PhD Candidacy Exam:
The Demise and Resurrection of Analog Computing

Yipeng Huang
{*yipeng@cs.columbia.edu*}

Department of Computer Science
Columbia University

April 8, 2015

Abstract

Until the 1970s, analog computers were anticipated to coexist with digital computers, participating in computations that were suited to the continuous model of computation, compared to the discrete one of digital computers. The past 50 years have proven those predictions wrong: the digital von Neumann architecture has become synonymous with computing machinery. Recently, however, voltage scaling has ended in digital computers, and new workloads have reignited interest in analog computers.

A rigorous approach to understanding how analog computers may support modern workloads, and where analog computers excel in relation to existing machines, is needed. In this candidacy syllabus, I begin by examining techniques for characterizing workloads, necessary for identifying problems suitable for analog computation. I will examine some of the seminal work in parallel CPUs, GPUs, and accelerators that have given digital computers high scalability and programability. Finally I will study contemporary approaches to analog computing, in particular from the perspective of neural networks.

1 Purpose

"The candidacy exam is an oral exam based on a syllabus prepared jointly by the student and his/her candidacy committee. Admission to candidacy (passing the exam) certifies that the student has demonstrated a depth of scholarship in the literature and the methods of the student's chosen area of research, and has demonstrated a facility with the scholarly skills of critical evaluation and verbal expression." – http://www.cs.columbia.edu/education/phd/requirements/candidacy

2 Candidate Research Area Statement

Emerging workloads in robotics and machine learning increasingly have inputs and desired outputs that are approximate and relatively low precision. The possibility of approximate computation opens the door to highly efficient computing, which may be supported by radically different machinery such as analog computers. I have worked on building a prototype analog computer, which has been fabricated in 65nm IC, and created the programming interface and architecture to expose the efficient analog computing core. My ongoing work in this research include finding the right programming model for scalable analog computing, and making an adequate comparison against existing digital techniques, which would allow us to understand the advantages of analog computing.

3 Faculty Committee Members

The candidate respectfully solicits the guidance and expertise of the following faculty members and welcomes suggestions for other important papers and publications in the exam research area.

1. Prof. Simha Sethumadhavan
2. Prof. Martha Kim
3. Prof. Luca Carloni

4 Exam Syllabus

The papers have broad coverage of seminal and/or recent work in computer architecture, which is needed to make a fair assessment of the advantages of analog computation. I start with benchmark suites that capture contemporary workloads for a variety of architectures [2] [3] [9], along with the metrics and techniques for studying these benchmarks [25] [10] [15].

Looking to understand limitations of digital computing, I will cover strengths and limitations of single-core [14] [18] [20] and multicore hardware architectures [1] [16] [6] [19], along with the accompanying parallel software model [5] [21].

I will examine extensions to digital architectures, which include GPGPU [24] and accelerator [13] [17] techniques, including specific case studies in physics and graphics [28] [12] [11] [22].

Moving toward an increasingly analog model of computation, I look at how approximate computing can be achieved in GPUs and accelerators [23] [27]. Finally, I will cover recent work in analog computing, particularly from the perspective of large-scale neural computation [8] [7] [26] [4].

References

[1] Balfour, J., and Dally, W. J. Design tradeoffs for tiled cmp on-chip networks. In *Proceedings of the 20th Annual International Conference on Supercomputing* (New York, NY, USA, 2006), ICS '06, ACM, pp. 187–198.

[2] Bienia, C., Kumar, S., Singh, J. P., and Li, K. The parsec benchmark suite: Characterization and architectural implications. In *Proceedings of the 17th International Conference on Parallel Architectures and Compilation Techniques* (New York, NY, USA, 2008), PACT '08, ACM, pp. 72–81.

[3] Che, S., Sheaffer, J., Boyer, M., Szafaryn, L., Wang, L., and Skadron, K. A characterization of the rodinia benchmark suite with comparison to contemporary cmp workloads. In *Workload Characterization (IISWC), 2010 IEEE International Symposium on* (Dec 2010), pp. 1–11.

