The ACM IEEE International Symposium on Computer Architecture (ISCA) is the premier forum for new ideas and experimental results in computer architecture. In 2010, the 37th edition of ISCA was held in Saint Malo, France. The conference received 245 submissions, and accepted 44 of them (an acceptance rate of 18%).

ISCA has a long tradition of having strong workshops and tutorials associated with the conference. Thanks to Yanos Sazeides, the Workshop/Tutorial Chair in 2010, several high-quality workshops and tutorials were again collocated with the conference. They were very much appreciated by the attendees, and they had an important contribution to the overall success of ISCA 2010.

In 2010, ISCA featured 4 half-day tutorials, as well as 13 workshops on topics ranging from novel memory architectures to emerging application design and performance analysis. This proceedings volume gathers the valuable scientific contributions from five of these workshops:

1. **A4MMC**—The First Workshop on Applications for Multi- and Many-Core Processors: Analysis, Implementation, and Performance—focuses entirely on application case studies. With A4MMC, the organizers provided a forum where multi- and manycore application designers could exchange knowledge, insights and discoveries, and discuss their latest research advances. Further, by collocating A4MMC with ISCA, the application design and development community was able to directly expose its findings, requirements, and problems to a select audience of top computer architecture researchers. This workshop offered an ideal opportunity for software and hardware researchers to communicate and debate on how to find the right balance between these two sides of the “multicore revolution.”

2. **AMAS-BT**—The Third Workshop on Architectural and Micro-Architectural Support for Binary Translation—is motivated by the large-scale use of binary translation and on-the-fly code generation, which are becoming pervasive as enablers for virtualization, processor migration and also as processor implementation technology. AMAS-BT brought together researchers and practitioners with the aim of stimulating the exchange of ideas and experiences on the potential and limits of architectural and microarchitectural support for binary translation (hence the acronym AMAS-BT). The key focus is on challenges and opportunities for such assistance and opening new avenues of research. A secondary goal is to enable dissemination of hitherto unpublished techniques from commercial projects.

3. **EAMA**—The Third Workshop on Emerging Applications and Many-core Architecture—is equally motivated by the emerging workloads that bring new challenges for developing future computer architectures, and the breakthroughs in computer system design, which enable new application domains.
As recent development trends suggest that industry is moving to many-core architectures to better manage trade-offs among performance, energy efficiency, and reliability in deep-submicron technology nodes, many opportunities for developing new classes of applications have opened. Such computationally intensive tasks include real-time ray-tracing, multi-modal data mining, physical simulation, financial analytics, or virtual worlds. EAMA brought together application domain experts and computer architects to discuss emerging applications in these novel fields, as well as their implications on current- and next-generation many-core architectures.

4. **WEED**—The Second Workshop on Energy-Efficient Design—provided a forum for the exchange of ideas on research on critical areas relating to energy-efficient computing, including energy-aware design techniques for systems (large and small), energy management policies and mechanisms, and standards for evaluating energy efficiency. It was well attended with a good mix of researchers and practitioners from industry and academia.

5. **WIOSCA**—The 6th Annual Workshop on the Interaction Between Operating System and Computer Architecture—focused on characterizing, modeling and optimizing the interaction between OS and hardware in the light of emerging architecture paradigms (e.g., multi-core processors), workloads (e.g., commercial and server workloads) and computing technology (e.g., virtualization). The WIOSCA workshop provided an active forum for researchers and engineers from academia and industry to discuss their latest research in computer architecture and system software.

