

A Low Power Design Methodology Based on High Level Models

Jalel KTARI, Mohamed ABID

CES-National Engineering School of Sfax, Sfax, Tunisia
jalel.ktari@enis.rnu.tn, mohamed.abid@enis.rnu.tn

Abstract— Most actual electronic circuit and system design are confronted with the problem of delivering high performance with a limited consumption of electric power. High performance is required by the increasingly complex applications that are running even on portable devices. Low-power consumption is required to achieve acceptable autonomy in battery-powered systems. In this paper, we are interested in exploring low power architectures at system level in order to find solutions that satisfy the constraints. We propose high performances models as well as a low power exploration technique. This approach allows considering many algorithmic and architectural parameters on power consumption. A complete model is proposed in order to compute total performances of the system, which will be used during exploration thanks to a technique based on simulated annealing.

Keywords: Low power, Design space exploration, High level models.

I. INTRODUCTION

The complexity of the embedded systems is in full growth in order to reach the performance criteria of the new applications especially multimedia applications and 3D games. This makes the design increasingly difficult by integrating a multitude of functionalities while respecting the constraints. Moreover, energy consumption has become one of the principal constraints since those applications are now running on small and mobile battery-operated systems. As the battery capacity growth is too slow (Eveready's law), it becomes crucial for those systems to achieve high performance and low power consumption at the same time. Reducing energy consumption permits also to minimize the thermal dissipation which increases the system reliability and avoids the use of noisy and cumbersome cooling systems.

To achieve these antagonist goals, several low power methodologies were established. The researchers deal with the energy consumption optimization problem at several levels. The efforts were often focused on specific components like hardware, or software, or communication or memory, but seldom on the architecture as a whole. However, the designer needs a more global methodology that offers more efficient low power exploration. He/she needs also methods and tools for estimating the system performance in order to extract the most promising architectural solutions and

those which respect the constraints. However, a minority of works deal with this problem. In this paper, we present a low power exploration methodology. It is based on parametric models of performance and energy consumption. These models are exploited by an extensible exploration environment based on the simulated annealing heuristics. This environment makes it possible to extract an adequate solution respecting the various constraints. Thanks to the approach suggested, we have provided the designer with a decision-making system where he/she is guided in the choice of the architectural solution and the application parameters.

The paper is organized as follows. The next section presents the related works. The methodology and the approach are addressed in section 3. The exploration environment and the MPEG 2 results are presented in section 4.

II. RELATED WORK

During the last years, energy consumption has become an essential constraint, which is added to the real time when designing an embedded system. These systems are often based on mixed architectures: software/hardware communicating via buses and shared memories. Several works treat the low power design of software or hardware targets separately. Indeed, in [1], a method for low power design for the processors is presented. [2] and [3] focus on DSPs, [4] focuses on the memories, [5] and [6] on FPGAs, [7] and [8] on communication buses. In those various works the authors showed that there is an important profit of time to market for the designer, through the developed high-level models. Unfortunately, most of those works do not treat the architecture in its entirety, which can be made up of various software and/or hardware resources. Moreover, considering the diversity of the possible architectural solutions for an application, the designer needs some methodologies of low power exploration and estimation models of the overall system consumption. These models must take account of the various parameters which influence the performance.

Indeed, an application can have various performances on a given target by varying the algorithmic or architectural parameters. In addition, the majority of the existing partitioning software/hardware approaches do not consider the supplied maximum energy in the choice of the architectural solution [9].

Such a constraint can influence the system. Moreover, since the partitioning and the scheduling are dependent, neglecting the available energy can engender a non-schedulable solution. We thus need an estimate technique for the temporal performances, consumption and cost of the solution to do a low power design space exploration. The works presented in this paper treat more particularly an approach of low consumption exploration. The study focuses on four exploration tools which treat time, area and consumption constraints. These tools are Mogac [10], Cosyn-LP [11], Ghali [12], Codef-LP [13]. The following table summarizes the characteristics of each tool (table 1).

Table I : Low power exploration tools

Tools	Consumption			
	Hard	Soft	Communication	Total
Ghali	Xpower	Simple power	x	Each case evaluation
Codef-LP	Watt Watcher	Vestim Joule track	x	Sum
Mogac	Available models	Available models	Consumption /packet	Sum
Cosyn-LP	Available models	Available models	Available models	Sum

The tools presented in this overview are not usually available, so we require the development of a low power exploration environment that integrates the models of performances. Thus the tool can be extended and enriched with models according to the needs. Moreover, most of these approaches explore predefined or monoprocessor architectures. This is the case in the Codef tool and the Ghali methodology. In addition, with the Codef-LP tool, we can explore the solutions area but without taking account of the communication consumption which can be significant.

