Research Narrative

David Brooks
School of Engineering and Applied Sciences
Harvard University

November 18, 2008

Before introducing my research accomplishments, I briefly discuss the motivation behind the research in three major challenge areas which define the scope of my research to date. I then outline my general research approach, before discussing the details of my major research directions across each of the research challenges.

1 Challenges

My group focuses on three critical technology challenges that threaten continued scaling of computing systems: power dissipation and thermal limitations, process variations in nanoscale technology nodes, and efficient power delivery and voltage control [1].

The first challenge, power dissipation, became evident at the beginning of this decade, and the importance has only grown in magnitude in the last several years. Power dissipation manifests as a problem in three fundamental ways: 1) high power dissipation is the primary performance limiter for mainstream high-performance microprocessors — it is simply not cost-effective to cool and deliver power to microprocessors with high power dissipation; 2) power dissipation limits the capabilities of battery-powered mobile devices and "off-grid" computers such as wireless sensor nodes and computing devices in the developing world; and 3) energy costs dominate the cost of modern internet datacenters; 80% of the costs of Microsoft's recent 200MW datacenter in Chicago are energy related. The environmental impact of increased energy dissipation in information technology devices must also be considered.

The second challenge, process variations, is an old challenge, but device miniaturization has drastically changed the scale and scope of the problem. For chips with 100M+ transistors, it is impossible to fabricate all devices to have identical transistor performance due to imperfections in the manufacturing process. These device variations lead to two major ramifications in the high-level design process. First, SRAM memories used in the construction of on-chip cache memories are particularly sensitive to device variations, and empirical evidence suggests that memory cell sizes have not shrunk at the pace anticipated by Moore's Law in recent technology nodes. Given that cache memories make up nearly 50% of the die area of most microprocessors, a slow-down in memory cell scaling is tantamount to the end of Moore's Law. The second major ramification is that chip clock frequency and supply voltage are determined by the slowest path across a synchronous design. If process variations slow down even one critical path in a system, the entire chip will have to run at this slow clock frequency. Thus, process variations can have a large impact on overall chip performance and power dissipation.

The third challenge, efficient power delivery and voltage control, is directly related to the first two challenges. As designers seek to reduce the power dissipation of microprocessors, many low-power design

techniques are employed. For example, designers implement techniques such as clock and power gating which reduce the power consumption of the chip during idle periods. However, these gatings lead to current swings due to finite impedance in the power delivery network, resulting in voltage fluctuations. This voltage noise problem is compounded by efforts to run microprocessors at reduced supply voltages, also to reduce power consumption. Voltage noise can exceed 15% of nominal supply voltage, which leads to large timing safety margins in order to guarantee correctness under essentially worst-case behavior. In addition to maintaining a steady power supply, power management strategies require hardware and software mechanisms to control chip voltage/frequency settings. These control mechanisms are becoming especially complex in the era of chip-multiprocessors.

While discussed as separate challenges, the three problems are heavily interlinked, and solutions must provide balance. For example, one can reduce the impact of process variations by having larger design margins and increasing supply voltage, but this has a negative impact on power dissipation and performance. It is important to note that the ultimate goal of overcoming each of these challenges is to provide continued performance scalability and/or lower power consumption for future computer designs.

2 Approach

My research focuses on these design issues across a range of computing application domains, from embedded devices such as wireless sensor nodes to high-performance servers. This research is connected by several common themes. First, the best solutions to these problems will likely span the boundary of hardware and software. While my research is primarily in computer architecture, solutions to these low-level problems require an in-depth understanding of circuit implementation and VLSI-CAD issues. At the same time, many of the solutions to these problems will heavily depend on application characteristics and leveraging sophisticated control mechanisms only possible through runtime system software. Thus, my research efforts have been cross-disciplinary. A second theme to my research is that in order to make headway on the design of novel solutions to these low-level problems, we must develop architectural tools to accurately model power, temperature, and power supply noise across a wide range of design parameters. In this context, we are developing a technology-scalable modeling infrastructure for power and related issues that is essential for research in these domains. This modeling work forms the underpinnings of each of the major projects ongoing within my research group and our modeling tools and approaches are actively used by many other research groups in academia and industry.

3 Power-Aware Computing

Moore's Law provides increasingly abundant resources for chip architects to deliver performance but such performance must be power efficient. My research in this vein falls into three distinct categories. First, we must model and understand sources of power dissipation, particularly in the early (pre-RTL) stages of the design process and at the software layers. Second, we leverage this information to drive designs to optimal parts of power-performance design spaces. This analysis presents greater challenge in the multi-core era as the space of viable designs explodes. Finally, we are developing ultra-low-power computing devices such as wireless sensor nodes where power dissipation and device lifetime are paramount, driving us to explore new classes of computing architectures. The remainder of this subsection will describe each of these aspects of my research in turn.

3.1 High-Level Power Modeling

Part of my thesis work at Princeton involved developing a methodology for power modeling at the architectural design level (the Wattch toolkit) [2]. Wattch couples a standard architectural performance simulator providing activity information with analytical models for the capacitance (and hence dynamic power dissipation) of key nodes within typical microarchitectural structures. The original tool is still heavily used for both research and teaching (an optional registration form recorded almost 250 downloads of the tool in the past 12 months as of October 2008).

My post-Ph.D. work at IBM in 2001-02 applied aspects of this approach to the early-stage power analysis of parts of the IBM Cell processor and future IBM server class microprocessors. We also developed a power-performance modeling infrastructure (PowerTimer) for a POWER4-like microprocessor core [3, 4]. PowerTimer differs from Wattch in that the power models are derived from empirical measurements of complete circuits from the POWER4 processor design. While PowerTimer currently requires a license from IBM, it is in use by several universities including the University of Illinois, the University of Virginia, and my research group at Harvard. Nearly all architectural-level power-related research in both academia and industry uses derivatives of either the Wattch or PowerTimer approach.

