No-Cost and Low-Cost Methods of Reducing Floating Point Error in Sums

Vanessa Job

LA-UR-20-25608

Because floating point addition on finite precision machines is not associative, mathematically equivalent floating-point summations can yield different computational results. Depending on the ordering and grouping chosen, rounding errors propagated across timesteps can be substantial, leading to significant inaccuracy in final results.

We examine sums in two codes, an adaptive mesh refinement hydrocode and a chemical reaction network. We reduce error by generating proper ordering and grouping of the sums and verify on typical simulation runs. Our techniques show accuracy comparable to Kahan sums without extra overhead.

We present heuristics that could improve accuracy for many codes. With minimal effort, researchers can apply these techniques and see improvement in accuracy with little or no overhead and minimal disruption to the code base.