Contribution ID: 16

Type: not specified

## **Emulating Matrix Multiplication Using Mixed-Precision Computation**

Wednesday 2 July 2025 16:00 (30 minutes)

This talk introduces a method for emulating matrix multiplication through mixed-precision computation. As exemplified by the Matrix Engine on GPUs, low-precision arithmetic can be performed significantly faster than conventional FP32 or FP64 operations. We present Ozaki Scheme I and II, which leverage low-precision arithmetic to achieve accuracy comparable to standard FP64, and discuss their numerical performance.

Presenter: OZAKI, Katsuhisa