

Greg Hamerly

Department of Computer Science
Baylor University
One Bear Place #97356
Waco, TX 76798-7356, USA

+1-254-710-6846
hamerly@cs.baylor.edu
<http://cs.baylor.edu/~hamerly/>

1 Personal information

- Date of birth: November 16, 1977
- Place of birth: Rochester, New York, USA
- Sex: Male
- Nationality: United States of America
- Marital status: Married

2 Degrees

- Ph.D. Computer Science, University of California, San Diego (June 2003)
- M.S. Computer Science, University of California, San Diego (June 2001)
- B.S. (Magna cum laude) Computer Science, California Polytechnic State University, San Luis Obispo (June 1999)

3 Research experience

- Postdoctoral researcher in machine learning at the Katholieke Universiteit, Leuven (Belgium) (2003-2004)
- Ongoing research with Brad Calder and other colleagues on the SimPoint project (UCSD) for allowing fast and accurate simulation of computer processors (2002-2005)
- Research with Charles Elkan (UCSD) on developing high-quality data clustering algorithms (2000-2003)
- Research assistant for Scott Baden (UCSD) for developing a visualization system for scientific simulations (1999)

4 Professional employment

- Assistant professor, Baylor Department of Computer Science (2004 - present)
- Machine learning researcher and software developer for Mohomine, Inc. (2/2000-8/2000, 6/2001-9/2001)
- Graphics programmer and software developer for Netscape Communications Corporation (6/1997-9/1997)

- Graphics programmer and software developer for DigitalStyle Corporation (6/1996-9/1996)
- Web and software developer for ElectriCiti Corporation (6/1995-9/1995)

5 Honors

- Two-time world finalist in the ACM international collegiate programming contest (2000, 2001)

6 Grants

- Intel Corporation: \$50000 (2007-2008)
- Baylor Young Investigator Development Program: \$25000 (2006-2007)
- Belgian American Educational Foundation (BAEF) fellowship: \$17000 (2003-2004)
- Sloan foundation research fellowship through the UCSD CMRR: \$3000 (2001)
- J.L. Moore Ph.D. fellowship: \$40000 (1999-2003)
- Powell fellowship: \$15000 (1999)

7 Professional activities

I have been a program committee member for the following conferences:

- ACM SIG Knowledge Discovery in Databases (2008)
- Conference on Information and Knowledge Management (2006-2008)
- International Conference on Machine Learning (2004)

I have been a reviewer for the following conferences and journals:

- Journal of Machine Learning Research (JMLR)
- International Conference on Machine Learning (ICML)
- Neural Information Processing Systems (NIPS)
- AAAI Conference
- IEEE Transactions on Knowledge and Data Engineering (TKDE)
- IEEE Transactions on Parallel and Distributed Systems (TPDS)
- IEEE Transactions on Computers (TC)
- Journal of Intelligent and Fuzzy Systems (JIFS)
- International Conference on Discovery Science

I have also been involved with the following events:

- Organizer and coach for the ACM programming contest at Baylor (2004-present)
- Organizer and coach for the ACM programming contest at UCSD (2000-2003)
- Coordinator for UCSD Artificial Intelligence Seminar (2000-2002)

8 Teaching

At Baylor University:

- CSI 4v96 (competitive learning) (Each semester Fall 2005 - Spring 2008)
- CSI 4336 (theory of computation) (Each fall 2005-2007)
- CSI 5v93/5325 (machine learning) (Each spring 2005-2008)
- CSI 3334 (data structures) (Fall 2004, each semester Fall 2005-Spring 2008)

At UCSD:

- TA for CSE 134A (web programming and databases), taught by Charles Elkan (Fall 2002)
- TA for CSE 130 (programming languages), taught by Charles Elkan (Spring 2002)
- TA for CSE 127 (computer security), taught by Bennet Yee (Winter 2002)
- TA for CSE 134A (web programming and databases), taught by Charles Elkan (Fall 2001)
- TA for CSE 240 (graduate computer architecture), taught by Scott Baden (Fall 2000)

At Cal Poly, San Luis Obispo:

- Lecturer for CSc 101 (introduction to programming) (Summer 2002)

9 Software written

1. I am the primary author and maintainer of the SimPoint toolkit, which is used by hundreds of researchers in industry (Intel, IBM, etc.) and academia as the standard for identifying program phase behavior. The software is available for free from <http://www.cs.ucsd.edu/users/calder/simpoint/>.
2. I have written Matlab code for several novel clustering algorithms which I make available to the research community (G-means, PG-means, and variants on k -harmonic means).