So, the objective of this work is to define a low power exploration approach. The work consists in:

- Multi-granularity specification: it is a question of specifying the application and the constraints on several granularity levels. This permits more efficient solution exploration.
- Parametric consumption models set up: it consists in establishing parametric models of power which include the consumption of the whole system {hardware+software+ communication}. In fact, the application and the architecture parameters will be considered in the performances models in order to have rich models.
- Effective low power architecture exploration: it is about being able to choose the adequate architectural solution where the number of resources to be exploited is not well known. Indeed, the designer can be confronted with the problem concerning the choice of the resources numbers, for instance whether the application needs two or three DSPs.

The next section treats the methodology of elaborating power and performance models as well as the cost model. We also introduce the estimation method and the design space exploration technique.

III. GENERAL APPROACH

A. Task graph and target description

- The specification model must allow a functional description of the whole application while being independent from its final implementation. Concerning the application specification, it is often represented by a task graph [10] [13] [9] [14] [15]. This representation makes it possible to model the tasks as well as the inter-tasks dependencies of the application. In the suggested approach, we start from an application specification in the form of a directed acyclic graph of tasks (DAG). In this graph, the nodes represent the tasks T_i of the system and the dependences between them are represented by arcs. We associate to each arc A_{ij} of the graph the quantity of data which the task T_i must transfer to the T_j task.

- The architecture will be heterogeneous (mainly software (TI DSP C6201, C55, C67) and hardware: FPGA) in the form of discrete components (DSPs & FPGA) communicating via a bus and having a shared memory.

B. Methodology

Our methodology is presented in Fig.1. The approach rests on a task graph specification of the application (Fig.1-A). The parameters and performances knowledge is necessary for every task present in the application specification. Concerning the tasks power estimation (Fig.1-B), every task is evaluated by measurement or by estimation tools in terms of time and consumption according to the target (each DSP and FPGA) and according to its parameters. The following stage (Fig.1-C) consists in elaborating a library of performance, power models of the application for the various tasks on various targets. Examples of performances library models are presented in our previous works [16]. These models can be recovered manually through direct measurement or through simulation using software [9] and/or hardware [6] estimation (Fig.1-B) tools.

The following stage has to do with choosing the architectural solutions (Fig.1-D). It is based on the analysis of the available solutions and the retrieval of the adequate solution. The solution analysis consists in assessing every solution and estimating its performance and its consumption (Fig.1-E). Following the analysis of various solutions, it is necessary to retrieve the adequate solution (Fig.1-F) which minimizes the cost and respects the constraints. Moreover, one or several solution(s) can be eligible and respect the real time constraints, area and consumption. It is at this moment that the exploration algorithm intervenes to choose the “adequate” solution.

