**Math 1310 – Technical Math for IT**

**ASSIGNMENT 3** Name:\_\_\_\_\_Markus Afonso\_\_\_\_\_

**Due:** October 10th at 11:59 PM (all three sets)

Online submission, ONE pdf file ID :\_\_\_\_ A01333486\_\_\_\_\_\_

**After completing all the questions on separate paper, place all answers on THIS sheet. Be sure to attach your work showing all intermediate steps in a clear and well organized fashion for full credit.**

1. [2]

Convert the following single precision IEEE floating point number to decimal.

|  |  |  |
| --- | --- | --- |
| 1 | 1011 1011 | 01011010000000000000000 |

A picture containing application

Description automatically generated

1. [3] The following sequence of 32 bits is stored in memory:

1010 1110 1111 1011 0100 0000 0000 0000

What is the decimal value of the number stored if the binary string given represents a number in: a) Unsigned binary form?



* 1. Twos-complement binary form?



* 1. IEEE-754 single precision floating point form?



Text

Description automatically generated

|  |  |
| --- | --- |
| **Mini-Standard Floating Point Rep** | **Decimal Number** |
| (a) 1 0000 00000 |  |
| (b) 1 1111 00000 |  |
| (c) 0 0101 01110 |  |
| (d) 0 0010 01110 |  |
| (e) 0 0000 00110 |  |
| (f) 1 1111 11101 |  |
| (g) 1 1110 01010 |  |

1. [7] Find the decimal number corresponding to each of the mini-standard floating point representations in the table to the right.

Be sure to check special cases!

# Math 1310 – Technical Math for IT

Graphical user interface, text, application

Description automatically generatedGraphical user interface, application

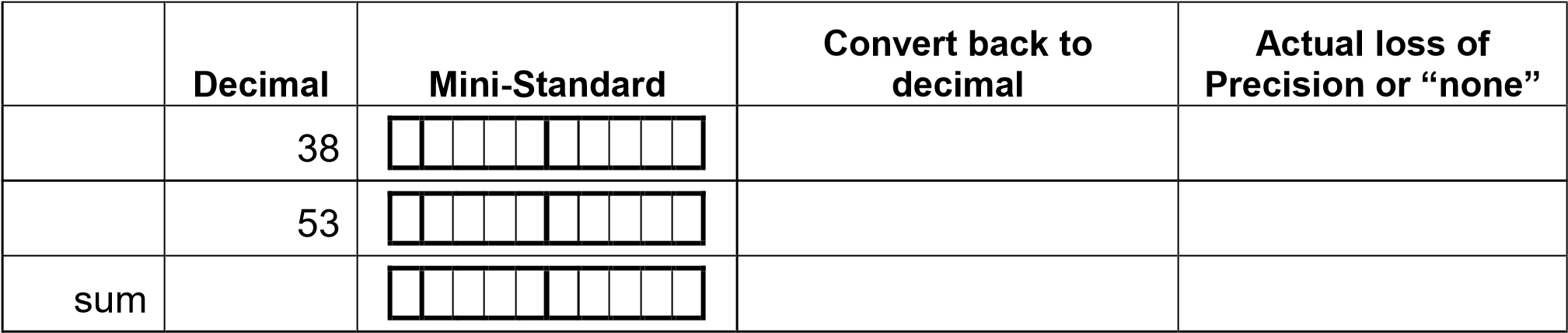
Description automatically generated

1. [12 marks] Use the **mini-standard** floating point representation (1 sign bit, 4 exponent bits, and 5 mantissa bits assuming the hidden bit, with the exponent recorded in bias 7) to perform the arithmetic operations below. Each of the following steps should be performed for each problem (use the template to record the results, but attach sheets with detailed work to support these results):

i. Both of the given numbers should first be coded in themini-standard.

ii. the numbers in the mini-standard should converted back to decimal form and the precise loss of precision recorded.

