

# Bibliography for CS/ECE 670 Fall 2013

Sanjay Rajopadhye, [Sanjay.Rajopadhye@colostate.edu](mailto:Sanjay.Rajopadhye@colostate.edu)

CS/ECE 670 is a course on Special Topics in Architecture, and the syllabus and focus may change from year to year. Students will build important research skills by reading papers, writing reviews, presenting papers to the class, and writing and presenting a mini-research exam. Another major component of the course will be a research project of the student's choosing. During the course, we will cover how to write a research proposal, practice presenting intermediate research project results in writing and with a presentation, and writing a research paper.

In Fall 2013, we will explore the evolution of modern computer architectures to exascale systems, where power and energy dominate most concerns, and where we are moving to a massive degree of parallelism. The topics that will be of special concern to us are

- General overview (e.g., “dark silicon,” and its implications) [10, 18, 26].
- Non-programmable (fixed function) Accelerators & their design/synthesis [6, 7, 24, 28, 29, 37].
- Memory system architectures especially for multi- and many-cores [2, 9, 12, 14, 30, 35, 38],
- Tradeoffs [11, 19] such as energy vs. delay, “performance” vs. flexibility.
- High level synthesis (especially for “domain specific” systems) [1, 3, 8, 25, 27, 34].
- Coarse-Grain Reconfigurable Architectures [17, 20, 33].
- Programmable accelerators: GPUs and massively vectorized systems [4, 23, 32, 36].
- Others [5, 13, 15, 16, 21, 22, 31]

Papers will be drawn from the following tentative bibliography. If you have strong preferences for specific papers in the list (or others not on the list) please email [Sanjay.Rajopadhye@colostate.edu](mailto:Sanjay.Rajopadhye@colostate.edu).

## References

- [1] J. Auerbach, D. F. Bacon, I. Burcea, P. Cheng, S. J. Fink, R. Rabbah, and S. Shukla. A compiler and runtime for heterogeneous computing. In *Proceedings of the 49th Annual Design Automation Conference, DAC '12*, pages 271–276, San Francisco, California, June 2012. ACM.
- [2] R. Ausavarungnirun, K. K.-W. Chang, L. Subramanian, G. H. Loh, and O. Mutlu. Staged memory scheduling: Achieving high performance and scalability in heterogeneous systems. In *ISCA '12 Proceedings of the 39th Annual International Symposium on Computer Architecture*, pages 416–427, Portland, OR, June 2012.
- [3] J. Bachrach, H. Vo, B. Richards, Y. Lee, A. Waterman, R. Avižienis, J. Wawrzynek, and K. Asanović. Chisel: Constructing hardware in a Scala embedded language. In *Proceedings of the 49th Annual Design Automation Conference, DAC '12*, pages 1216–1225, San Francisco, California, June 2012. ACM.

