It is useful to express the power radius of a cell in its own size. This is 10$\times$ the height for a full-adder or 30$\times$ its width. The power radius of a multiplier is of the order of 35$\times$ its height and width. Note that the power radius gives useful information about the amount of overhead due to placement.

K. Data storage

Analog circuits have a poor data retention when compared to their digital counterparts. This is mainly due to charge which leaks over diffusion areas from the relatively small capacitors to ground. Large amounts of digital values can be stored in a RAM. The effect of leakage is in most cases negligible. Instead it is the *transport of the data* which consumes energy.
A RAM with N words of width $\mathcal{W}$, uses $\mathcal{A} =^2 \log(N)$ address bits. It contains $\mathcal{W}N$ Ram cells with diameter: $\mathcal{D}_{cell} = \mathcal{H}_{cell} + \mathcal{W}_{cell}$. The diameter of a RAM is hence: $\mathcal{D}_{ram} = \mathcal{H}_{ram} + \mathcal{W}_{ram} = \mathcal{D}\sqrt{\mathcal{W}N}$. The amount of energy needed to transport $\mathcal{W}$-bits wide data words over a distance $\mathcal{D}$ram is:

$$\mathcal{E}_{ram} = \frac{\mathcal{W}\sqrt{\mathcal{W}N}\mathcal{D}_{cell}\mathcal{E}_{/m}}{\theta_{acc}\theta_{ov}}$$

With an access efficiency: $\theta_{acc} \approx 18$ for an SRAM architecture, and an overhead efficiency: $\theta_{ov} = \mathcal{W}/(\mathcal{W}+\mathcal{A})$.

III. Energy Management in Design

A. The problem

The increased density of VLSI realizations makes it possible to integrate increasingly larger systems on a single chip. Increased functionality at the same clock-rate will not affect the overall energy consumption provided that the area is kept constant. It is however attractive to utilize the increased speed capability of the new process as well. This results in an increased power density for digital designs, which limits our capability to fully exploit new processes. The creation of designs which consume a minimal amount of energy for a given function is hence an issue of major importance. This objective can be met with the energy management methodology:
Energy management is the part of the design process in which the predicted energy consumption of the design under construction, is related to the lower bound of the energy consumption, with the aim to approximate the lower bound as good as possible.
It is likely that a solution rather close to the fundamental limit, can be costly in terms of the effort needed, to reach it. The general design issue is hence not to come as close to the fundamental limit, but to come sufficiently close without high incurred costs. This objective can only be met in an effective way when all design libraries used offer a solution which comes sufficiently close to the fundamental limit.
Any design effort which starts an optimization based on a sub-optimal library will likely result in a sub-optimal partitioning of the design as well.
This has as effect that a subsequent introduction of an energy efficient library may not contribute to a better re-implementation of the design at hand.
This is not only an issue of concern for random access memories, which sometimes come with an access efficiency of 1/800 instead of the 'fair' value of 1/8, it is quite important for A/D and D/A converters too, as they have an even higher energy overhead

B. Methodology

A good methodology starts with the calculation of a parameterized survey in the form of a pie-chart for the energy consumption of a feasible design for the problem at hand. The basic contributions are best expressed in terms of a fundamental limit with additional overhead. Design alternatives which do not change the energy consumption are considered next. For instance a digital part of a design in which all operands represent random data can be repartitioned in space and time, without affecting its energy consumption, as long as any extra registers needed to store intermediate results require negligible energy. The same is true for an analog design provided that the S/N ratio is kept invariant. Sub-circuits like for instance the quantizers used within an A/D converter which produce N-bits and an analog remainder, for which a realization close to5 the fundamental limit is not known, should be analyzed up to their primitives, in order to be able to categorize the energy consumption as fundamental, useful to reach 'additional' specifications such as linearity, monoticity etc.

and superfluous, to indicate details which may need a redesign.

