Foreword to a Resolution on Computer Arithmetic

The following IMACS-GAMM Resolution on Computer Arithmetic is based on unpleasant experiences made with existing vector processors. Most vector processors provide so-called elementary compound operations. These are heavily pipelined and greatly accelerate the computation. Usually, they are automatically inserted into a user's algorithm by the vectorizing compiler. If these operations are not carefully implemented, the user loses all control of the computation. The IMACS-GAMM Resolution is intended to influence and put pressure on manufacturers to implement these operations with extreme care.

The "Resolution" was initiated by GAMM, and was unanimously approved by the GAMM Vorstandsrat (Steering Committee) at the GAMM meeting in 1987.

Its proponents felt that more weight would be given to it if it were also supported and approved by IMACS. It was thus submitted to **the IMACS Board** of Directors who also approved it unanimously in 1988.

The text of this resolution which is thus an official IMACS-GAMM document follows:

IMACS - GAMM Resolution on Computer Arithmetic

The elementary floating-point operations +, -, *, / in electronic computers are currently required to be of highest machine accuracy: For any choice of operands, the computed result must coincide with the rounded exact result of the Operation, rounded according to the rounding mode in use (if no overflow occurs). For reference, sec the IEEE Arithmetic Standards 754 (binary floating-point arithmetic) and 854 (general floating-point arithmetic).

In recent years there has been a significant shift of numerical computation from general-purpose Computers towards vector and parallel Computers - so-called supercomputers. Along with the 4 elementary operations +, -, *, /, these Computers usually offer compound operations as additional elementary operations. This leads to an increase in computing speed. Some of these elementary compound operations are:

- multiply and add: a * b + c
- multiply and subtract: a * b - c
- accumulate: computes the sum of the components of a vector
- multiply and accumulate: computes the inner (or scalar) product of two vectors and others.

IMACS and GAMM require that all elementary compound operations be implemented by the manufacturer in such a way that guaranteed bounds are delivered for the deviation of the floating-point result from the exact result. It is desirable and usually achievable that for all possible data the computed result of such a compound floating-point operation agrees with the result that would be obtained if the exact result were computed and then rounded by the rounding in use (if no overflow occurs). In this case no explicit error bounds need be delivered. The user should not be obliged to perform an error analysis every time an elementary compound Operation, predefined by the manufacturer, is employed.

All elementary compound operations should also be provided with directed roundings, a feature needed both for fast computation of reliable and narrow bounds in numerical algorithms and for verification of the correctness of computed results. It must be ensured that the final floating-point result can differ from the exact result only in the direction defined by the rounding in use. This is already required of the elementary floating-point operations by the arithmetic standards mentioned above.

IMACS = International Association for Mathematics and Computers in Simulation
GAMM = Gesellschaft für Angewandte Mathematik und Mechanik