

# A Floating-Point Unit for Numerical Calculations

Jeff Walden  
6.111 Fall 2006

Final Project Presentation

---

## The Goal

---

- A partially IEEE 754-compatible FPU
- Addition/subtraction, multiplication, exceptions, and comparisons
- “Enough” functionality
- Hardware-based speedup

The goal of my project is to partially implement a floating-point unit, by which I mean a hardware device capable of performing a set of basic mathematical operations on numbers which can contain fractional parts. The intent is that the final product be usable within the context of some other project — a device which must manipulate floating-point numbers, such as a CPU or a GPU, could use the unit to handle many common floating-point calculations.

Time and complexity limitations restrict the functionality which will be implemented to addition, subtraction, multiplication, comparison values, and support for IEEE exceptions; floating-point arithmetic is a surprisingly complex topic. The goal is “enough” functionality to do most of what you might want to do with floating-point numbers in most situations — with the speedup afforded by implementation in hardware as a motivation.

---

## A Brief Diversion

---

- An IEEE 754 floating-point number (Wikipedia)



- IEEE 754
  - Floating-points,  $+/-0$ ,  $+/-\text{Infinity}$ , NaN
  - Exceptions: invalid operation, divide by zero, overflow/underflow, inexact

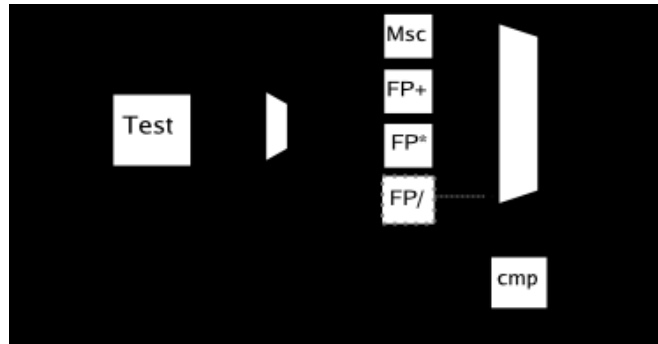
- Floating-point exactness and rounding

---

## FPU Overview

---

- Test, adder, multiplier, comparison modules



---

## User Interface

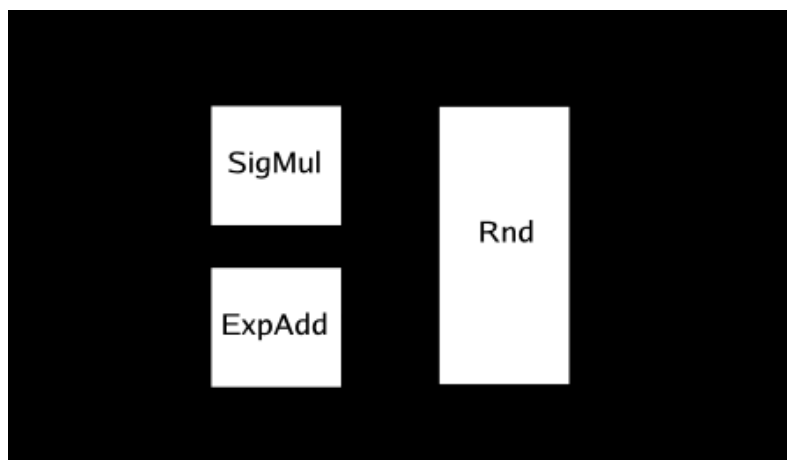
---

- Inputs: buttons and state machine used to modify input floating-points
  - evaluating other interfaces for use after initial mockup
- Outputs: sign bit, exponent, and significand on labkit LEDs
  - Time-permitting, hope to implement a VGA-based display

---

## The Multiplier

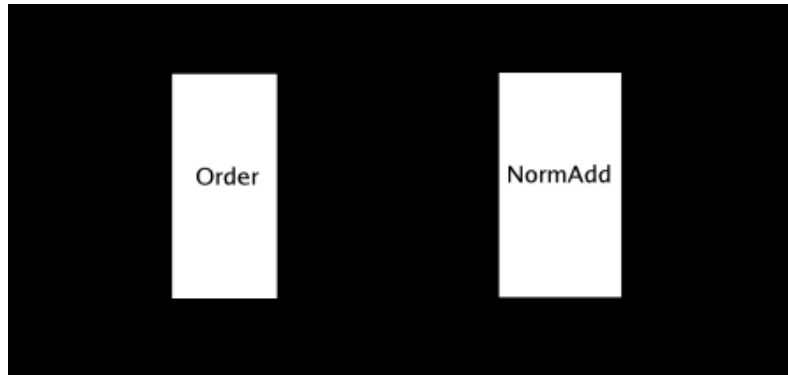
---



---

## The Adder

---



---

## Implementation Steps

---

1. Multiplier module
2. Concurrently:
  - FPU container module (partial functionality)
  - User interface, first mockup
3. Concurrently:
  - Improved user interface
  - Adder module
4. *Other FPU operations (division, etc.)*

---

## Questions?

---