Closed-Loop Control for Power and Thermal Management in Multi-Core Processors: Formal Methods and Industrial Practice

(Extended Abstract)

Ibrahim (Abe) M. Elfadel
Masdar Institute of Science and Technology
Abu Dhabi, UAE
ielfadel@masdar.ac.ae

Radu Marculescu
Carnegie Mellon University
Pittsburg, PA, USA
radum@ece.cmu.edu

David Atienza
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
david.atienza@epfl.ch

The need to use feedback to come up with context-dependent and workload-aware strategies for runtime power and thermal management (PTM) in high-end and mobile processors has been advocated since the early 2000. Two seminal papers that appeared in 2002 [1], [2] defined a framework for the use of feedback mechanisms for power and temperature control. In [1], the focus was on power management with the goal being to extend battery life on the AMD Mobile Athlon. This was one of the earliest papers to use DVFS settings as actuators to guarantee a given energy level in the battery at the end of a given time interval. The controller was implemented using a combination of OS files and Linux kernel modules. Almost simultaneously, [2] posed the dynamic thermal management task as a formal control-theoretic problem requiring the thermal modeling of the processor and the use of the established control structures of classical feedback theory. Some of the defining features of [2] include the development of layout-based thermal RC models for the processor; the use of an architecturally-driven control mechanism, namely, the instruction fetching rate; and the use of the SPEC2000 benchmarks to illustrate temperature control action under various workloads. The controller used in [2] is a Proportional-Integral-Differential (PID) structure whose input is the deviation of the sensed temperature from the target temperature and whose output is the toggle rate of the instruction fetching mechanism.

These two early papers were concerned with runtime power and thermal feedback control in a single-core processor with [2] noticing the stringent requirements imposed by the looming "power wall." Rather than lessening the dependence on feedback for power and thermal management, the advent of dual and multi-core processors to mitigate the impact of the "power wall" has in fact motivated experts in industry and academia to rely even more on feedback mechanisms for power and thermal management. A case in point is Intel's Foxton Technology for power and thermal control in the Itanium family of dual core processors [3]. This technology was one of the earliest industrial attempts at using on-chip hardware micro controllers for power and thermal control of multi-core processors. The technology featured per-core power meters and temperature sensors with the possibility to modulate both supply voltage and clock rate. Reported measurements indicated 30% reduction in power for a 10% reduction in frequency. Another dual core processor that also featured a feedback mechanism for chip-level power budgeting and thermal capping was IBM's Power 6 [4].

The formal analysis, design and verification of feedback structures for multi-core processors and multi-processor System-on-Chip (MPSoC) require the development of rigorous dynamical models linking the controlled variables to the controlling ones. Modern control theory uses the framework of state-space representations for such models. Table I is a summary of a variety of state-space models that have been used for PTM since the early 2000 along with their controller and actuator types. These state-space models may be categorized into three different categories: power-based, queue-utilization-based, and temperature-based.

A. Power-Based Models

These are state-space models where the state-space variable is the power of each core in the processor. Typically, these power-based models are linear models with the control input being the effective per-core frequency of the DVFS policies. The power-based models originated in the power-capping problem for a single server [5], have been extended to power capping in server clusters and data centers [6], and the extended models have been used for power management in multi-core processors [7]. In the latter case, the power budget of the full processor has to be partitioned among the various cores according to a processor figure of merit. An early open-loop policy for core power budgeting is the MaxBIPS algorithm which used processor throughput maximization as the criterion for assigning power states to the various cores [8]. One challenge of per-core DVFS closed-loop MPSoC power management is the sensing of the per-core power in real time. Core power proxies based on hardware activity monitors have been proposed as on-line power estimators. The EnergyScale™ power management platform of IBM's eight-

978-3-9815370-0-0/DATE13/©2013 EDAA

TABLE I
MODELS AND CONTROLLERS: P (PROPORTIONAL), LQR (LINEAR QUADRATIC REGULATOR), PID (PROPORTIONAL-INTEGRAL-DIFFERENTIAL), MPC (MODEL PREDICTIVE CONTROLLER), DVFS (DYNAMIC VOLTAGE AND FREQUENCY SCALING)

State Variable	State Space Equations	Controller	Actuator	References
Power	Linear	P	DVFS	[1]
Power	Linear	P	DVFS	[5]
Power	Linear	MPC	DVFS	[7]
Queue	Linear(ized)	PID	DVFS	[10]
Queue	Linear	LQR	DVFS	[11], [12]
Queue	Fractional	LQR	DVFS	[13]
Temperature	Linear	PID	Toggle rate	[2]
Temperature	Linear	LQR	DVFS	[14]
Temperature	Linear	MPC	DVFS	[15], [16]

core Power7 processor uses such proxies for processor power budgeting [9].