10 Refereed publications

1. Greg Hamerly and Charles Elkan. Bayesian approaches to failure prediction for disk drives. In *Proceedings of the eighteenth international conference on machine learning (ICML)*, June 2001.
2. Tim Sherwood, Erez Perelman, Greg Hamerly, and Brad Calder. Automatically characterizing large scale program behavior. In *Proceedings of the tenth international conference on architectural support for programming languages and operating systems (ASPLOS-X)*, October 2002.
3. Greg Hamerly and Charles Elkan. Alternatives to the k -means algorithm that find better clusterings. In *Proceedings of the ACM conference on information and knowledge management (CIKM)*, November 2002.
4. Greg Hamerly. *Learning structure and concepts in data through data clustering*. PhD thesis, University of California San Diego, June 2003.
5. Erez Perelman, Greg Hamerly, Michael Van Biesbrouck, Tim Sherwood, and Brad Calder. Using SimPoint for accurate and efficient simulation. In *Proceedings of the international conference on measurement and modeling of computer systems (SIGMETRICS)*, June 2003.
6. Erez Perelman, Greg Hamerly, and Brad Calder. Picking statistically valid and early simulation points. In *Proceedings of the international conference on parallel architectures and compilation techniques (PACT)*, September 2003.
7. Tim Sherwood, Erez Perelman, Greg Hamerly, Suleyman Sair, and Brad Calder. Discovering and exploiting program phases. *IEEE Micro: Micro's top picks from computer architecture conferences*, 23(6), November–December 2003.
8. Greg Hamerly and Charles Elkan. Learning the k in k -means. In *Proceedings of the seventeenth annual conference on neural information processing systems (NIPS)*, December 2003.
9. Greg Hamerly, Erez Perelman, and Brad Calder. How to use SimPoint to pick simulation points. *ACM SIGMETRICS Performance Evaluation Review*, 31(4), March 2004.
10. John Seng and Greg Hamerly. Exploring perceptron-based register value prediction. In *Proceedings of the second value-prediction and value-based optimization workshop*, pages 10–16. International conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS-XI), October 2004.
11. Jeremy Lau, Erez Perelman, Greg Hamerly, Tim Sherwood, and Brad Calder. Motivation for variable length intervals to find hierarchical phase behavior. In proceedings of the IEEE international symposium on performance analysis of systems and software (ISPASS), 2005.
12. Jeremy Lau, Jack Sampson, Erez Perelman, Greg Hamerly, and Brad Calder. The strong correlation between code signatures and performance. In the IEEE international symposium on performance analysis of systems and software (ISPASS), 2005.
13. Greg Hamerly, Erez Perelman, Jeremy Lau, and Brad Calder. SimPoint 3.0: Faster and more flexible program analysis. In proceedings of the Workshop on Modeling, Benchmarking and Simulation (MoBS), June 2005.
14. Greg Hamerly, Erez Perelman, and Brad Calder. Comparing multinomial and k -means clustering for SimPoint. In proceedings of the IEEE international symposium on performance analysis of systems and software (ISPASS), 2006.

15. Greg Hamerly, Erez Perelman, Jeremy Lau, and Brad Calder. SimPoint 3.0: Faster and more flexible program phase analysis. In the Journal of Instruction-Level Parallelism (JILP), vol. 7, September 2005.
16. Greg Hamerly, Erez Perelman, Jeremy Lau, Tim Sherwood, and Brad Calder. Using machine learning to guide architecture simulation. In the Journal of Machine Learning Research (JMLR), vol. 7, pages 343–378, February 2006.
17. Yu Feng and Greg Hamerly. PG-means: learning the number of clusters in data. In proceedings of the twentieth annual conference on neural information processing systems (NIPS), December 2006.
18. Erez Perelman, Jeremy Lau, Harish Patil, Aamer Jaleel, Greg Hamerly, and Brad Calder. Cross Binary Simulation Points. In proceedings of the International Symposium on Performance Analysis of Systems and Software (ISPASS-2007), March 2007.
19. Joshua Johnston and Greg Hamerly. Improving SimPoint accuracy for small simulation budgets with EDCM clustering. In proceedings of the SMART workshop at the HiPEAC conference, January 2008.

11 Book chapters

1. Brad Calder, Timothy Sherwood, Greg Hamerly, and Erez Perelman. SimPoint: Picking Representative Samples to Guide Simulation. Chapter 7 in the book Performance Evaluation and Benchmarking, edited by Lizy Kurian John and Lieven Eeckhout; published by CRC Press.

12 Graduate students advised

- Joshua Johnston, M.Sc. in Computer Science, May 2007. Thesis title: “Clustering in high dimension and choosing cluster representatives for SimPoint.”
- Yu ‘Frank’ Feng, M.Sc. in Computer Science, August 2006. Thesis title: “PG-means: learning the number of clusters in data.”

13 Professional references

- Brad Calder, calder@cs.ucsd.edu
- Charles Elkan, elkan@cs.ucsd.edu
- Jeff Donahoo, donahoo@cs.baylor.edu