A portion of my NSF CAREER award proposes to develop a new methodology for power and delay modeling that leverages insights gained from the development of both of these infrastructures. PowerTimer's underlying power models are very accurate because they are based on empirical measurements of complete circuits from the POWER4 processor design. On the other hand, it can be difficult to use PowerTimer to model new microarchitectures, because it is difficult to parameterize many of these underlying models. Wattch's analytical models are very flexible, but are not as accurate and neglect leakage power, an important source of power dissipation in modern process technologies. Our new methodology leverages aspects of both of these approaches. We use detailed empirical characterization of key circuit building blocks and develop an analytical approach to combine these building blocks. Our results show great promise for SRAM and CAM-based memory structures [5], and we are currently working to model complete microarchitectural structures and other microprocessor components. Kristen Lovin, an undergraduate working in my group (Harvard College, AB/SM 2008), applied a similar modeling methodology to model retention time and access time characteristics of dynamic memories. Another aspect of our power modeling research has been to understand the impact of underlying circuit implementation on our power modeling assumptions [6].

3.2 Power-Performance Optimal Design Space Exploration

Power efficiency has, in part, driven the transition to chip multiprocessors. The power-performance modeling infrastructure that has been developed over the past decade allows us to determine optimal designs for these complex systems. For example, homogeneous multiprocessors might consist of many small, simple cores or a few large, complex cores. Core heterogeneity, combining general-purpose cores and application-specific accelerators, further increases degrees of freedom in the design. Designers must assess the relative performance and power characteristics of these very diverse trajectories into the multi-core era.

Power-Performance Optimal Pipeline Depth

Choosing the pipeline depth of a microprocessor is one of the most critical design decisions that computer architects must make in the concept phase of a microprocessor design. Pipeline depth directly impacts the final clock frequency of the machine – very deep pipelines will have higher clock frequency, but lower instructions-per-cycle (due to increased branch resolution delays) than shorter pipelines. Because computer performance is a product of clock frequency and instructions-per-cycle, there is a classic optimization that must be made to select the proper pipeline depth. Deep pipelines will also incur significantly higher power

dissipation due to the higher clock frequency and additional pipeline latches that must be clocked. This research is especially relevant in light of product cancellations from Intel due to power concerns in very deeply pipelined CPU designs in the 2004 timeframe.

During my post-Ph.D. tenure at IBM Research, I extended PowerTimer to model the power and performance of pipeline depth [7,8]. Our work was the first to show that when power considerations are taken into account, the optimal pipeline depth of a microprocessor can shift to considerably shorter pipelines. This analysis has had a major influence on industrial design teams at Intel and IBM, as witnessed by the shift in recent years to machines with much shorter pipelines. Our paper, published in the International Symposium on Microarchitecture (MICRO), was selected from 158 submissions by IBM Research authors as one of the best papers in Computer Science, Electrical Engineering and Mathematical Sciences published in 2002 (now referred to as the Pat Goldberg Memorial Best Paper Award). I have received feedback from industrial researchers that this paper shaped the thoughts of computer designers within IBM, Intel, and AMD about future designs.

This work lead to one of my initial research directions at Harvard: finding optimal power-performance designs from the larger space of fundamental architectural parameters (such as pipeline width and memory hierarchy). This work eventually lead to Benjamin Lee's Ph.D. thesis as described below.

Power-Efficiency of Thread-Level Parallelism

Chip architects have embraced thread-level parallelism (TLP) in the form of simultaneous multi-threading (SMT) and chip multi-processing (CMP) as the primary driver of future performance. At the same time, we have reached a point where energy and thermal issues can no longer be abstracted from high-performance architecture design. In collaboration with IBM and the University of Virginia, we extended the PowerTimer infrastructure to model multi-threaded architectures and thermal issues. Along with Prof. Kevin Skadron, I co-advised Yingmin Li, a Ph.D. student at the University of Virginia, from 2003 until his graduation in 2006, which included weekly telecon meetings, in-person visits, and frequent email exchanges. Our initial work in this area studied the energy and thermal tradeoffs between two prevalent multi-threading architectures: SMT in a single core and chip-multiprocessing with multiple cores. The SMT analysis was published in the International Symposium on Low-Power Electronics and Design (ISLPED) in 2004 and the SMT/CMP comparison work was published in the International Symposium on High-Performance Computer Architecture (HPCA) in 2005 [9,10]. A key insight from this work is that designers of new CMP architectures must be cognizant of global chip heat-up mechanisms due to the large power output of multiple cores.

During the spring of 2004, Benjamin Lee and I worked on quantifying the power-performance of pipeline complexity in multiple dimensions (both pipeline depth and width) [11]. Yingmin and Benjamin collaborated to fold this research into the multi-threaded simulation environment discussed above, allowing us to understand the complex tradeoff between pipeline complexity (e.g. pipeline depth and width) and the number of cores in future, large scale chip multi-processors [12]. This work, presented at the HPCA conference in 2006, was the first to consider so many design metrics (performance, energy, thermal considerations) across such as a wide design space (number of cores, complexity of cores, size of L2 cache, area budget, and heat-sink configuration).

This work represented our first steps towards the long-term goal of understanding the large design space of TLP-oriented machines. However, our modeling approach of using detailed cycle-accurate simulation was clearly not suitable for studying more than a handful of design parameters simultaneously. Thus, Benjamin and I began to work on a new approach for large-scale architectural design space studies (the remainder of the work was independent of UVA).

Applying Statistical Inference

Microarchitects use detailed simulation to drive core research and development. Simulation costs are

prohibitively expensive for comprehensive design space exploration, forcing designers to ad hoc and highly constrained design optimization. Prior efforts reduce time per simulated design to tens of minutes but have no effect on the number of simulations required for design space exploration. This limitation is problematic since the design space and, consequently, the number of potential simulations scale exponentially with the number of design parameters. Thus, simulation costs quickly become intractable for any comprehensive analysis of a large and diverse design space.

