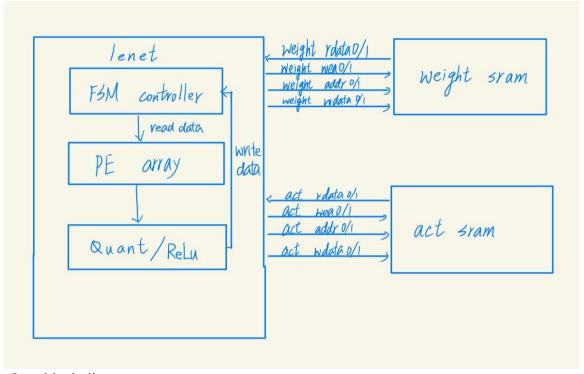
Homework 3 Report

Student ID: 112061533 Name: 潘金生

- 1. Design concept (You may write the report in Chinese):
 - Explanation of the overall hardware architecture and block diagram of each component.

本次架構的 block diagram 如下圖,SRAM_weight 是存放權重的資料,SRAM_activation 是存放 input activation 的和寫入 output activation 資料,而 Lenet 用來做 FC1 的運算,Lenet 大致上分成 4 個模塊,其中一個模塊是 FSM controller,用來控制何時要讀取 weight sram, act sram 資料和寫入 act sram 資料,並且把讀取的資料給 PE array 去做運算,另一個模塊是 PE array 負責來計算乘加,這裡採用 8 個乘加運算,因為一次可以讀取 8 筆 data,接著 quant 跟 Relu 用來處理算出來後的資料,進行量化和 relu function,在運算完後,會傳給 FSM controller,在由 controller 判斷何時需要寫入 act sram。



圖一 block diagram

• State diagram and its detailed description (if any).

首先我分成 4 個 State,分別為 IDLE, FC1_READ, FC1_WRITE, DONE 去實作,下面我會一一解釋這 4 種 state,分別在做什麼事。

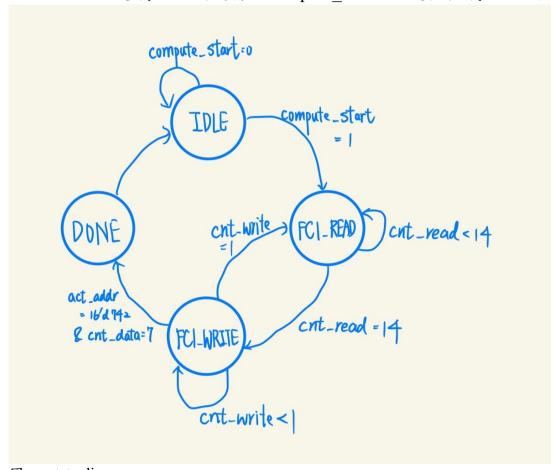
IDLE: 判斷 Compute_start 的到來,如果收到 Compute_start 訊號,則會開始讀取地址,因此會跳到 FC1 READ state,否則一直保持在 IDLE state。

FC1_READ: 這邊會開始給要讀取位置的地址, weight 的地址從 13020 開始, 15539 結束, act 的地址從 692 開始, 721 結束。我在這裡的做法是

先給定 15 次的地址,由於每個 clk 讀取資料總共可以讀 8 筆資料(dual port sram),因此在經過 15 個 clock 後,則會讀到 120 筆的資料,在 FC1 可以輸出一筆資料,因此,在經過 15 個 clock 後, state 會跳到 FC1 WRITE state。

FC1_WRITE: 這邊會開始寫資料,在給定地址後,由於記憶體讀取的關係,在讀取資料上,會慢 FC1_READ 1.5 個 cycle,因此,在進來 FC1_WRITE 後,會花 2 個 clock 時間,在兩個 clock 時間後,資料被正確運算完且寫入,在返回 FC1_READ state,直到寫到 act_addr = 742 和最後一筆後,最後才會到 DONE state。

DONE: 完成 FC1 的運算, compute finish 訊號被拉成 1, 宣告完成。



圖二 state diagram

• Explanation of the dataflow.

在 start 訊號進來後,會先透過 sram 讀取 data,由於 sram 是 dual port sram 因此在讀取資料時,可以一次讀取 8 筆資料,也就是說 1 個 clk 可以進行 8 筆資料運算,由於 FC1 的資料量為(1,120)*(120,84),所以我們必須讀取權重和 activation 總共 120 筆,所以算完一筆 data 需要花 15 個 clk 進行,除此之外,讀取資料會慢 1.5 個 clk,因此在寫入資料時,需要有 2 個 clk 的延遲,才會算出對的值,接著才開始寫入到 sram,直到 84 筆 data 寫完後,才完成本次的 Lenet。

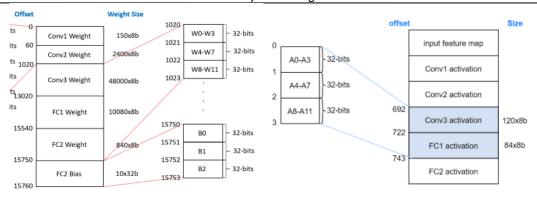


Figure 1: Layout of weight SRAM.

Figure 2: Layout of the activation SRAM.

圖三 SRAM 讀取資料

2. Result

Paste the screenshot of your RTL simulation result.

圖三 RTL simulation pass



圖四 waveform Write data

3. Others (optional)

謝謝老師用心的講解課程和助教熱心的回覆討論區。