February 2011

Ana Lucia Varbanescu
ISCA Workshops Committees

A4MMC: First Workshop on Applications for Multi- and Many-Cores

Workshop Chairs and Organizers
Henri Bal Vrije Universiteit Amsterdam, The Netherlands
Henk Sips Delft University of Technology, The Netherlands
Ana Lucia Varbanescu Delft University of Technology, The Netherlands
Anca Molnos Delft University of Technology, The Netherlands
Rob van Nieuwoort Vrije Universiteit Amsterdam, The Netherlands

Program Committee
John Romein ASTRON, The Netherlands
Sorin Cotofana Delft University of Technology, The Netherlands
Joerg Keller FernUniversität Hagen, Germany
Christoph Kessler Linköping University, Sweden
Rosa Badia Barcelona Supercomputing Center, Spain
Xavier Martorell Universitat Politecnica de Catalunya, Spain
Paul Kelly Imperial College London, UK
Anton Lokhmotov ARM, UK
Raymond Namyst University of Bordeaux, France
David Bader GeorgiaTech, USA
Michael Perrone IBM T.J. Watson Research Center, USA
Virat Agarwal IBM T.J. Watson Research Center, USA

AMAS-BT: Third Workshop on Architectural and Micro-Architectural Support for Binary Translation

Workshop Chairs and Organizers
Mauricio Breternitz AMD
Robert Cohn Intel
Erik Altman IBM
Youfeng Wu Intel
**Program Committee**

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erik Altman</td>
<td>IBM</td>
</tr>
<tr>
<td>Guido Araujo</td>
<td>UNICAMP</td>
</tr>
<tr>
<td>Edson Borin</td>
<td>UNICAMP</td>
</tr>
<tr>
<td>Mauricio Breternitz</td>
<td>AMD</td>
</tr>
<tr>
<td>Mark Charney</td>
<td>Intel</td>
</tr>
<tr>
<td>Josep M. Codina</td>
<td>Intel</td>
</tr>
<tr>
<td>Robert Cohn</td>
<td>Intel</td>
</tr>
<tr>
<td>Andy Glew</td>
<td>Intel</td>
</tr>
<tr>
<td>Kim Hazelwood</td>
<td>University of Virginia</td>
</tr>
<tr>
<td>David Kaeli</td>
<td>Northeastern University</td>
</tr>
<tr>
<td>Chris J. Newburn</td>
<td>Intel</td>
</tr>
<tr>
<td>Suresh Srinivas</td>
<td>Intel</td>
</tr>
<tr>
<td>Chenggang Wu</td>
<td>CAS, China</td>
</tr>
<tr>
<td>Youfeng Wu</td>
<td>Intel</td>
</tr>
</tbody>
</table>

**EAMA: Third Workshop for Emerging Applications and Many-Core Architectures**

**Workshop Chairs and Organizers**

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrea Di Blas</td>
<td>Oracle and UC Santa Cruz</td>
</tr>
<tr>
<td>Engin Ipek</td>
<td>University of Rochester</td>
</tr>
<tr>
<td>Victor Lee</td>
<td>Intel Corporation</td>
</tr>
<tr>
<td>Philipp Slusallek</td>
<td>Intel and Saarland University</td>
</tr>
</tbody>
</table>

**Program Committee**

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Olivier Temam</td>
<td>INRIA</td>
</tr>
<tr>
<td>David August</td>
<td>Princeton University</td>
</tr>
<tr>
<td>David Holmes</td>
<td>Mayo Clinic</td>
</tr>
<tr>
<td>Ravi Murthy</td>
<td>Oracle</td>
</tr>
<tr>
<td>Milos Prvulovic</td>
<td>Georgia Institute of Technology</td>
</tr>
<tr>
<td>Jose Renau</td>
<td>UC Santa Cruz</td>
</tr>
<tr>
<td>Eric Sedlar</td>
<td>Oracle</td>
</tr>
</tbody>
</table>

**WEED: Second Workshop on Energy-Efficient Design**

**Workshop Chairs and Organizers**

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Carter</td>
<td>IBM</td>
</tr>
<tr>
<td>Karthick Rajamani</td>
<td>IBM</td>
</tr>
</tbody>
</table>
Program Committee

Pradip Bose IBM
David Brooks Harvard
Kirk Cameron Virginia Tech
John Carter IBM
Jichuan Chang Hewlett-Packard
Babak Falsafi EPFL
Sudhanva Gurumurthi University of Virginia
Fernando Latorre Intel - UPC
Jie Liu Microsoft
Onur Mutlu Carnegie-Mellon University
Karthick Rajamani IBM
Karsten Schwan Georgia Tech
Farhana Sheikh Intel
Thomas Wenisch University of Michigan

