Proposed Guidelines for Reporting Performance

March 4, 1993

These guidelines are not presented as mandatory, inflexible requirements. Clearly in a fast-moving field such as ours, it would be unwise to do so. However, if a paper or presentation does present results that significantly deviate from these guidelines, we feel the author has an obligation to justify these deviations.

1. If results are presented for a well-known benchmark, comparative figures should be truly comparable, and the rules for the particular benchmark should be followed. For example, Linpack 1000 results should not be compared against Linpack 100 results.

2. Only actual performance rates should be presented, not projections or extrapolations. For example, performance rates should not be extrapolated to a full system from a scaled-down system. Comparing extrapolated figures with actual performance figures, such as by including both in the same table, is particularly inappropriate.

3. Comparative performance figures should be based on comparable levels of tuning. For example, results tuned with months of effort on one system should not be directly compared with results on a well-known vector system where no attempt has been made to even fully vectorize the code.

4. Direct comparisons of run times are preferred to comparisons of megaflops rates or the like. Whenever possible, timings should be true elapsed time-of-day measurements (this might not be possible in some "production" environments).

5. Megaflops figures should not be presented for any comparative purpose unless they are computed from consistent flop counts, preferably flop counts based on efficient implementations of the best practical serial algorithms. One intent here is to discourage the usage of numerically inefficient algorithms, which may exhibit artificially high performance rates on a particular parallel system.

6. If speedup figures are presented, the single processor rate should be based on a reasonably well tuned program without multiprocessing constructs. If the problem size is scaled up with the number of processors, then the results should be clearly cited as "scaled speedup" figures, and details should be given explaining how the problem was scaled up in size.

7. Any ancillary information that would significantly affect the interpretation of the performance results should be fully disclosed. For example, if the results are for 32-bit rather than for 64-bit data, or if assembly-level coding was employed, or if only one processor of a conventional system is being used for comparison, these facts should be clearly stated.

8. Due to the natural prominence of abstracts, figures and tables, special care should be taken to insure that these items are not misleading, even if presented alone. For example, if significant performance claims are made in the abstract of the paper, any important qualifying information should also be included in the abstract.

9. Whenever possible, the following should be included in the text of the paper: the hardware, software and system environment; the language, algorithms, the datatypes and programming techniques employed; the nature and extent of tuning performed; and the basis for timings, flop counts and speedup figures. The goal here is to enable other scientists to accurately reproduce the performance results presented in the paper.