Chapter 13
Reliability–Aware Proactive Energy Management in Hard Real–Time Systems: A Motivational Case Study

Satyakiran Munaga
IMEC/SSET and K.U. Leuven/ESAT, Belgium

Francky Catthoor
IMEC/SSET and K.U. Leuven/ESAT, Belgium

ABSTRACT

Advanced technologies such as sub-45nm CMOS and 3D integration are known to have more accelerated and increased number of reliability failure mechanisms. Classical reliability assessment methodology, which assumes ad-hoc failure criteria and worst-case for all influencing dynamic aspects, is no longer viable in these technologies. In this paper, the authors advocate that managing temperature and reliability at run-time is necessary to overcome this reliability-wall without incurring significant cost penalty. Nonlinear nature of modern systems, however, makes the run-time control very challenging. The authors suggest that full cost-consciousness requires a truly proactive controller that can efficiently manage system slack with future in perspective. This paper introduces the concept of “gas-pedal,” which enhances the effectiveness of the proactive controller in minimizing the cost without sacrificing the hard guarantees required by the constraints. Reliability-aware dynamic energy management of a processor running AVC motion compensation task is used as a motivational case study to illustrate the proposed concepts.

INTRODUCTION

Technology and System Reliability Wall

Advanced technologies such as sub-45nm CMOS and 3D integration are known to have more accelerated and increased number of reliability failure mechanisms. Lifetime reliability of a system-in-package is a complex function of fabrication process variability, system activity and thermal environment. Classical reliability qualification methodology assesses reliable lifetime of a chip at design-time using ad-hoc failure criteria and
Reliability-Aware Proactive Energy Management in Hard Real-Time Systems

assumes worst-case for electrical stress and thermal stress and process variability (Groeseneken, 2005). Such assessment indicates that advanced technologies can no longer meet lifetime requirements (Srinivasan et al., 2004; Groeseneken, 2005). Closer look reveals that material degradation will result in significant variation over time in parametric behaviour of transistors and wires (such as leakage and resistance) before resulting in the malfunction (Groeseneken, 2005). With appropriate circuit design techniques these device-level variations can be tolerated without violating the functional correctness but resulting in parametric variations (such as delay and energy) of components (such as functional units, memories and communication links) (Wang et al., 2005; Cosemans et al., 2007). These are traditionally hidden under worst-case clock period and made invisible to higher abstraction levels. With suitable techniques, e.g. delayed clocking or measurement-driven adaptive clocking, this dynamism can be propagated through abstraction layers and made visible to higher levels leading to dynamic hardware interface with a positive consequence of reduced guard-bands to ensure circuit-level timing correctness (Papanikolaou et al., 2005; Papanikolaou et al., 2006; Ernst et al., 2003).

The electrical and thermal stresses to which the devices are subjected to are dependent on the system usage which is highly dynamic in modern real-time systems. User inputs (e.g., switching from GPS navigation to video streaming, changes in requested frame-rate/resolution) result in a dynamically changing set of applications and their quality constraints. Smart algorithms of future systems which adapt themselves to environment changes (e.g., going from a Wi-Fi hotspot area to WiMax coverage area, fluctuating wireless channel state due to fading effects) result in a dynamic workload (computation/communication/storage resource usage). The input data being processed by the algorithm significantly influences the workload such as in a video codec or graphics rendering. Handling of all such sources of dynamism is a major design challenge of portable embedded systems where the system-level quality constraints (such as QoE, timing, temperature, lifetime) are stringent and cost (both design-time costs such as area & package and run-time costs such as energy & quality-based revenue loss) sensitivities are high. By definition, constraint must be guaranteed always, whereas “cost” should be minimized. Some quality metrics have both constraint aspect, namely minimum acceptable level, and cost aspect, namely better to have levels.

Dynamism affects, ideally, almost all design abstraction levels including hardware implementation, algorithm design, platform resource capacity sizing, package design, and mapping. Unfortunately, conventional system design approaches, knowingly or unknowingly, (try to) avoid/hide (and limit the propagation of) dynamism using worst-case abstractions. Worst-case design leads to highly sub-optimal systems when evaluated under real usage conditions - the higher the number of sources and amount of dynamism, the higher the inefficiency. Although better-than-worst-case design is an active research area, in prevailing real-time systems the maximum temperature limit and minimal reliable lifetime are still ensured at design-time using worst-case values for all influencing dynamic system aspects. Unfortunately, this is no longer economically viable in advanced technologies.

**Mixed Design-Time and Run-Time**

It is widely being accepted that cost-effective handling of dynamism requires continuous run-time system adaptation and optimization with the help of “knobs” spanning all abstraction levels, leading to the so-called self-adaptive systems (TCM-DT01, http://www.ics.uci.edu/~forge/index.html, http://rsim.cs.uiuc.edu/grace). Knobs offer run-time trade-offs between various metrics of interest and thus enable translation of variations in one system aspect into gains in another
Related Content

Activity Theory
[www.igi-global.com/chapter/activity-theory/39672?camid=4v1a](www.igi-global.com/chapter/activity-theory/39672?camid=4v1a)

Prognostics and Health Management of Choke Valves Subject to Erosion: A Diagnostic-Prognostic Frame for Optimal Maintenance Scheduling
[www.igi-global.com/chapter/prognostics-health-management-choke-valves/69685?camid=4v1a](www.igi-global.com/chapter/prognostics-health-management-choke-valves/69685?camid=4v1a)

A Task Context Aware Physical Distribution Knowledge Service System
[www.igi-global.com/article/task-context-aware-physical-distribution/53042?camid=4v1a](www.igi-global.com/article/task-context-aware-physical-distribution/53042?camid=4v1a)

Compensating Resource Fluctuations by Means of Evolvable Hardware: The Run-Time Reconfigurable Functional Unit Row Classifier Architecture
[www.igi-global.com/article/compensating-resource-fluctuations-means-evolvable/74364?camid=4v1a](www.igi-global.com/article/compensating-resource-fluctuations-means-evolvable/74364?camid=4v1a)