題目: 32-Bit Fused Multiply-Add(FMA)

學號: 7110064481

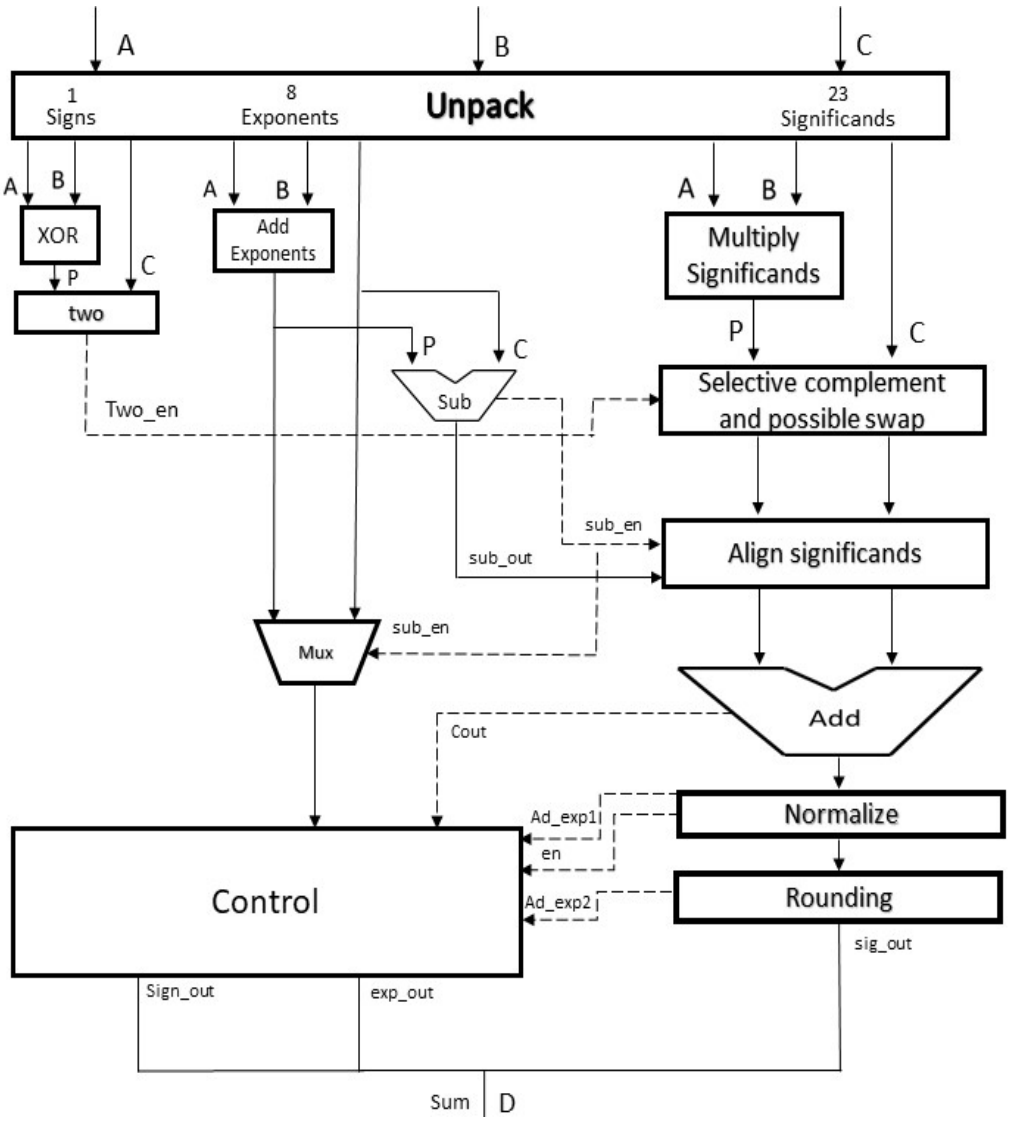
姓名: 紀政均

1. 簡介

設計一個浮點數乘加器，採用三個 32 位單精度浮點輸入，

執行 A*B+C 的運算

2. 系統方塊圖



3. 功能敘述

將 A、B、C 的值讀進來後經過 unpack 分成 sign、exp、sig(左側多補一個 0 改成有號數)，一開始先做乘法，將 A、B 的 sign 做 XOR，exp 相加(這裡相加後的結果沒有加 bias)，sig 相乘(50bits)，接著利用 sign 判斷是否為負數，將負數做二補數運算，之後進行 alignment，先將 P、C 的 exp 相減，得以比較大小，再將較小的數右移相減得到的結果，訊號也同時給到 mux 將較大的數的 exp 往後送，經過 alignment 之後的兩個 sig 進行相加，這裡取相加後最左邊的 bit 給 control，若是 0 即為正數，若是 1 即為負數並做二補數運算，因為此架構中相加後的結果並不會進位超過 50bits，故以此來判斷結果的正負，之後就進行 normalize(LOD)、rounding，將有移位的 bits 數給 control 來修正 exp，最後在 control 內計算出運算結果的 sign、exp、sig，pack 後即為輸出。

