



BIRKHÄUSER



# HANDBOOK OF FLOATING-POINT ARITHMETIC



JEAN-MICHEL MULLER

NICOLAS BRISEBARRE

FLORENT DE DINECHIN

CLAUDE-PIERRE JEANNEROD

VINCENT LEFÈVRE

GUILLAUME MELQUIOND

NATHALIE REVOL

DAMIEN STEHLÉ

SERGE TORRES





# HANDBOOK OF FLOATING-POINT ARITHMETIC

Floating-point arithmetic is by far the most widely used way of implementing real-number arithmetic on modern computers. Although the basic principles of floating-point arithmetic can be explained in a short amount of time, making such an arithmetic reliable and portable, yet fast, is a very difficult task. From the 1960s to the early 1980s, many different arithmetics were developed, but their implementation varied widely from one machine to another, making it difficult for nonexperts to design, learn, and use the required algorithms. As a result, floating-point arithmetic is far from being exploited to its full potential.

This handbook aims to provide a complete overview of modern floating-point arithmetic, including a detailed treatment of the newly revised (IEEE 754-2008) standard for floating-point arithmetic. Presented throughout are algorithms for implementing floating-point arithmetic as well as algorithms that use floating-point arithmetic. So that the techniques presented can directly be put into practice in actual coding or design, they are illustrated, whenever possible, by a corresponding program.

## KEY TOPICS AND FEATURES INCLUDE:

- Presentation of the history and basic concepts of floating-point arithmetic and various aspects of the past and current standards
- Development of smart and nontrivial algorithms, and algorithmic possibilities induced by the availability of a fused multiply-add (fma) instruction, e.g., correctly rounded software division and square roots
- Implementation of floating-point arithmetic, either in software—on an integer processor—or hardware, and a discussion of issues related to compilers and languages
- Coverage of several recent advances related to elementary functions: correct rounding of these functions and computation of very accurate approximations under constraints
- Extensions of floating-point arithmetic such as certification, verification, and big precision

*Handbook of Floating-Point Arithmetic* is designed for programmers of numerical applications, compiler designers, programmers of floating-point algorithms, designers of arithmetic operators, and more generally, students and researchers in numerical analysis who wish to better understand a tool used in their daily work and research.

**Birkhäuser**  
www.birkhäuser.com