B. Queue-Utilization-Based Models

Per-core DVFS opens up new possibilities for managing power in MPSoC made of multiple, asynchronous voltage-frequency islands (VFI's). One such possibility is to adopt a network view of the MPSoC and to consider the network traffic among the multiple VFI's as the object of state-space modeling. An early representative of such network-based approach to state-space modeling and power-performance management is [10] where the states of the model are the utilizations of the FIFO queues at the interfaces between the different VFI's. The controller used in [10] is a classical PID structure with two versions: distributed and local. The Network-on-Chip (NoC) perspective is fully described along with several outstanding research problems and an extensive literature survey up until 2008 in [17]. In [11], the work of [10] has been extended to include a full accounting of the routing infrastructure, the leakage power, and the power overhead resulting from adding more VFI's. Furthermore, the control design uses full-state feedback with a gain matrix generated using eingenvalue placement methods. One drawback of the full-state feedback controller is its centralized nature, which makes it difficult to scale up with the number of VFI's. This drawback has been addressed in [12] where a greedy algorithm for the sparsification of the gain matrix is proposed. Such algorithm allows the exploration of the full feedback connectivity space from the purely local controller to the purely centralized. A very important feature of this class of control structures for power management is that they enable the modeling of the fractal nature of on-chip network traffic. Such modeling is achieved using fractional state-space models [13]. There is also a fully developed fractional optimal control theory that can be used to derive feedback control algorithms adapted to the network workloads [13].

C. Temperature-Based Models

While the previous two classes of models are devoted to power management, temperature-based models are specifically used for thermal management, temperature emergency monitoring, and hot spot mitigation. They are typically based on a layout-based, distributed thermal RC model of the MPSoC. An important tool for the generation of such models is HotSpot [18]. The thermal RC model is a linear state-space model with the input being the power map resulting from the workload of the MPSoC. It is in fact a 3D model as it not only accounts for the horizontal diffusion of heat among the various units of the MPSoC, but also it accounts for the vertical heat diffusion from the substrate to the heat sink. An accurate thermal RC model may have a large number of states and may thus be amenable to reduced-order modeling [14]. Several feedback control strategies for thermal management have been proposed that go far beyond the early PID controller of [2]. They are based on modern control and optimization techniques and are surveyed in [19]. The advent of 3D integrated circuits along with their new and challenging problems of thermal management has required a fresh look at the existing cooling techniques which appear to be inadequate for 3D MPSoC. Liquid-Cooled 3-D stacked architectures have been proposed and thermal control strategies evaluated to achieve energy-efficient thermal management for 3-D MPSoC [20]. The evaluation of such thermal control strategies requires of course the modeling of the full system, including the cooling subsystem and the feedback subsystem. Furthermore, the design and verification of such complex thermal systems will usher new methods for fast thermal analysis similar to the one proposed in [21].

It is important to note that the above three classes of frameworks are not independent. As previously noted, the power map is an input to the temperature state-space model. Similarly, in [13], the finite-horizon objective function used to define the optimal control policy of the fractional model includes the equivalent of a power term. In [22], an optimization framework is used to define policies for power management in 3-D multi-core architectures under both peak power and temperature constraints. Including constraints on state and control variables is one of the main motivations for using model predictive control (MPC) formulations as is the case in [23], [15], [16].

Conclusions

In summary, this embedded tutorial will bring DATE attendees to the forefront of the latest academic research and industrial practice in the area of closed-loop PTM for MPSoC. Starting with power capping techniques based on classical control theory, the tutorial will cover the more advanced techniques of optimal control, model predictive control, and adaptive control. Practical issues such as power and thermal proxies, power and thermal sensors, and various actuation techniques will be surveyed. Furthermore, the tutorial will address optimal power control techniques in NoC architectures, with particular attention to methods for handling multiple voltage and frequency domains under variable workloads. Finally, it will cover recent techniques for pro-active and reactive closed-loop temperature control for 2D and 3D MPSoC, including

the handling of emerging inter-tier liquid cooling techniques. Important emerging problems such as heterogeneity of the computational fabric and scalability of the control methods will also be discussed along with their emerging solutions.

Acknowledgement

Abe Elfadel would like to acknowledge SRC support under Contract 2011-TJ-2192. Radu Marculescu would like to acknowledge NSF support under Contract CNS-1128624 and GRC support under Contracts 2011-HJ-2168 and 2012-HJ-2232. David Atienza would like to acknowledge the support of the Nano-Tera RTD project CMOSAIC (ref.123618), financed by the Swiss Confederation and scientifically evaluated by SNSF, and the support of the Spanish Ministry of Economy and Competitiveness under Contract TEC2012-33892.

REFERENCES

[1] R.J. Minerick, V.W. Freeh, and P.M. Kogge. Dynamic power management using feedback. 2002.

[2] K. Skadron, T. Abdelzaher, and M.R. Stan. Control-theoretic techniques and thermal-rc modeling for accurate and localized dynamic thermal management. In *Eighth International Symposium on High-Performance Computer Architecture (HPCA 2002)*, pages 17–28, 2002.

[3] R. McGowen, C.A. Poirier, C. Bostak, J. Ignowski, M. Millican, W.H. Parks, and S. Naffziger. Power and temperature control on a 90-nm itanium family processor. *IEEE Journal of Solid-State Circuits*, 41(1):229 – 237, jan. 2006.

