實驗:簡單神經網路使用Vitis HLS合成，並放進IP驗證成功

本次實驗與上次不同之處在於本次傳輸介面使用Stream，並且不需要輸入weight、bias等等，只需輸入圖片資料，輸出向量就是判斷結果。

# Python軟體

與上次架構相同。

Dataset :MNIST

Accuracy:96.9

Neural network structure:



# Vitis HLS

本次使用stream傳輸介面，有兩種使用AXI4-Stream的方式

1. AXI4-Stream without side channel
2. AXI4-Stream with side channel

Axit4 stream傳輸介面被定義在ap\_axi\_sdata.h中，包含T,WUser,WId,WDest。

|  |  |
| --- | --- |
| T | Stream data type |
| WUser | Width of the TUSER signal |
| WId | Width of the TID signal |
| WDest | Width of the TDest signal |

Table axis interface

當沒有使用side channel時，ap\_axiu<32,0,0,0>資料不會包含WUser, WId, and WDest等等。

WUser,WId,WDest這些訊號是能被使用在RTL level中的stream傳輸訊號使用。

在本次使用時，我只需要注意T(stream data type) ,就能完成本次實驗，無須用到其他side channel訊號，所以本次實驗使用的是1.stream without side channel。

Stream data type若傳輸的是整數，則會有以下格式可以做選擇

1. Singned :hls::axis<ap\_int<WData>, WUser, WId, WDest>
2. Unsigned: hls::axis<ap\_uint<WData>, WUser, WId, WDest>

本次使用2. unsigned 的資料傳輸格式。

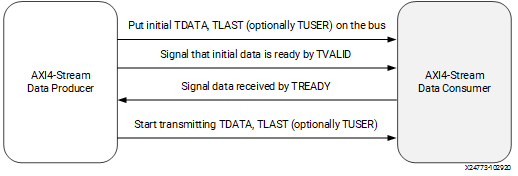


Figure Stream handshake protocol

Stream傳輸資料時，一定包含Data producer 傳輸的TVAID,TDATA,TLAST以及consumer回傳的TREADY。

在實作中，傳輸資料時TVALID,TDATA等資料會由host幫忙發送與傳遞，若使用without side channel時，則必須要自己加上TLAST在傳遞訊息中，否則硬體沒有接受到TLAST永遠不會完成資料傳輸。

# 硬體架構

由於float point尚未明白如何傳輸，本次實驗所有資料為32bit int。

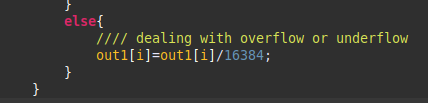
並且由於硬體計算時會遇到overflow問題，在做完第一層layer運算時，我將所有bit向右移14bit，除以16384。對準確度有一點影響。

Figure dealing with overflow

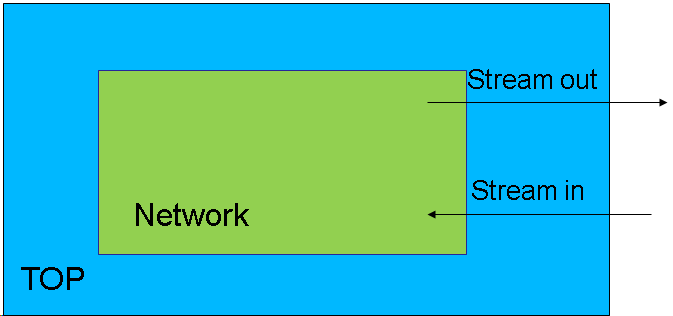


Figure Hardware Block Diagram

僅有Stream in, stream out 兩個port。

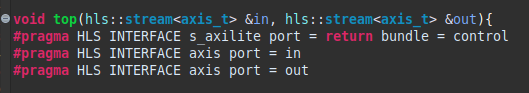


Figure stream, in/out port

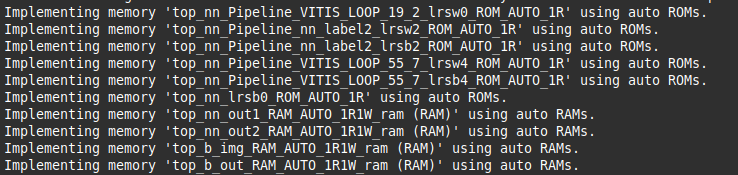
在硬體合成時，wieght,bias等等已知的資料被合成為ROM，而在硬體中的累加器被合成為RAM。

