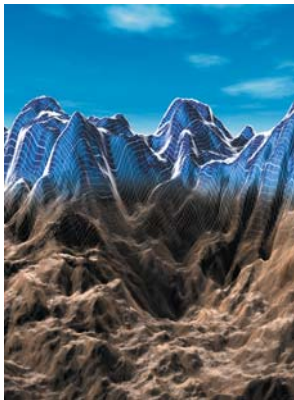


TOP PICKS FROM THE COMPUTER ARCHITECTURE CONFERENCES OF 2007



Sarita V. Adve
University of Illinois at
Urbana-Champaign

David Brooks
Harvard

Craig Zilles
University of Illinois at
Urbana-Champaign

..... This special issue represents the fifth anniversary of an important tradition in the computer architecture community: *IEEE Micro's* Top Picks from the Computer Architecture Conferences. This tradition, initiated by *IEEE Micro's* former Editor in Chief, Pradip Bose, attempts to share with the *IEEE Micro* readership a sampling of the top papers published in computer architecture in the past year. In choosing these papers, we ask ourselves, "For our colleagues who were not able to attend any conferences this year, which few papers would we recommend that they read?" We select these papers on the basis of the novelty of the work and its potential impact on industry—either short-term or long-term. As always, it is a difficult task to select only 10 from the many high-quality papers that have already been distinguished by being published in the proceedings of our field's top conferences. This year was no different.

The review process

For each submission, we requested, in addition to a copy of the conference paper, a three-page summary that highlighted the novelty of the work and argued its relevance for architects and designers of current- and future-generation computing systems. We received 70 submissions.

To review the submissions, we assembled a program committee of 33 highly respected architects from industry and academia:

- Dennis Abts, Cray
- David Albonese, Cornell University
- Erik Altman, IBM
- David August, Princeton University
- Todd Austin, University of Michigan, Ann Arbor
- Bryan Black, AMD
- Pradip Bose, IBM
- Doug Burger, University of Texas at Austin
- Calin Cascaval, IBM
- Yen-Kuang Chen, Intel
- Lieven Eeckhout, Ghent University
- Krisztián Flautner, ARM
- Rajiv Gupta, University of California, Riverside
- Mark D. Hill, University of Wisconsin–Madison
- Steve Keckler, University of Texas at Austin
- Alvin Lebeck, Duke University
- Mikko Lipasti, University of Wisconsin–Madison
- Margaret Martonosi, Princeton University
- Chuck Moore, AMD
- Vijaykrishnan Narayanan, Pennsylvania State University
- Chris J. Newburn, Intel
- Mark Oskin, University of Washington
- Vijay Pai, Purdue University
- Ravi Rajwar, Intel
- Partha Ranganathan, HP

- Eric Rotenberg, North Carolina State University
- Amir Roth, University of Pennsylvania
- André Seznec, IRISA/INRIA
- Balaram Sinharoy, IBM
- Josep Torrellas, University of Illinois at Urbana-Champaign
- Uri Weiser, Technion—Israel Institute of Technology
- Chris Wilkerson, Intel
- David Wood, University of Wisconsin—Madison

Each paper was assigned and received five reviews, all from program committee members. In assigning the reviews, we ensured that every paper had at least two reviews by industry representatives. On October 23, we held a day-long program committee (PC) meeting in Chicago, at which we had 30 of the 33 program committee members in attendance. (Three PC members took part by phone, due to unexpected obligations to product teams and a family crisis.)

Prior to this meeting, each PC member had read all of the reviews for their assigned papers and, taking these other viewpoints into account, entered a final score for each paper. For those papers lacking consensus in this final scoring, committee members discussed the papers by e-mail in advance of the PC meeting to try to reach an agreement. We found this process to be invaluable in making the PC meeting run smoothly, as few unexpected issues came to light at the meeting.

In anticipation of the difficulty of selecting only 10 papers from the many high-quality submissions, we used a two-pass acceptance process: On the initial pass, we voted each paper into one of three categories—*definite accept*, *accept-if-space*, and *reject*. Then, at the end of the day, we voted on which of the *accept-if-space* papers we should include. This approach addressed the concern of many PC members about committing to accept a paper before they had seen the whole field, but avoided tabling an unduly large number of papers.

Each paper discussion was open to all committee members except those who had a conflict of interest with the work under consideration; members were requested to

leave the room during the discussion and decision process for papers with which they had conflicts. Papers that included a program chair among their authors were handled completely outside the purview of the program chairs; David Albonesi, Editor in Chief of *IEEE Micro*, coordinated reviewer assignment, collected reviews by e-mail, and led the discussion for these submissions.

The papers

For this year's Top Picks issue, the committee selected 10 articles that demonstrate the breadth of ongoing computer architecture research (see the sidebar, "Top Picks of 2007").

The first two articles consider the role that compilers will play in future architectures. The first article, by Bridges et al., demonstrates that there may yet be opportunities for automatically parallelizing non-numeric programs, especially if programmers use language extensions that permit programs to have a range of legal outcomes to provide flexibility to the compiler. The second, by Neelakantam et al., shows that hardware atomicity is a simple, but powerful primitive that greatly simplifies the implementation of speculative compiler optimizations, which can improve single-thread performance while reducing power.

