COMPUTATIONAL RESEARCH in BOSTON and BEYOND SEMINAR

Mixed-precision algorithms: from ML to HPC

PIOTR LUSZCZEK

University of Tennessee, Knoxville

ABSTRACT:

IEEE 754 standard for floating-point precisions has served the scientific computing community for decades. The advent of deep learning marked the race to the bottom for the forever smaller bit representations that could still be successfully used in training of neural networks of increasing size and complexity. Yet shorter bit representations are applied for inference tasks that, unlike training, come at the speed of internet search queries. This first expanded and subsequently fractured the landscape of floating-point representations, reinvigorated the hardware startup scene, and applied pressure on the community to develop new mixed precision algorithms and their analysis methods. In this talk, I will focus on my contributions to the algorithmic side of this burgeoning new field. I will cover a range of numerical methods, their convergence properties, and performance considerations when applied to HPC benchmarking.

FRIDAY, JUNE 2, 2023
12:00 Noon – 1:00 PM

https://math.mit.edu/sites/crib/

https://mit.zoom.us/j/96155042770

Massachusetts Institute of Technology