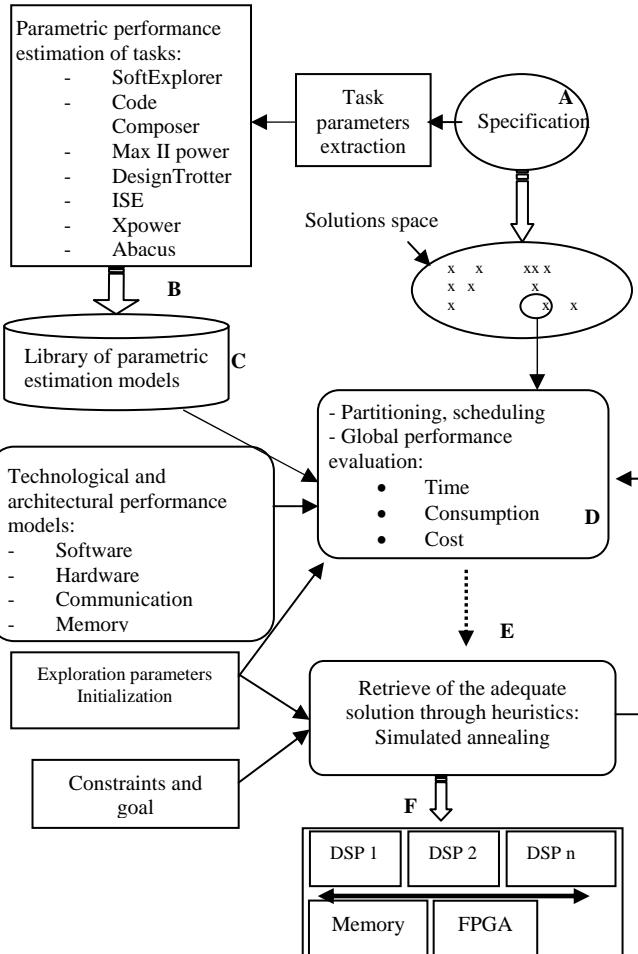


Fig.1: Low power exploration methodology

C. Energetic, cost and performance models

Nowadays, various works treat this problem for multiprocessors architectures (see [17] [18] [19]). In this study, the exploitation of a ready and validated scheduler is a possible solution and will be used.

Concerning the temporal performances, we introduce the partitioning and the scheduling in order to extract the temporal model. In fact, the partitioning and the scheduling of tasks are two recurrent problems in real time systems. Concerning consumption, we propose in table 2 the parametric consumption models.

Table II: Parametric consumption models

Target	Model
DSP	$P_{Idle}*(T_{exe_total}(DSP)) - \sum_{task(i)} T_{exe}(i) + \sum_{Task(i)} T_{exe}(Ti)*P(Ti)$
FPGA	$\sum_{Task_active(i)} P_{dynamic_Task}(i)*T_{exe}(i) + \sum_{all_Task} P_{stat}*T_{exe_total}$
Memory	$T_{exe_total}*P_{stat} + \sum_{N_accés} P_{access}(R/W)*T_access$
Buses	$1/2*C_{bus}*V^{2*N_bits}*N_Words_s*(\frac{2*N_data}{N_bit_bus}+2)/F$
System	$\sum_{Target(i)} Energy_Target(i) + Energy_buses + Energy_memory$

Due to the dominant presence of the software part (DSPs), the cost will be a secondary constraint which we cannot modify only by modifying the DSP number. In addition, due to the diversity of the technologies, the cost of each resource is balanced by a cost coefficient (Equation 1). (1 mm² of DSP area can be less expensive than that of an FPGA).

$$Cost_{tot} = \sum_{Ressources} ai*Area(i) \quad (1)$$

D. Estimation method

From the application functional analysis, the FLPA [7] methodology makes it possible to develop a parametric model which represents the target consumption behaviour. This methodology is composed of four stages:

- Functional analysis which determines the parameters influencing the power model.
- The characterization of each parameter is done to qualify its influence on the application power consumption.
- The global model is established according to the available parameters.
- The model validation by measurements.

Thus, we can take account of the algorithmic characteristics, in order to evaluate the consumption at the algorithmic level according to the variation parameters. In fact, this methodology starts from the extraction of the algorithmic, architectural and technological parameters, which have a direct influence on the application consumption (image size, image resolution, computing precision, DSP target, and frequency).

IV. Exploration tool and results

We present in this section the exploration environment which rests on two tools. The first one is useful for the task graph specification and performances capture. The second one is dedicated for the evaluation and the solutions space exploration in order to extract the adequate solution (Fig. 2).

The necessary information for the exploration includes the various possible implementations of each task. Since a task can have several performances according to these parameters, each implementation takes account of algorithmic and architectural parameters during the data capture of the execution time, the average power, the maximum power and the resulting data size.

This description is managed by a graphic interface written in java. It allows generating a text file containing this information. This textual description of the application will be the principal entry for computing under the Matlab environment.

