Brakefield Research = *Optimal Architecture* www.brakefieldresearch.com
1214 Vista Del Rio, San Antonio, Texas 78216

Various Floating-Point Formats
James Brakefield ©2001[1,2]

FP(radix, # bits, exponent, mantissa)
FSM: fraction sign-magnitude, ISM: integer sign-magnitude, US: unsigned
FTC: fraction 2's complement, FPOC: floating-point one's complement

Radix 16 Mantissa:

IBM 360 FP(16, 32, US(2, 7)-64, FSM(16, 25))
m_s e e e e e e e.m m m m m m m m m m m m m m m m m m m m m m m m
Data General FP(16, 32, US(2, 7)-64, FSM(16, 25))
m_s e e e e e e e.m m m m m m m m m m m m m m m m m m m m m m m m
SEL FP(16, 32, US(2, 7)-64, FSM(16, 25))
m_s e e e e e e e.m m m m m m m m m m m m m m m m m m m m m m m m
Xerox Sigma 5 FP(16, 32, US(2, 7)-64, FSM(16, 25))
m_s e e e e e e e.m m m m m m m m m m m m m m m m m m m m m m m m

Radix 8 Mantissa:

Bendix G20 FP(8, 29, ISM(2, 7), ISM(8, 22))
* * $s/_d$ m_s e_s e e e e e e m m m m m m m m m m m m m m m m m m m m m.
Burroughs 5500 FP(8, 47, ISM(2, 7), ISM(8, 40))
* m_s e_s e e e e e e m m m m m m m . . . . . m m m m.

Radix 10 Mantissa:

IBM 1620 FP(10, 5n+10, ISM(10, 10, FSM(10, 5n))
.M M M . . . M M E E
Burroughs 2500 FP(10, 4n+16, ISM(10, 12), ISM(10, 4n+4))
E_s E E M_s M M . . . M M.

Radix 2 Mantissa:

16-bit Word Size:

Pixar FP(2, 16, US(2,5)-15, FSM(2, 11))
m_s e e e e e 1.m m m m m m m m m m

32-bit Word Size:

PDP 11 FP(2, 32, US(2, 8)-128, FSM(2, 24)+1)
m_s e e e e e e e e1.m m m m m m m m m m m m m m m m m m m m m m m
HP 3000 FP(2, 32, US(2, 9)-256, FSM(2, 23)+1)
m_s e e e e e e e e e1.m m m m m m m m m m m m m m m m m m m m m m m
IEEE-754 FP(2, 32, US(2, 8)-127, FSM(2, 24))
m_s e e e e e e e e1.m m m m m m m m m m m m m m m m m m m m m m m

[1] Derived from an earlier list: Brakefield, J.C.; Quin, M.J. 1977. Variable length data formats. Data Management Symposium; Huntsville, AL; Oct 1977 Proceedins p. 243-253.
[2] **2010**: http://en.wikipedia.org/wiki/IEEE_754-2008 has latest version of IEEE standard

36-bit Word Size:

PDP 10 FP(2, 36, US(2, 8)-128, FUS(2, 27))
m_s e e e e e e e e.m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m
GE 635 & Honeywell 6000 FP(2, 36, ITC(2, 8), 2*FTC(2, 28))
e_s e e e e e e e m_s m.m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m
Univac 1100's FPOC(2, 36, US(2, 8)-128, US(2, 27))
m_s e e e e e e e e.m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m
IBM 7090 FP(2, 36, US(2, 8)-128, FSM(2, 28))
m_s e e e e e e e e.m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m

48-bit Word Size:

Burroughs 8500 FP(2, 48, ISM(2, 12), ISM(2, 36))
m_s e_s e e e e e e e e e e e e m m m m m m m . . . . . . m m m m.
CDC 1604 FP(2, 48, US(2, 11)-1024, FUS(2, 36))
m_s e e e e e e e e e e e e.m m m m m m m m . . . . . m m m m
Harris Datacraft FP(2, 47, ITC(2, 8), FTC(2, 39)) (two words of 24-bits each)
m_s.m m m . . . m m m, 0 m m m m m m m m e_s e e e e e e e
Philco 213 FP(2, 48, ITC(2, 12), FTC(2, 36))
m_s.m m m . . . . . m m m m e_s e e e e e e e e e e e e

60-bit Word Size:

CDC 6600 FPOC(2, 60, US(2, 11)-1024, US(2, 48))
m_s e e e e e e e e e e e e.m m m m m m m m . . . . . m m m m

64-bit Word Size:

CRAY 1 FP(2, 64, US(2, 15)-16384, FSM(2, 49))
m_s e e e e e e e e e e e e e e e e e e.m m m m m m m m . . . . . m m m m
CDC STAR FP(2, 64, ITC(2, 16), ITC(2, 48))
e_s e e e e e e e e e e e e e e e e e e m_s m m m m m m m . . . . . m m m m.
IBM 7030 FP(2, 60, ISM(2, 11), FSM(2, 49))
* e e e e e e e e e e e_s.m m m m . . . . . m m m m_s * * *
ILLIAC IV FP(2, 64, US(2, 14)-8192, ISM(2, 50))
m_s e e e e e e e e e e e e e e e e e e m m m m . . . . . m m m m m.
IEEE-754 FP(2, 64, US(2, 11)-1023, FSM(2, 53))
m_s e e e e e e e e e e e 1.m m m m m m m . . . . . m m m m m

Miscellaneous:

Interleaved Exponent & Mantissa (proposed[3]):

Variable Length FP(2, 4n, ISM(2, n), FSM(2, 3n)+1.)
m m m e m m m e . . . . . m m m e m m m m_s e_s

Decimal Exponent

Microsoft C# decimal FP(2, 128, ITC(2, 5), US(2, 96))
(1-2*s) * m * 10^e, where e between 0 and -28

[3] Brakefield, J.C. 1972. An Optimal Floating Point Format. ACM SIGARCH 1:4 pg 16-17, Oct. 1972

Semi-Logarithmic

FP(2, 32, ITC(2, 19)-128, US(2, 12)) (nominal)

m_s e e e e e e e e **.** e e e e e e e e e e e e e m m m m m m m m m m m

(e has a fraction component, mantissa has hidden leading 1.00000000000)

Konrad Zuse: http://irb.cs.tu-berlin.de/~zuse/Konrad_Zuse/en/

Z3 FP(2, 21, ITC(2, 7)-64, FSM(2, 15))

m_s e e e e e e e1.m m m m m m m m m m m m m m m