Is Mixed Precision Computing really the Top Priority?

Hartwig Anzt*[1]

[1]Technical University of Munich – Germany

Abstract

The hardware architectures are turning into low precision matrix multiplication units with some auxiliary functionality like scalar operations. High precision computations are increasingly realized via emulation and the compute power grows much faster than the memory bandwidth, the latency is barely reduced from one hardware generation to the next. For the scientific computing community, all these trends are challenging the progression in advancing science. But what are the most promising strategies? In this talk, we discuss the challenges and strategies for overcoming these.

*Speaker