WIOSCA

Workshop Chairs and Organizers

Tao Li University of Florida
Onur Mutlu Carnegie Mellon University
James Poe Miami Dade College

Program Committee

Brad Beckmann AMD
Evelyn Duesterwald IBM Research
Alexandra Fedorova Simon Fraser University
Nikos Hardavellas Northwestern University
Jim Larus Microsoft Research
Shan Lu University of Wisconsin
Chuck Moore AMD
Nacho Navarro UPC
Lu Peng Louisiana State University
Partha Ranganathan Hewlett Packard Labs
Ben Sander AMD
Yanos Sazeides University of Cyprus
Per Stenstrom Chalmers University
Osman Unsal Barcelona Supercomputing Center
Kushagra Vaid Microsoft
Wei Wu Intel
Zhao Zhang Iowa State University
Multi- and manycore processors are here to stay. And this is no longer an academic rumor, but a reality endorsed and enforced by all processor vendors. However, both academia and industry agree that although the novel processors may drill through the power and performance walls, they also open up a new and wide programmability gap. As technology advances, the software seems to lag behind more and more. In fact, the multicore world is witnessing a technology push from the hardware side.

We believe that developing novel hardware, and also software stacks, tools and programming models is going nowhere if the requirements of the applications are not taken into account, or if these platforms are simply too difficult to program (efficiently). It is, after all, a matter of economics: if we do not focus on productivity and efficiency, the software development cost per “performance unit” might become so high that the next generations of multi- and manycores will simply be unsuccessful.

The best way to initiate productivity and efficiency analyses is to collect a large enough pool of representative, variate, real-life applications that make use of multi-/manycore architectures. Hence this workshop, “Applications for Multi- and Many-Core Processors” (A4MMC), which focuses entirely on application case studies. With A4MMC, we aimed to provide a forum where multi- and manycore application designers can exchange knowledge, insights and discoveries, and discuss their latest research advances. Further, by collocating A4MMC with ISCA, we aimed to directly expose the software community’s findings, requirements, and problems to a select audience of top computer architecture researchers. This workshop provides room for the pull from the software side, and offers an ideal opportunity for software and hardware researchers to communicate and debate on how to find the right balance between these two sides of the “multicore revolution.”

Our final goal is to build a pool of real-life multicore applications, backed up by performance studies and potential hardware add-ons. Such a collection will be useful to both the hardware and software developers, and it will be a good starting point for the tools and programming models communities in their work toward more effective models and methods. We strongly believe this is the most efficient way to bridge the multicore programmability gap in a systematic way.

Ana Lucia Varbanescu
Anca Molnos
Rob van Nieuwpoort
Keynote: Many-Core Processing for the LOFAR Software Telescope, by Rob van Nieuwpoort from ASTRON and Vrije Universiteit Amsterdam, The Netherlands

Abstract. This talk provides an overview of the many-core work carried out at ASTRON, The Netherlands foundation for radio astronomy. ASTRON is currently constructing LOFAR, a revolutionary new radio telescope that uses tens of thousands of small antennas instead of a traditional steel dish. This telescope is the first of its kind, and will be the largest radio telescope in the world. The data rate that LOFAR generates is about 14 times higher than that of the Large Hadron Collider in Cern. LOFAR uses software to combine the antenna signals into one large virtual instrument.

The presentation focuses on the investigation of the use of many-core hardware (multi-core CPUs, GPUs from NVIDIA and ATI, and the Cell) for several important algorithms that LOFAR uses. Further, it includes extensive performance evaluation and comparisons, in terms of both computational and power efficiency. Finally, it presents several reflections on the many-core hardware properties that are important for the field of radio astronomy, and looks toward how many-cores can help build even larger instruments.