To address these challenges, we proposed a simulation paradigm in which inferential models are constructed empirically from detailed simulation and are used as surrogates for simulators. This paradigm consists of (1) comprehensive design spaces, (2) spatial sampling, and (3) statistical inference. First, the paradigm defines a comprehensive design space of microarchitectural design parameters, spanning billions of diverse core designs. By considering many parameters simultaneously, the paradigm exposes interactions between them. Second, the paradigm sparsely samples designs from the space for detailed simulation. These simulations map design parameters to metrics, such as performance and power. Third, statistical inference constructs spline-based (e.g., piecewise polynomial) regression models derived from these sparse simulations.

Inferential models accurately capture performance and power trends so that design space exploration uses regression models, not further simulation. Furthermore, regression models are computationally efficient. Model training is a linear solve and model prediction is a matrix multiply. Thus, regression models can produce thousands of performance and power predictions per second while reducing the number of simulations required in design space exploration by many orders of magnitude (e.g., 500 samples from a multi-billion-point space).

A preliminary version of this work was published in the Modeling, Benchmarking and Simulation (MOBS) workshop at ISCA in 2006, and the full version appeared in the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS) later that year [13, 14]. A follow-on paper in IEEE Micro provided additional detail to enable other researchers to use the methodology [15]. We also teamed up with a group at Lawrence Livermore National Laboratories and Cornell University to compare the prediction capabilities of regression-based approaches with artificial neural networks [16].

New Analysis Capabilities

Statistical inference enables qualitatively new capabilities in performance and power characterization. Instead of exploring a few hundred designs in simulation, microarchitects can explore hundreds of billions of designs, characterizing performance and power trade-offs. Pareto optima, designs that minimize delay for a given power target or minimize power for a given delay target, can be identified quickly for comprehensive design spaces. Contour maps, which illustrate non-linearity, can be drawn quickly for performance and power trends. Furthermore, the degree of non-linearity can be quantified accurately by computing numerical derivatives or integrals for regression equations. More generally, microarchitects can use regression equations to leverage the wealth of history in classical analysis and optimization, which often require these derivatives and integrals. Such characterization is intractable using detailed microarchitectural simulation.

Iterative optimization heuristics (e.g., gradient ascent, genetic algorithms) are required for multi-billion point spaces. Such heuristics become computationally tractable after replacing simulation with regression equations within the iterative loop. These heuristics enable qualitatively new capabilities in optimization for emerging design priorities.

We explored these new design space analysis capabilities enabled by statistical inference approaches in a series of papers that make up the core of Benjamin Lee's Ph.D. thesis. The first paper, published in the HPCA conference in 2007, performed a series of case studies to illustrate the utility of the approach [17]. First, it

expands upon the power-performance optimal pipeline depth problem by relaxing other parameters of the machine while searching for the optimal pipeline depth. The paper then tackles the more computationally difficult problem of developing power-performance Pareto-frontiers for microarchitectural design spaces with over one billion points. This also allowed us to perform a design space study for asymmetric (or heterogeneous) multi-core processors from this large space.

In exploring these design spaces, we often came to the question of whether microarchitectural design spaces have many peaks and valleys or whether they are smooth and flat. This led Ben to develop a metric for the "roughness" of a design space, which can be used to gauge differences in the number of peaks and valleys in a design space. This roughness metric has many interesting applications, particularly when performing optimizations within a design space. Ben's analysis of roughness and its applications appeared in HPCA in 2008 [18].

Seeking an even greater challenge, Ben set out to answer a question that many researchers have studied over the years without arriving at a conclusive answer. While the domain of adaptive microarchitectures has been a fertile field for research for the past 5-7 years, fundamental questions about the potential benefits from microarchitecture adaptivity had remained unanswered. Fundamentally, this involves finding the optimal point in a multi-billion point design space for each phase of work, where we would like to explore how small the phase needs to be to achieve saturating benefits. Answering these types of questions can only be accomplished using these approaches. Ben's analysis finds that much additional potential benefit can be tapped through finer-grain adaptivity in both a temporal and spatial dimension. This work appeared at ASPLOS in 2008 [19].

Ben took an internship at Intel's Microprocessor Research Lab (MRL) from June to December of 2007. This internship allowed him to study two problems: first, he was able to validate the regression modeling approach using two of Intel's product simulators with a much wider range of interesting workloads; second, he studied the problem of modeling multi-core chips and developed a technique that partially decouples processor core simulation from interconnect and shared memory simulation. This allows much more tractable analysis of multi-core design spaces. This work will appear at the International Symposium on Microarchitecture (MICRO) in 2008 [20].

3.3 Ultra-Low-Power Architectures

Most of the above work has focused on power-issues in high-performance systems. However, energy-constrained device lifetimes are one of the most important design requirements in wireless and mobile embedded systems. As part of the Hourglass project, Prof. Gu-Yeon Wei and I have led an effort to develop an ultra low power computing platform for nodes in wireless sensor networks. Ultimately, we would like to develop programmable, wireless sensor nodes that operate on energy scavenged from the ambient environment. This represents an exciting new domain for computer architects – a domain where *power* is the primary design requirement and performance needs are negligible [21].

Mark Hempstead is the primary student lead on this project, and work began in spring of 2004. Our initial focus was to understand the wireless sensor network application space to develop ideas for our system architecture [22]. We have developed a novel, application-driven system architecture for wireless sensor nodes. This architecture provides explicit interrupt processing hardware support to match the event-driven nature of sensor network applications. Another key architectural feature is a modularized design that supports software-controlled, fine-grain power management including functionality to gate the supply voltage of on-chip modules to reduce idle power. In order to study the features of this architecture, we developed a detailed SystemC model to simulate the performance and activity of the architecture. Employing architectural power modeling approaches in conjunction with this model (similar to my prior work in modeling

power in higher performance processors), we estimated the dynamic and leakage power of our system with various power down modes and low-leakage techniques. We published our architecture approach to sensor network computing at the International Symposium on Computer Architecture (ISCA) in 2005 [23].