[4] Chen, Y., Luo, T., Liu, S., Zhang, S., He, L., Wang, J., Li, L., Chen, T., Xu, Z., Sun, N., and Temam, O. Dadiannao: A machine-learning supercomputer. In *Microarchitecture (MICRO), 2014 47th Annual IEEE/ACM International Symposium on* (Dec 2014), pp. 609–622.

[5] Culler, D., Karp, R., Patterson, D., Sahay, A., Schauser, K. E., Santos, E., Subramonian, R., and von Eicken, T. Logp: Towards a realistic model of parallel computation. In *Proceedings of the Fourth ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming* (New York, NY, USA, 1993), PPOPP '93, ACM, pp. 1–12.

[6] Esmaeilzadeh, H., Blem, E., St. Amant, R., Sankaralingam, K., and Burger, D. Dark silicon and the end of multicore scaling. In *Proceedings of the 38th Annual International Symposium on Computer Architecture* (New York, NY, USA, 2011), ISCA '11, ACM, pp. 365–376.

[7] Esser, S., Andreopoulos, A., Appuswamy, R., Datta, P., Barch, D., Amir, A., Arthur, J., Cassidy, A., Flickner, M., Merolla, P., Chandra, S., Basilico, N., Carpin, S., Zimmerman, T., Zee, F., Alvarez-Icaza, R., Kusnitz, J., Wong, T., Risk, W., McQuinn, E., Nayak, T., Singh, R., and Modha, D. Cognitive computing systems: Algorithms and applications for networks of neurosynaptic cores. In *Neural Networks (IJCNN), The 2013 International Joint Conference on* (Aug 2013), pp. 1–10.

[8] Farabet, C., Martini, B., Corda, B., Akselrod, P., Culurciello, E., and LeCun, Y. Neuflow: A runtime reconfigurable dataflow processor for vision. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on* (June 2011), pp. 109–116.

[9] Ferdman, M., Adileh, A., Kocberber, O., Volos, S., Alisafaee, M., Jevdjic, D., Kaynak, C., Popescu, A. D., Ailamaki, A., and Falsafi, B. Clearing the clouds: A study of emerging scale-out workloads on modern hardware. *SIGPLAN Not. 47*, 4 (Mar. 2012), 37–48.

[10] Fields, B., Bodik, R., and Hill, M. Slack: maximizing performance under technological constraints. In *Computer Architecture, 2002. Proceedings. 29th Annual International Symposium on* (2002), pp. 47–58.

[11] Grossman, J. P., Kuskin, J. S., Bank, J. A., Theobald, M., Dror, R. O., Ierardi, D. J., Larson, R. H., Schafer, U. B., Towles, B., Young, C., and Shaw, D. E. Hardware support for fine-grained event-driven computation in anton 2. In *Proceedings of the Eighteenth International Conference on Architectural Support for Programming Languages and Operating Systems* (New York, NY, USA, 2013), ASPLOS '13, ACM, pp. 549–560.

[12] Hameed, R., Qadeer, W., Wachs, M., Azizi, O., Solomatnikov, A., Lee, B. C., Richardson, S., Kozyrakis, C., and Horowitz, M. Understanding sources of inefficiency in general-purpose chips. In *Proceedings of the 37th Annual International Symposium on Computer Architecture* (New York, NY, USA, 2010), ISCA '10, ACM, pp. 37–47.

[13] Hartenstein, R. A decade of reconfigurable computing: A visionary retrospective. In *Proceedings of the Conference on Design, Automation and Test in Europe* (Piscataway, NJ, USA, 2001), DATE '01, IEEE Press, pp. 642–649.

[14] Hind, M. Pointer analysis: Haven't we solved this problem yet? In *Proceedings of the 2001 ACM SIGPLAN-SIGSOFT Workshop on Program Analysis for Software Tools and Engineering* (New York, NY, USA, 2001), PASTE '01, ACM, pp. 54–61.

[15] Hoste, K., and Eeckhout, L. Microarchitecture-independent workload characterization. *Micro, IEEE 27*, 3 (May 2007), 63–72.

