1.4 Write the number 45.0 in the following forms (in part (*c*), follow the IEEE-754 standard):
(*a*) Binary form. (*b*) Base 2 floating point representation. (*c*) 32-bit single-precision string.

Solution

(*a*) To convert 45 to binary form, first find the largest power of 2 that can be divided into 45. This is $2^5 = 32$. Note that $2^6 = 64$, which is larger than 45. Next, subtract $45-2^5 = 13$. Now, find the largest power of 2 that will divide into 13, without exceeding it. This would be $2^3 = 8$. Subtract $13-2^3 = 5$, and so on. Thus,

$$45 = 32+8+4+1 = 1\times2^5+1\times2^3+1\times2^2+1\times2^0$$
$$= 1\times2^5+0\times2^4+1\times2^3+1\times2^2+0\times2^1+1\times2^0 = 101101$$

Thus, the number 45 in binary form is 101101.

(*b*) From part (*a*), the largest power of two that divides into 45 is 2^5. The binary floating point representation is obtained as follows:

$$\frac{45}{2^5}\times2^5 = \frac{45}{32}\times2^5 = 1.40625\times2^5$$

(*c*) According to the IEEE-754 standard, 81 in single precision form is as follows:

- Since the number is positive, the first bit is 0.
- From part (*b*), the exponent is 5. Adding a bias of 127, the value of the exponent that must be stored is 5+127 = 132. The number 132 in binary form is $132 = 1\times2^7+1\times2^2 = 128+4$. Thus the number 132 in binary form is 10000110. In single precision, 8 bits can be used to store the exponent so that 132 is stored as 10000101 without the need for rounding or chopping.
- Next, the mantissa 0.40625 is converted to binary form: $1\times2^{-2}+1\times2^{-3}+1\times2^{-5}$ or 0.01101
- Since 23 bits are allocated for the mantissa, the binary number stored is 01101000000000000000000.

Thus, the number 45 in single precision is stored as: |0|10000110|01101000000000000000000|