A Cross-layer Framework for Temporal Supply Noise Prediction

Yaguang Li. Advisor: Jiang Hu.
ACM Number: 4606439. Category: Graduate.
Department of Electrical & Computer Engineering
Texas A&M University, College Station, USA, 77843
E-mail: liyg@tamu.edu

I. Problem and Motivation

Supply noise margin in current and future processor chips has been significantly reduced as the supply voltage level continues to decrease [1]. On the other hand, multicore chips see larger supply noise variations in both temporal and spatial domains due to the heterogeneous workloads executed on different cores concurrently [2]. All these facts threaten the chip reliability in the form of supply emergency [3]–[5]. To avoid functional errors induced by supply emergency, supply noise prediction is needed for processor chips. In the past decades, people developed voltage sensor based methods to detect supply noise on chips [4], [6]–[8]. But those sensor based methods fail to detect/predict system level workload behaviors and ignore back-end power delivery details. Thus to better understand the correlation between system-level workload and circuit-level supply noise fluctuation, an efficient cross-layer noise emergency prediction method is desired.

II. Background and Related Work

In the literature, people employ voltage sensors [6]–[8] to monitor the supply noise at certain positions in a processor chip [9]–[12]. Their idea is to correlate the supply noise spatially by either statistic or machine learning methods, then predicts supply noise at target positions by observing supply noise at voltage sensors. But their methods are limited by the following facts: 1). Their methods fail to detect/predict system level workload behaviors. 2). Their methods typically have large time overhead to detect the emergencies and recover the system to the correct status. [4] has noticed the disadvantage of time overhead in those methods. It proposes a signature-based method to predict the supply emergency at the system-level. Based on the embedded voltage sensors on chip, they consistently extracts and records the signature patterns (i.e., micro-architecture-level events) that result in voltage emergencies. Although [4] has overcome the disadvantage of time overhead, all the methods mentioned above highly rely on voltage sensors and ignore the back-end power delivery details.

In practice, time-varying workload dominates power consumption of processor chips while the supply noise is mainly determined by the consumed power. Temporal supply noise prediction for on-the-fly chips is a cross-layer problem, as workload is estimated/monitored at system level but power and supply noise are conventionally considered as back-end metrics. Conventionally, workload can be simulated by an architectural simulator with a speed of millions instructions per second, the latter two need micro-architectural level and layout level simulators to achieve at a rate of one instruction per minute or even lower. To achieve temporal supply noise prediction, we have to bridge the front-ends and back-ends into one same platform.

III. Approach and Uniqueness

In this work, we propose an system-level framework for temporal supply noise prediction, as shown in Fig. 1. The framework consists two stages:

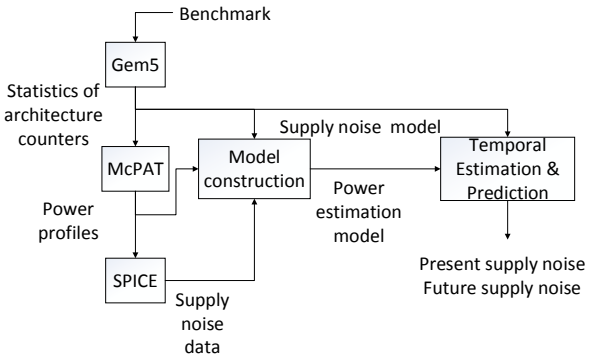

Fig. 1: Overview of the proposed framework.

- **Model Construction**: In this stage, an system level power estimation model [13] and a proposed circuit level supply noise model [14] are constructed. The data profiles to construct those models are obtained from architecture simulator GEM5 [15], power simulator McPAT [16] and SPICE.
- **Temporal Supply Noise Prediction**: In this stage, the framework predicts supply noise by collaborating an ARIMA-based power prediction model [14] with aforementioned constructed models. It should be noted the ARIMA-based prediction model is constructed/updated dynamically according to the running benchmarks on process chips.

