ISTeC
The Information Science and Technology Center

COLORADO STATE UNIVERSITY

Distinguished Lectures

Spring 2024

Dr. P. Sadayappan

Professor
Kahlert School of Computing
University of Utah

Advancing Machine Learning in the End-of-Moore Era: Challenges and Opportunities

Monday, February 19, 2024
Reception with Refreshments: 10:30 a.m.
Lecture: 11:00 - 12:00 a.m.
LSC University Ballroom

Compiler Optimization of Tensor Computations

Tuesday, February 20, 2024
Lecture: 9:30-11:00 a.m.
Lory Student Center 322

Sponsored by
Colorado State University's Information Science and Technology Center (ISTeC)

In conjunction with the Department of Computer Science and Department of Electrical and Computer Engineering Seminar Series

Abstracts

Advancing Machine Learning in the End-of-Moore Era: Challenges and Opportunities

Although artificial neural networks were invented over fifty years ago, it was not until computers were sufficiently powerful that the deep learning revolution could get started earlier in this century. Thanks to steady increase in computational power over the last 15 years, increasingly complex and powerful deep learning models have had a transformative impact on virtually all aspects of society. However, it is getting more and more difficult to maintain the dramatic rates of increase in compute power made possible by smaller transistor sizes with successive generations of VLSI technology, because we are approaching transistor feature sizes close to atomic limits. Therefore other avenues will become more critical in enabling more powerful machine learning models in the future. In this talk, we will discuss approaches to make more effective use of bounded hardware resources, including exploitation of sparsity, algorithm-architecture co-design, and the use of machine learning to improve the performance of machine learning applications.

Compiler Optimization of Tensor Computations

Production compilers like clang and gcc are extremely effective in generating very compact machine code from high-level C/C++ programs, i.e., they are very effective in minimizing the number of executed instructions. However, the dominant cost (both in terms of energy and execution time) on all computer systems today is not that of execution of the needed arithmetic operations but of the movement of data, between processors of a parallel system and through the memory hierarchy at each processor. Despite significant research advances in compiler optimization for affine computations, such as the powerful polyhedral model for dependence analysis and loop transformation, it remains extremely challenging for any compiler today to generate optimized code (for either multicore CPUs or GPUs) that achieves performance comparable to manually architected vendor libraries or autotuning optimizers like TVM. In this talk, we will discuss challenges and opportunities for compiler optimization of tensor computations, including design-space exploration, performance modeling, and algorithm-architecture co-design.

Speaker Biography

Dr. Sadayappan is a Professor in the Kahlert School of Computing at the University of Utah, with a joint appointment at Pacific Northwest National Laboratory. His primary research interests center around compiler optimization for high-performance computing, with an emphasis on performance optimization of matrix/tensor computations arising in applications from scientific computing and machine learning. He is the recipient of an ACM SIGPLAN Most Influential PLDI Paper award. Sadayappan is an IEEE Fellow.

To arrange a meeting with the speaker,
please contact Prof. Sanjay Rajopadhye <Sanjay.Rajopadhye@colostate.edu>.

ISTeC (Information Science and Technology Center) is a university-wide organization for promoting, facilitating, and enhancing CSU's research, education, and outreach activities pertaining to the design and innovative application of computer, communication, and information systems. For more information please see ISTeC.ColoState.edu.