

ISO/IEC/IEEE 60559:2011-06 (E)

Information technology - Microprocessor Systems - Floating-Point arithmetic

Contents		Page
1.	Overview	1
1.1	Scope	1
1.2	Purpose	1
1.3	Inclusions	1
1.4	Exclusions	2
1.5	Programming environment considerations	2
1.6	Word usage	2
2.	Definitions, abbreviations, and acronyms	3
2.1	Definitions	3
2.2	Abbreviations and acronyms	5
3.	Floating-point formats	6
3.1	Overview	6
3.2	Specification levels	7
3.3	Sets of floating-point data	7
3.4	Binary interchange format encodings	9
3.5	Decimal interchange format encodings	10
3.6	Interchange format parameters	13
3.7	Extended and extendable precisions	14
4.	Attributes and rounding	15
4.1	Attribute specification	15
4.2	Dynamic modes for attributes	15
4.3	Rounding-direction attributes	16
5.	Operations	17
5.1	Overview	17
5.2	Decimal exponent calculation	18
5.3	Homogeneous general-computational operations	19
5.4	formatOf general-computational operations	21
5.5	Quiet-computational operations	23
5.6	Signaling-computational operations	24
5.7	Non-computational operations	24
5.8	Details of conversions from floating-point to integer formats	26
5.9	Details of operations to round a floating-point datum to integral value	27
5.10	Details of totalOrder predicate	28
5.11	Details of comparison predicates	29
5.12	Details of conversion between floating-point data and external character sequences	30
6.	Infinity, NaNs, and sign bit	34
6.1	Infinity arithmetic	34
6.2	Operations with NaNs	34
6.3	The sign bit	35
7.	Default exception handling	36
7.1	Overview: exceptions and flags	36
7.2	Invalid operation	37
7.3	Division by zero	37
7.4	Overflow	37
7.5	Underflow	38
7.6	Inexact	38

8.	Alternate exception handling attributes	39
8.1	Overview	39
8.2	Resuming alternate exception handling attributes	39
8.3	Immediate and delayed alternate exception handling attributes	40
9.	Recommended operations	41
9.1	Conforming language- and implementation-defined functions	41
9.2	Recommended correctly rounded functions	42
9.3	Operations on dynamic modes for attributes	46
9.4	Reduction operations	46
10.	Expression evaluation	48
10.1	Expression evaluation rules	48
10.2	Assignments, parameters, and function values	48
10.3	preferredWidth attributes for expression evaluation	49
10.4	Literal meaning and value-changing optimizations	50
11.	Reproducible floating-point results	51
	Annex A (informative) Bibliography	53
	Annex B (informative) Program debugging support	55
	Index of operations	57