[4] MS Floyd, S. Ghiasi, TW Keller, K. Rajamani, FL Rawson, JC Rubio, and MS Ware. System power management support in the ibm power6 microprocessor. *IBM Journal of Research and Development*, 51(6):733–746, 2007.

[5] C. Lefurgy, Xiaorui Wang, and M. Ware. Server-level power control. In *Autonomic Computing, 2007. ICAC '07. Fourth International Conference on*, page 4, june 2007.

[6] Xiaorui Wang, Ming Chen, C. Lefurgy, and T.W. Keller. Ship: A scalable hierarchical power control architecture for large-scale data centers. *IEEE Transactions on Parallel and Distributed Systems*, 23(1):168 –176, jan. 2012.

[7] Kai Ma, Xue Li, Ming Chen, and Xiaorui Wang. Scalable power control for many-core architectures running multi-threaded applications. In *38th Annual International Symposium on Computer Architecture (ISCA 2011)*, pages 449 –460, june 2011.

[8] C. Isci, A. Buyuktosunoglu, C.Y. Cher, P. Bose, and M. Martonosi. An analysis of efficient multi-core global power management policies: Maximizing performance for a given power budget. In *39th Annual IEEE/ACM International Symposium on Microarchitecture, 2006, MICRO-39.*, pages 347–358, 2006.

[9] M. Floyd, M. Ware, K. Rajamani, T. Gloekler, B. Brock, P. Bose, A. Buyuktosunoglu, J. C. Rubio, B. Schubert, B. Spruth, J. A. Tierno, and L. Pesantez. Adaptive energy-management features of the ibm power7 chip. *IBM Journal of Research and Development*, 55(3):8:1 –8:18, may-june 2011.

[10] Q. Wu, P. Juang, M. Martonosi, and D.W. Clark. Formal online methods for voltage/frequency control in multiple clock domain microprocessors. *ACM SIGARCH Computer Architecture News*, 32(5):248–259, 2004.

[11] U.Y. Ogras, R. Marculescu, D. Marculescu, and E.G. Jung. Design and management of voltage-frequency island partitioned networks-on-chip. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 17(3):330–341, 2009.

[12] S. Garg, D. Marculescu, and R. Marculescu. Custom feedback control: enabling truly scalable on-chip power management for mpsocs. In *ACM/IEEE International Symposium on Low-Power Electronics and Design (ISLPED 2010)*, pages 425–430, 2010.

[13] P. Bogdan, R. Marculescu, S. Jain, and R.T. Gavila. An optimal control approach to power management for multi-voltage and frequency islands multiprocessor platforms under highly variable workloads. In *Sixth IEEE/ACM International Symposium on Networks on Chip (NoCS 2012)*, pages 35–42, 2012.

[14] F. Zanini, D. Atienza, and G. De Micheli. A control theory approach for thermal balancing of mpsoc. In *Asia and South Pacific Design Automation Conference (ASP-DAC 2009)*, pages 37–42, 2009.

[15] F. Zanini, D. Atienza, L. Benini, and G. De Micheli. Multicore thermal management with model predictive control. In *European Conference on Circuit Theory and Design (ECCTD 2009)*, pages 711–714, 2009.

[16] F. Zanini, C.N. Jones, D. Atienza, and G. De Micheli. Multicore thermal management using approximate explicit model predictive control. In *IEEE International Symposium on Circuits and Systems (ISCAS 2010)*, pages 3321–3324, 2010.

[17] R. Marculescu, U.Y. Ogras, L.S. Peh, N.E. Jerger, and Y. Hoskote. Outstanding research problems in noc design: system, microarchitecture, and circuit perspectives. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 28(1):3–21, 2009.

[18] W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M.R. Stan. Hotspot: A compact thermal modeling methodology for early-stage vlsi design. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 14(5):501–513, 2006.

[19] F. Zanini, D. Atienza, C.N. Jones, L. Benini, and G. De Micheli. Online thermal control methods for multiprocessor systems. *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 18(1):6, 2012.

[20] M.M. Sabry, A.K. Coskun, D. Atienza, T.S. Rosing, and T. Brunschwiler. Energy-efficient multi-objective thermal control for liquid-cooled 3d stacked architectures. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 30(12):1883 – 1896, Dec. 2011.

[21] A. Sridhar, A. Vincenzi, M. Ruggiero, and D. Atienza. Neural network-based thermal simulation of integrated circuits on gpus. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 31(1):23–36, 2012.

[22] K. Kang, J. Kim, S. Yoo, and C.M. Kyung. Runtime power management of 3-d multi-core architectures under peak power and temperature constraints. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 30(6):905–918, 2011.

[23] X. Wang, K. Ma, and Y. Wang. Adaptive power control with online model estimation for chip multiprocessors. *IEEE Transactions on Parallel and Distributed Systems*, 22(10):1681–1696, 2011.