[16] Kim, C., Sethumadhavan, S., Govindan, M. S., Ranganathan, N., Gulati, D., Burger, D., and Keckler, S. W. Composable lightweight processors. In *Proceedings of the 40th Annual IEEE/ACM International Symposium on Microarchitecture* (Washington, DC, USA, 2007), MICRO 40, IEEE Computer Society, pp. 381–394.

[17] Lee, Y., Avizienis, R., Bishara, A., Xia, R., Lockhart, D., Batten, C., and Asanović, K. Exploring the tradeoffs between programmability and efficiency in data-parallel accelerators. In *Proceedings of the 38th Annual International Symposium on Computer Architecture* (New York, NY, USA, 2011), ISCA '11, ACM, pp. 129–140.

[18] Leverich, J., Arakida, H., Solomatnikov, A., Firoozshahian, A., Horowitz, M., and Kozyrakis, C. Comparing memory systems for chip multiprocessors. In *Proceedings of the 34th Annual International Symposium on Computer Architecture* (New York, NY, USA, 2007), ISCA '07, ACM, pp. 358–368.

[19] Lotfi-Kamran, P., Grot, B., Ferdman, M., Volos, S., Kocberber, O., Picorel, J., Adileh, A., Jevdjic, D., Idgunji, S., Ozer, E., and Falsafi, B. Scale-out processors. *SIGARCH Comput. Archit. News 40*, 3 (June 2012), 500–511.

[20] McFarlin, D. S., Tucker, C., and Zilles, C. Discerning the dominant out-of-order performance advantage: Is it speculation or dynamism? *SIGPLAN Not. 48*, 4 (Mar. 2013), 241–252.

[21] Pingali, K., Nguyen, D., Kulkarni, M., Burtscher, M., Hassaan, M. A., Kaleem, R., Lee, T.-H., Lenharth, A., Manevich, R., Méndez-Lojo, M., Prountzos, D., and Sui, X. The tao of parallelism in algorithms. *SIGPLAN Not. 46*, 6 (June 2011), 12–25.

[22] Qadeer, W., Hameed, R., Shacham, O., Venkatesan, P., Kozyrakis, C., and Horowitz, M. A. Convolution engine: Balancing efficiency & flexibility in specialized computing. *SIGARCH Comput. Archit. News 41*, 3 (June 2013), 24–35.

[23] Samadi, M., Lee, J., Jamshidi, D. A., Hormati, A., and Mahlke, S. Sage: Self-tuning approximation for graphics engines. In *Proceedings of the 46th Annual IEEE/ACM International Symposium on Microarchitecture* (New York, NY, USA, 2013), MICRO-46, ACM, pp. 13–24.

[24] Seiler, L., Carmean, D., Sprangle, E., Forsyth, T., Abrash, M., Dubey, P., Junkins, S., Lake, A., Sugerman, J., Cavin, R., Espasa, R., Grochowski, E., Juan, T., and Hanrahan, P. Larrabee: A many-core x86 architecture for visual computing. In *ACM SIGGRAPH 2008 Papers* (New York, NY, USA, 2008), SIGGRAPH '08, ACM, pp. 18:1–18:15.

[25] Sherwood, T., Perelman, E., Hamerly, G., and Calder, B. Automatically characterizing large scale program behavior. In *Proceedings of the 10th International Conference on Architectural Support for Programming Languages and Operating Systems* (New York, NY, USA, 2002), ASPLOS X, ACM, pp. 45–57.

[26] St. Amant, R., Yazdanbakhsh, A., Park, J., Thwaites, B., Esmaeilzadeh, H., Hassibi, A., Ceze, L., and Burger, D. General-purpose code acceleration with limited-precision analog computation. *SIGARCH Comput. Archit. News 42*, 3 (June 2014), 505–516.

[27] Yeh, T., Faloutsos, P., Ercegovac, M., Patel, S., and Reinman, G. The art of deception: Adaptive precision reduction for area efficient physics acceleration. In *Microarchitecture, 2007. MICRO 2007. 40th Annual IEEE/ACM International Symposium on* (Dec 2007), pp. 394–406.

[28] Yeh, T. Y., Faloutsos, P., Patel, S. J., and Reinman, G. Parallax: An architecture for real-time physics. *SIGARCH Comput. Archit. News 35*, 2 (June 2007), 232–243.