CSE 306

Computer Architecture Sessional

Assignment on **Floating Point Adder**

In this assignment, you are required to design a floating point adder circuit which takes two floating points as inputs and provides their sum, another floating point as output. Each floating point will be 16 bits long with following representation:

Sign **1 bit**	*Exponent* **4 bit**	*Fraction* **11 bit**

You have to implement your design in any simulator software of your choice. Please note that, if your chosen simulator does not provide support for 16-bit ALU, you can construct one by cascading number of smaller ALUs. Moreover, since construction of ALU is not major focus of this assignment, you can take help from the Internet or other sources (or even use someone else's implementation) for the 16 bit ALU part only. The rest of the circuit design and implementation must be done by yourselves.

For this assignment, you will work in a group (same as the group for assignment on ALU). The report should contain introduction, problem specification, flowchart of the addition/subtraction algorithm, high-level block diagram of the architecture, detailed circuit diagram of the important blocks, ICs used with count as a chart, simulator used along with the version number, Discussion.

- Input is in IEEE 754 Standard format of binary representation of the floating point number (sign bit, exponent bit, fraction bits).
- The shifter circuit provided by the simulator software tool may be used. Alternatively, you may implement them by yourselves.
- Implementations of Overflow/Underflow flags are necessary. Rounding is also necessary. There is no need to involve clock pulses if you do one-step rounding. Alternatively, you may decide to do it in one clock pulse.