While the key architectural concepts formed the basis of the ISCA paper, we needed to prove that the ideas translated into real energy savings. For this we needed to move beyond simulations and design real hardware. Mark taped out two chip designs to prove these ideas. The first test chip was fabricated in IBM's $.18\mu m$ CMOS technology and was part of the Semiconductor Research Corporation (SRC) 2004-06 SoC Design Challenge, receiving 1st Prize (out of 39 entrants) in the challenge. Our second design was completed in 130nm UMC technology. We have recently completed measurements on the chip which demonstrate the energy saving capabilities of the architecture [24].

Another element of this research effort has been to explore the optimal process technology for compute elements in power-constrained, low-throughput applications. This is especially important when idle power dominates total energy consumption, as is the case for WSN applications. Process technology selection provides a natural optimization problem, as chip designs achieve an active power reduction, but an idle power increase by moving to more advanced process technologies. We performed a detailed study across a range of process technology parameters and performance targets. This study suggests that process technology for embedded processors for low-throughput applications like WSNs cannot efficiently scale below $.18\mu m$ without aggressive Vdd-gating or other leakage control support. This work appeared at the International Conference on Compilers, Architecture, and Synthesis for Embedded Systems (CASES) in 2006 [25].

While the bulk of our research in the wireless sensor domain has considered accelerator-based systems, we have also explored the power-performance efficiency advantages of heterogeneous multiple-ISA microcontrollers. I advised Lukasz Strozek (Harvard College, AB/SM 2006) on this project for his senior thesis. He created an environment that allowed simultaneous construction of new microcontroller designs and the required code translation (from an intermediate language) to each of these new designs. This work showed that significant efficiency advantages are possible by tuning the microcontroller ISA to application requirements. A conference paper summarizing Lukasz's senior thesis appeared in CASES in 2006 [26] and an extended version has been accepted to appear in ACM Transactions on Architecture and Code Optimization [27].

Our experience with design of the wireless sensor node demonstrated the huge potential for hardware acceleration in power-efficient systems. Hardware acceleration of common application tasks can improve performance and reduce energy consumption. Processor architectures that utilize application specific computational blocks have been proposed for a variety of domains including mobile computing, personal computing, embedded systems, high performance computing and wireless sensor networks. Most designs include large memories within each hardware accelerator that contribute to leakage power, require significant area, and add additional overhead when transferring state between accelerators.

To address these concerns, we are exploring accelerator-centric computing architectures. Mike Lyons, a third-year Ph.D. student, is developing such an architecture called SMASH – a shared memory framework for hardware accelerator based architectures. At the heart of SMASH is the accelerator store, a centralized memory structure that maintains state for every accelerator. The accelerator store implements complex memory and power management functionality that can simplify the design of complex accelerators. The accelerator store also allows efficient sharing of state between hardware accelerators, providing explicit support for the producer/consumer programming model and reducing the need to transfer data between blocks. Inactive memory blocks are automatically Vdd-gated, providing large energy savings.

4 Process Variations

Variations in transistor device characteristics threaten to severely slow the pace of advancement in future CMOS technology. This challenge, known as *manufacturing process variations*, has primarily been explored by circuit and CAD researchers. Over the past three years my research has increasingly focused on architecture and system design approaches that are resilient to device variations. These efforts can broadly be split into variation-tolerant pipeline design and variation-tolerant memory design.

4.1 Variation-Tolerant Pipeline Design

Starting in the summer of 2005, I began to think about addressing variations from a system perspective with Xiaoyao Liang, a new PhD student. Our initial research efforts in the area of architectural strategies to mitigate manufacturing process variations explored both static design-time approaches and test-time microarchitecture reconfiguration to address process variations. Our first effort explored the interaction between architectural techniques to achieve higher instructions-per-cycle (IPC) and the potential costs in clock frequency under process variation. The paper proposes a methodology that allows computer designers to explore tradeoffs between expected clock frequency under the impact of process variability and IPC benefits for relevant architectural optimizations such as the selection of latencies and sizings for caches, register files, and queues. These results appear in the International Conference on Computer-Aided Design (ICCAD) in 2006 [29].

Our MICRO 2006 paper is the first paper in a major computer architecture research conference to address process variation [30, 31]. Interest in variation-tolerant system design is rapidly growing – in fact, the MICRO conference in 2008 now has a separate track on variation-tolerant architectures. This paper provides a detailed study of the microarchitectural ramifications and circuit implementation of variable-latency operations in the integer and floating-point register files and execution units within a microprocessor. The key concept is that variable-latency microarchitectures are able to cope with device variability by judiciously extending the latency of slow registers and/or execution units in individual structures within chips that experience variability. Using detailed Monte-Carlo circuit simulations of the variable latency structures under random and systematic variations, the paper demonstrates that the approach can make the register files and execution units tolerant of a large amount of variation with minimal impact on system-level performance (IPC).

