CS5120 Homework Assignment 04

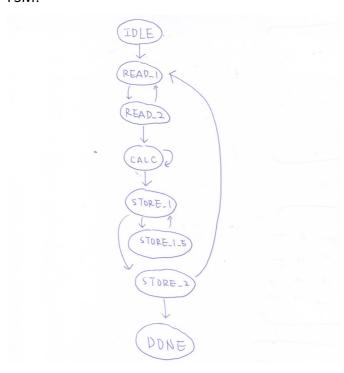
Student ID: 107065507 Name: 盧允凡

1. Design Concept

Goal: 設計一個以 systolic array 的方式完成 4x4 matrix multiplication,並且用 sram 存放。

 $C = A \times B$

FSM:



IDLE: 待 start signal 開始

READ_1: 從 sram 讀取 A 和 B 矩陣的值存下來

READ_2:確認A與B都存取完畢,進入下個state

CALC: 開始計算,以 systolic 方式,需要 8 cycles

STORE_1: 將 C 矩陣的值存入 sram

STORE_1_5: 緩衝用的 state

STORE_2: 確認是否完成,如完成,進入 DONE,尚未完成回到 READ_1 計算下一個測資

DONE: 完成

Processing Elements (PE):

PE 是一個用來計算的單元, 4x4 matrix 總共用了 16 個 PE

NTHU CS 1/4



其中

clk: clock signal

r: reset signal

out_a = in_a

out_b = in_b

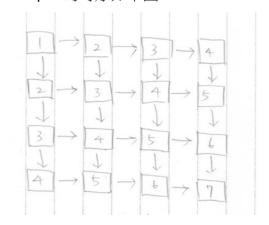
out_c = out_c + in_a*in_b

需要注意的是每個 PE 的 reset signal 不一,所以不能吃同一個 signal,需要再另外宣告 7 個 reset signal (下部分解釋)

Systolic array:

一矩陣乘法 $C = A \times B$

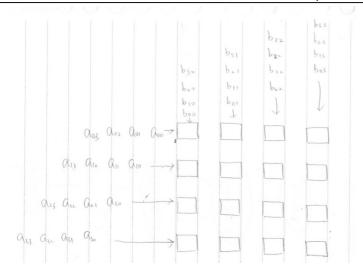
PE input 的時序如下圖



PE 編號是 row-major,所以 pe0 的 reset signal 為 r0,pe1 和 pe4 的 reset signal 為 r1,跟上圖的時間序一致,以此類推。

共需要 7 cycles 再加上一個 input 皆為 0 的 cycle, 共 8 個。 從外界 input 的 signal 有 a0, a1, a2, a3 以及 b0, b1, b2, b3。

NTHU CS 2/4



每個 cycle 的 input 皆不同。

第一個 cycle 為 a0=A[0][0], a1=0, a2=0, a3=0, b0=B[0][0], b1=0, b2=0, b3=0, b4=0 第二個 cycle 為 a0=A[0][1], a1=A[1][0], a2=0, a3=0, b0=B[1][0], b1=B[0][1], b2=0, b3=0 以此類推…

SRAM:

此次 lab 使用的 sram 為 16*128*3 =6144,每個 8-bit (1 byte)的記憶體。

且為 single port sram。

每一筆測資需花費 16*2*2 cycles 讀取。總共花費 128*16*2*2 cycles 讀取。

每一筆測資需花費 16*2 cycles 存入。這裡會需要 2 cycles 是因為我要把答案印出來寫在一個文字檔上,要讓 out_valid 這個 signal 有 rise and down 的行為,所以多花了一個 cycle,可能有更好的方式不用多花費。

SRAM area estimation:

According to the following information,

"The area of a single-port 8192x8 SRAM is 347,200 um²."

The area of a single-port 16*128*3 SRAM is 3703466 um².

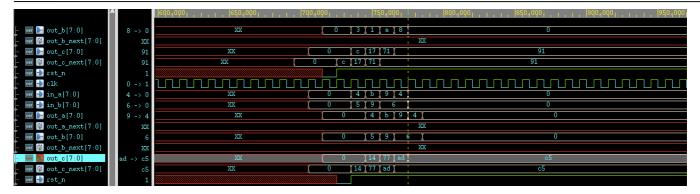
2. Simulation and Discussion

Latency:

128 筆測資總共花費 134,435,000 ns 的時間完成。

第一筆測資的完成間。

NTHU CS 3 / 4



把答案印出在一個.txt 檔,跟 golden output 檔案做比較。

答案皆正確。

3. Summary

上一份 lab 做過 convolution 的練習,所以這份作業寫起來較得心應手,也學習到了 systolic array 的設計技巧。

在存入答案到 sram 的過程用了 2 cycles,為的是要寫入 output file,可能有更好的寫法只用 1 cycle。

NTHU CS 4 / 4