SANE: Accuracy and Performance Information

Revised: 6/24/90
Security: Everyone

SANE: Accuracy and Performance Information

==========

This article last reviewed: 6 March 1990

TOPIC ----------

I am interested in the accuracy of SANE when working with numbers that take SANE to its maximum accuracy potential. Also, what is the performance difference when using the most accurate SANE number format compared to using less accurate formats?

DISCUSSION ----------

Though different in detail, the MC68881 and SANE packages are both implementations of the IEEE Standard 754. Following are the range and precision of the data types supported by SANE:

	Single	Double	Comp	Extended
Size (bytes:bits)	4:32	8:64	8:64	10:80
Range of binary exponents				
Minimum	-126	-1022	-	-16383
Maximum	127	1023	-	16383
Significand precision				
Bits	24	53	63	64
Decimal digits	7-8	15-16	18-19	19-20
Decimal range approximate				
Maximum positive	3.4E+38	1.7E+308	9.2E18	1.1E+4932
Minimum positive norm	1.2E-38	2.3E+308		1.7E-4932
Minimum positive denorm	1.5E-45	5.0E-324		1.9E-4951
Maximum negative denorm	-1.5E-45	-5.0E-324		-1.9E-4951
Maximum negative norm	-1.2E-38	-2.3E-308		-1.7E-4932
Minimum negative	-3.4E+38	-1.7E+308	-9.2E18	-1.1E+4932

If the compiler is making direct calls to the MC68881, accuracy may suffer in some situations. For elemental functions, both SANE and the MC68881 have errors in the least-significant bits of the fraction part of extended-format results, but SANE's errors rarely exceed the last bit, whereas the MC68881 errors may extend to as many as the last five bits. The MC68881 is much more likely to return an error in a double-precision result than SANE is.

Due to the architecture of the Macintosh, and how SANE is implemented on the Macintosh, the extended-precision number format not only provides the most accuracy, but also provides better performance than the computational or double-number formats. The Macintosh converts double and computational numbers into extended number before using them, and thus takes more time than just using the extended-number format.

Single precision provides the most speed, but is hardly ever used because most high-level language compilers today generate code using the extended number format. You can use the single number format, but the code probably would have to be written in assembly language.

For more complete and detailed information, we suggest the "Apple Numerics Manual, Second Edition" from Addison-Wesley.

Tech Info Library Article Number:5286