

Recognized as an American National Standard (ANSI)

IEEE Std 754-1985

An American National Standard

IEEE Standard for Binary Floating-Point Arithmetic

Sponsor
**Standards Committee
of the
IEEE Computer Society**

Approved March 21, 1985
Reaffirmed December 6, 1990

IEEE Standards Board

Approved July 26, 1985
Reaffirmed May 21, 1991

American National Standards Institute

© Copyright 1985 by

The Institute of Electrical and Electronics Engineers, Inc

345 East 47th Street, New York, NY 10017, USA

Second printing 5 May 2006. To obtain errata information,
please go to <http://standards.ieee.org/reading/ieee/updates/errata/index.html>.

No part of this publication may be reproduced in any form, in an electronic retrieval system or otherwise, without the prior written permission of the publisher.

IEEE Standards documents are developed within the Technical Committees of the IEEE Societies and the Standards Coordinating Committees of the IEEE Standards Board. Members of the committees serve voluntarily and without compensation. They are not necessarily members of the Institute. The standards developed within IEEE represent a consensus of the broad expertise on the subject within the Institute as well as those activities outside of IEEE which have expressed an interest in participating in the development of the standard.

Use of an IEEE Standard is wholly voluntary. The existence of an IEEE Standard does not imply that there are no other ways to produce, test, measure, purchase, market, or provide other goods and services related to the scope of the IEEE Standard. Furthermore, the viewpoint expressed at the time a standard is approved and issued is subject to change brought about through developments in the state of the art and comments received from users of the standard. Every IEEE Standard is subjected to review at least once every five years for revision or reaffirmation. When a document is more than five years old, and has not been reaffirmed, it is reasonable to conclude that its contents, although still of some value, do not wholly reflect the present state of the art. Users are cautioned to check to determine that they have the latest edition of any IEEE Standard.

Comments for revision of IEEE Standards are welcome from any interested party, regardless of membership affiliation with IEEE. Suggestions for changes in documents should be in the form of a proposed change of text, together with appropriate supporting comments.

Interpretations: Occasionally questions may arise regarding the meaning of portions of standards as they relate to specific applications. When the need for interpretations is brought to the attention of IEEE, the Institute will initiate action to prepare appropriate responses. Since IEEE Standards represent a consensus of all concerned interests, it is important to ensure that any interpretation has also received the concurrence of a balance of interests. For this reason IEEE and the members of its technical committees are not able to provide an instant response to interpretation requests except in those cases where the matter has previously received formal consideration.

Comments on standards and requests for interpretations should be addressed to:

Secretary, IEEE Standards Board
345 East 47th Street
New York, NY 10017
USA

Foreword

(This Foreword is not a part of ANSI/IEEE Std 754-1985, IEEE Standard for Binary Floating-Point Arithmetic.)

This standard is a product of the Floating-Point Working Group of the Microprocessor Standards Subcommittee of the Standards Committee of the IEEE Computer Society. This work was sponsored by the Technical Committee on Microprocessors and Minicomputers. Draft 8.0 of this standard was published to solicit public comments.¹ Implementation techniques can be found in *An Implementation Guide to a Proposed Standard for Floating-Point Arithmetic* by Jerome T. Coonen,² which was based on a still earlier draft of the proposal.

This standard defines a family of commercially feasible ways for new systems to perform binary floating-point arithmetic. The issues of retrofitting were not considered. Among the desiderata that guided the formulation of this standard were

- 1) Facilitate movement of existing programs from diverse computers to those that adhere to this standard.
- 2) Enhance the capabilities and safety available to programmers who, though not expert in numerical methods, may well be attempting to produce numerically sophisticated programs. However, we recognize that utility and safety are sometimes antagonists.
- 3) Encourage experts to develop and distribute robust and efficient numerical programs that are portable, by way of minor editing and recompilation, onto any computer that conforms to this standard and possesses adequate capacity. When restricted to a declared subset of the standard, these programs should produce identical results on all conforming systems.
- 4) Provide direct support for
 - a) Execution-time diagnosis of anomalies
 - b) Smoother handling of exceptions
 - c) Interval arithmetic at a reasonable cost
- 5) Provide for development of
 - a) Standard elementary functions such as exp and cos
 - b) Very high precision (multiword) arithmetic
 - c) Coupling of numerical and symbolic algebraic computation
- 6) Enable rather than preclude further refinements and extensions.

¹Computer Magazine vol 14, no 3, March 1981.

²Computer Magazine vol 13, no 1, January 1980.

