Aalborg Universitet

Follow-up Modelling for Wireless Personal Communication Systems

Le Moullec, Yannick; Ben Amor, Nader; Diguet, Jean-Phlilippe; Koch, Peter

Citation for published version (APA):
Le Moullec, Y., Ben Amor, N., Diguet, J.-P., & Koch, P. (2004). Follow-up Modelling for Wireless Personal Communication Systems. In *WPMC 2004* (Vol. CDROM, pp. 255-259). WMPC. http://www.wpmc-archives.com/site/user_login.php?page_no=1&menu_kind=ref&ref_mode=keyword&ref_head=&ref_year=&within_key=&dtl_ref_title=&dtl_ref_keyword=&dtl_ref_author=moullec&dtl_ref_belong=&dtl_start_year=2004&dtl_end_year=2005&menu_state=inp&essay_id=139

Follow-up Modeling for Wireless Personal Communication Systems

Yannick Le Moullec, Peter Koch
Center for Indlejrede Software Systemer
(CISS)
Aalborg University
Fr. Bajers Vej 7B, 9220 Aalborg Ø,
Denmark
moullec@cs.auc.dk, pk@kom.auc.dk

Nader Ben Amor, Jean-Philippe Diguet
LESTER
Université de Bretagne Sud
BP 92116, 56321 Lorient Cedex, France
diguet@iuplo.univ-ubs.fr, benamor@iuplo.univ-ubs.fr

Abstract— **We present a framework for the modeling and the design of adaptive wireless systems such as software radio or multimedia applications. These applications use adaptive algorithms to minimize the required processing power and to take into account the system environment. We propose a framework based upon a model that can predict the behavior of a fluctuant system in real-time and that can be used to dynamically perform the application parameterization in order to optimize the trade-off between power consumption, QoS and performances. One of the originality of our approach is the use of the theory of follow-up systems for discrete processes.**

Keywords— **adaptive systems, wireless communications, HW/SW co-design**

I. INTRODUCTION

The design of wireless personal communication systems is a challenging task. There is an obvious mismatch between the required performances for the applications, the evolution of VLSI technology and economically viable architectural solutions. Future telecommunication applications such as 4G imply computational power that does not compare favorably with the energy autonomy required. In theory, a dedicated, tailored architecture would therefore be a must. However, the increasing circuit costs implies a maximum re-use of hardware and software IPs. Economic constraints push on flexibility while efficiency requirements often tend toward the opposite direction. This trend is intensified by the technological evolution which results in circuits dominated by interconnections, which are required for flexibility in terms of hardware (e.g., FPGAs) as well as software (e.g., processor control unit). Moreover, the power consumption issue is not simple: on the one hand the reduction of transistor sizes enables the reduction of the supply voltage and therefore of the power consumption (since it's proportional to the squared voltage). On the other hand this size reduction increases leakage currents and therefore of the static consumption (which used to be neglected). Finally, new telecommunication (UMTS, 4G) applications target large data transfers requiring efficient memory resources and coding/decoding architectures, which increases the complexity of the systems.

A. Related work

Some solutions have been proposed to tackle the trade-off between flexibility and performances. Those operating at the system level have been proven to be efficient since they rely on a global approach of the problem while exploiting the configuration capabilities of the target architecture. Two main approaches have been recently explored.

The first one is the dynamic scheduling of tasks under power consumption and real-time constraints, usually by means of voltage/frequency control [1].

The main idea used in the second approach is to shut down the power of unused resources by estimating sufficient inactivity periods while respecting the time constraints [2]. An extension of these techniques is proposed in [3] through a mode approach which copes with both real time and power constraints. Unfortunately this is a temporary solution since the voltage downscaling is approaching physical limits.

Recently QoS has been considered conjointly with power savings, in [4] the voltage management is combined with a QoS–based dynamic tasks scheduler and in [5] an interesting study is performed about the aggregation of various known techniques used at different levels (architecture, MiddleWare, API) of a end-to-end media delivery QoS/power optimization. A relevant QoS management based on control theory is introduced in [6], the method only addresses the QoS through the real time scheduling issue by priority adaptation.

