

DVFS in Presence of Process Variations

Diana Marculescu

▶ To cite this version:

Diana Marculescu. DVFS in Presence of Process Variations. ISCA tutorial on "Multi-domain Processors: Challenges, Design Methods, and Recent Developments", Jun 2010, Saint Malo, France. inria-00493908

HAL Id: inria-00493908 https://inria.hal.science/inria-00493908

Submitted on 21 Jun2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Multi-Core, Variations, and Power Management

• Chip-multiprocessor is here \rightarrow New use of Moore's Law: 2x number of cores every technology generation



- + Smaller cores (more cores per die)
- Increased process variability

45nm, 4 cores

Designs moving toward local clock/voltage control \rightarrow power management per core/DVFS per voltage-frequency island (VFI)

RGM2- ISCA'10

Design Variability and Power Management

- Existing power management techniques
 - Oblivious to manufacturing process induced uncertainty...
 - ... while relying exclusively on workload-induced variations, without considering design variations
- To be able to cope with increased design variability, power management mechanisms need to:
 - Incorporate reliable models for design variability early in the process
 - **Be able to work with variability models for** *system components* to determine system behavior
 - Allow for seamless adaptation to hardware characteristics, while relying on existing dynamic power management

Outline





Dynamic voltage/frequency scaling

- Reduces both dynamic and static power
- Control algorithm attempts to lower performance cost
- Current algorithms are unaware of variations
 - [Juang et al. ISLPED'05], [Isci et al. MICRO'06], [Herbert and Marculescu ISLPED'07]
 - Treat all cores as being identical
 - Result in wasted power
 - Single software-based exception [Teodorescu ISCA'08]
- Variability-aware DVFS attempts to improve energy-efficiency
 - Reduce power while maintaining performance [Herbert and Marculescu HPCA'09]

139

RGM2-ISCA'10

Possible Solutions

- Develop dynamic control algorithms from scratch
 - Characterize each die separately post-manufacturing
 - Include those characteristics in newly developed algorithms
 - + Attacks the problem with a specific solution
 - May make a wealth of already existing algorithms potentially unusable
 - Requires pre-characterization at test/post-manufacturing
- Adapt existing dynamic control algorithms to including variations
 - Modify existing algorithms to include per core/per VFI models for process variations
 - + Easy to reuse existing control algorithms
 - Some pre-characterization at test/post-manufacturing required

Physical Parameter Modeling

- Model of spatially-correlated intra-die V_{th} and L_{eff} variation
- Multivariate normal with spherical correlation function [Sarangi TSM'08]



Core Power/Performance Modeling

- Fit response surface models to SPICE data [Herbert and Marculescu HPCA'09,TCOMP'09]
 - 22nm hi-K metal gate PTM models [Zhao TED'06]
 - ▼ Lump L_{eff} and V_{th} variation into process variation parameter $p \rightarrow$ capture correlations
- Models for frequency and leakage
 - **As functions of** p, V_{dab} and temperature T
 - ▼ Model form is *e*^{polynomial}
 - Frequency model tracks 13-stage FO4 ring oscillator
 - **Leakage model tracks** I_{sd} of FETs with $|V_{ds}| = V_{dd}$ and $V_{qs} = 0$

Models fit to minimize maximum absolute percent error

Iterative pruning removes coefficients until error doubles

Core-to-core Variability Characterization



Exposing Variations to DVFS Algorithms

Variability model has many points per VFI, each with own p

- Need a simple way of aggregating and exposing this information at microarchitecture level
- *p_{eff}* is *p* value that best tracks VFI's leakage across *V_{dd}* and temperature
- Computed based on four measurements of VFI leakage
 - High and low temperature and V_{dd}
 - Can be done based on IDDQ test results

Negligible error compared to modeling many points per core

- Tested across 1,000 dies and 13,456 (V_{dd}, 7) pairs
- ▼ Single-VFI MAPE of 0.011% for single-core VFIs
- ▼ Single-VFI MAPE of 0.014% for quad-core VFIs

Exploiting Variability Information



Threshold DVFS Algorithm

- Keeps utilization in a specified range [Herbert and Marculescu ISLPED'07]
 - Utilization = (instructions retired / number of retire slots)
 - ◄ Higher VF level → lower utilization
 - Lower utilization target → higher VF level
- Threshold-unaware uses range of [0.2, 0.4] for all VFIs
- Threshold-aware sets threshold based on each VFI's variations