- [4] A. Bakhoda, J. Kim, and T. M. Aamodt. On-chip network design considerations for compute accelerators. In *Proceedings of the 19th International Conference on Parallel Architectures and Compilation Techniques*, PACT '10, pages 535–536, Vienna, Austria, 2010. ACM.
- [5] L. A. Bathen and N. Dutt. HaVOC: a hybrid memory-aware virtualization layer for on-chip distributed scratchpad and non-volatile memories. In *Proceedings of the 49th Annual Design Automation Conference*, DAC '12, pages 447–452, San Francisco, California, June 2012. ACM.
- [6] S. Cadambi, A. Majumdar, M. Becchi, S. Chakradhar, and H. P. Graf. A programmable parallel accelerator for learning and classification. In *Proceedings of the 19th International Conference on Parallel Architectures and Compilation Techniques*, PACT '10, pages 273–284, Vienna, Austria, 2010. ACM.
- [7] S. Chakradhar, M. Sankaradas, V. Jakkula, and S. Cadambi. A dynamically configurable co-processor for convolutional neural networks. In *Proceedings of the 37th Annual International Symposium on Computer Architecture*, ISCA '10, pages 247–257, Saint-Malo, France, June 2010. ACM.
- [8] J. Cong, P. Zhang, and Y. Zou. Optimizing memory hierarchy allocation with loop transformations for high-level synthesis. In *Proceedings of the 49th Annual Design Automation Conference*, DAC '12, pages 1233–1238, San Francisco, California, June 2012. ACM.
- [9] E. Donkoh, A. Lowery, and E. Shriver. A hybrid and adaptive model for predicting register file and SRAM power using a reference design. In *Proceedings of the 49th Annual Design Automation Conference*, DAC '12, pages 62–67, San Francisco, California, June 2012. ACM.
- [10] H. Esmailzadeh, E. Blem, R. St. Amant, K. Sankaralingam, and D. Burger. Dark silicon and the end of multicore scaling. In *ISCA '11 Proceedings of the 38th Annual International Symposium on Computer Architecture*, pages 365–376, San Jose, CA, June 2011. ACM.
- [11] K. Fan, M. Kudlur, G. S. Dasika, and S. A. Mahlke. Bridging the computation gap between programmable processors and hardwired accelerators. In *Proceedings of the 15th International Symposium on High Performance Computer Architecture*, pages 313–322, Raleigh, NC, June 2009.
- [12] D. Kim, S. Lee, J. Chung, D. H. Kim, D. H. Woo, S. Yoo, and S. Lee. Hybrid DRAM/PRAM-based main memory for single-chip CPU/GPU. In *Proceedings of the 49th Annual Design Automation Conference*, DAC '12, pages 888–896, San Francisco, California, June 2012. ACM.
- [13] M. A. Kim and S. A. Edwards. Computation vs. memory systems: Pinning down accelerator bottlenecks. In *Computer Architecture: ISCA 2010 International Workshops A4MMC, AMAS-BT, EAMA, WEED, WIOSCA, Revised Selected Papers*, pages 86–98, Saint-Malo, France, June 2012. Springer.
- [14] Y. Kim, V. Seshadri, D. Lee, J. Liu, and O. Mutlu. A case for exploiting subarray-level parallelism (SALP) in DRAM. In *ISCA '12 Proceedings of the 39th Annual International Symposium on Computer Architecture*, volume 40, pages 368–379, Portland, OR, June 2012. ACM.

- [15] A. Krishna, T. Heil, N. Lindberg, F. Toussi, and S. VanderWiel. Hardware acceleration in the IBM PowerEN processor: Architecture and performance. In *Proceedings of the 21st International Conference on Parallel Architectures and Compilation Techniques, PACT '12*, pages 389–400, Minneapolis, Minnesota, USA, 2012. ACM.
- [16] T. Krishna, L.-S. Peh, B. M. Beckmann, and S. K. Reinhardt. Towards the ideal on-chip fabric for 1-to-many and many-to-1 communication. In *MICRO-44 '11: Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture*, pages 71–82, Porto Alegre, Brazil, 2011. ACM.
- [17] V. Lari, S. Muddasani, S. Boppu, F. Hannig, M. Schmid, and J. Teich. Hierarchical power management for adaptive tightly-coupled processor arrays. *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 18(1):2:1–2:25, Jan. 2013.
- [18] V. W. Lee, C. Kim, J. Chhugani, M. Deisher, D. Kim, A. D. Nguyen, N. Satish, M. Smelyanskiy, S. Chennupaty, P. Hammarlund, R. Singhal, and P. Dubey. Debunking the 100x GPU vs. CPU myth: an evaluation of throughput computing on CPU and GPU. *ISCA '10 Proceedings of the 37th annual international symposium on computer architecture*, 38(3):451–460, June 2010.
- [19] Y. Lee, R. Avizienis, A. Bishara, R. Xia, D. Lockhart, C. Batten, and K. Asanović. Exploring the tradeoffs between programmability and efficiency in data-parallel accelerators. In *ISCA '11 Proceedings of the 38th Annual International Symposium on Computer Architecture*, pages 129–140, San Jose, CA, June 2011. ACM.
- [20] D. Liu, S. Yin, L. Liu, and S. Wei. Polyhedral model based mapping optimization of loop nests for CGRAs. In *Proceedings of the 50th Annual Design Automation Conference, DAC '13*, pages 19:1–19:8, Austin, TX, June 2013. ACM.
- [21] H. Ltaief, P. Luszczek, and J. Dongarra. Profiling high performance dense linear algebra algorithms on multicore architectures for power and energy efficiency. *Computer Science – R&D*, 27(2), Nov. 2012.
- [22] D. Melpignano, L. Benini, E. Flamand, B. Jogo, T. Lepley, G. Haugou, F. Clermidy, and D. Dutoit. Platform 2012, a many-core computing accelerator for embedded SoCs: Performance evaluation of visual analytics applications. In *Proceedings of the 49th Annual Design Automation Conference, DAC '12*, pages 1137–1142, San Francisco, California, June 2012. ACM.
- [23] J. Meng, D. Tarjan, and K. Skadron. Dynamic warp subdivision for integrated branch and memory divergence tolerance. In *ISCA '10 Proceedings of the 37th Annual International Symposium on Computer Architecture*, volume 38, pages 235–246, Saint-Malo, France, June 2010. ACM.
- [24] P. Milder, F. Franchetti, J. C. Hoe, and M. Püschel. Computer generation of hardware for linear digital signal processing transforms. *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 17(2):15:1–15:33, Apr. 2012.