The results from our initial studies were promising, but the analysis was entirely based on simulations. Given that we are studying a low-level phenomenon such as device variation, we were very interested in proving the ideas through a silicon implementation. Xiaoyao and I began to collaborate with Prof. Gu-Yeon Wei on a prototype floating-point unit that implemented the variable-latency technique and a new technique that we developed called *voltage interpolation*. Voltage interpolation allows individual parts of a chip to have a fine-grain "effective voltage" that can be used to reduce the impact of variations with little power overhead. This is in contrast to raising the global voltage, which would cause huge power overheads. The chip prototype design results were presented at the International Solid-State Circuits Conference (ISSCC) in 2008 [32] and the architectural investigation of the two techniques was presented at the International Symposium on Computer Architecture (ISCA) in June 2008 [33]. This work has also been selected as one of the "Top Picks in Computer Architecture in 2008" for IEEE Micro magazine [34]. We have also explored power-saving techniques for deeply pipelined architectures using this prototype chip [35]

While our chip implementation successfully demonstrated the concept of variable latency and voltage interpolation, there are many questions related to the design of systems that utilize voltage interpolation. Kevin Brownell, a second-year Ph.D. student, started to answer some of these questions as part of his

initial research explorations. Specifically, Kevin's work explores tradeoffs in the synthesis flow for VI-based systems, by investigating design choices related to the number and distribution of voltage domains. These choices impact the delay tuning range of the technique and static power overheads at domain boundaries. This work will appear at the International Conference on Computer Aided Design (ICCAD) in November 2008 [36].

4.2 Replacing SRAM with DRAM

Process variation is particularly problematic for memory designs. Traditional six transistor (6T) SRAM designs are highly sensitive to device mismatch and under variations suffer inherent stability problems and delay sensitivity. We have explored novel dynamic memory technologies to replace static 6T RAM cells. In particular, we are exploring the benefits of dynamic 3T1D memory cells which do not suffer stability issues as they have a single dynamic storage node. The key insight is that all sources of variations can be lumped into a single parameter, the data retention time for the 3T1D cell, and variations in the retention time can be managed by architectural refresh and data placement mechanisms. We have explored the benefits and challenges of utilizing 3T1D DRAM technology within architectural structures such as register files and data caches. Our results show significant advantages in both tolerance to process variability and energy efficiency for these architectural structures. We presented our findings at the 2007 International Symposium on Microarchitecture (MICRO) [37] and the paper was subsequently published in the IEEE Micro magazine [38] as one of the "Top Picks" from 2007 computer architecture conferences. This work was also one of three papers nominated by ACM SIGMICRO for consideration for publication in the Communications of the ACM (CACM). We are in the process of extending this work to incorporate more memory blocks within a microprocessor.

Tuning techniques are capable of adapting the microarchitecture to mitigate the impact of variations at post-fabrication testing time. Most of the existing techniques ignore testing cost or simply assume a naive exhaustive testing scheme. But testing has associated costs, which might be prohibitively expensive for a large space of post-fabrication tuning configurations. Our recent work explores a new post-fabrication testing framework that accounts for testing costs. This framework uses on-chip canary circuits to capture systematic variation while using statistical analysis to estimate random variation. A regression model is applied to predict the chip performance and power for different configurations. These techniques comprise an integrated framework that identifies the most energy efficient post-fabrication tuning configuration for each chip.

5 Power Delivery and Voltage Control

Reliable and efficient power delivery is critical to all types of computing systems. As designers seek to reduce the power consumption of systems by reducing the supply voltage, systems will begin to experience power supply fluctuations due to the finite impedance of the power supply network. These supply fluctuations, referred to as voltage emergencies, must be managed by the system to provide correctness. Our research seeks ways to handle these *alarm* conditions through a combined hardware/software approach. In addition to handling these alarm conditions, we consider the problem of voltage selection for power management. Energy-constrained systems typically will employ dynamic voltage and frequency scaling to match workload behavior to required performance levels – setting the voltage/frequency to the correct level to match performance needs provides the best energy efficiency. The advent of chip-multiprocessors greatly complicates traditional dynamic voltage/frequency scaling, and our recent research focuses on hardware/software approaches to leverage DVFS in multicore systems.

5.1 Alarm-based Computing

I began to study power delivery in my post-Ph.D. position at IBM in 2002. Russ Joseph, a Ph.D. student in my former group at Princeton, was an intern in my research group and we developed a control-theoretic feedback approach to handle voltage emergencies [40]. The approach ensures correctness via a voltage sensor that detects voltage fluctuations that pass through a soft threshold. After detecting a droop, an actuation mechanism throttles processor activity to reduce power consumption and allows the supply voltage to recover. The paper demonstrates that the technique can work reasonably well for a wide range of applications.

My work at IBM motivated me to continue studying voltage emergencies at Harvard, and I was particularly interested in understanding whether these emergencies could be linked to high-level application behavior. Kim Hazelwood and I began to explore voltage emergencies with the eventual goal of using a dynamic compiler to optimize troublesome code sequences as indicated by sensors in the underlying architecture [41]. This work provided two major contributions which motivated us to continue work in this area: 1) voltage emergencies regularly repeat throughout an application's runtime; and 2) the number of static code locations where these emergencies occur is quite small. This proof-of-concept work provided initial motivation for our NSF-funded "Alarms Project" with Prof. Michael Smith and Prof. Gu-Yeon Wei. Meeta Gupta has been the primary PhD student working on this project, and VJ Reddi, a second-year PhD student, has recently joined this project as well.

Our first project involved developing an architectural model for the power delivery network that can be used to understand the architectural implications of power supply noise. The model incorporates a grid-based model for the on-chip power delivery network allowing exploration of the impact of fine-grain events in microprocessors; the model is particularly relevant for understanding how the varying utilization of different cores in a chip-multiprocessor system will impact power delivery noise. Our 2007 paper in the Design, Automation, and Test Conference in Europe (DATE) describes the model and explores various scenarios in chip-multiprocessor systems that can magnify power delivery issues [42].