Biography. Rob V. van Nieuwpoort is a postdoc at the Vrije Universiteit Amsterdam and ASTRON. His current research interests focus on many-core systems and radio astronomy. He got his PhD at the Vrije Universiteit Amsterdam on efficient Java-centric grid computing. He has designed and implemented the Manta, Ibis, Satin, and JavaGAT systems, and worked on the GridLab and Virtual Labs for E-science projects. At ASTRON, he works on the central, real-time data processing of the LOFAR software telescope, the largest telescope in the world. His research interests include high-performance computing, parallel and distributed algorithms, networks, programming languages, and compiler construction.
EAMA Foreword

This workshop aims to bring together application domain experts and computer architects to discuss emerging applications as well as their implications on current- and next-generation many-core architectures.

There has always been a close connection between the emergence of new usage models and new computer architectures. Only a decade ago, a typical desktop PC user may have cared a great deal about speeding up an Excel calculation. Today, users may care more about the computer’s ability to play media files downloaded from the Internet, as well as their experience in on-line virtual worlds. New, emerging workloads bring about new challenges for developing future computer architectures. At the same time, breakthroughs in computer system design enable new application domains. Recent development trends suggest that industry is moving to many-core architectures to better manage trade-offs among performance, energy efficiency, and reliability in deep-submicron technology nodes. This industry-wide movement toward many-core architectures opens up many opportunities for developing new classes of applications. Such computationally intensive tasks as real-time ray-tracing, multi-modal data mining, physical simulation, financial analytics, or virtual world that were not possible just a few years ago due to a lack of adequate computing power are now being realized on emerging many-core platforms.

We encouraged authors to submit papers focusing on emerging application domains (such as recognition/mining/synthesis (RMS), medical imaging, bioinformatics, visual computing, Web3D, datacenter workloads, business analytics, virtual worlds, etc.) and architectural implications of emerging applications. Besides the technical papers, EAMA 2010 included four invited talks.

John Carter
Karthick Rajamani
Application of Many-Core Architecture to Virtual World Workloads (invited talk)
by John Hurliman, Huaiyu (Kitty) Liu, and Mic Bowman, from Intel Corporation

Abstract. More and more individuals and organizations are using virtual worlds for training, corporate collaboration, collective design, and sharing experiences in ways that are only possible in a rich 3D environment. To meet the increasing demand of rich user experiences, high level of realism, and new usages such as experiencing a major-league baseball game virtually, virtual worlds need to scale up in several aspects: the number of simultaneously interacting users, the scene complexity, and the fidelity of user interactions.

Typically, a virtual world is composed of the scene graph that describes the world and its content plus a set of heterogeneous actors (physics engine, scripts, clients, etc.) operating on the scene graph. Yet most state-of-art virtual worlds are based on a homogeneous simulator-centric architecture, which treats virtual world operations as a set of homogeneous simulators, each owning a portion of the scene graph and the complete simulation and communication work inside the portion of the scene. Our work reveals that this architecture has inherent scalability barriers: a quadratic increase in communication overhead when the number of concurrent client connections increases, high overhead of workload migration, inefficient workload partitioning during load balancing, and a limited ability to provide low-detail aggregate views of large portions of the world.

The disconnect between homogeneous simulators and heterogeneous actors presents a major barrier to scalability. For instance, our measurements show the script engine has a much more scattered memory usage (higher rates of L1 cache misses and resource stalls), while the physics engine has a high locality of memory access. The homogeneous simulator-centric architecture, however, limits a virtual world’s ability to apply “appropriately configured” hardware to the heterogeneous compute and communication tasks and to scale flexibly with the dynamic addition of hardware.

We propose a new architecture, called “distributed scene graph” (DSG), which externalizes the scene graph and uses it as a communication bus to connect the heterogeneous actors. It enables the workload of each actor to be independently load balanced while mapping the actors to appropriate hardware that fits their compute characteristics. As an example, we discuss how the actors could be detached from the scene graph and appropriately mapped to small-core, wide vector processing units or large-core, large cache processors.