Our framework models the interaction among system workload, micro-architecture power and circuit level supply noise altogether, and no voltage sensors are required. The entire

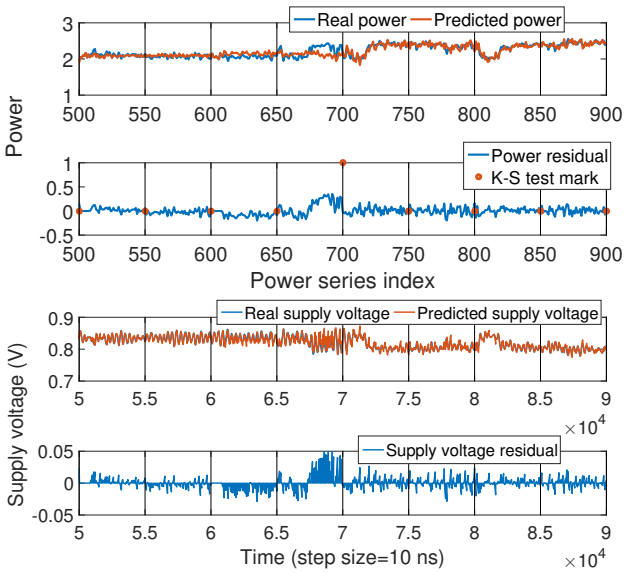

Fig. 2: Power and Supply noise prediction samples

framework only takes system level performance counters as inputs, then power/supply noise can be continually predicted. This feature is supported by:

- **System Level Power Estimation Model**: This model provides the framework with fast mapping capability from performance counters to chip power, which helps to estimate the current power behaviors of the running benchmarks on chips.
- **Micro-architecture-level ARIMA-based Power Prediction Model**: This model extracts the characteristic of estimated power series. Then it further reports the future power behavior and updates itself if necessary. With this model, we predict the future power behaviors of the running benchmarks.
- **Circuit Level Supply Noise Model**: Taking both the estimated and predicted power as inputs, this model reports the future supply noise behaviors. This model is based on the characteristic of Linear Time Invariant (LTI) system and the convergence of LTI's response. Compared with traditional MNA method [17], this model provides a new fast simulation approach for RLC network.

IV. Results and Contributions

Fig. 2 shows how our framework predicts the future power and supply noise behaviors. Each node of power series is estimated/predicted every 1 μs, ie., each power index represents 1 μs. For power series with index $\in [500, 650]$, the framework predicts the future power accurately. When the index comes between $[650, 700]$, predicted power mismatches (up tp 8%) with real power but the framework detects the mismatches (as indicated by K-S test mark, which is triggered to 1). Then the framework updates the ARIMA-based power prediction model and continues to predict accurately for the rest of power series. Moreover, the framework predicts the supply noise concurrently. A mismatch (up to 0.05V) in time window $[6.5\times10^4 ns, 7\times10^4 ns]$ can be observed, which is due to the power phase change. We test 14 standard benchmarks of SPEC CPU2006 [18] on a processor chip with an area of 17.91 mm^2. Experiment results shows the framework accomplishes supply noise prediction with 1.58%/1.17% relative error under X86/ARM architecture. Meanwhile, the proposed ARIMA-based power prediction model improves up to 37.5%/63% performance than prior methods in [19], under X86/ARM architecture.

Our framework provides a solution for run time supply noise prediction. The framework includes three cross-layer models, namely power estimation/prediction and supply noise model. The solution helps to make system level decisions for reliability concerns, especially in the form of power and supply noise. Meanwhile this framework are effective for early stage design, which not only helps to explore design space among system, micro-architecture and circuit level, but also definitely speed up the entire design cycle.

References

[1] J. Fang, S. Gupta, S. Kumar, S. Marella, V. Mishra, P. Zhou, and S. Sapatnekar, "Circuit reliability: From physics to architectures (embedded tutorial)," in *IEEE/ACM International Conference on Computer-Aided Design*, 2012, pp. 243–246.

[2] M. S. Gupta, J. L. Oatley, R. Joseph, G. Y. Wei, and D. M. Brooks, "Understanding voltage variations in chip multiprocessors using a distributed power-delivery network," in *Design, Automation and Test Conference in Europe*, 2007, pp. 624–629.