The architectural power delivery model allowed us to focus on understanding the root cause of power supply noise in microprocessor systems. The source of noise is a complex interaction of the power delivery network and the load characteristics of the system. Because of low-power design techniques like clock gating, the load is very dependent on characteristics of the microarchitecture and running applications. Our ISLPED paper in 2007 defines "voltage emergencies" as noise events that cause the power supply voltage to fall below a threshold level, and the paper identifies specific microarchitectural events that lead to changes in the power draw of the processor as the causes of these voltage emergencies [43]. Understanding the root cause of voltage emergencies has allowed us to develop approaches to detect at runtime when they will occur by leveraging information about the control flow of a program and microarchitectural statistics. We have recently developed a *voltage emergency predictor* that uses this runtime information to predict voltage emergencies up to 16 cycles in advance with greater than 90% accuracy. These predictions can be used to much more effectively guide throttling approaches that avoid emergencies. This work has recently been accepted to the HPCA conference in 2009 and has been nominated for the best paper award at the conference [44].

The end goal of the modeling and characterization work is to understand architectural and software mechanisms that can be used to reduce the system-level impact of power supply noise. We have begun to explore two such approaches. First, we proposed a "delayed commit" mechanism that allows microprocessors to guarantee system-level correctness in the presence of voltage noise. This mechanism potentially allows the circuits in a microprocessor to operate with much smaller voltage margins because correctness is guaranteed at the system-level. This work was published in HPCA in 2008 [45]. Second, we have begun to explore software and compiler optimizations that smooth the current load of the processor core(s) [46].

We anticipate that the combination of system-level correctness guarantees and software-based current load smoothing provide an effective high-level approach to handle power supply noise and could have a significant influence on the way that the industry designs these systems.

5.2 Runtime Voltage/Frequency Scaling

Dynamic runtime systems provide many opportunities for energy savings due to the potential for exploiting slack within program execution by applying techniques such as dynamic voltage/frequency scaling (DVFS). For example, memory-bound loops within workloads provide an opportunity to reduce the frequency and voltage of the CPU with minimal performance impact due to the presence of significant stall behavior. In collaboration with Princeton, Intel, and the University of Colorado, we developed an approach to perform DVFS within a runtime optimization infrastructure based on PIN. Our DVFS control algorithm led to very impressive results for a pure software approach (over 20% energy-delay product improvements for many benchmarks). This work was published in the International Symposium on Microarchitecture [47]. This paper was selected as the best paper of the conference and was also selected to appear in a special issue of IEEE Micro for the top papers that appeared in computer architecture conferences in 2005 [48].

The effectiveness of DVFS schemes such as the one discussed above is hampered by slow voltage transitions that occur over tens of microseconds. In addition, the recent trend towards chip-multiprocessors (CMP) executing multi-threaded workloads with heterogeneous behavior motivates the need for core-level DVFS control mechanisms. Wonyoung Kim, a third-year Ph.D. student, and Meeta recently explored the costs and benefits of integrated, on-chip voltage regulators [49, 50]. Integrating voltage regulators onto the same chip as the microprocessor core provides the benefit of both nanosecond-scale voltage switching and per-core voltage control. Our work shows that these characteristics provide significant energy-saving opportunities compared to traditional off-chip regulators. However, the implementation of on-chip regulators presents many challenges, including regulator efficiency and output voltage transient characteristics, which are significantly impacted by the system-level application of the regulator. Wonyoung's current efforts include the implementation of two prototype regulator designs that will allow a detailed analysis of the costs of on-chip regulators.

In addition to traditional DVFS mechanisms, we have been exploring an alternate approach to power management that is applicable to large multi-core systems. This approach, called *thread motion* takes the approach that it may be easier to move a thread of computation (including relevant program state) to a core that is running at the appropriate voltage-frequency level, rather than trying to change the voltage of the core that the thread is running on. Given the expected large number of cores in future systems, thread management of this kind can provide effective power management, with low design cost compared to building on-chip regulators for each processor core. Our work shows that with only two voltage/frequency settings, power savings that approach those observed for fast per-core DVFS can be achieved with minimal overheads. This effort is lead by Krishna Rangan, a third-year graduate student.

6 Future research efforts

The research area of architectural design under technology considerations (e.g. power, resiliency, variability) is still in its infancy. Many technologists project that Moore's Law will continue for at least the next two decades, with primarily CMOS-based systems. I believe there is still significant research to be done, and these challenges will only increase in importance as we move to process technologies in the nanoscale. It is likely that future CMOS technology nodes will begin to incorporate other nanoscale materials. In this area, I have recently been collaborating with Prof. Soha Hassoun at Tufts University to explore bundled carbon

nanotubes to assist with power delivery and heat removal in standard CMOS-based microprocessors. There are numerous interesting technology innovations (e.g. on-chip nanophotonics, phase-change memories, etc) that will lead to new system-level research directions.

Another major research direction is to further explore accelerator-centric architectures, expanding upon our initial efforts in the domain of wireless sensor devices. Given power limitations on the number of simultaneous switching devices, scalable systems will increasingly need to perform more useful work per switching event. Our vision is that future systems will include a modest number of homogeneous general-purpose cores, coupled with a large set of highly diverse hardware accelerators, of which only a small fraction are useful for a given workload. Accelerators can provide 10-1000x efficiency gains when utilized and inactive accelerators are aggressively power-gated when not in use. In this model, general-purpose cores are only needed for portions of applications that do not map well to the palette of accelerators or as "glue-logic" to connect accelerators. These accelerators can take many forms. Existing media processing accelerators used in modern embedded systems (such as Apple's iPhone) utilize System-on-a-Chip (SoC) designs that provide monolithic, fixed-purpose accelerators. However, fine-grain accelerators provide more flexibility. As an example, rather than constructing a JPEG encoder accelerator, it may be preferable to divide the block into minimally programmable accelerator kernels for DCT, quantization, and Huffman encoding. Accelerator kernels could then be shared by other high-level applications, and power-gated with fine resolution. Hardware specialization meshes with development of higher productivity hardware description languages such as Liquid Metal and Bluespec and the rise of fully-synthesized design for nearly all non-memory structures in high-performance CPUs. This research has significant challenges in multiple areas to enable accelerator-centric architectures: the design of an architectural framework that eases the development of accelerators, development of programming language and compiler support for these systems, and automated tools to discover potential accelerators.