Our preliminary work has demonstrated the great potential of DSG on virtual worlds well beyond the capabilities of their current architecture. By detaching the client managers and running them on separate hardware provisioned for supporting a massive number of client connections, we have demonstrated orders of magnitude increase in the number of concurrent client connections comparing to previous best over-the-network performance.
Computational Challenges in the Operating Room: A Study of Transurethral Imaging and Technologies (invited talk)
by David Holmes from the Mayo Clinic

Abstract. Over the past 20 years, there have been radical changes to the surgical treatment of disease. New approaches can eradicate disease and spare viable tissue. The primary premise of these new approaches is to use small incisions and complex instrumentation. As a result, the field of minimally invasive and robotic surgery now dominates several medical disciplines. Instead of cracked bones and large incisions, patients are walking away from surgery with 1-inch scars and shorter recoveries. These new procedures are helping to reduce the healthcare burden of patients and societies.

With such a dramatic change in operating room procedures, surgeons and interventionalists are faced with a challenge: providing the same level of service with fewer resources. During a procedure, the surgeons frame of reference is now limited. Instead of viewing the entire organ of interest, the surgeon is afforded only a limited view either from an endoscopic camera or an x-ray fluoroscope. Rather than a large open cavity, the surgeon is left with a small area to work in. To overcome the challenges of minimally invasive procedures, surgeons are now relying heavily on technology to enhance the surgical experience. Imaging is a crucial technology as it provides detailed information about the patient’s body—both healthy and diseased. Moreover, when properly integrated into a procedure, imaging can provide real-time feedback about the therapy. The use of technology—imaging and otherwise—requires adequate computational resources to process, analyze, and visualize the data.

This talk presents a review of the basics of minimally invasive image-guided procedures, providing several specific examples of technology in the Operating Room. I highlight where computational tools have worked—and where they have not. For example, the use of transurethral ultrasound will likely change the way prostate disease is treated, but only if it is correctly married to the right computational tools. Finally, I look at the computational needs for future procedures.
Emerging Applications and the Macrochip (invited talk),
by Herb Schwetman from Sun Labs

Abstract. Emerging applications share some common trends: the need for faster response times, the need for an ever-increasing number of cycles, the need for faster access to main memory and remote processors.

The first two trends are being addressed by systems with more chips and more cores/chip. This means that as core and chip counts increase, the underlying system must supply more inter-chip communications bandwidth in order to show an improvement in system performance. The third trend is being addressed by systems with an increased number of nodes (a node has a processor and memory) and by larger memory parts. In both cases, the bandwidth requirements are increasing again.

This talk introduces the macrochip, a multi-site node with an embedded silicon-photonic interconnection network. Here, a site refers to a stack consisting of a processor chip, a memory chip and an interface bridge chip. This network uses high-density, energy-efficient optical communications to support high on-node bandwidth and low site-to-site message latencies. The talk then shows how the technologies being developed for the macrochip can address the needs of emerging applications. The talk concludes with some thoughts on future work in this area.
Accelerating the Future of Many-Core Software with the Single-Chip Cloud Computer: A 48-Core Research Microprocessor (invited talk), by Matthias Gries from Intel Germany Research Lab

Abstract. We present the design of the experimental single-chip cloud computer (SCC) by Intel Labs. The SCC is a research microprocessor containing the most Intel architecture cores ever integrated on a single silicon chip: 48 cores. We envision SCC as a concept vehicle for research in the areas of parallel computing including system software, compilers and applications. It incorporates technologies intended to scale multi-core processors to 100 cores and beyond including an on-chip network, advanced power management technologies, new data-sharing options using software-managed memory coherency or hardware-accelerated message passing, and intelligent resource management.

SCC is implemented in a 45-nm process integrating 1.3-B transistors. It is based on a tiled architecture with each tile containing two Pentium class cores, private L1 and L2 caches, and one mesh router. All 24 tiles have access to four DDR3 memory channels. These channels can provide up to 64-GB of main memory to the system. The on-die communication is organized in a regular 6x4 mesh of tiles using 16-B-wide data links. The SCC contains one frequency domain for each tile and eight voltage domains: two for on and off chip I/O and six for the cores. Each tile contains sensors to monitor the thermal state. SCC has a NUMA architecture including local caches and on-die distributed memory for low latency, hardware-assisted message passing or scratchpad use as well as an abundant external DRAM bandwidth and capacity. Thus, the processor can be used as a proxy for future manycore platforms by running several independent applications and operating systems concurrently on dedicated resources while applying fine-grain voltage and frequency scaling for best energy efficiency.

