AMIN FIROOZSHAHIAN

Phone: (650) 714-0592 Mountain View, CA 94040
aminf13@gmail.com
www.firoozshahian.com

EXPERIENCE

Intel Corporation, Santa Clara, CA Aug. 2025 – Present

Senior Director – AI SoC (Architecture)

- Driving AI accelerator SoC architecture definition and execution

Rain AI, San Franciso, CA May 2024 – June 2025

Lead Architect

- Led architecture team
- Drove architecture development of the next generation energy efficient AI accelerators
  - Oversaw architectural specification development
  - Developed functional and cycle accurate simulators
  - Led workload mapping and performance analysis

Meta Platforms Inc., Menlo Park, CA Feb. 2019 – May 2024

Research Scientist, Infrastructure

- Leading architect, core ASIC architecture team
  - Led a team of experts, defining, specifying and documenting next generation deep learning accelerator architectures
- Led architectural model development
  - Functional models used for software prototyping
  - Cycle-accurate models used for performance evaluation and design verification
- Led vendor interaction for processor (RISC-V) IP procurement
  - Specified IP requirements and expectations
  - Configured, customized and deployed IPs internally
  - Handled internal maintenance and support

Intel Corporation, Santa Clara, CA Nov. 2014 – Feb. 2019

Senior Staff - Silicon Architecture Engineer, Product Architecture Group

- Led architecture development and software enablement of novel technology for Xeon family of products
  - Drove definition of the architecture and hardware/software interfaces
  - Collaborated with various software teams for technology enablement
  - Closely cooperated with the micro-architecture and design teams for technology implementation and validation
  - Lead performance evaluation and benchmarking of various aspects of the technology
- Member of the AIPG performance evaluation team
  - Developed cycle-accurate performance simulator of the Crest product line for deep learning acceleration

- Member of the Architecture Patent Committee
    - Evaluated and voted on architecture related patent proposals

Hicamp Systems Inc, Menlo Park, CA Oct. 2008 – Nov. 2014

Member of Technical Staff

- One of the first two employees joining the company
- Key member of the technical team developing and prototyping novel technology
    - Owned and delivered two major blocks of the architecture
- Responsibilities included:
    - Functional simulation
    - Architecture definition
    - Micro-architecture design
    - RTL coding and implementation
    - Test and verification
    - FPGA synthesis and timing closure
    - System debug and bring up
- Performed benchmarking and performance evaluation of the technology on the prototype system
- ***Acquired by Intel in 2014***

Stanford University, Stanford, CA Jan. 2001 – Oct. 2008

Research Assistant, Computer Systems Laboratory

- Architect for Smart Memories Project
- Designed a reconfigurable memory system for the test chip, with support for cache coherence, streaming and transactional memory models
- Developed a functional simulator to evaluate performance of the memory system architecture
- Designed micro-architecture and developed RTL implementation of a reconfigurable protocol controller for supporting cache coherent, streaming and transactional memory models
- Synthesized, and preformed pre-silicon test and verification of Smart Memories multiprocessor system
- Designed, implemented and tested an 8×8 packet switch with support for eight virtual channels and broadcasting for connecting Smart Memories test chips on a multi-chip board

Talesh Electronic Company (TEC), Tehran, Iran 1996 – 2001

Hardware/System Engineer

- Involved in development of a number of projects, including industrial automation, data acquisition and microprocessor-based systems

EDUCATION

PhD Stanford University, Electrical Engineering Department Jan. 2009
Stanford, CA, 94305
Dissertation: "Smart Memories: A Reconfigurable Memory System Architecture"

MS University of Tehran, Faculty of Engineering Jan. 2001
Tehran, Iran

BS Sharif University of Technology, Computer Engineering Department — Jun. 1998
Tehran, Iran

AWARDS AND HONORS

Winner of the Best Paper Award — 2012
26th International Conference on Supercomputing (ICS '12)

Winner of the DAC/ISSCC Student Design Contest — 2009
Smart Memories Polymorphic Chip Multiprocessor

SELECTED PUBLICATIONS

Journal Papers

A. Firoozshahian, J. Coburn, A. Punj, A. Sukumaran Rajam, C. Boyer, R. Nattoji, M. Bathla, B. Dreyer, S. Srinivasan, H. Pilla, M. Rotzin, S. Rajupalem, K.R. Jagannath, K. Noru, H. Reddy, Ch. Yang, Ch.H. Su, Ch. Cheng, "Comparative Analysis of Loosely and Tightly Coupled Accelerator Architectures for Machine Learning," IEEE Micro, volume 45, issue 5, Sep.-Oct. 2025, pp. 79-93.