Exponential scaling applied

- ▼ VFI *i* thresholds set to [e^{-p}eff,i · T_{down}, e^{-p}eff,i · T_{up}]
- Target band [0.279, 0.558] for core in 5th percentile of *p* distribution (most leaky)
- Target band [0.139, 0.278] or core in 95th percentile of *p* distribution (least leaky)

Greedy DVFS Algorithm

- Attempts to minimize power/throughput (energy per instruction)
 - [Herbert and Marculescu ISLPED'07] based on [Magklis et al. ISLPED'06]
 - Performs greedy search to find the optimal VF level
 - Upon overshooting, moves back to optimal and holds for *H* intervals
- *Greedy-aware* scales energy per instruction (EPI) to bias VFIs
- Exponential scaling applied
 - **VFI** *i* metric for interval *n* is $SEPI_{n,i} = EPI_{n,i} \cdot (e^{p_{eff,i}})^{L_n}$

	VF level, <i>L</i>	V _{dd}	<i>SEPI</i> scaling for 5 th percentile (leakier)	<i>SEPI</i> scaling for 95 th percentile (less leaky)
	0	0.9 V	1.00	1.00
	1	0.8 V	0.70	1.39
	2	0.7 V	0.49	1.94
	3	0.6 V	0.34	2.71
RGM2-ISCA'10				147

Custom Solution: Linear Optimization (LinOpt)

• Linear programming:

- Maximize objective function: $f(x_1,...,x_n)$, with $x_1,...,x_n$ independent
- Subject to constraints such as: g(x1,...,xn) < C
- f,g are linear functions
- Unknowns: voltages V₁,...,V_n for all cores
- Objective function: maximize CMP throughput
 - Throughput (MIPS) = Frequency X IPC = f(V₁,...,V_n)
- Constraint: keep power under Ptarget
 - Power = g(V)
- Possible issue: functions *f* and *g* are NOT linear; worst-case complexity is exponential [Teodorescu et al. ISCA'08]

Experimental Setup



Power-Performance



Low throughput loss

- Worst-case throughput loss under 4%
- Significant improvement in power/throughput
 - 8.0%/6.5% for fine-/coarse-grained VFIs on Threshold
 - **15.4%/9.2% for fine-/coarse-grained VFIs on** *Greedy*

Comparison at Iso-throughput



Lose some benefit for fine-grained VFIs, gain for coarse-grained

Var.-aware Threshold or Greedy vs. LinOpt



- LinOpt power budget set equal to power used by alternate scheme
- Variability-awareness mean benefits up to 10%
- *LinOpt* performance highly dependent on linearizing accuracy

Power Profile Results



Summary

Adaptive Body-Biasing (ABB)

Powerful technique to mitigate variations in leakage power dissipation [Tschanz et al. JSSC'02]



How About Combining Body Biasing & DVFS?

- V_{DD} and body biasing provide two power/performance knobs
 - Nominal V_{DD} with RBB may not yield best power for given frequency
 - V_{DD} has stronger effect on dynamic power, BB has stronger effect on static power
- Independent implementation of DVFS and body biasing
 - DVFS algorithms may stay unchanged
 - Reclaim frequency margin by adjusting body bias until frequency is exactly met
- Integration of DVFS and body biasing
 - The system specifies desired frequency
 - Goal is to chose the V_{DD} / BB combination with the lowest total power

RGM2-ISCA'10



Main Motivation

- Large variation in leakage as percent of total power
- Leads to large variation in optimal V_{DD} / BB combination
 - ▼ Lower % leakage needs lowest V_{DD} for minimum power
 - ◄ Higher % leakage needs highest V_{DD} for minimum power



Proposed Approach

Delay the { V_{DD}, f } mapping until test-time [Bonnoit et al. ISLPED09]

Exploit variability information for each core (or VFI)

- Available frequency levels are chosen as usual
- V_{DD} for each f is chosen at test based on a single leakage measurement
- Function mapping leakage to V_{DD} for each f defined once per technology / product

159

Body biasing used to reclaim performance margin



Example Mapping



Example Schemes

- ZBB: Traditional DVFS with no body-biasing
- Independent: Independent ABB and DVFS
- *Minimum*: ABB and DVFS, run each core at its minimum V_{DD}
 Assume this could be found at test (even though cost would be prohibitive)
- Ratio: V_{DD} and BB set to meet f, achieve target ratio of switching to total power [Nomura JSSC'06]
 - Ratio chosen to minimize total power across all frequency levels in Monte Carlo
- TTVS: Proposed scheme