C. A mixed A/D example

An example of a circuit with analog as well as digital parts is the multistage A/D converter of Figure 1. Such a circuit was considered in [13] for A/D conversion in GSM base stations. Each quantizer converts an

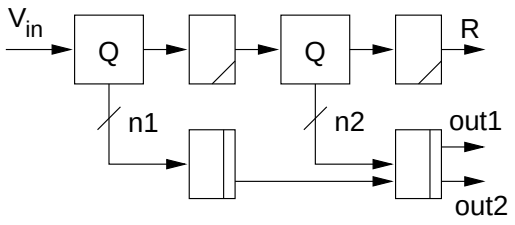

Fig. 1. Multistage A/D convertor

analog input signal (bold-line) into a digital quotient represented with Ni bits (thin-line) and an analog remainder (bold). The converter can be pipelined using analog D flip-flops (indicated with a corner marker) as well as digital D flip-flops (indicated with a line marker). A quantizer Q which is able to convert N bits at once, is called a flash converter. It consists of a voltage divider and 2^N identical comparators on the analog side and a conversion function on the digital side which converts a thermometer code into an n-bits code. The analog input signal is represented by the concatenation of the delayed bits produced by individual quantizers. A n-bit flash A/D converter has an energy consumption which is, due to the duplication, 2^N times the energy consumption for the conversion of a single bit. The minimal amount of energy needed for conversion of a single bit with an error margin of $n\sigma$ is of the order: $4\ n\ K\ T = n\ 1.6610^{-20}$.

A practical converter, with corrections for distortion due to second harmonics (costfactor 2) & monotonicity (costfactor 2), realized in CMOS technology should ideally dissipate an amount of energy per sample, for 5σ accuracy: $\approx 5\ 100\ K\ T\ 2\ 2 = 8.3\ 10^{-18}$ An N-bit converter should handle voltages at the first comparator stage of an amplitude which is 2^N times the quantization step. We have to take the triangular shape introduced in the quantization process into account as well. Hence the total amount of energy needed for conversion is: $\mathcal{E} = n\ 100\ K\ T\ 4\ 1.5\ 2^{2N} = 815\ fJ$ for $N = 8$.

The actual energy consumption for the converter published in [9] is 1nJ per sample. This implies that either the converter is implemented less ideally and that the voltage levels used on the data-path, which are of the order of a few volts, are about 35 times larger than predicted by the model.

A piece-wise linear converter characteristic can be guaranteed in a configuration which uses two quantizers, one which calculates the quotient q and a positive remainder r, and one which calculates the quotient $q1 = q + 1$ and a negative remainder $r1$, such that $r - r1$ is the unit for the next stage.

Let us now come to points of overall importance in A/D converter design. Any quantizer with a high radix has an high energy overhead due to the exponential factor 2^{Ni}. A multistage A/D converter, like the one shown in Figure 1 consumes $1/2^{Ni}$ times the amount of energy in subsequent stages due to the reduced requirements for the signal to noise ratio in the next stage, provided that an interstage amplifier is used. I.e. a radix 2 converter, uses $1 + \frac{1}{2} + \frac{1}{4} \rightarrow$ twice the basic amount dissipated in the first stage. Elimination of even distortion terms and circuitry which guarantees piece-wise linearity each introduce another factor of two energy-overhead.

Is a fully digital sigma delta approach a solution? The answer is no, in that case there is a need for 2^N subsequent comparisons. This is as bad as a full flash converter.

Let me make some final remarks about the pipelining, which extends the maximal speed of the converter of Figure 1 with a factor of two. Note that the discussion was not how fast the converter would be, but how much energy would be used. It is quite well possible, i.e. without inclusion of (analog & digital) pipeline registers, to obtain the same speed through *duplication* of A/D converter circuitry. Note that the energy consumption of analog pipeline registers is $1/2^N$ the basic energy consumption need of the first stage when a radix n stage is used. Note that analog pipeline registers should be duplicated whenever the internal circuitry is duplicated, so their relative cost is independent of any concern for linearity and monoticity.

The energy consumption of digital pipeline registers is proportional with the number of bits stored in the register. This cost scales with $V_{DD}^2/\lambda 2$. I.e. the energy consumption for digital operations on signals is mainly proportional with the number of bits. This makes it attractive, from a point of energy consumption, to realize circuitry which corrects the overall converter characteristic in the digital domain.

So far, an helicopter view has been used to look at the problem of A/D converter design. When it comes to the realization of a specific converter it is important to study its internal construction as well. For instance the actual division of the input signal in equal parts, within the quantizer can be performed using a voltage or current mode divider with matched resistors or transistors, on one hand, or with charge sharing techniques on the other hand. Note that the first two options *consume energy even when there is no conversion going on*. Energy consumption in the steady state is however avoided in the last option.