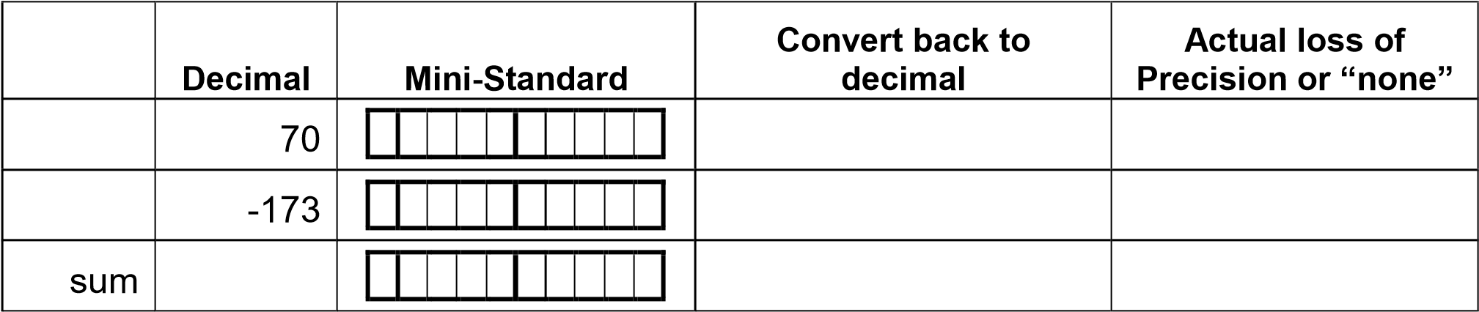
1. The addition or subtraction should be performed in standardized form with **only five bits to the right of the radix point**.
2. The result should then be recorded in the normalized mini form, **if possible**.
3. Finally, the result should be interpreted as a **decimal** floating point number**.**
4. Any **further** loss of precision (resulting from the standardization process or from renormalization) should be noted (yes/no).
5. 38 + 53

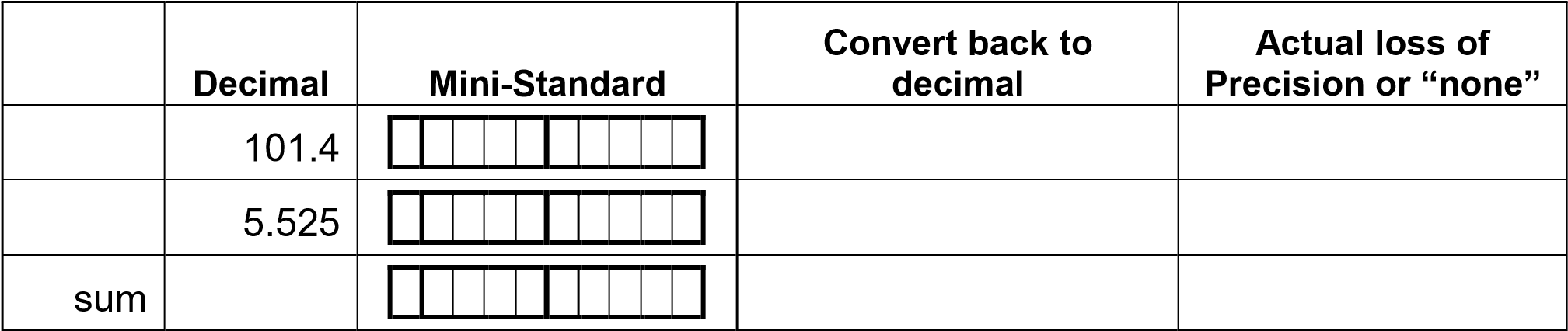




1. 70 - 173





1. 101.4 + 5.525



Graphical user interface, text

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generatedGraphical user interface, text, application

Description automatically generated

A picture containing text

Description automatically generated

Graphical user interface, text

Description automatically generated