

EX.0.3.6, Sauer

Do the following sums by hand in IEEE double precision computer arithmetic, using the Rounding to Nearest Rule.

a. $(1 + (2^{-51} + 2^{-52} + 2^{-54})) - 1$

b. $(1 + (2^{-51} + 2^{-52} + 2^{-60})) - 1$

$$a. (1 + (2^{-51} + 2^{-52} + 2^{-54})) = +1. \boxed{00 \dots 011} 01 \times 2^0$$

using the Rounding to Nearest Rule,

$$fl(1 + (2^{-51} + 2^{-52} + 2^{-54})) = +1. \boxed{00 \dots 011} \times 2^0$$

$$\begin{aligned} \text{Therefore, } fl((1 + (2^{-51} + 2^{-52} + 2^{-54})) - 1) &= . \boxed{00 \dots 011} \times 2^0 \\ &= 2^{-51} + 2^{-52} \end{aligned}$$

$$b. (1 + (2^{-51} + 2^{-52} + 2^{-60})) = +1. \boxed{00 \dots 011} 00000001 \times 2^0$$

using the Rounding to Nearest Rule,

$$fl(1 + (2^{-51} + 2^{-52} + 2^{-60})) = +1. \boxed{00 \dots 011} \times 2^0$$

$$\begin{aligned} \text{Therefore, } fl((1 + (2^{-51} + 2^{-52} + 2^{-60})) - 1) &= . \boxed{00 \dots 011} \times 2^0 \\ &= 2^{-51} + 2^{-52} \end{aligned}$$