References

[1] D. Brooks, R. Dick, R. Joseph, and L. Shang, "Power, thermal, and reliability modeling in nanometer-scale microprocessors," *IEEE Micro*, May 2007.

[2] D. Brooks, V. Tiwari, and M. Martonosi, "Wattch: A framework for architectural-level power analysis and optimizations," in *International Symposium on Computer Architecture (ISCA-27)*, June 2000.

[3] D. Brooks, P. Bose, S. E. Schuster, H. Jacobson, P. N. Kudva, A. Buyuktosunoglu, J.-D. Wellman, V. Zyuban, M. Gupta, and P. W. Cook, "Power-aware microarchitecture: Design and modeling challenges for next-generation microprocessors," *IEEE Micro*, vol. 20, pp. 26–44, Nov/Dec 2000.

[4] D. Brooks, P. Bose, V. Srinivasan, M. Gschwind, P. G. Emma, and M. G. Rosenfield, "New methodology for early-stage, microarchitecture-level power-performance analysis of microprocessors," *IBM Journal of Research and Development*, vol. 47, pp. 739–51, October/November 2003.

[5] X. Liang, K. Turgay, and D. Brooks, "Architectural power models for SRAM and CAM structures based on hybrid analytical/empirical techniques," in *International Conference on Computer-Aided Design (ICCAD)*, Nov. 2007.

[6] Y. Li, M. Hempstead, P. Mauro, D. Brooks, and Z. H. K. Skadron, "Power and thermal effects of SRAM vs. latch-mux design styles and clock gating choices," in *International Symposium on Low-Power Electronics and Design (ISLPED)*, August 2005.

[7] V. Srinivasan, D. Brooks, M. Gschwind, P. Bose, V. Zyuban, P. N. Strenski, and P. G. Emma, "Optimizing pipelines for power and performance," in *International Symposium on Microarchitecture (MICRO-35)*, Nov. 2002. Selected as one of the four Best IBM Research Papers in Computer Science, Electrical Engineering and Math published in 2002.

[8] V. Zyuban, D. Brooks, V. Srinivasan, M. Gschwind, P. Bose, P. Strenski, and P. Emma, "Integrated analysis of power and performance for pipelined microprocessors," *IEEE Transactions on Computers*, vol. 53, pp. 1004–1016, August 2004.

[9] Y. Li, D. Brooks, Z. Hu, and K. Skadron, "Performance, energy, and thermal considerations for SMT and CMP architectures.," in *International Conference on High-Performance Computer Architecture (HPCA-11)*, Feb. 2005.

[10] Y. Li, D. Brooks, Z. Hu, K. Skadron, and P. Bose, "Understanding the energy efficiency of simultaneous multithreading," in *International Symposium on Low-Power Electronics and Design (ISLPED)*, August 2004.

[11] B. Lee and D. Brooks, "Effects of pipeline complexity on SMT/CMP power-performance efficiency," in *Workshop on Complexity Effective Design (WCED), held in conjunction with ISCA-32"*, June 2005.

[12] Y. Li, B. Lee, D. Brooks, Z. Hu, and K. Skadron, "CMP design space exploration subject to physical constraints," in *International Conference on High-Performance Computer Architecture (HPCA-12)*, Feb. 2006.

[13] B. Lee and D. Brooks, "Statistically rigorous regression modeling for the microprocessor design space," in *Workshop on Modeling, Benchmarking, and Simulation (MoBS), held in conjunction with ISCA-33*, June 2006.

[14] B. Lee and D. Brooks, "Accurate and efficient regression modeling for microarchitectural performance and power prediction," in *International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, October 2006.

[15] B. C. Lee and D. Brooks, "A tutorial in spatial sampling and regression strategies for microarchitectural analysis," *IEEE Micro*, May 2007.

[16] B. Lee, D. Brooks, B. de Supinski, M. Schulz, K. Singh, and S. McKee, "Methods of inference and learning for performance modeling of parallel applications," in *Symposium on Principles and Practice of Parallel Programming (PPoPP)*, March 2007.

[17] B. Lee and D. Brooks, "Illustrative design space studies with microarchitectural regression models," in *International Symposium on High-Performance Computer Architecture (HPCA-13)*, Feb. 2007.

[18] B. Lee and D. Brooks, "Roughness of microarchitectural design topologies and its implications for optimization," in *International Symposium on High-Performance Computer Architecture (HPCA-14)*, Feb. 2008.

[19] B. Lee and D. Brooks, "Efficiency trends and limits from comprehensive microarchitectural adaptivity," in *International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, March 2008.

[20] B. Lee, J. Collins, H. Wang, and D. Brooks, "CPR: composable performance regression for scalable multiprocessor models," in *International Symposium on Microarchitecture (MICRO-41)*, December 2008.

[21] M. Hempstead, M. J. Lyons, D. Brooks, and G.-Y. Wei, "Survey of hardware systems for wireless sensor networks," *ASP Journal of Low Power Electronics*, 2008.

[22] M. Hempstead, D. Brooks, and M. Welsh, "TinyBench: The case for a standardized benchmark suite for TinyOS based wireless sensor network devices.," in *IEEE Workshop on Embedded Networked Sensors (EmNets'04)*, Nov. 2004.

[23] M. Hempstead, N. Tripathi, P. Mauro, G.-Y. Wei, and D. Brooks, "An ultra low power system architecture for sensor network applications.," in *International Symposium on Computer Architecture (ISCA-32)*, June 2005.

[24] M. Hempstead, D. Brooks, and G.-Y. Wei, "An accelerator-based wireless sensor network processor in 130nm CMOS," Under preparation for submission.

[25] M. Hempstead, G.-Y. Wei, and D. Brooks, "Architecture and circuit techniques for low throughput, energy constrained systems across technology generations," in *International Conference on Compilers, Architecture, and Synthesis for Embedded Systems*, Oct. 2006.