These approaches do not address the issue of algorithmic choices and propose a restricted online power model. Our view could somewhat be seen as a combination of the global view proposed in [5], the feedback priority management detailed in [6] and an online power estimator driving application configuration ranking.

B. Contribution

We propose a framework based upon a model that can predict the behavior of a fluctuant system in real-time and can be used to dynamically perform the application parameterization in order to optimize the trade-off between power consumption, QoS and performances. The main purpose of the system is to dynamically update some algorithmic configuration parameters which follow-up the effects of the user's requirements, the environment and the algorithmic configuration choices. A key feature of our work is the use of a follow-up system for discrete processes. Each "application" (thread, process, task…) is identified by an interface which specifies the user requirements in terms of energy (battery life-time…) and QoS (framerate, picture quality…). Relevant data is then monitored on the system and fed-back to the follow-up task. Moreover the architecture has also a set of parameters used

for the regulation process: control of the frequency/voltage supply, on/off of the peripherals. Also, some data is collected on the architecture and fed back to the follow-up system: utilization of input/output buffers (i.e., data-rates), energy level, …

As compared to existing works we consider simultaneously the following elements:

- User/System constraints: battery life-time, peak power, tasks QoS…,

- Operating environment : network type, location…,

- Context and energy aware regulation by means of data-metrics and HW/SW reconfigurations (algorithmic and architectural reconfigurations).

The second contribution of this work is the consideration of a multi-threaded embedded processor. Wireless applications are inherently multi-threaded (user and data communication control, data adaptation, packet processing, security…). Recently, a new class of multi-threaded processors for network and wireless applications has been introduced [7][8]; in this work we use one of these processors as part of a wireless system case study.

The rest of the paper is organized as follows: section 2 details the model that we propose. In section 3 experimental results illustrating our approach are presented. Finally we conclude in section 4.

II. MODEL

C. Principle

The basic idea of the proposed method is to maintain a systematic arranging of algorithmic configurations which enables the selection, for a given average power, of the most suited solution while taking into account the task priorities assigned by the user as well as his requirements in terms of energy (battery life-time…) and QoS (frame-rate, picture quality…).

The first step in our approach is to define the interfacing between the application and the follow-up system. For that purpose it is necessary to identify the input/output parameters which enable the application configuration and the measure of the obtained QoS, respectively.

The second step consists in devising a state model which can be used to predict the evolution of the parameters influencing the system (data-rate, power consumption, QoS of channel decoding, ...). The resulting state matrix accounts for interactions between parameters and between application configurations.

Finally, in a third step a follow-up system is devised to track as precisely as possible the user's commands in terms of e.g., battery life-time and QoS. This follow-up system is implemented as a task and is responsible for the algorithmic configuration of the other tasks of the system. The core feature of the system is a graph enabling the classification of the algorithmic configurations according to the task priorities and power consumption.

There is a trade-off between a fixed pre-defined systematic arranging and the handling of hazards. Indeed a certain number of configurations are foreseen according to i) the variation parameters of each task (or thread, process); ii) the value of critical data and iii) the number of possible algorithmic versions for a given task (or thread, process).

However, the (possibly) random nature of the system would require to foreseen an infinite number of cases to take into account all the possible hazards. As it is almost impossible (and time consuming) to model everything we propose to limit on the one hand the number of possible cases. On the other hand our approach handles data versatility by means of update of the systematic arranging of the configurations. The update process is possible if there is access to the relevant data (such as the energy level of the battery) on the system.

D. Follow-up model elements

In our approach we consider the following elements (note that we do not always use them all):

- TAC[k]: Task Algorithmic Configuration. Array indicating the current algorithmic configurations of the system tasks at time k;

- DM[k]: Data Metrics. Array indicating the value, at time k, of metrics characterizing meaningful data for each task. This point is a non-trivial issue since it requires the designer to analyze his system and select meaningful data-metrics, which enables the characterization of the configurations. For this purpose the designer may use DesignTrotter [9] which is currently being updated to perform such dynamic characterization;

- CA[.]: Configuration Associations. Matrix representing allowed/forbidden associations between configurations according to DM values (e.g., no motion correction in MPEG is picture type is I);

