Seminar

A CASE STUDY OF 64-BIT FLOATING POINT EMULATION IN QUANTUM ESPRESSO

The evolution of computer architecture, driven by AI applications, has significantly enhanced the reduced- and mixed-precision computing capabilities of GPUs, offering various low-precision representations such as FP16, BF16, and FP8. There is growing interest in utilizing these capabilities through mixed-precision algorithms and emulation techniques to boost scientific computing performance without compromising accuracy.

This presentation explores the potential of leveraging lower precision Tensor Cores for scientific applications by emulating FP64 and FP32 matrix multiplications without sacrificing accuracy, thereby achieving improved performance per watt.

We will discuss a novel library developed at NVIDIA, which transparently accelerates DGEMM and ZGEMM calls using INT8 Tensor Cores. Results from several use cases in Quantum Espresso will be presented, highlighting the effectiveness of this approach.

3 April 2025
11:00 CEST

MHPC Lecture hall,
Main Building
SISSA, ICTP campus
via Beirut (TS)

Speaker

Massimiliano Fatica

Director, NVIDIA

MAX DRIVING THE EXASCALE TRANSITION

QUANTUMESPRESSO

ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

SISSA

QUANTUM ESPRESSO FOUNDATION

Consiglio Nazionale delle Ricerche

IOM ISTITUTO OFFICINA DEI MATERIALI

Master in High Performance Computing