In this talk we review the chip’s architecture and highlight different system configurations that enable the exploration of compute, memory or communication limited workloads. We show the emulation-based design flow that enabled us to build the SCC with a relatively small design team while keeping high confidence in the quality of the design. This approach allowed us to boot an OS and begin system software design before production. We give an overview of the system software and prototype API that comes with SCC in order to access on-die resources. Finally, we describe an SCC co-traveler research program where Intel will collaborate with dozens of academic and industry research partners. We expect that this program will significantly accelerate the evolution and adoption of manycore software technologies at all levels of the manycore software stack. To highlight the potential of this program, we share some initial experiences and results from the SCC research community.
AMAS-BT Foreword

Long employed by industry, large scale use of binary translation and on-the-fly code generation is becoming pervasive both as an enabler for virtualization, processor migration and also as processor implementation technology. The emergence and expected growth of just-in-time compilation, virtualization and Web 2.0 scripting languages brings to the forefront a need for efficient execution of this class of applications. The availability of multiple execution threads brings new challenges and opportunities, as existing binaries need to be transformed to benefit from multiple processors, and extra processing resources enable continuous optimizations and translation.

The main goal of this half-day workshop was to bring together researchers and practitioners with the aim of stimulating the exchange of ideas and experiences on the potential and limits of architectural and microarchitectural support for binary translation (hence the acronym AMAS-BT). The key focus was on challenges and opportunities for such assistance and opening new avenues of research. A secondary goal was to enable dissemination of hitherto unpublished techniques from commercial projects.

The workshop scope includes support for decoding/translation, support for execution optimization and runtime support. It set a high scientific standard for such experiments, and requires insightful analysis to justify all conclusions. The workshop favors submissions that provide meaningful insights, and identify underlying root causes for the failure or success of the investigated technique. Acceptable work must thoroughly investigate and communicate why the proposed technique performs as the results indicate.

Mauricio Breternitz
Robert Cohn
Erik Altman
Youfeng Wu
WEED Foreword

The Second Workshop on Energy-Efficient Design provided a forum for the exchange of ideas on research on critical areas relating to energy-efficient computing, including energy-aware design techniques for systems (large and small), energy management policies and mechanisms, and standards for evaluating energy efficiency. It was well attended with a good mix of researchers and practitioners from industry and academia.

WEED 2010 built on the success of its predecessor by encouraging lively exchanges of ideas on energy-efficient computing, including energy-aware design techniques for systems and datacenters, power management techniques and solutions, and standards to promote energy-efficient computing. The discussion reflected the ideas of a broad mix of researchers from both academia and industry, as reflected in the composition of the technical program, panel speakers, and Program Committee.

We received a strong collection of papers describing research and practices in many areas of energy-efficient computing. All technical submissions were reviewed by at least four Program Committee members. We held an energy-efficient Program Committee meeting via teleconference and ultimately selected nine papers for inclusion in the program. We rounded out the program with a keynote talk by Parthasarathy Ranganathan, Saving the World, One Server at a Time.

We would like to thank the Program Committee for all of their hard work and the authors for their excellent submissions.

John Carter
Karthick Rajamani
Keynote: Saving the World, One Server at a Time!
by Parthasarathy Ranganathan from Hewlett Packard Research Labs

Abstract. Power and energy management, and more recently sustainability, are emerging to be critical challenges for future IT systems. While there has been extensive prior work in this space, a lot more needs to be done. In this talk, I discuss the challenges and opportunities in rethinking how we study and reason about energy efficiency for future systems. Specifically, I talk about how the confluence of emerging technology and industry trends offers exciting opportunities to systematically rethink the “systems stack” for the next orders of magnitude improvements in energy efficiency.

