Divya Kiran Kadiyala

dkadiyala3.github.io | dkadiyala3@gatech.edu | +1 (480) 577 - 0493 |

SUMMARY

I am a computer systems researcher at Georgia Tech focused on developing tailored memory system architectures to enhance the performance of parallel, scientific, and deep learning AI/ML workloads in resource-constrained and memory bandwidth-intensive environments. My research draws on insights from computer architecture, memory system design, distributed AI/ML systems, and CXL technologies. I am currently seeking full-time opportunities where I can apply my academic expertise and industry experience to drive innovation in next-generation memory and computer system architectures.

EDUCATION

EBOOMION
Georgia Institute of Technology (Georgia Tech)
Arizona State University
KL University, Guntur, A.P., India
RESEARCH & WORK EXPERIENCE
School of Computer Science, Graduate Research Assistant, Georgia Tech
 Working on novel memory system optimizations and CXL-enabled architectures to accelerate parallel, bandwidth intensive, and deep learning workloads in datacenter and cloud environments. Pioneered a cluster design methodology that co-optimizes parallelization strategies and resource provisioning for efficient distributed deep learning training Developed novel hardware-software co-design techniques to mitigate capacity aborts in commercial hardware transactional memory (HTM) systems achieving an 8.7× improvement over baseline HTM performance.

Hewlett Packard Labs, Milpitas, CA, Research Associate Intern Summer 2024 – Present Project: Improving memory efficiency and scalability of AI/ML Training leveraging CXL memory Supervisors: Dr. Puneet Sharma & Dr. Lianjie Cao

Advisor: Prof. Lawrence T. Clark

TEACHING EXPERIENCE

High Performance Computer Architecture (CS 6290 / ECE 6100), Georgia Tech,	Summer 2023
High Performance Computer Architecture (CS 6290 / ECE 6100), Georgia Tech,	Fall 2021
VLSI and Advanced Digital Deisgn (ECE 3150), Georgia Tech	Spring 2020
Digital System Design (ECE 2020), Georgia Tech	Fall 2019

PUBLICATIONS

- 1. Harvesting idle I/O resources for boosting memory bandwidth [under peer review]
 D. K. Kadiyala, and A. Daglis
- 2. Enabling Flexible and Composable AI Systems via Memory Disaggregation [under peer review]
 D. K. Kadiyala, L. Cao, P. Sharma, S. Sury, and A. Daglis
- 3. COMET: A Comprehensive Cluster Design Methodology for Distributed Deep Learning Training

D. K. Kadiyala, S. Rashidi, T. Heo, A. R. Bambhaniya, T. Krishna, and A. Daglis preprint arXiv, 2022

4. Safety Hints for HTM Capacity Abort Mitigation

A. Jain*, <u>D. K. Kadiyala</u>*, and A. Daglis

High-Performance Computer Architecture (HPCA), 2023. Acceptance rate: 25.0%

* Equal Contribution

- 5. Exploring Memory Expansion Designs for Training Mixture-of-Experts Models
 - T. Heo, S. Rashidi, C. Man, <u>D. K. Kadiyala</u>, W. Won, S. Srinivasan, M. Elavazhagan, M. Kumar, A. Daglis, and T. Krishna

Workshop on Hot Topics in System Infrastructure, (HotInfra), June 2023

- 6. Physically Unclonable Functions Using Foundry SRAM Cells
 - L. T. Clark, S. B. Medapuram, D. K. Kadiyala, and J. Brunhaver *IEEE Transactions on Circuits and Systems I (TCAS)*, 2019. Acceptance rate: 30.0%
- 7. SRAM Circuits for True Random Number Generation Using Intrinsic Bit Instability

L. T. Clark, S. B. Medapuram, and D. K. Kadiyala

IEEE Transactions on Very Large Scale Integration Systems, (TVLSI), 2018. Acceptance rate: 37.3%

TECHNICAL SKILLS

Programming Languages : C, C++, CUDA, Perl, Python, System Verilog, Bash Scripting Performance modeling : ZSim, ASTRA-Sim, DRAMSim, gem5, SESC, Garnet2.0

SPOKEN LANGUAGES

Telugu (native), Hindi (fluent), and English (fluent)