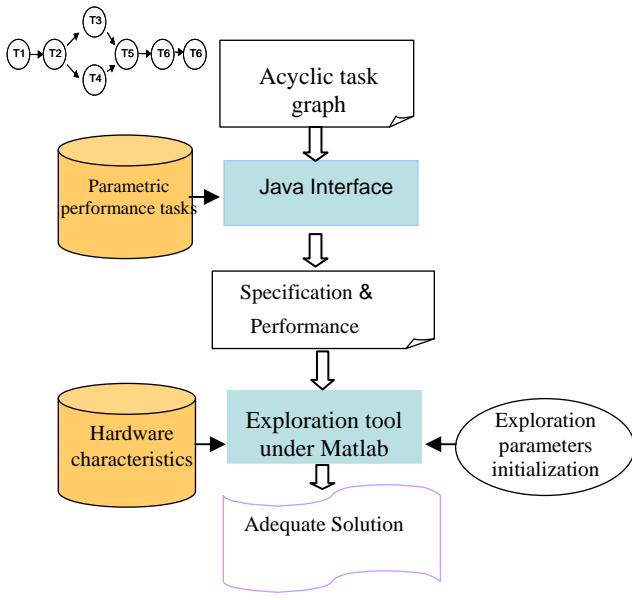


Fig. 2: Exploration environment

A. Heuristics

In order to extract an adequate solution among those present in the space of solutions which respect the system constraints, a meta-heuristics is necessary in order to solve this optimization of an NP hard problem [20]. With meta-heuristic (simulated annealing, taboo method), it is a question of authorizing a temporary degradation of the solution, during the current configuration change. A control mechanism of degradations allows avoiding the process divergence. Consequently, it becomes possible to be extracted from the trap which represents a local minimum, to leave and to explore another more promising "zone". During this work, the simulated annealing heuristic is exploited. The advantage of this method is its aptitude to extract an adequate solution. Moreover, it is a general method; it is applicable and easy to program for most of the problems, which concern the techniques of iterative optimization.

B. Implementation

In order to implement the tool with Matlab, a mono-objective exploration is done. It is allowed according to designer choices: to reach low consumption architectural solutions under real time constraints. As the designer has the possibility of imposing only the maximum number of processors in architecture without fixing the exact number, this number will not be fixed. The algorithm will explore the most promising solutions among those which respect the real time constraints and the maximum number of processors. It is the tool, which extracts the number of useful processors as well as the adequate architectural mapping and the performances of the whole system. This parameter setting of the number of the treatment units will allow the designer on the one hand, not to limit himself/herself to a unique architecture when designing the product and to be

guided by the tool when choosing the hardware solution on the other hand.

This method is based on:

- Parametric tasks performances present in the textual description file: With the diversity of the existing performance values according to the algorithmic and architectural established parameters, a model library can be analyzed by the tool. This allows the performance evaluation of each task and the adjustment of its parameters according to the objective.
- Target technology: The characteristics of each target technology are necessary in order to be able to consider the total performance of the whole system. Among these characteristics, we can mention: the Vdd, the frequency, buses size, idle power, etc ...
- Constraints: the designer defines the application constraints which will be provided to the tool in order to accept or refuse the solutions extracted during the exploration.

C. MPEG-2 Results and analysis

An initialization of the solution space exploration parameters is necessary. The user or the designer has the possibility to choose a random initial solution or a particular solution according to his/her knowledge on the application behaviour. The system constraints will be considered during each solution evaluation in order to satisfy the constraints. In this case, the designer will impose the real time constraint as well as the maximum number of resources to be exploited. In addition, the exploration algorithm makes it possible to extract the most promising solution according to the objective. Thus, we extract the necessary resources number to exploit in order to implement the application. The tool thus guides the designer when choosing the target architecture in terms of number and type of resources on a high abstraction level.

We present here some results of our methodology. We have considered the MPEG2 application. The most important tasks are: motion estimation, prediction (MPC), DCT, Inverse DCT, Quantization, Inverse Quantization and VLC (Variable Length Coder). (Fig.3) [21]

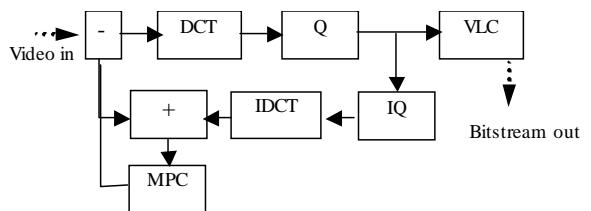


Fig. 3: MPEG2 Encoder tasks

Fig. 4 shows the exploration results in a virtual space composed of 6 DSPs (2*C5510, 2*C6701 and 2*C6201) communicating via a shared PCI buses with an objective to extract a low consumption solution respecting a strict real time constraint. Up to now, hardware modules are not included in the exploration. The algorithm converges "quickly" towards solutions using only three processors (500 iterations). It is thanks to the simulated annealing heuristics that the space

complexity problem is reduced. Thus, the user can know the adequate DSPs number for the application and also the architectural mapping which minimizes the whole system consumption while respecting the constraints. (Table III)