To conclude this example it may be clear that energy management techniques can be used to analyze the basic features of a design at hand. A pie-chart, in which the lower bounds for all contributions to the energy consumption are given can be used to identify the most critical design issues. Once this has been identified it becomes important to formulate the problem once again with rather abstract alternatives, like 'voltage/current' division versus 'charge sharing' and to work on the most promising option. It is good to study realizations of semiconductor circuits with energy consumption as the main figure of merit. This has resulted in a 20-fold reduction in the energy consumption of an FFT processor [4,10]. In addition it gives new insights in the desired architecture of field programmable arrays which could replace full custom chips.

The situation is quite extreme when A/D conversion is taken as an illustrative example, as practical realizations are almost never as much a factor 19,000 from the optimum. Nevertheless it could be shown that transformations in which this factor is taken constant are quite useful. The extreme overhead factor in the analog domain can be attributed to the fact that signal handling in the presence of a noisy substrate is an art for the analog domain, whereas it is solved due to properly defined noise margins in the digital domain. Nevertheless it are the mixed applications which deserve much attention during the specification of abstract systems.

At this point *education of designers of chips as well as standards* is highly needed. Simple concepts like the need for calibration of the energy consumption of basic building blocks and the notion of the power radius to be able to reason about placement in digital designs are crucial. It has been shown that these concepts differ in the analog and the digital domain. This difference can be traced back to the consequences of the signal to noise ratio in analog subsystems.

References

[1] Smit, J., Bosma, M.K., On the Energy Complexity of Algorithms realized in CMOS, a Graphics Example Proceeding of the 11th Eurographics Workshop on Graphics Hardware, pp. (93-101) Potiers, France.

[2] Smit J., Bosma M. K., Graphics Algorithms on Field Programmable Function Arrays, (1996, August 26), Proceeding of the 11th Eurographics Workshop on Graphics Hardware, pp. (103-108) Potiers, France.

[3] Donnay S., G. Gielen, W. Sansen, "High-level analog/digital partitioning in low-power signal processing applications. Proceedings of the Seventh International Workshop PATMOS-'97. pp.(47-56) Louvain La Neuve, Belgium, ISBN: 2-87200-025-9.

[4] Huisken, J.A., Bekooij, M.J.G. Gruijters, P.W.F and Welten F.P.J. A power efficient single-chip OFDM demodulator and channel decoder for Multi Media Broadcasting. Submitted for ISSCC 98.

[5] Smit J., Design Techniques for Low Power Multipliers (1996, September 23) Proceedings of the Sixth International Workshop PATMOS-'96. pp.(9-23) Bologna, Italy, ISBN: 88-371-0868-O

[6] Laurent, B. and Saucier,G. Performance/Power Tradeoffs in ASIC Multipliers. (1997, September 8-10) Proceedings of the Seventh International Workshop PATMOS-'97. pp.(15-24) Louvain La Neuve, Belgium, ISBN: 2-87200-025-9

[7] Smit J., Energy Complexity & Architecture. (1997, September 8-10) Proceedings of the Seventh International Workshop PATMOS-'97. pp.(357-368) Louvain La Neuve, Belgium, ISBN: 2-87200-025-9

[8] Gray, P.R. and Meyer, R. G. Analysis and Design of Analog Integrated Circuits. John Wiley & Sons.

[9] Venes, A.G., and van der Plassche R.J.,An 80-mW 8-b CMOS Folding A/D Converter with distributed Track-and-Hold Preprocessing, IEEE Journal of Solid-State Circuits, vol. 31, pp. 1846-1853, Dec. 1996.

[10] Smit J., and Huisken J.H. On the energy complexity of the FFT. (1995) Proceedings of the Fifth International Workshop PATMOS-'95. pp.(119-132) Oldenburg, Germany, ISBN: 3-8142-0526-X..

[11] Smit. J, Remmelink, G. B., Implementation of Graphics Algorithms on Field Programmable Function Arrays. (1996, November 27). Proceedings of the ProRISC/IEEE Workshop on Circuits, Systems and Signal Processing], pp. (331,341) Mierlo: ISBN: 90-73461-09-X

[12] Pelgrom, M.J.M, Duinmaijer, A.C.J., Welbers, A.P.G. Matching properties of MOS Transistors, IEEE Journal of Solid State Circuits, Vol. 24, No 5, pp. 1433-1440, Okt 1989.

[13] Ploeg, H van der, A CMOS signal conversion circuits for GSM base-stations, M.Sc. Report, University of Twente, BSC 005N96