- [25] A. A. Nacci, V. Rana, F. Bruschi, D. Sciuto, I. Beretta, and D. Atienza. A high-level synthesis flow for the implementation of iterative stencil loop algorithms on FPGA devices. In *Proceedings of the 50th Annual Design Automation Conference, DAC '13*, pages 52:1–52:6, Austin, TX, June 2013. ACM.
- [26] K. Palem and A. Lingamneni. What to do about the end of Moore’s law, probably! In *Proceedings of the 49th Annual Design Automation Conference, DAC '12*, pages 924–929, San Francisco, California, June 2012. ACM.
- [27] A. Papakonstantinou, D. Chen, W.-M. Hwu, J. Cong, and Y. Liang. Throughput-oriented kernel porting onto FPGAs. In *Proceedings of the 50th Annual Design Automation Conference, DAC '13*, pages 11:1–11:10, Austin, TX, June 2013. ACM.
- [28] A. Pedram, S. Z. Gilani, N. S. Kim, R. A. van de Geijn, M. J. Schulte, and A. Gerstlauer. A linear algebra core design for efficient Level-3 BLAS. In *IEEE 23rd International Conference on Application-Specific Systems, Architectures and Processors (ASAP)*, pages 149–152, Delft, the Netherlands, July 2012.
- [29] A. Pedram, R. A. van de Geijn, and A. Gerstlauer. Codesign tradeoffs for high-performance, low-power linear algebra architectures. *IEEE Transactions on Computers*, 61(12):1724–1736, Dec. 2012.
- [30] S. Rixner, W. J. Dally, B. Khailany, P. R. Mattson, U. J. Kapasi, and J. D. Owens. Register organization for media processing. In *Proceedings of the Sixth International Symposium on High-Performance Computer Architecture*, pages 375–386, Toulouse, France, Jan. 2000.
- [31] G. Russel, C. Riley, N. Henning, U. Dolinsky, A. Richards, A. F. Donaldson, and A. S. van Amesfoort. The impact of diverse memory architectures on multicore consumer software: an industrial perspective from the video games domain. In *Proceedings of the 2011 ACM SIGPLAN Workshop on Memory Systems Performance and Correctness, MSPC '11*, pages 37–42, San Jose, California, 2011. ACM.
- [32] G. Tan, L. Li, S. Triechle, E. Phillips, Y. Bao, and N. Sun. Fast implementation of DGEMM on Fermi GPU. In *SC'11: Proceedings of the 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, SC '11*, pages 35:1–35:11, Seattle, WA, 2011. ACM.
- [33] P. Vaidya and J. J. Lee. A novel multicontext coarse-grained reconfigurable architecture (CGRA) for accelerating column-oriented databases. *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*, 4(2):13:1–13:30, May 2011.
- [34] Y. Wang, P. Li, P. Zhang, C. Zhang, and J. Cong. Memory partitioning for multidimensional arrays in high-level synthesis. In *Proceedings of the 50th Annual Design Automation Conference, DAC '13*, pages 12:1–12:8, Austin, TX, June 2013. ACM.

- [35] T. Yamauchi, L. Hammond, and K. Olukotun. The hierarchical multi-bank DRAM: A high-performance architecture for memory integrated with processors. In *Proceedings of the Seventeenth Conference on Advanced Research in VLSI*, pages 303–319, Ann Arbor, MI, Sept. 1997.
- [36] G. L. Yuan, A. Bakhoda, and T. M. Aamodt. Complexity effective memory access scheduling for many-core accelerator architectures. In *Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*, pages 34–44, New York, NY, Dec. 2009.
- [37] W. Zhang, V. Betz, and J. Rose. Portable and scalable FPGA-based acceleration of a direct linear system solver. *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*, 5(1):6:1–6:26, Mar. 2012.
- [38] H. Zheng, J. Lin, Z. Zhang, and Z. Zhu. Decoupled DIMM: Building high-bandwidth memory system using low-speed DRAM devices. *Proceedings of the 2009 International Conference on Computer Architecture*, 37(3):255–266, June 2009.