**Exercises**

**Floating Point**

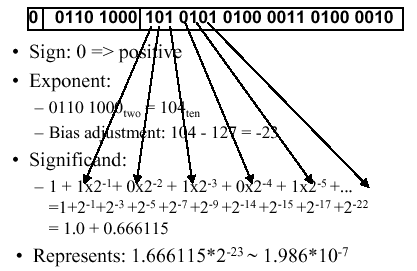
The IEEE 754 standard defines a binary representation for floating point values using three fields:

• The *sign* determines the sign of the number (0 for positive, 1 for negative)

• The *exponent* is in **biased notation** with a bias of 127

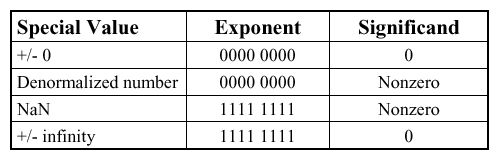
• The *significand* is akin to unsigned, but used to store a fraction instead of an integer.

**1.**The below numbers shows the bit breakdown for the single precision (32-bit) representation:



Convert the numbers from binary to decimal.

**2.** There is also a double precision encoding format that uses 64 bits. Convert the above numbers from a single precision binary to a double precision binary.

**3.As Table.** Special values in the IEEE 754 standard.  


1). What is the largest finite positive value that can be stored using a single precision float?

2). What is the smallest positive value that can be stored using a single precision float?

3). What is the smallest positive normalized value that can be stored using a single precision float?

4). Convert the following numbers from binary to decimal or from decimal to binary:

0x00000000 8.25 0x00000F00 39.5625 0xFF94BEEF -∞

**4.**Write the following numbers as powers of 2: 2 Ki, 256 Pi, 512 Ki, 64 Gi, 16 Mi, 128 Ei.

**5**.assume an 8 bit integer and answer each one for the case of a two's complement number, and unsigned number, indicating if it cannot be answered with a specific representation.

1). What is the largest integer? The largest integer + 1?（unsigned number, wo's complement number）

,

2). How do you represent the numbers 0, 1, and -1 with unsigned number and a two's complement number?

3). How do you represent 17, -17 with unsigned number and a two's complement number?