A Control Theoretic Approach to Run-Time Energy Optimization of Pipelined Processing in MPSoCs

Andrea Alimonda¹, Andrea Acquaviva³, Salvatore Carta¹, and Alessandro Pisano²

¹Dept. of Mathematics and Informatics, University of Cagliari, Cagliari, Italy

²Dept. of Electrical and Electronic Engineering (DIEE), University of Cagliari, Cagliari, Italy

³ Center for Applied Information Science and Technology (STI), University of Urbino, Urbino, Italy

Abstract

In this work we take a control-theoretic approach to feedbackbased dynamic voltage scaling (DVS) in Multi Processor System on Chip (MPSoC) pipelined architectures. We present and discuss a novel feedback approach based on both linear and non-linear techniques aimed at controlling interprocessor queue occupancy. Theoretical analysis and experiments, carried out on a cycleaccurate multiprocessor simulation platform, show that feedbackbased control reduces energy consumption with respect to standard local DVS policies and highlight that non-linear strategies allows a more flexible and robust implementation in presence of variable workload conditions.

1. Introduction

The design and implementation of Multi-Processor Systemon-Chip (MPSoC) architectures is characterized by conflicting requirements in terms of performance demand and stringent power budgets. Dynamic voltage scaling (DVS) is a well-known technique to address power minimization of digital CMOS systems. It allows the clock speed and supply voltage to be adjusted on-line over a range of feasible voltage/frequency pairs [6]. Run-time voltage and frequency scaling techniques have been extension with a feasible voltage for a speed of the second

Run-time voltage and frequency scaling techniques have been extensively studied for single processor systems with soft realtime constraints (see [1] for an overview). Even if DVS for multiprocessor systems has also been studied (see [4, 5, 3]), control theoretic approaches to DVS have been previously proposed only for single processor systems [10, 9, 8, 7].

The main contributions of the present note are: i) the control theoretic model of a multi-processor system using interprocessor queues for communication; ii) the design and implementation of linear and non-linear feedback control policies for frequencies adaptation using queues occupancy; iii) their comparative experimental evaluation on a cycle-accurate, energy-aware, simulator running streaming benchmarks.

In experiment section we shall compare the proposed feedback strategies with local DVS and shutdown policies.

2. Queue-Based Control-Theoretic DVS Approach

We consider MPSoC systems in pipelined multi-layer configuration. Each layer is represent by a single processor (P_i) which processes incoming stream and communicates results to the following layer through data buffers (Q_j) . The schematic representation in Fig. 1 refers to an M-layer architecture. The core frequency of the last stage P_M is constrained by application throughput specifications. We are aimed at designing feedback policies for adjusting core frequencies of processors P_1, \ldots, P_{M-1} . The occupancy levels of queues are chosen as the feedback signals driving the adaptation policies.



Figure 1. M-layer pipelined architecture

2.1 System modeling

Based on the assumption that the "data-rate" \mathcal{D}_i of the j - th processor, i.e. the number of frames processed in the unit of time, is proportional to its core frequency f_j , an approximate model of an M-layer pipeline can be expressed as follows:

$$\dot{Q}_{j}(t) = \mathcal{D}_{j} - \mathcal{D}_{j+1} = k_{Oj}f_{j} - k_{I(j+1)}f_{j+1}$$
(1)

where Q_j represents the occupancy of the *j*-th buffer $(1 \le j \le M - 1)$, f_j is the current clock frequency of the *j*-th processor $(1 \le j \le M)$ while coefficients $k_I j$ and $k_O j$ define respectively the ratio between input data rate and current processor frequency and the ratio between output data rate and current processor frequency. In real-life systems the range of available frequencies f_j 's is discrete, leading to non-trivial control policy design trade-offs.

2.2 Linear Analysis and Design

Let Q^* be the set-point for the queue occupancy levels. The dynamics of the (M-1)-th queue, together with the control system feedback architecture, can be represented by a standard block-diagram (see Fig. 2).

The "outcoming data-rate" $k_M f_M$ can be considered as a constant "disturbance" acting on the input channel. Classical linear



Figure 2. Feedback control system for the (M-1)-th queue

analysis tells us that a type-II control system (i.e. a control system containing two integrative actions in the forward path) can guarantee the zeroing of the error e_{M-1} whatever k_M , f_M , k_{M-1} are. Thus, the choice of a proportional/integrative (PI) controller (see Fig. 2) is motivated. Vanishing of e_{M-1} implies a constant setting of f_{M-1} in steady state. The same feedback control strategy can be iterated backward and applied in sequence to each previous stage.

2.3 Nonlinear Analysis and Design

The dynamical system (1) under investigation is also suitable for the application of the following nonlinear integral-type adaptation policy

$$\dot{f}_i = -Gsiqn(e_i) - Gsiqn(\dot{e}_i), \tag{2}$$

with G > 0 a sufficiently large controller parameter. The sign of \dot{e} can be approximated by the sign of the difference between the current and past sample of e. The above control law can be seen as a special realization of the "Twisting" algorithm [12] and belongs to the class of control algorithms referred to as "secondorder sliding-mode controllers", nonlinear control laws endowed by superb robustness properties against modeling errors, disturbances and non-idealities of various kind [12]. A positive (negative) "command signal" f_j can be understood as the requirement of increasing (decreasing) the frequency.