- UC[.]: User/System Constraints. Matrix storing the user constraints (e.g., desired battery life-time, maximum instantaneous power…) and the associated tolerances (expressed in percentage);

- TP[]: Task Priorities. Array indicating the task priorities;

- P^[k+1]: Estimation at time k of the average power at time k+1;

- ASG: Algorithmic Selection Graph. Graph-based representation of the systematic arranging of the configurations. This representation is used to store, update and extract configurations for a given power value. A node represents a specific configuration for a given task associated to compatible DMs. The graph is ordered using a left-edge algorithm such as the left branch corresponds to the minimum consumption and the right branch to the maximum consumption. An example is depicted in Fig.1.
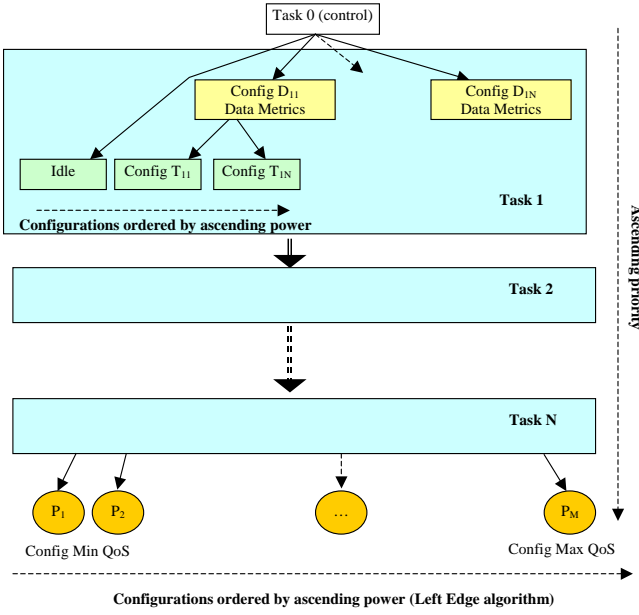
Fig.1. Graph-based representation of the systematic arranging of the configurations.

same interval and iii) to estimate E[k+1]. For that purpose several solutions can be considered:

- Kalman filter if E[k] is assumed to be a random noisy signal,
- In a regular gas-gauge circuit the energy signal is filtered, in that case a linear extrapolation of the evolution of P can be approximated such as *P[k+1] =Ak+B with A = P'[k] and B=P[k].*

We assume that the evolution of P is locally linear, from where:

$$(1) \quad \hat{E}[k+1] - E[k] = \int_{kT}^{(k+1)T} P(t)dt = \int_{kT}^{(k+1)T} (at+b)dt$$

$$= \frac{1}{2}P'[k]T^2 + P[k]T$$

$$where \quad a = P'[k] \quad and \quad b = P[k]$$

$$(2) \quad \hat{P}[k+1] = P'[k]T + P[k]$$

$$(3) \quad P\max[k+1] = \frac{E[k] - E_0}{D_{\min} - kT}$$

E. Follow-up algorithm

The follow-up mechanism (Fig.2) is implemented as a task on the system. This task is executed with a period T. The period T depends on a trade-off between the accuracy of the follow-up and the overhead (in terms of time, power…) due to the execution of the follow-up task. The algorithm is as follows:

1. Read (via a gas-gauge circuit for example) or estimate the battery level E[k], read data metrics DM

2. Update the content of current active node in the Algorithmic Selection Graph (ASG)

3. Re-position the node if necessary (i.e., if the current node power is greater than those of its left neighbors)

4. Compute P'[k]. If P'[k] is not equal to zero then continue

5. Compute P^[k+1]

6. Extract the configuration complying with P^[k+1]

F. Estimation of P[k+1]

Knowing E[n] at time k for n = k-1, k-2, …k-M it is possible to i) compute P[k], i.e., the average power consumed in the interval [k-1 ; k], ii) compute P'[k] the discrete derivative of P[k] in the
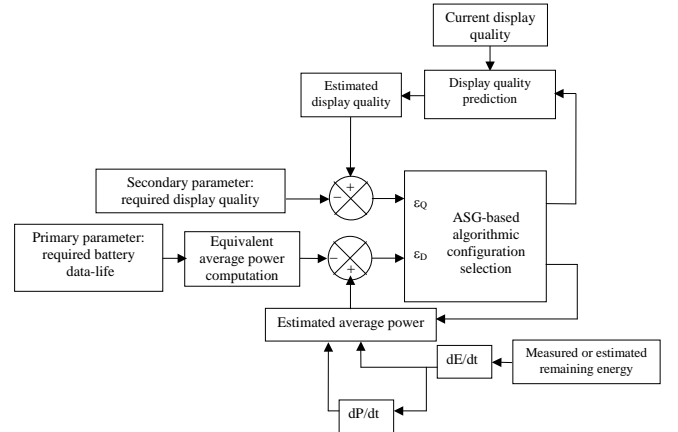

Fig.2. Follow-up model considering two parameters: battery data-life and quality of display. Here the algorithmic configurations are selected according to ε_Q and ε_D which represent the difference between the required and the estimated quality and average power respectively.

III. EXPERIMENTAL RESULTS

We are currently considering the system described in Fig.3 as a case study. The system is composed of a network processor (XINC XC1100) and a media processor (LEON core + HW accelerators). The XINC processor features hardware-multithreading capabilities making it quite efficient in terms of equivalent MIPS.

In order to classify the power consumption of the different algorithmic configurations we are devising a functional power model of the XINC processor [10]. The other part of the system, the media processor, is designed using the methodology that we have developed in [11]. We use as a test vehicle a set of applications from the domains of network processing: NpBench [12]. In this paper we present preliminary results based on experiments conducted with a modified version of the media-transcode module of NpBench, implemented on the Xinc processor.
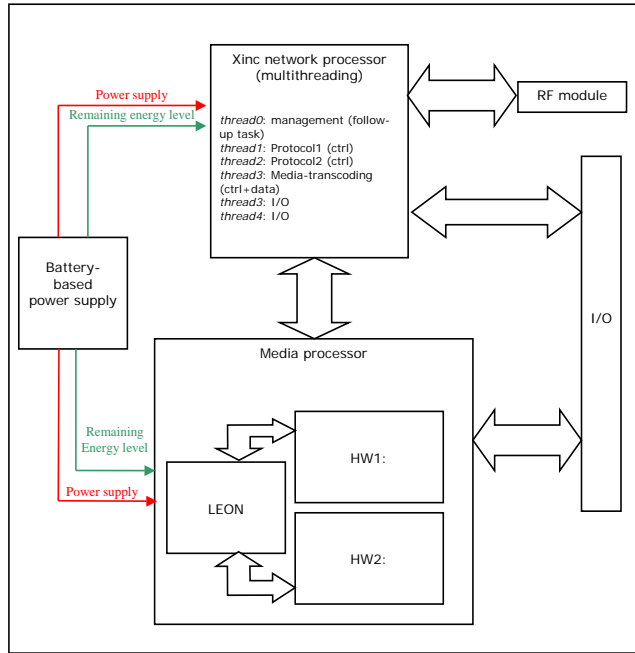

Fig.3. Case study SoC. The XINC is a multi-threaded processor where Thread0 implements the follow-up task described in Fig.2. The other threads implement the application itself (here network protocols and media-transcoding).

G. Media-transcode description

In this section we briefly present the media-transcode module of NpBench. The objective of this module is to adapt the images contained in web-pages to improve their delivery to a wide range of clients, including portable wireless devices. The module is made of two main steps. The first one is control-plane oriented and is responsible for taking transcoding decisions according to parameters such as image size, type and purpose, available bandwidth, client display capability… The second step is data-plane oriented and performs the actual transformations of the data (image re-sizing, color conversion…).

H. Experiments

Since both the transcoding level and the input data are adjustable it is possible to generate several algorithmic configurations for the media-transcoding module and therefore to explore the QoS/performance/energy trade-off.

The transcoding level is adjustable between 0 (no-transcoding) and 5 (maximum transcoding). When the transcoding level is set to 0 the content is most likely to be un-viewable on a portable device with limited display capabilities. On the other hand, setting the level to 5 will increase the quality at the cost of computation time and energy consumption.

