# The float data types

By: Yehia M. Abu Eita

# Outlines

- Introduction
- The float data type
- Floating numbers into memory
- The double data type
- The double data type into memory

# Introduction

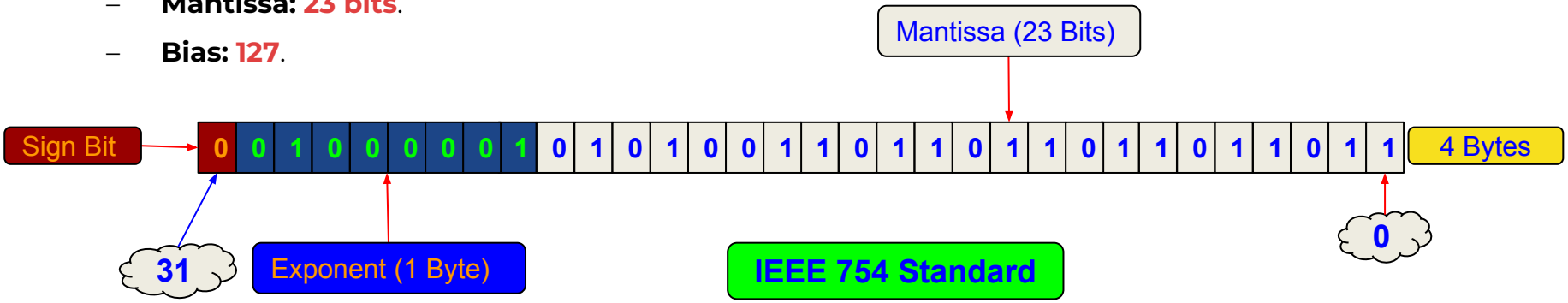
- The floating point data types are used to store numbers with **fraction** into the memory.
- **Two** data types are representing the floating point numbers into the memory.
- The **float** data type and the **double** data type.
- The difference between them is in their **sizes** and **precisions**.

# The float data type

- The **float** data type has the following properties:
  - **Size in memory: 4 bytes**.
  - **Precision: 6 decimal points**.
  - **Minimum positive value: +1.17549e-38**.
  - **Maximum positive value: +3.40282e+38**.
  - **Minimum negative value: -1.17549e-38**.
  - **Maximum negative value: -3.40282e+38**.

# Floating numbers into memory

- The **float** data type has the following bit groups into memory:
  - **Sign bit:** 1 bit.
  - **Exponent:** 8 bits.
  - **Mantissa:** 23 bits.
  - **Bias:** 127.



# Floating numbers into memory

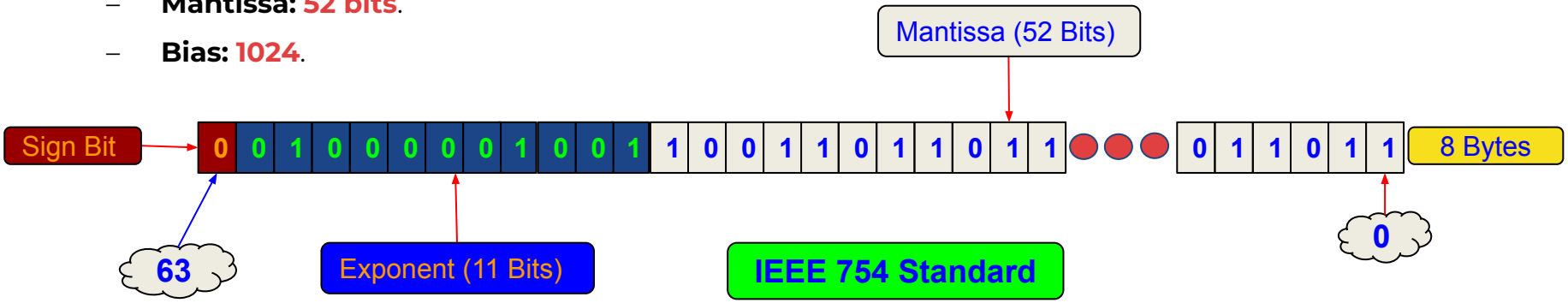
- Five steps to get the binary representation of a float number into memory:
  - Convert the **float value to binary**.
  - Convert the binary value to **scientific notation** form **1.xxxx * $2^e$**.
  - If the value is **negative put 1 in the sign bit otherwise put 0**.
  - The **exponent** bits are the **binary representation of the Bias + e**.
  - **The Mantissa is the xxxx binary value**.

# The double data type

- The **double** data type has the following properties:
  - **Size in memory:** **8 bytes**.
  - **Precision:** **15 decimal points**.
  - **Minimum positive value:** **+2.22507e-308**.
  - **Maximum positive value:** **+1.79769e+308**.
  - **Minimum negative value:** **-2.22507e-308**.
  - **Maximum negative value:** **-1.79769e+308**.

# The double data type into memory

- The **double** data type has the following bit groups into memory:
  - **Sign bit:** 1 bit.
  - **Exponent:** 11 bits.
  - **Mantissa:** 52 bits.
  - **Bias:** 1024.



# The double data type into memory

- Five steps to get the binary representation of a double number into memory:
  - Convert the **double value to binary**.
  - Convert the binary value to **scientific notation** form **1.xxxx * $2^e$**.
  - If the value is **negative put 1 in the sign bit otherwise put 0**.
  - The **exponent** bits are the **binary representation of the Bias + e**.
  - **The Mantissa is the xxxx binary value**.

# Summary

- You are able to differentiate between float and double data types
- You are able to represent any float number into the memory
- You have learned about the difference between storing floats and integers into the memory and bugs appeared when reading float numbers as integers