The next pair of articles brings new insights to the field of transactional memory (TM). Demonstrating that the manner in which any hardware TM system resolves conflicts can have a first-order impact on its performance, Bobba et al. describe several pathologies that can occur and propose approaches for their mitigation. Ramadan et al. bring TM research into a new domain—operating systems. They demonstrate how key parts of the Linux operating system can be converted to use transactions and present extensions to TM semantics that simplify this conversion.

Reliability and variability issues threaten to slow nanoscale technology scaling, and designers are increasingly looking for architectural solutions. Meixner, Bauer, and Sorin present a novel error detection mechanism that breaks the work of the processor into four separate tasks and detects errors by determining whether each

Top Picks of 2007

Compiler-architecture interactions

- Revisiting the Sequential-Programming Model for the Multicore Era
- Hardware Atomicity: An Effective Abstraction for Reliable Software Speculation

Transactional memory

- Performance Pathologies in Hardware Transactional Memory
- MetaTM/TxLinux: Transactional Memory for an Operating System

Reliability and variability

- Argus: Low-Cost, Comprehensive Error Detection in Simple Cores
- Replacing 6T SRAMs with 3T1D DRAMs in the L1 Data Cache to Combat Process Variability

On-chip interconnects

- Architecting Efficient Interconnects for Large Caches with CACTI 6.0
- Toward Ideal On-Chip Communication Using Express Virtual Channels

Memory system design

- Set-Dueling-Controlled Adaptive Insertion for High-Performance Caching
- Virtual Hierarchies

of these tasks is performed correctly. Liang et al. address process variability in on-chip cache memories by proposing to replace traditional static memory cells with high-performance dynamic memory cells.

As we enter the era of multimegabyte shared on-chip caches and many-core systems, interconnect design is increasingly important. Muralimanohar, Balasubramanian, and Jouppi present the latest version of the CACTI cache modeling tool, with improved support for modeling interconnects between large cache arrays. Kumar et al. consider the problem of interconnects in many-core chips, with a novel virtual-channel approach that bypasses intermediate routers for long-distance interconnects.

Finally, as we integrate more cores onto a chip, memory and coherence systems need to become more efficient to effectively utilize pin and internal interconnect bandwidths. Qureshi et al. demonstrate that a least-recently-used (LRU) policy is frequently ineffective for second-level caches, because the data working set exceeds the

cache size. They show a simple, effective, and robust extension that improves performance in these cases by adaptively setting the recency of new blocks. Marty and Hill propose a technique targeting server consolidation workloads that permits the localization of coherence traffic within a virtual server on a many-core chip, improving performance and performance isolation between virtual machines.

We hope you enjoy reading these articles, and we encourage you to read the original works as well. We welcome your feedback on this year's Top Picks and any suggestions on how to improve the special issue for the future. MICRO

Acknowledgments

Many people contributed to this special issue of *IEEE Micro*. We must first acknowledge the outstanding support from Editor in Chief David Albonesi, former Editor in Chief Pradip Bose, and the *IEEE Micro* staff—especially Robin Baldwin, Cheryl Baltes, Lindsey Buscher, Thomas Centrella, Noel Deeley, Carmen Garvey, Patricia Hildenbrand, Keith Parsons, Joan Taylor, Margaret Weatherford, and Alkenia Winston. We particularly thank the members of the program committee for their hard work and outstanding professionalism throughout the process. A special thanks goes to Edward S. Lee of University of Illinois, who went beyond the call of duty to ensure that the submission and review process went smoothly. Finally, we thank all authors who submitted papers; the depth of quality present in our top conferences makes selecting so few papers very difficult, but we wouldn't want it any other way.

Sarita V. Adve is a professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign. Her research interests are in computer architecture and systems, with a current focus on reliability and multicore programming. She has a PhD in computer science from the University of Wisconsin-Madison.

David Brooks is an associate professor of computer science at Harvard University. His research interests include architectural and software approaches to address power, thermal, and reliability issues for embedded and high-performance computing systems. Brooks has a PhD in electrical engineering from Princeton University.

Craig Zilles is an assistant professor in computer science at the University of Illinois at Urbana-Champaign. His research interests include interaction between compilers and computer architecture, especially in the context of managed and dynamic

languages. Zilles has a PhD in computer science from the University of Wisconsin-Madison. He is a member of IEEE and the ACM.

Direct questions and comments about this special issue to the guest editors: Sarita Adve, sadve@cs.uiuc.edu; David Brooks, dbrooks@eecs.harvard.edu; Craig Zilles, zilles@cs.uiuc.edu.

For more information on this or any other computing topic, please visit our Digital Library at <http://computer.org/csdl>.

IEEE
micro
The magazine for chip and silicon systems designers

**EDITORIAL
CALENDAR**

March-April Hot Chips 19

May-June Interaction of Computer Architecture and Operating Systems in the Many-Core Era

July-August Accelerator Architectures