Results by Speed Bin and DVFS Granularity



Results by DVFS Time Distribution



- Same trends hold across various distributions of time per voltage level
 - Uniform → the amount of time spent at each voltage level is uniformly distributed
 - Skewed-high, low or mid \rightarrow time spent at each level is proportional to the level *i*, *n*-*i*, or min (*n*, *n*-*i*) (*n* = total number of levels)

Outline



Technology-driven Limits



Time-optimal Control



Process Variations

 Impact of process variations on DVFS controller performance [Garg et al. DAC09]

Chip-to-chip variations in the voltage-frequency curves of VFIs



- Assuming maximum supply voltage is fixed, *F_{max}* will be different for each manufactured die
- Probability of Controllability (PoC): % of manufactured die that meet performance specification
- RGM2-ISCA'10
 Computed using Monte-Carlo simulations

Explicit Energy Minimization

 Objective: minimize energy spent by controller to reach the steady state

$$E_{total} = \sum_{i=1}^{M} \sum_{k=1}^{J} TC_i V_i(k)^2 f_i(k)$$

Approximate E_{total} only as a function of f(k) since constraints are also functions of the VFI frequencies



$$E_{total} \approx \sum_{i=1}^{M} \sum_{k=1}^{J} \alpha_i f_i(k)^2 + \beta_i f_i(k) + \lambda_i$$

Previous Linear Programming (LP) problem converted to a Quadratic Program (QP)

171

RGM2-ISCA'10

Experimental Results



- Controller response for MPEG-2
 - Both queues empty
 - ▼ *J = 10*

$$\checkmark \gamma = f_{max}/f_{nom} = 1.5$$

$$f_{step} = 10\% f_{nom}$$

 Observed about 9% energy savings for minimum energy controller

Experimental Results



Summary

- Including process variation information in power control algorithms is essential
 - Lose significant power savings at iso-throughput or performance at iso-power
 - Customized algorithms suboptimal → best bet are variationaware versions of already existing dynamic control algorithms
- Scalability and controllability of dynamic power management algorithms are key
 - Especially with increased number of cores ...
 - ... and increased impact of variations multi-core systems

Conclusions

<list-item><list-item><list-item><list-item><list-item><list-item><list-item>

References relevant to this tutorial (General)

- G. De Micheli, L. Benini, 'Networks on Chips: Technology and Tools,' Morgan Kaufmann, 2006.
- W. J. Dally and B. Towles, Principles and Practices of Interconnection Networks. San Mateo, CA: Morgan Kaufmann, 2004
- J. Duato, S. Yalamanchili, and L. Ni, Interconnection Networks: An Engineering Approach. San Mateo, CA: Morgan Kaufmann, 2002
- R. Marculescu, et al., 'Computation and Communication Refinement for Multiprocessor SoC Design: A System-Level Perspective,' in ACM TODAES, Vol.11, No.3, July 2006.
- R. Marculescu, et al., 'Outstanding Research Problems in NoC Design: System, Microarchitecture, and Circuit Perspectives', in IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems (TCAD), vol. 28, no. 1, pp. 3-21, Jan. 2009
- T. Bjerregaard and S. Mahadevan, "A survey of research and practices of network-on-chip," ACM Comput. Surv., vol. 38, no. 1, pp. 1–51, Mar. 2006
- J. Henkel, W. Wolf, and S. Chakradhar, "On-chip networks: A scalable, communication-centric embedded system design paradigm," in Proc. VLSI Des., Jan. 2004, pp. 845–851
- Milos Krstic, et al, "Globally Asynchronous, Locally Synchronous Circuits: Overview and Outlook", IEEE Design and Test of Computers, Sept.-Oct. 2007
- A large selection of NoC papers is available at http://www.cl.cam.ac.uk/~rdm34/onChipNetBib/browser.htm http://www.ocpip.org/university/biblio_main/comparison/

References (Part III)