[26] L. Strozek and D. Brooks, "Efficient architectures through application clustering and architectural heterogeneity," in *International Conference on Compilers, Architecture, and Synthesis for Embedded Systems*, Oct. 2006.

[27] L. Strozek and D. Brooks, "Efficient architectures through application clustering and heterogeneity," *ACM Transactions on Architecture and Code Optimization*, Accepted for publication. 2009.

[28] X. Liang and D. Brooks, "Microarchitecture parameter selection to optimize system performance under process variation," in *International Conference on Computer-Aided Design (ICCAD)*, Nov. 2006.

[29] X. Liang and D. Brooks, "Latency adaptation for multiported register files to mitigate the impact of process variations," in *Workshop on Architectural Support for Gigascale Integration (ASGI), held in conjunction with ISCA-33*, June 2006.

[30] X. Liang and D. Brooks, "Mitigating the impact of process variations on cpu register file and execution units," in *International Symposium on Microarchitecture (MICRO-39)*, December 2006.

[31] X. Liang, D. Brooks, and G.-Y. Wei, "A process-variation-tolerant floating-point unit with voltage interpolation and variable latency," in *International Solid-State Circuits Conference (ISSCC)*, Feb. 2008.

[32] X. Liang, G.-Y. Wei, and D. Brooks, "ReVIVaL: A variation tolerant architecture using voltage interpolation and variable latency," in *International Symposium on Computer Architecture (ISCA-35)*, June 2008.

[33] X. Liang, G.-Y. Wei, and D. Brooks, "ReVIVaL: Variation tolerant microarchitecture," *IEEE Micro Top Picks*, Feb. 2009.

[34] G.-Y. Wei, D. Brooks, A. D. Khan, and X. Liang, "Instruction-driven clock scheduling with glitch mitigation," in *International Symposium on Low Power Electronics and Design (ISLPED)*, August 2008. Nominated for Best Paper Award.

[35] K. Brownell, G.-Y. Wei, and D. Brooks, "Evaluation of voltage interpolation to address process variations," in *International Conference on Computer-Aided Design (ICCAD)*, Nov. 2008.

[36] X. Liang, R. Canal, G.-Y. Wei, and D. Brooks, "Process variation tolerant 3T1D-based cache architectures," in *International Symposium on Microarchitecture (MICRO-40)*, Dec. 2007. Nominated for CACM special issue consideration by SIGMICRO. Selected as one of the Top Picks in Computer Architecture in 2007.

[37] X. Liang, R. Canal, G.-Y. Wei, and D. Brooks, "Replacing 6T SRAMs with 3T1D DRAMs in the L1 data cache to combat process variability," *IEEE Micro Top Picks*, Feb. 2008.

[38] R. Joseph, D. Brooks, and M. Martonosi, "Control techniques to eliminate voltage emergencies in high performance processors.," in *International Conference on High-Performance Computer Architecture (HPCA-9)*, February 2003.

[39] K. Hazelwood and D. Brooks, "Eliminating voltage emergencies via microarchitectural voltage control feedback and dynamic optimization," in *International Symposium on Low-Power Electronics and Design (ISLPED)*, August 2004.

[40] M. S. Gupta, J. L. Oatley, R. Joseph, G.-Y. Wei, and D. Brooks, "Understanding voltage variations in chip multiprocessors using a distributed power-delivery network," in *Design, Automation, and Test in Europe Conference (DATE-10)*, April 2007.

[41] M. S. Gupta, K. K. Rangan, M. D. Smith, G.-Y. Wei, and D. Brooks, "Towards a software approach to mitigate voltage emergencies," in *International Symposium on Low Power Electronics and Design (ISLPED)*, Aug. 2007.

[42] V. Reddi, M. Gupta, G. Holloway, M. D. Smith, G.-Y. Wei, and D. Brooks, "Voltage emergency prediction: A signature-based approach to reducing voltage emergencies," in *International Conference on High-Performance Computer Architecture (HPCA-15)*, Feb. 2009. Nominated for Best Paper Award.

[43] M. S. Gupta, K. K. Rangan, M. D. Smith, G.-Y. Wei, and D. Brooks, "DeCoR: A delayed commit and rollback mechanism for handling inductive noise in processors," in *International Symposium on High-Performance Computer Architecture (HPCA-14)*, Feb. 2008.

[44] V. Reddi, M. Gupta, G.-Y. Wei, and D. Brooks, "An event-guided approach to handling inductive noise in processors," in *Design, Automation, and Test in Europe Conference (DATE-12)*, April 2009.

[45] Q. Wu, V. Reddi, Y. Wu, J. Lee, D. Connors, D. Brooks, M. Martonosi, and D. W. Clark, "Dynamic compilation framework for controlling microprocessor energy and performance," in *International Symposium on Microarchitecture (MICRO-38)*, Nov. 2005. Best Paper Award. Selected as one of the Top Picks in Computer Architecture in 2005.

[46] Q. Wu, V. Reddi, Y. Wu, J. Lee, D. Connors, D. Brooks, M. Martonosi, and D. W. Clark, "Dynamic compiler driven control for microprocessor energy and performance," *IEEE Micro Special Issue: Top Picks form Computer Architecture Conferences*, Jan/Feb 2006.

[47] W. Kim, M. S. Gupta, G.-Y. Wei, and D. Brooks, "Enabling on-chip switching regulators for multi-core processors using current staggering," in *Workshop on Architectural Support for Gigascale Integration (ASGI), held in conjunction with ISCA-34*, June 2007.

[48] W. Kim, M. S. Gupta, G.-Y. Wei, and D. Brooks, "System level analysis of fast, per-core DVFS using on-chip switching regulators," in *International Symposium on High-Performance Computer Architecture (HPCA-14)*, Feb. 2008.