Table III: Bests low power solutions

MPEG2 - 1 GOP (Group of Picture)			
Architecture	2*C5510 & C6701	1*C5510 & 2*C6701	2*C5510 & C6701
Execution time (mS)	70.61	53.36	65.67
Average power (W)	1.13	1.61	1.22
Energy (mJ)	86.57	85.91	80.12
Tasks mapping /DSP	MPC / (C) DCT / (A) IDCT / (B) Quant / (A) IQuant / (B) VLC / (B)	MPC / (C) DCT / (D) IDCT / (D) Quant / (A) IQuant / (A) VLC / (D)	MPC / (C) DCT / (B) IDCT / (B) Quant / (A) IQuant / (C) VLC / (A)
<i>1st</i> C5510 : (A) <i>2nd</i> C5510 : (B) <i>1st</i> C6701 : (C) <i>2nd</i> C6701: (D)			

In Fig. 4-A, we present the result of the solution space exploration based on the simulated annealing heuristics. The tool explores the space through this heuristics and converges towards the solutions whose consumptions are less than 100 mJ/GOP (Group of Pictures). Moreover, in the Fig. 4-B, we present as an indication the random global solution space exploration. With this “intelligent” heuristics, a time saving in the adequate solution retrieve has been proven. Furthermore, we can conclude from the table III that the C6201 is not adequate for the low power design. For more legibility, the Fig. 4-D shows the area and the consumption evolution for various architectural solutions whose number of processing elements is variable: from two to six processors.

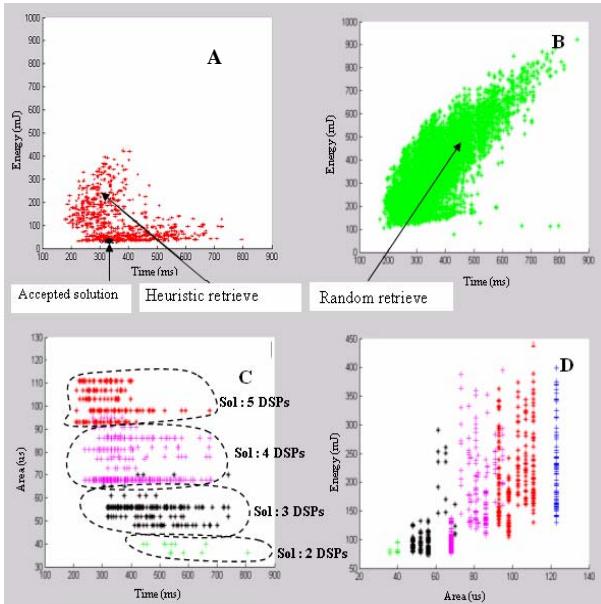


Fig. 4: Exploration results

Fig. 4 C-D shows also the energy evolution according to the area and/or time, thus allowing a detailed knowledge concerning the consumption variation space according to the solution number of processing units. Thus, the designer has the possibility to extract adequate target architecture for his/her product with minimum parametric information on a high level of abstraction. In addition, the tool proposes an adequate mapping of the tasks graph in order to have a system which answers the objective and the constraints.

V. CONCLUSION

In this paper, the low power design in embedded systems is treated. A methodology and an environment of low consumption design space exploration are proposed. The developed environment exploits a rich performance model of time and energy that takes account of many algorithmic and architectural parameters. This allows us to establish the characteristics and the mechanisms necessary in order to extract an architectural solution, which meets the needs. The key points of this problem are approached through a parametric analysis method and a heuristics based on the simulated annealing. In the future works, it is interesting to exploit this environment to explore the solutions space of a more significant application like H264 in order to validate the approach. Moreover throughout this work, parametric models of H264 are established on various levels of granularity. In addition, in the work already made, the maximum power constraint supported by the target architecture is not considered yet during the exploration. It will be integrated in the future version.

