Contribution ID: 163 Type: not specified

Lightning Talk: Accelerating machine learning workloads using new GCC built ins

Wednesday 26 August 2020 08:00 (10 minutes)

Basic Linear Algebra Subprograms (BLAS) are used everywhere in machine learning and deep learning applications today. OpenBLAS is an optimized BLAS open source library used widely in AI workloads that implement algebraic operations for specific processor types.

This talk covers recent optimization in the OpenBLAS library for the POWER10 processor. As part of this optimization, assembly code for matrix multiplication

kernels in OpenBLAS is converted to C code using new compiler builtins. A sample optimization for matrix multiplication for POWER hardware in OpenBLAS will be used to explain how builtins are used and show the impact of application performance.

I agree to abide by the anti-harassment policy

I agree

Primary author: S, Rajalakshmi

Presenter: S, Rajalakshmi

Session Classification: GNU Tools Track

Track Classification: GNU Tools Track