[3] P. Zhou, K. Sridharan, and S. S. Sapatnekar, "Congestion-aware power grid optimization for 3D Circuits Using MIM and CMOS decoupling capacitors," in *The IEEE Asia-South Pacific Design Automation Conference*, 2009, pp. 179–184.

[4] V. J. Reddi, M. S. Gupta, G. Holloway, and G. Y. Wei, "Voltage emergency prediction: Using signatures to reduce operating margins," in *IEEE International Symposium on High Performance Computer Architecture*, 2009, pp. 18–29.

[5] T. Wang, P. W. Luo, Y. S. Su, L. C. Cheng, D. M. Kwai, and Y. Shi, "Capturing the phantom of the power grid - on the runtime adaptive techniques for noise reduction," in *ACM/EDAC/IEEE Design Automation Conference*, 2012, pp. 640–645.

[6] C. Lefurgy, A. Drake, M. Floyd, M. Allen-Ware, B. Brock, J. Tierno, J. Carter, and R. Berry Jr, "Active guardband management in power7+ to save energy and maintain reliability," *IEEE Micro*, vol. 33, no. 4, pp. 35–45, 2013.

[7] K. Bowman, C. Tokunaga, J. Tschanz, and A. Raychowdhury, "Dynamic variation monitor for measuring the impact of voltage droops on microprocessor clock frequency," in *IEEE Custom Integrated Circuits Conference*, 2010, pp. 1–4.

[8] N. James, P. Restle, J. Friedrich, and B. Huott, "Comparison of split-versus connected-core supplies in the power6 microprocessor," in *IEEE International Solid-State Circuits Conference*, 2007, pp. 298–604.

[9] T. Wang, C. Zhang, J. Xiong, and Y. Shi, "Eagle-eye: A near-optimal statistical framework for noise sensor placement," in *IEEE/ACM International Conference on Computer-Aided Design*, 2013, pp. 437–443.

[10] X. Liu, S. Sun, X. Li, H. Qian, and P. Zhou, "Machine learning for noise sensor placement and full-chip voltage emergency detection," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 36, no. 3, pp. 421–434, 2017.

[11] X. Liu, S. Sun, P. Zhou, and X. Li, "A statistical methodology for noise sensor placement and full-chip voltage map generation," in *ACM/EDAC/IEEE Design Automation Conference*, 2015, pp. 1–6.

[12] T. Wang, C. Zhang, J. Xiong, and Y. Shi, "On the deployment of on-chip noise sensors," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 33, no. 4, pp. 519–531, 2014.

[13] G. Contreras and M. Martonosi, "Power prediction for intel xscale® processors using performance monitoring unit events," in *IEEE International Symposium on Low Power Electronics and Design*, 2005, pp. 221–226.

[14] Y. Li, C. Zhuo, and P. Zhou, "A cross-layer framework for temporal power and supply noise prediction," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 1–14, 2018 (To appear).

[15] N. Binkert, B. Beckmann, G. Black, A. Saidi, A. Saidi, A. Basu, J. Hestness, D. R. Hower, T. Krishna, and S. Sardashti, "The gem5 simulator," *ACM Sigarch Computer Architecture News*, vol. 39, no. 2, pp. 1–7, 2011.

[16] S. Li, J. H. Ahn, R. D. Strong, J. B. Brockman, D. M. Tullsen, and N. P. Jouppi, "McPAT: An integrated power, area, and timing modeling framework for multicore and manycore architectures," in *IEEE/ACM International Symposium on Microarchitecture*, 2010, pp. 469–480.

[17] C. W. Ho, A. E. Ruehli, and P. A. Brennan, "The modified nodal approach to network analysis," *IEEE Trans Circuits Syst*, vol. 22, no. 6, pp. 504–509, 1975.

[18] http://www.spec.org/cpu2006/, "[online]."

[19] E. Duesterwald, C. Cacaval, and S. Dwarkadas, "Characterizing and predicting program behavior and its variability," in *International Conference on Parallel Architectures and Compilation Techniques*, 2003, pp. 220–231.