Table.1 shows the power required by some of the algorithmic configurations on the xinc processor.

T1: Media-transcoding configurations	
Version	Power (mW)
a	284,022
b	284,200
c	285,147
d	286,065
e	286,443
f	287,013

Tab.1. Power consumption for some algorithmic configurations implemented on the Xinc processor. T1 configurations are related to the transcoding level.

We have experimented with two scenarios. In the first one the user decides to maximize the readability of the content, while in the second one he decides to maximize the battery data-life (i.e., to reduce the energy consumption).

The effects of the dynamic reconfiguration by the mean of the proposed follow-up system are summarized in Table.2. In the first scenario the main objective is to well adapt the content to the display capabilities of the device. When enabling the follow-up system, the number of data processed increases by 10% while the consumed energy only increases by 4,66%.

In the second scenario the main objective is to minimize the energy consumption. By using the proposed methodology, the energy consumption can be reduced by 15,8% while the number of data processed only decreases by 8,2%.

Finally the follow-up task overhead is quite minimal: from a timing point of view it is almost negligible since it is implemented as a thread on the multithreaded Xinc processor. From a power point of view the overhead due to the follow-up task is about +0,72%.

	No follow-up	Follow-up
Max QoS	1242,02 µJ (per packet)	1300 µJ(per packet)
Max battery life-time	55202,08 µJ (per packet)	46921,768 µJ (per packet)

Tab.2 Comparison of the energy consumption for two scenarios (max QoS and max battery life-time) with and without the follow-up technique.

IV. CONCLUSION

In this paper we have presented a framework for the modeling and the design of adaptive wireless systems. We have proposed a framework based upon a model that can predict the behavior of a fluctuant system in real-time and can be used to dynamically perform the application parameterization in order to optimize the trade-off between power consumption, QoS and performances. The originality of our work is the simultaneous consideration of the user constraints (battery life-time, tasks QoS), the operating environment (network type, location) and a context and energy aware regulation by means of data-metrics and HW/SW reconfigurations (algorithmic and architectural reconfigurations). The proposed methodology has been evaluated using a study case where the experimental results illustrate the interest of this work. Future work includes the implementation of the methodology on the OMAP platform from Texas Instrument.

REFERENCES

[1] K. Choi, K. Dantu, W.C. Cheng and M. Pedram, "Frame-Based Dynamic Voltage and Frequency scaling for a MPEG Decoder", ICCAD, 2002

[2] C.H. Hwang and A. Wu, "A Predictive System Shutdown Method for Energy Saving of Event-driven Computation, DAC 1997

[3] P.H. Chou and N. Bagherzadeh, "Mode selection and mode-dependency modeling for power-Aware Embedded Systems", ASP-DAC, 2002

[4] J.L. Wong, G. Qu. M. Potkonjak, "An On-line Approach for Power Minimization in QoS Sensitive Systems", ASP-DAC, 2003

[5] H.V. Antwerpen, N. Dutt, R. Gupta, S. Mohapatra, C. Pereira, N. Venkatasubramanian, R.von Vignau, "Energy-Aware System Design for Wireless Multimedia", DATE, 2004

[6] J. A. Stankovic, C. Lu, S. H. Son and G. Tao, "The Case for Feedback Control Real-Time Scheduling", 11th EuroMicro Conference on Real-Time Systems, 1999

[7] XINC XC100, www.elevenengineering.com

[8] UBICOM IP3023, www.ubicom.com

[9] Y. Le Moullec, N. Ben Amor, J-Ph. Diguet, J-L. Philippe and M. Abid, "Multi-granularity Metrics For The Era Of Strongly Personalized SOCs", DATE 2003

[10] E. Baud and C. Leroux, "FLPA Model of the XINC Processor", internal report, Aalborg University, Denmark, 2004

[11] N. Ben Amor, Y. Le Moullec, I. Maalej, J-Ph. Diguet, M. Abid and J-L. Philippe," Design of a multimedia processor based on metric computation", SSD 2003

[12] B. K. Lee and L. K. John, "NpBench: A Benchmark Suite for Control plane and Data plane Applications for Network Processors", ICCD 2003