REFERENCES

- [1] P. Pakdeepaiboonpol, S. Kittitornkun, “Low energy optimization for MPEG-4 video encoder on ARM-based mobile phones”, 1st International Symposium on Wireless Pervasive Computing, Thailand, 2006.
- [2] J. Ktari, J. Laurent, M. Abid, N. Julien, “Estimation de la consommation logicielle dans un système embarqué : Etude de cas”, Proceedings FTFC’2005, pp. 55-59, May 18-19, 2005, France
- [3] J. Laurent, N. Julien, E. Senn, E. Martin, “Functional Level Power Analysis: An Efficient Approach for Modeling the Power Consumption of Complex Processors”, DATE 2004, pp.666-667
- [4] F. Marteil, “High Level Memory Hierarchy Optimisation”, In EDAA Ph.D. Forum at DATE, 6-10 March, Munich 2006.
- [5] A. Garcia, L. Gonzales, R. Felix, “Power consumption management on FPGAs”, 15th International Conference on Electronics, Communication and Computers, March 1-2, Mexico, 2005.
- [6] D. Elleouet, Y. Savary, N. Julien, D. Houzet, “A FPGA Power Aware Design Flow”, Patmos06, September 13-15, France ,2006
- [7] K. Lahiri, A. Raghunathan, “Power analysis of system-level on-chip communication architectures”, International Conference on Hardware/Software Codesign and System Synthesis, CODES + ISSS, September 8-10, Sweden, 2004.
- [8] M Caldari, M. Conti, M. Coppola, P. Crippa, S. Orcioni, L. Pieralisi and C. Turchetti, “System-level power analysis methodology applied to the AMBA AHB bus”, DATE03, 2003.

[9] V. Kappagantula, N. Mahapatra, "PAP: PowerAware Partitioning of Reconfigurable Systems", HPCA9/SSRS '03, California, USA, 2003.

[10] P. Robert Dick, K. Niraj, "MOGAC: A Multiobjective Genetic Algorithm for Hardware-Software Co-Synthesis of Distributed Embedded Systems", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 1998.

[11] P. Bharat, P. Dave, G. Lakshminarayana and K. Jha, "COSYN: Hardware-Software Co-Synthesis of Heterogeneous Distributed Embedded Systems", IEEE Transaction (VLSI) Systems, VOL. 7, NO. 1 , 1999.

[12] K. Ghali, O. Hammami, I. Hermann, "Multiobjective Design of Embedded Processors on FPGA Platforms", ICDCS Workshops, 2004.

[13] P. Guittou-Ouhamou, C. Belleudu, M. Auguin, "Energy Optimization in Hw/Sw Tool: Design of Low Power Architecture System", IEEE International Workshop on System on Chip for Real-Time Systems - IWSOC'2003, Canada, 2003.

[14] H. Tmar, J.P. Diguet, A. Azzedine, J-L. Philippe, M. Abid. "RTDT : a Static QoS Manager, RT Scheduling, HW/SW Partitioning CAD Tool". Microelectronics Journal, 2007.

[15] A. Azzedine, J-P Diguet, J.L Pillippe, "Large exploration for HW/SW partitioning of multirate and aperiodicreal-time systems", CODES 2002. Proceedings of the Tenth International Symposium on on Hardware/Software Codesign, 2002 Page(s):85 – 90, Estes Park, Colorado, USA,

[16] J. Ktari, M. Abid, N. Julien, Johann Laurent, "Power Consumption and Performance's Library on DSPs: Case Study MPEG2", Journal of Computer Science Vol.3, N°3: 168-173, 2007 Sciences Publication, ISSN: 1549-3636, New York, USA.

[17] J. Pinot, S. Bhattacharyya, A. Edward, "A Hierarchical Multiprocessor Scheduling System for DSP Applications", IEEE Proceedings of ASILOMAR-29, 1996.

[18] T. Bandyopadhyay, B. Susnata, B. Swapan, "Multi Processor Scheduling Algorithm for tasks with Precedence Relation", TENCON proceedings analog and digital techniques in electrical engineering, Thailand , 2004.

[19] Z. Lichen, H. Jiwu, Z. Yi, "Scheduling Algorithms For Multiprocessor Real-Time Systems", International Conference on Information, Communications and Signal Processing, ICICS '97

[20] N. Chabini, "A Heuristic for Reducing Dynamic Power Dissipation in Clocked Sequential Designs". PATMOS 2007, 64-74, Sweden.

[21] J. Sohn, H. Kim, J. Jeong, E. Jeong, S. Lee, "A low power multimedia SoC with fully programmable 3D graphics and MPEG4/H.264/JPEG for mobile devices", ISLPED 2007, 238 – 243, USA

