Fatih Yildiz

Q 2.4) We use the formula $X = 0{,}m \cdot 3^{e}$ to calculate the numbers we represent. In order to be able to represent $-\infty$, $+\infty$, NaN, we have to reserve some values. I decided to reserve the exponent configuration "$\underbrace{+ + + + +}_{5 \text{ bits}}$" for this purpose. (so we can use the "- - - - -" configuration for values very close to 0)

| exponent | mantissa | |
|---|---|---|
| + + + + + | + + + + + + + + + + + + + + | ⇒ $+\infty$ |
| + + + + + | - - - - - - - - - - - - - - | ⇒ $-\infty$ |
| + + + + + | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | ⇒ NaN |
| 0 0 0 0 0 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | ⇒ 0 ⇒ This design avoids having signed zeros. |

So we have a unique representation for zero. Also, eliminates the need for subnormal numbers to represent zero.

Then,

maximum positive number

exp: $\underbrace{+ + + + 0}_{+120}$, m: $\underbrace{+ + + + + + + + + + + + + +}_{\approx +0{,}5}$ ⇒ $X = 0{,}5 \cdot 3^{120}$ (much bigger than $2^{128}$ in binary f.p. repr)

minimum positive number

exp: $\underbrace{- - - - -}_{-121}$, m: $\underbrace{+ - - - - - - - - - - - - -}_{\approx +0{,}166}$ ⇒ $X = 0{,}166 \cdot 3^{-121}$ (closer to zero than $2^{-126}$ in binary f.p. repr)

minimum negative number

exp: $\underbrace{+ + + + 0}_{+120}$ m: $\underbrace{- - - - - - - - - - - - - -}_{\approx -0{,}5}$ ⇒ $X = -0{,}5 \cdot 3^{120}$ (much smaller than $-(2^{128})$ in binary f.p. repr)

maximum negative number

exp: $\underbrace{- - - - -}_{-121}$ m: $\underbrace{- + + + + + + + + + + + + +}_{\approx -0{,}166}$ ⇒ $X = -0{,}166 \cdot 3^{-121}$ (closer to zero than $-(2^{-126})$ in binary f.p. repr)

⇒ So the absolute range for normalized numbers: $0{,}166 \cdot 3^{-121} < |X| < 0{,}5 \cdot 3^{120}$

As we can see, with this design, we can represent a wider range of values (also the ones closer to 0) than binary floating point format with high precision.