Members of the Floating-Point Working Group of the Microprocessor Standards Subcommittee and those who participated by correspondence were as follows:

David Stevenson, *Chair*

Andrew Allison
William Ames
Mike Arya
Janis Baron
Steve Baumel
Dileep Bhandarkar
Joel Boney
E.H. Bristol
Werner Buchholz
Jim Bunch
Ed Burdick
Gary R. Burke
Paul Clemente
W.J. Cody
Jerome T. Coonen
Jim Crapuchettes
Itzhak Davidesko
Wayne Davison
R.H. Delp
James Demmel
Donn Denman
Alvin Despain
Augustin A. Dubrulle
Tom Eggers
Philip J. Faillace
Richard Fateman
David Feign
Don Feinberg
Smart Feldman
Eugene Fisher

Paul F. Flanagan
Gordon Force
Lloyd Fosdick
Robert Fraley
Howard Fullmer
Daniel D. Gajski
David M. Gay
C.W. Gear
Martin Graham
David Gustavson
Guy K. Haas
Kenton Hanson
Chuck Hastings
David Hough
John Edward Howe
Thomas E. Hull
Suren Irukulla
Richard E. James III
Paul S. Jensen
W. Kahan
Howard Kaikow
Richard Karpinski
Virginia Klema
Les Kohn
Dan Kuyper
M. Dundee Maples
Roy Martin
William H. McAllister
Colin McMaster
Dean Miller
Webb Miller

John C. Nash
Dan O'Dowd
Cash Olsen
A. Padegs
John F. Palmer
Beresford Parlett
Dave Patterson
Mary H. Payne
Tom Pittman
Lew Randall
Robert Reid
Christian Reinsch
Frederic N. Ris
Stan Schmidt
Van Shahan
Robert L. Smith
Roger Stafford
G.W. Stewart
Robert Stewart
Harold S. Stone
W.D. Strecker
Robert Swarz
George Taylor
James W. Thomas
Dar-Sun Tsien
Greg Walker
John Steven Walther
Shlomo Waser
P.C. Waterman
Charles White

When the IEEE Standards Board approved this standard on March 21, 1985, it had the following membership:

John E. May, *Chair*

John P. Riganati, *Vice Chair*

Sava I. Sherr, *Secretary*

James H. Beall
Fletcher J. Buckley
Rene Castenschiold
Edward Chelotti
Edward J. Cohen
Paul G. Cummings
Donald C. Fleckenstein

Jay Forster
Daniel L. Goldberg
Kenneth D. Hendrix
Irvin N. Howell, Jr
Jack Kinn
Joseph L. Koepfinger*
Irving Kolodny
R.F. Lawrence

Lawrence V. McCall
Donald T. Michael*
Frank L. Rose
Clifford O. Swanson
J. Richard Weger
W.B. Wilkens
Charles J. Wylie

*Member emeritus

Acknowledgements

Appreciation is expressed to the following companies and organizations for contributing the time, talent, and resources of their employees to make possible the development of this standard:

Altos Computer Systems
Analytical Mechanics
Apple Computer
Argonne National Laboratory
Beckman Instruments
Bell AT&T Laboratories
Computervision
Data General Corp
Digital Equipment Corp
ELXSI
ESL Inc
Fairchild Microprocessor
Four Phase Systems
The Foxboro Company
Hewlett Packard Co
IBM Corp
Intel Corp
Itty Bitty Computers
Kylex
Lawrence Livermore Laboratory
Leibniz-Rech/Bay. Akad. Wiss.
Lockheed Research Laboratory
M & E Associates
Massachusetts Institute of Technology
Menlo Computer Associates
MMI
Motorola, Inc
National Semiconductor, Corp
Oregon Software
Parasitic Engineering
Prime Computer
Signetics Corp
Sperry Univac
Stanford Linear Accelerator Center
Stewart Research Ent.
The Systems Group
Tandem
University of Arizona
University of California
University of Colorado
University of Illinois
University of Maryland
University of Massachusetts
University of Ottawa
University of Toronto
Volition Systems
Zilog

CLAUSE	PAGE
1. Scope	1
1.1 Implementation Objectives	1
1.2 Inclusions	1
1.3 Exclusions	1
2. Definitions.....	2
3. Formats.....	3
3.1 Sets of Values.....	3
3.2 Basic Formats.....	4
3.3 Extended Formats	5
3.4 Combinations of Formats.....	5
4. Rounding	5
4.1 Round to Nearest.....	5
4.2 Directed Roundings.....	5
4.3 Rounding Precision	6
5. Operations	6
5.1 Arithmetic	6
5.2 Square Root.....	6
5.3 Floating-Point Format Conversions	6
5.4 Conversion Between Floating-Point and Integer Formats	7
5.5 Round Floating-Point Number to Integer Value	7
5.6 Binary <----> Decimal Conversion.....	7
5.7 Comparison	8
6. Infinity, NaNs, and Signed Zero	9
6.1 Infinity Arithmetic	9
6.2 Operations with NaNs	10
6.3 The Sign Bit	10
7. Exceptions	10
7.1 Invalid Operation	11
7.2 Division by Zero	11
7.3 Overflow	11
7.4 Underflow	11
7.5 Inexact.....	12
8. Traps.....	12
8.1 Trap Handler	12
8.2 Precedence	13
Annex A (informative) Recommended Functions and Predicates	14