

Matrix Computations & Scientific Computing Seminar

Organizer: James Demmel & Ming Gu

Wednesday, 11:00AM–12:00Noon, 380 Soda

Sept. 11 **William Kahan**, UC Berkeley

Floating-Point Tricks to Solve Boundary-Value Problems Faster

Some old tricks are resuscitated to accelerate the numerical solution of certain discretized boundary-value problems. Without the tricks, half the digits carried by the arithmetic can be lost to roundoff when the discretization's grid-gaps get very small. The tricks can procure adequate accuracy from arithmetic with "float" variables 4-bytes wide instead of "double" variables 8-bytes wide that move slower through the computer's memory system and pipelines. Tricks are tricky for programs written in MATLAB 7+, JAVA, FORTRAN and post-1985 ANSI C. For the original Kernighan-Ritchie C of the late 1970s, and for a few implementations of C99 that fully support IEEE Standard 754 for Binary Floating-Point, the tricks are easy or unnecessary. Examples show how well the tricks work.

For details, see www.eecs.berkeley.edu/~wkahan/Math128/FloTrik.pdf