Biography. Partha Ranganathan is currently a distinguished technologist at Hewlett Packard Labs. His research interests are in systems architecture and manageability, energy efficiency, and systems modeling and evaluation. He is currently the principal investigator for the exascale datacenter project at HP Labs that seeks to design next-generation servers and datacenters and their management. He was a primary developer of the publicly distributed Rice Simulator for ILP Multiprocessors (RSIM). Partha received his BTech degree from the Indian Institute of Technology, Madras, and his MS and PhD from Rice University, Houston. Partha’s work has been featured in various venues including the Wall Street Journal, Business Week, San Francisco Chronicle, Times of India, slashdot, youtube, and Tom’s hardware guide. Partha has been named one of the world’s top young innovators by MIT Technology Review, and is a recipient of Rice University’s Outstanding Young Engineering Alumni award.
Welcome to the proceedings of the 6th Workshop on the Interaction between Operating System and Computer Architecture (WIOSCA). This workshop focuses on characterizing, modeling and optimizing the interaction between OS and hardware in light of emerging architecture paradigms (e.g., multi-core processors), workloads (e.g., commercial and server workloads) and computing technology (e.g., virtualization). The WIOSCA workshop aims at providing a forum for researchers and engineers from academia and industry to discuss their latest research in computer architecture and system software.

All submitted papers were reviewed by the Program Committee members. At least four reviews were written for each paper. In the end, the Program Committee decided to accept seven high-quality papers for this year’s workshop. By doing so, we hope that WIOSCA will provide an excellent forum for researchers to present and get feedback on their on-going, high-quality research.

This sixth edition of the WIOSCA workshop was held in conjunction with the 2010 International Symposium on Computer Architecture (ISCA 37). We therefore would like to thank the ISCA General Chair Andr Seznec, Program Chairs Uri Weiser and Ronny Ronen, and Workshop Chair Yanos Sazeides for accepting this workshop as part of the ISCA program. This workshop would not be possible without the help and hard work of many people. We would like to thank all the members of the Program Committee who spent considerable time reviewing the manuscripts. We also would like to thank all of the authors for their excellent submissions.

Tao Li
Onur Mutlu
James Poe
Table of Contents

A4MMC: Applications for Multi- and Many-Cores

Accelerating Agent-Based Ecosystem Models Using the Cell Broadband Engine ................................................................. 1
   Michael Lange and Tony Field

Performance Impact of Task Mapping on the Cell BE Multicore Processor ....................................................... 13
   Jörg Keller and Ana Lucia Varbanescu

Parallelization Strategy for CELL TV ........................................ 24
   Motohiro Takayama and Ryuji Sakai

Towards User Transparent Parallel Multimedia Computing on GPU-Clusters ......................................................... 28
   Ben van Werkhoven, Jason Maassen, and Frank J. Seinstra

Implementing a GPU Programming Model on a Non-GPU Accelerator Architecture ..................................................... 40
   Stephen M. Kofsky, Daniel R. Johnson, John A. Stratton,
   Wen-mei W. Hwu, Sanjay J. Patel, and Steven S. Lumetta

On the Use of Small 2D Convolutions on GPUs ...................... 52
   Shams A.H. Al Umairy, Alexander S. van Amesfoort,
   Irwan D. Setija, Martijn C. van Beurden, and Henk J. Sips

Can Manycores Support the Memory Requirements of Scientific Applications? ..................................................... 65
   Milan Pavlovic, Yoav Etsion, and Alex Ramirez

Parallelizing an Index Generator for Desktop Search ................. 77
   David J. Meder and Walter F. Tichy

AMAS-BT: 3rd Workshop on Architectural and Micro-Architectural Support for Binary Translation

Computation vs. Memory Systems: Pinning Down Accelerator Bottlenecks ................................................................. 86
   Martha A. Kim and Stephen A. Edwards