M. Wachs, O. Sacham, Z. Asgar, A. Firoozshahian, S. Richardson and M. Horowitz, "Bringing Up a Chip on the Cheap," IEEE Design and Test of Computers, vol. 29, no. 6, Dec. 2012, pp. 57-65.

O. Sacham, O. Azizi, M. Wachs, W. Qadeer, A. Asgar, K. Kelley, J.P. Stevenson, S. Richardson, M. Horowitz, B. Lee, A. Solomatnikov and A. Firoozshahian, "Rethinking Digital Design: Why Design Must Change," IEEE Micro, volume 30, issue 6, Nov.-Dec. 2010, pp. 9-24 ***(Invited Paper)***.

J. Leverich, H. Arakida, A. Solomatnikov, A. Firoozshahian, M. Horowitz and C. Kozyrakis, "Comparative Evaluation of Memory Models for Chip Multi-Processors," ACM Transactions on Architecture and Code Optimization (TACO), vol. 5, no. 3, Nov. 2008, Article no. 12.

Conference Papers

J. Coburn, C. Tang, S. Abu Asal, N. Agrawal, R. Chinta, H. Dixit, B. Dodds, S. Dwarakapuram, A. Firoozshahian, C. Gao, K. Gondkar, T. Graf, J. Hu, J. Huang, S. Hughes, A. Hutchin, B. Jakka, G.J. Chen, I. Kalyanaraman, A. Kamath, P. Kansal, E. Kazi, R. Levenstein, M. Maddury, A. Mastro, S. Medaiyese, P. Modi, J. Montgomery, S. Nadathur, A. Nagpal, A. Narasimha, M. Naumov, E. Ozer, J. Park, P. Ramani, H. Reddy, D. Reiss, D. Roy, S. Sekar, A. Sharma, P. Shetty, A. Sukumaran-Rajam, E. Tal, M. Tsai, S. Varshini, R. Wareing, O. Wu, X. Xie, J. Yang, H. Yu, T. Zargar, Z. Zeng, F. Zhang, A. Matthews, X. Jiao, J. Zhang, E. Menage, T.E. Stokke, M. Sourouri, "Meta's Second Generation AI Chip: Model-Chip Co-Design and Productization Experiences," Proceedings of the 52nd Annual International Symposium on Computer Architecture (ISCA '25), June 21-25, 2025, Tokyo, Japan.

A. Firoozshahian, J. Coburn, R. Levenstein, R. Nattoji, A. Kamath, O. Wu, G. Grewal, H. Aepala, B. Jakka, B. Dreyer, A. Hutchin, U. Diril, K. Nair, E. K. Aredestani, M. Schatz, Y. Hao, R. Komuravelli, K. Ho, S. Abu Asal, J. Shajrawi, K. Quinn, N. Sreedhara, P. Kansal, W. Wei, D. Jayaraman, L. Cheng, P. Chopda, E. Wang, A. Bikumandla, A. K. Sengottuvel, K. Thottempudi, A. Narasimha, B. Dodds, C. Gao, J. Zhang, M. Al-Sanabani, A. Zehtabioskuie, J. Fix, H. Yu, R. Li, K. Gondkar, J. Montgomery, M. Tsai, S. Dwarakapuram, S. Desai, N. Avidan, P. Ramani, K. Narayanan, A. Mathews, S. Gopal, M. Naumov, V. Rao, K. Noru, H. Reddy, P. Venkatapuram, A.

Bjorlin, "MTIA: First Generation Silicon Targeting Meta's Recommendation Systems," Proceedings of the 50th Annual International Symposium on Computer Architecture (ISCA '23), June 17-21, 2023, Orlando, FL.