Experiments 3.

We carried out experimental analysis within a SystemC-based, cycle accurate and energy-aware simulation platform [11]

Multi-layer pipelined applications based on standard signal processing algorithms (FIR filtering) and cryptography (DES encryption-decryption) have been used as benchmarks. Workload variability has been emulated through "dummy" loops with random length.

A detailed comparative analysis has been performed on pipelined MPSoCs architectures with different number of stages. We compared our PI and Twisting controllers with two local policies, namely ON-OFF (i.e. shut-down-when-idle without any volt-age scaling) and Vertigo (a standard DVS algorithm used in ARM IEM enabled systems [2] [13]).

For space limitations in this paper we report results for a 3-stage architecture only. Table 1 reports energy consumption and number of frequency switchings when applying different techniques to the benchmark application yielding a constant output throughput. Number of switching is a critical metric because of two reasons. First, frequency/voltage switching has a cost in terms of time and power. Second, frequency changes could be constrained by synchronization problems with other components of the chip.

Figure 3 reports the queues occupancy (the set point Q^* was chosen as 100) and the time varying profile of the processor frequencies obtained using PI control algorithm.

Feedback-based policies, especially nonlinear techniques, lead to considerable energy saving as compared with the local strategies. As outlined by several experiments, not reported for paper

3-stage pipeline - variable workload				
Technique	core1	core2	total	Ν
	energy	energy	energy	SW.
	(mJ)	(mJ)	(mJ)	
ON-OFF	574	430	2984	-
Vertigo	742	554	3248	-
PI	416	188	2644	716
Twisting	249	159	2438	22

Table 1. Control techniques comparison



Figure 3. Queues occupancy and processors frequencies using PI controllers

length limitations, nonlinear controllers are much less sensitive against operating conditions, leading to a more efficient tuning. Moreover, they allow for a considerable reduction of the number of frequency/voltage switchings.

Conclusions 4.

In this work we addressed the problem of run-time clock speed setting in MPSoCs pipeline stages by taking a feedback-based control-theoretic approach. We implemented and compared linear and non-linear feedback strategies, the latter resulting more effective. Next activities will be devoted to achieve deeper understanding of the inherent properties and capabilities of the proposed schemes.

REFERENCES 5.

- L. Benini, A. Bogliolo, and G. De Micheli. "A survey of design techniques for system-level dynamic power management". *IEEE Trans. on VLSI Systems*, pages 299–316, June 2000.
- K. Flautner, T.N. Mudge, ' Linux,", OSDI 2002. "Vertigo: Automatic Performance-Setting for

- Linux,"", OSDI 2002.
 D. Zhu, R. Melhem, and B. Childers. "Scheduling with dynamic voltage/speed adjustment using slack reclamation in multi-processor real-time systems". *IEEE Trans. on Parallel and Distributed Systems*, 14:686–700, July 2003.
 A. Andrei, M. Schmitz, P.Eles, Z.Peng, and B.M. Al-Hashimi. "Overhead-Conscious Voltage Selection for Dynamic and Leakage Energy Reduction of Time-Constrained Systems". *DATEO4*, pages 518–523, 2004.
 A. Andrei, M. Schmitz, P.Eles, Z.Peng, and B.M. Al-Hashimi. "Simultaneous Communication and Processor Voltage Scaling for Dynamic and Leakage Energy Reduction in Time-Constrained Systems". *ICCAD04*, pages 362–369, 2004. 2004

- 2004.
 [6] G. Qu. "What is the limit of energy saving by dynamic voltage scaling? ". *IEEE/ACM Int. Conf. on Computer Aided Design*, pages 560–563, 2001.
 [7] Z. Lu, J. Lach, M. Stan, "Reducing Multimedia Decode Power using Feedback Control," *ICCD03*, pages xx-xx, 2003.
 [8] C. Im, H. Kim, S. Ha, "Dynamic Voltage Scaling Technique for Low-Power Multimedia Applications using Buffers," *ISLPED01*, pages 34–39, 2001.
 [9] Y. Lu, L. Benini, G. De Micheli, "Dynamic Frequency Scaling with Buffer Insertion for Mixed Workloads," *IEEE Transactions on computer aided design of integrated circuits and systems*, 21(11), pages 1284–1305, 2002.
 [10] Z. Lu, J. Hein, M. Humphrey, M. Stan, J. Lach, K. Skadron, "Control Theoretic Dynamic Frequency and Voltage Scaling for Multimedia Workloads", *CASES02*, pages 156–163, 2002.
 [11] MPARM, http://www-micrel.deis.unibo.it/sitonew/research/mparm.html
- CASES02, pages 156–163, 2002.
 MPARM, http://www-micrel.deis.unibo.it/sitonew/research/mparm.html
 G. Bartolini, A. Pisano, A. Levant and E. Usai Higher-Order Sliding Modes for Output-Feedback Control of Nonlinear Uncertain Systems, in *Variable Structure Systems: Towards the 21-st century*, X. Yu and J, Xu (Eds.), Lecture Notes in Control and Information Sciences, Springer-Verlag, vol. 274, pp. 83-108, 2002.
 ARM Intelligent Energy Manager, "Dynamic Power Control for Portable Devices," www.arm.com/products/CPUs/cpu-arch-IEM.html, 2005.