

Mixed-precision Computing: High Accuracy with Low Precision

Erin Claire Carson

Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic
CARSON@KARLIN.MFF.CUNI.CZ

Abstract

Mixed-precision algorithms have launched an era in which efficiency and accuracy are no longer mutually exclusive. Rather than rely entirely on high-precision formats like double (64-bit) precision, mixed-precision algorithms apply lower precisions such as single (32-bit) or half (16-bit) precision whenever possible, reserving higher precision only for critical steps. Doing so can drastically reduce memory requirements, improve performance, and lessen energy consumption on modern computer hardware without sacrificing accuracy or stability. In this talk, we discuss the challenges of using low/mixed precision, and present five cases, common in scientific applications, where using mixed precision makes sense.