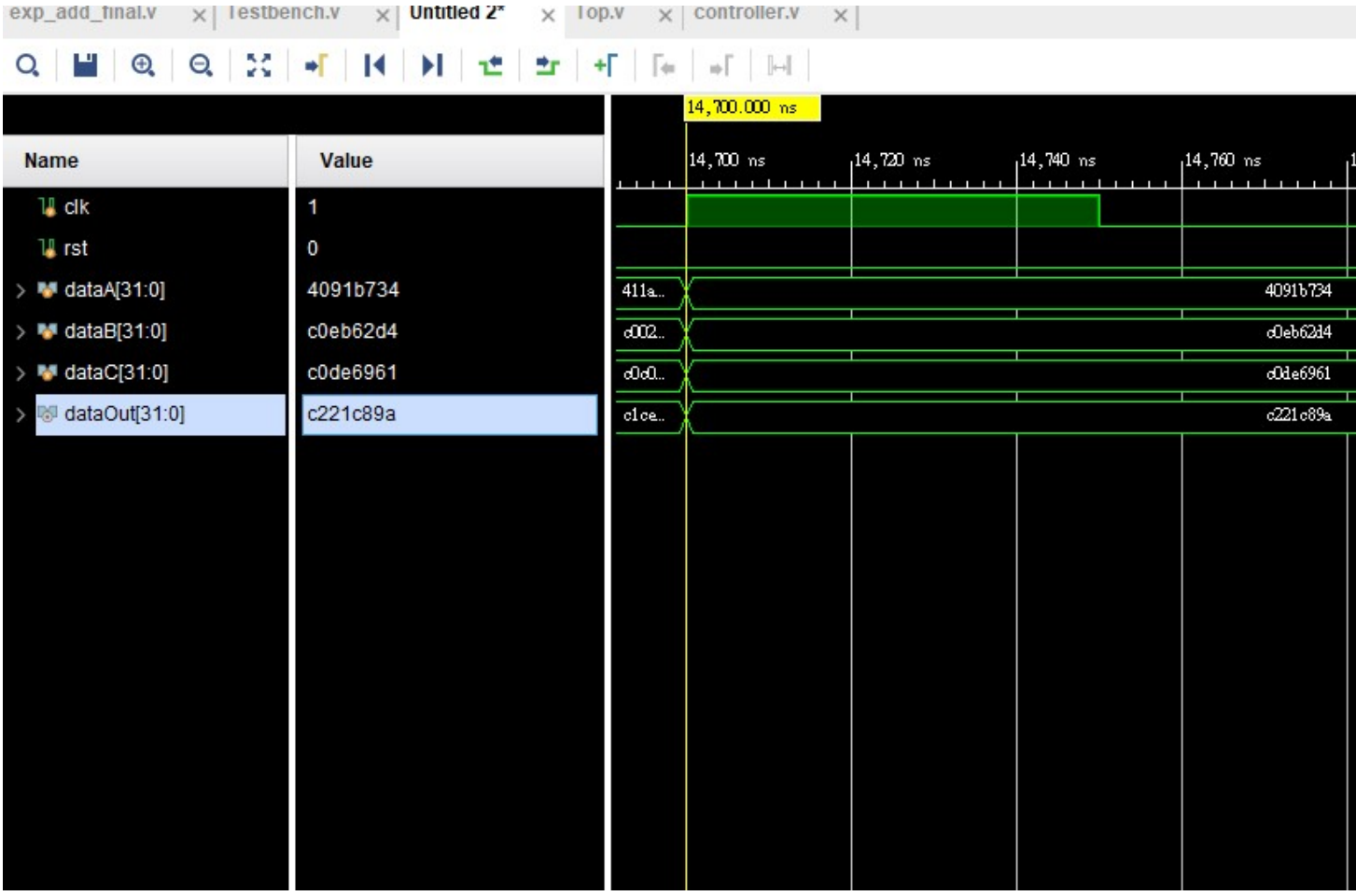
4. 輸入輸出介面

Signal	I/O	Width(bit)
dataA	Input	32
dataB	Input	32
dataC	Input	32
dataOut	Output	32

5. 驗證方式

將輸入轉為十進位做計算

以下圖為例



$$A = 4091b734_{16}$$
$$= 01000000100100011011011100110100_{IEE754}$$
$$= 4.55361366272_{10}$$
$$B = c0eb62d4_{16}$$
$$= 11000000111010110110001011010100_{IEE754}$$
$$= -7.3558139801_{10}$$

$$C = c0de6951_{16}$$
$$= 11000000110111100110100101010001_{IEE754}$$
$$= -6.95035600662_{10}$$

$$Output_{real}$$
$$= A * B + C = -40.44589105_{10}$$
$$= 11000010001000011100100010011000_{IEE754}$$
$$= c221c898_{16}$$
$$Output_{simulation}$$
$$= c221c89a_{16}$$
$$= 11000010001000011100100010011010_{IEE754}$$
$$= -40.4458999634_{10}$$

$$error = 2.20457 \times 10^{-5} \quad \%$$