Trace Execution Automata in Dynamic Binary Translation ............ 99
   João Porto, Guido Araujo, Edson Borin, and Youfeng Wu
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISAMAP: Instruction Mapping Driven by Dynamic Binary Translation</td>
<td>117</td>
</tr>
<tr>
<td>Maxwell Souza, Daniel Nicácio, and Guido Araújo</td>
<td></td>
</tr>
<tr>
<td>EAMA: 3rd Workshop for Emerging Applications and Many-Core Architectures</td>
<td>139</td>
</tr>
<tr>
<td>Matthew A. Goodrum, Michael J. Trotter, Alla Aksel, Scott T. Acton and Kevin Skadron</td>
<td></td>
</tr>
<tr>
<td>What Kinds of Applications Can Benefit from Transactional Memory?</td>
<td>150</td>
</tr>
<tr>
<td>Mark Moir and Dan Nussbaum</td>
<td></td>
</tr>
<tr>
<td>Characteristics of Workloads Using the Pipeline Programming Model</td>
<td>161</td>
</tr>
<tr>
<td>Christian Bienia and Kai Li</td>
<td></td>
</tr>
<tr>
<td>WEED: 2nd Workshop on Energy Efficient Design</td>
<td>172</td>
</tr>
<tr>
<td>Laura Keys, Suzanne Rivoire, and John D. Davis</td>
<td></td>
</tr>
<tr>
<td>KnightShift: Shifting the I/O Burden in Datacenters to Management</td>
<td>183</td>
</tr>
<tr>
<td>Sabyasachi Ghosh, Mark Redekopp, and Murali Annavaram</td>
<td></td>
</tr>
<tr>
<td>Guarded Power Gating in a Multi-core Setting</td>
<td>198</td>
</tr>
<tr>
<td>Niti Madan, Alper Buyuktosunoglu, Pradip Bose, and Murali Annavaram</td>
<td></td>
</tr>
<tr>
<td>Using Partial Tag Comparison in Low-Power Snoop-Based Chip Multiprocessors</td>
<td>211</td>
</tr>
<tr>
<td>Ali Shafiee, Narges Shahidi, and Amirali Baniasadi</td>
<td></td>
</tr>
<tr>
<td>Achieving Power-Efficiency in Clusters without Distributed File System Complexity</td>
<td>222</td>
</tr>
<tr>
<td>Hrishikesh Amur and Karsten Schwan</td>
<td></td>
</tr>
<tr>
<td>What Computer Architects Need to Know about Memory Throttling</td>
<td>233</td>
</tr>
<tr>
<td>Heather Hanson and Karthick Rajamani</td>
<td></td>
</tr>
<tr>
<td>Predictive Power Management for Multi-core Processors</td>
<td>243</td>
</tr>
<tr>
<td>William Lloyd Bircher and Lizy John</td>
<td></td>
</tr>
</tbody>
</table>

IOMMU: Strategies for Mitigating the IOTLB Bottleneck ........................................ 256
Nadav Amit, Muli Ben-Yehuda, and Ben-Ami Yassour

Improving Server Performance on Multi-cores via Selective Off-Loading of OS Functionality ......................................................... 275
David Nellans, Kshitij Sudan, Erik Brunvand, and Rajeev Balasubramonian

Performance Characteristics of Explicit Superpage Support .................... 293
Mel Gorman and Patrick Healy

Interfacing Operating Systems and Polymorphic Computing Platforms Based on the MOLEN Programming Paradigm .......................... 311
Mojtaba Sabeghi and Koen Bertels

Extrinsic and Intrinsic Text Cloning .......................................................... 324
Marios Kleanthous, Yiannakis Sazeides, and Marios D. Dikaiakos

A Case for Coordinated Resource Management in Heterogeneous Multicore Platforms ................................................................. 341
Priyanka Tembey, Ada Gavrilovska, and Karsten Schwan

Topology-Aware Quality-of-Service Support in Highly Integrated Chip Multiprocessors ................................................................. 357
Boris Grot, Stephen W. Keckler, and Onur Mutlu

Author Index ......................................................................................... 377