- U. Y. Ogras, et al., 'Design and Management of Voltage-Frequency Island Partitioned Networks-on-Chip,' in IEEE Trans. VLSI, March 2009.
- P. Choudhary, D. Marculescu, 'Power Management of Voltage/Frequency Island-Based Systems Using Hardware Based Methods,' in IEEE Trans. on VLSI, March 2009
- T. Simunic, S. P. Boyd, P. Glynn, 'Managing power consumption in networks on chips,' in IEEE Trans. on VLSI, Jan. 2004.
- A. Alimonda, et al., 'Feedback-Based Approach to DVFS in Data-Flow Applications,' in IEEE Trans. on CAD of Integrated Circuits and Systems 28(11): 1691-1704 (2009)
- E. Beigne, et al., 'Dynamic voltage and frequency scaling architecture for units integration within a GALS NoC,' in Proc. Int. Symp. Netw. Chip, 2008, pp. 129–138.
- C.-L. Chou and R. Marculescu, 'Energy- and performance-aware incremental mapping for networks on chip with multiple voltage levels,' IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol. 27, no. 10, pp. 1866–1879, Oct. 2008
- Q. Wu, et al, 'Formal Online Methods for Voltage/Frequency Control in Multiple Clock Domain Microprocessors', in Proc. ASPLOS 2004
- G. Semeraro, et al, 'Energy efficient processor design using multiple clock domains with dynamic voltage and frequency scaling,' in Proc. HPCA 2002
- U. Y. Ogras, et. al, 'NoC Prototyping Using FPGAs: Challenges and Promising Results in NoC Prototyping Using FPGAs, ' in IEEE Micro, September/October 2007
- Clermidy et al, 'A 477mW NoC-Based Digital Baseband for MIMO 4G SDR,' IEEE ISSCC, February 2010
- P. Juang, et al. 'Coordinated, distributed, formal energy management of chip multiprocessors,' Proc. ISLPED 2005...

179

References (Part IV)

- Y. Abulafia and A. Kornfeld. Estimation of FMAX and ISB in microprocessors. IEEE Trans. on VLSI Systems, 13(10), Oct 2006.
- A. Bonnoit, S. Herbert, D. Marculescu and L. Pileggi. Integrating Dynamic Voltage/Frequency Scaling and Adaptive Body Biasing using Test-time Voltage Selection. In Proc. of IEEE/ACM ISLPED, Aug. 2009.
- K. Bowman, S. Duvall, and J. Meindl. Impact of die-to die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration. IEEE Journal of Solid-State Circuits, 37(2), Feb 2002.
- S. Garg, D. Marculescu. System-Level Mitigation of WID Leakage Variations using Body-Bias Islands. In Proc. ACM/IEEE CODES+ISSS, Atlanta, GA, October 2008.
- S. Garg, D. Marculescu, R. Marculescu and U. Ogras. Technology-driven Limits on DVFS Controllability of Multiple Voltage-Frequency Island Designs. In Proc. of IEEE/ACM Design Automation Conference (DAC), Jul. 2009.
- S. Herbert and D. Marculescu. Analysis of dynamic voltage/frequency scaling in chip-multiprocessors. In ISLPED '07: Proc. of the 2007 ISLPED, 2007.
- S. Herbert and D. Marculescu. Variation-Aware Dynamic Voltage/Frequency Scaling. In Proc. of the 15th HPCA, Feb. 2009.
- 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 MICRO '06, 2006.
- S.R. Sarangi, B. Greskamp, R. Teodorescu, J. Nakano, A. Tiwari and J. Torrellas. VARIUS: A Model of Process Variation and Resulting Timing Errors for Microarchitects. IEEE Transactions on Semiconductor Manufacturing (IEEE TSM), February 2008.
- R. Teodorescu and J. Torrellas. Variation-aware application scheduling and power management for chip multiprocessors. In ISCA'08: Proc. of the 35th ISCA, 2008.
- J.Tschanz, J.T. Cao, S.G. Narendra, R. Nair, D.A. Antoniadis, A.P. Chandrakasan, V. De. Adaptive Body Bias for Reducing Impacts of Die-to-Die and Within-Die Parameter Variations on Microprocessor Frequency and Leakage. IEEE Journal of Solid-State Circuits, Vol. 37, No. 11, Nov. 2002.
- W. Zhao and Y. Cao. New generation of predictive technology model for sub-45nm early design exploration. IEEE Trans. Electron Devices, vol. 53, no. 11, pp. 2816--2823, Nov. 2006.
- This list of references is NOT exhaustive. There are many good contributions not mentioned here due to involuntary omissions or space limitations.
 180