L. Ke, U. Gupta, B. Y. Cho, D. Brooks, V. Chandra, U. Diril, A. Firoozshahian, K. Hazelwood, B. Jia, H.S. Lee, M. Li, B. Maher, D. Mudigere, M. Naumov, M. Schatz, M. Smelyanskiy, X. Wang, B. Reagen, C.J. Wu, M. Hempstead, X. Zhang, "RecNMP: Accelerating Personalized Recommendation with Near-Memory Processing," Proceedings of the 47th Annual International Symposium on Computer Architecture (ISCA '20), May 30th – June 3rd 2020, Valencia, Spain, pp. 790-803.

S. Ghorbani, Z. Yang, P.B. Godfrey, Y. Ganjali and A. Firoozshahian, "DRILL: Micro Load Balancing for Low-Latency Data Center Networks," Proceedings of the Conference of the ACM Special Interest Group on Data Communication (SIGCOMM '17), Aug. 21-25, 2017, Los Angeles, CA, pp. 225-238.

S. Ghorbani, B. Godfrey, Y. Ganjali and A. Firoozshahian, "Micro Load Balancing in Data Centers with DRILL," Proceedings of the 14th ACM Workshop on Hot Topics in Networks (HotNets XIV), Nov. 16-17, 2015, Philadelphia, PA, Article no. 17.

H. Litz, D. Cheriton, A. Firoozshahian, O. Azizi and J.P. Stevenson., "SI-TM: Improving Transactional Memory Abort Rates through Snapshot Isolation," Proceedings of the 19th international conference on Architectural support for programming languages and operating systems (ASPLOS 2014), Mar. 01-05, 2014, Salt Lake City, UT, pp. 383-398.

J.P. Stevenson, A. Firoozshahian, A. Solomatnikov, M. Horowitz and D. Cheriton, "Sparse Matrix-Vector Multiply on HICAMP Architecture," Proceedings of the 26th ACM international conference on Supercomputing (ICS '12), Jun. 25-29, 2012, Venice, Italy, pp. 195-204 ***(Winner of the Best Paper Award)***.

D. Cheriton, A. Firoozshahian, A. Solomatnikov, J.P. Stevenson and O. Azizi, "HICAMP: Architectural Support for Efficient Concurrency-Safe Shared Structured Data Access," Proceedings of the seventeenth international conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS XVII), Mar. 03-07, 2012, London, England, pp. 287-300.

A. Solomatnikov, A. Firoozshahian, O. Shacham, Z. Asgar, M. Wachs, W. Qadeer, S. Richardson and M. Horowitz, "Using a Configurable Processor Generator for Computer Architecture Prototyping," Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO 42), Dec. 12-16, 2009, New York, NY, pp. 358-369.

A. Firoozshahian, A. Solomatnikov, O. Shacham, Z. Asgar, S. Richardson, C., Kozyrakis and M. Horowitz, "A Memory System Design Framework: Creating Smart Memories," Proceedings of the 36th annual international symposium on Computer architecture (ISCA '09), June. 20-24, 2009, Austin, TX, pp. 406-417.

O. Shacham, M. Wachs, A. Solomatnikov, A. Firoozshahian, S. Richardson and M. Horowitz, "Verification of Chip Multiprocessor Memory Systems Using A Relaxed Scoreboard," Proceedings of the 41st annual IEEE/ACM International Symposium on Microarchitecture (MICRO 41), Nov. 08-12, 2008, Lake Como, Italy, pp. 294-305.

J. Leverich, H. Arakida, A. Solomatnikov, A. Firoozshahian, M. Horowitz and C. Kozyrakis, "Comparing Memory Systems for Chip Multi-Processors," Proceedings of the 34th annual international symposium on Computer architecture (ISCA '07), Jun. 09-13, 2007, San Diego, CA, pp. 358-368.

A. Firoozshahian, V. Manshadi, A. Goel and B. Prabhakar, "Efficient, Fully Local Algorithms for CIOQ Switches," Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM 2007), May. 06-12, 2007, Barcelona, Spain.