

Peer Review

Review of: "Evaluation of Bfloat16, Posit, and Takum Arithmetics in Sparse Linear Solvers"

Yohann Uguen¹

1. Intel, Santa Clara, United States

This work presents an evaluation of the precision of multiple data formats applied to different (LU, QR, MPIR, GMRES). The presented results are clear and well presented.

What is omitted from this work (and generally by related work on posits) is that the hardware cost for implementing operators for such formats is far greater than for uniform precision formats (such as IEEE floats and BFloat16). Therefore, making a comparison between posit8 and float8 has little value given that one can implement the dynamic range and extra precision of posit8s using a larger float, and still be cheaper and faster [Evaluating the hardware cost of the posit number system]. Given this property of tapered-precision numbers, data type widths should not be the criteria for grouping formats together (e.g., a posit16 adder costs as much as a float32 one). Ignoring the hardware cost and latency, one could compare takums to mpfr and find that mpfr is much more precise, yet it won't be implemented on hardware because it would be too costly and too slow.

Declarations

Potential competing interests: No potential competing interests to declare.