Figure Implementation ROM/RAM

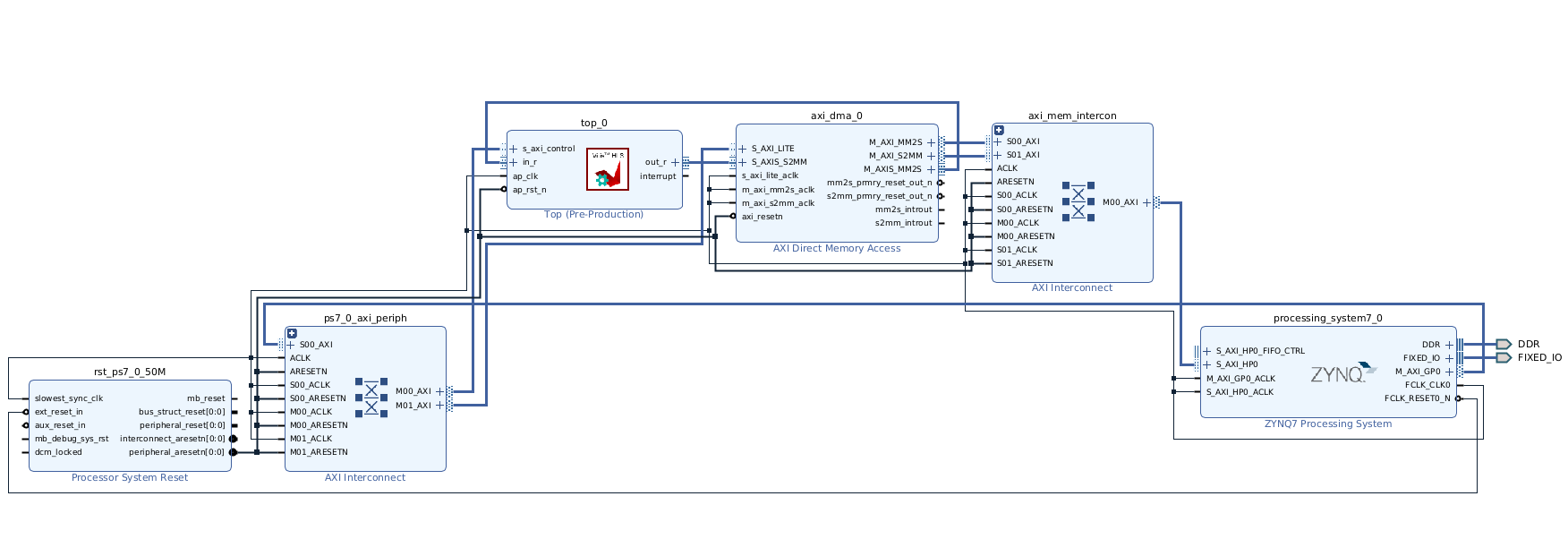
接著合成為IP。使用了一個DMA read/write處理資料in/out。

Figure Vivado Diagram

# Pynq驗證

拿1000張 test dataset中的圖片輸入觀察準確度96.6%

花費14.21秒。

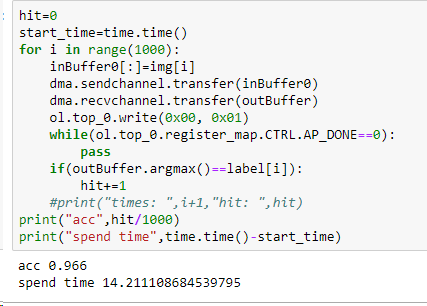


Figure Pynq outcome

# 問題與討論

1. float point data type

目前尚未了解float point如何再interface介面傳輸。

一開始做實驗時，我本來打算使用float point做為data type，但是合成為硬體後發現輸出資料是錯誤結果。

我後來將float point做成ROM如下圖。

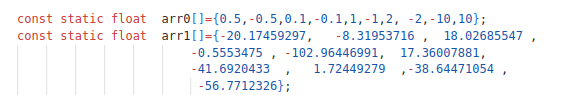


Figure C++ ROM

將arr0中10筆資料讀取，並加1得到1.e+00結果。小數點等資料都不見了。

將arr1中10筆資料單純讀取，僅得到nan。目前尚未明白float如何傳輸。

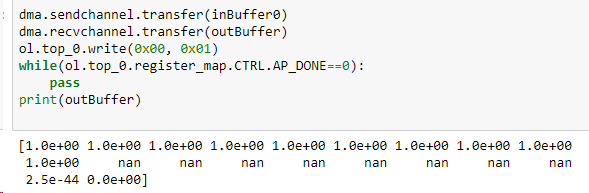


Figure Error outcome

# 結論

目前能夠使用stream 傳輸介面傳輸整數32bit資料，並實作簡單神經網路。

接下來:

* 1. 研究float point傳輸資料方式
  2. 嘗試AN code加入神經網路做訓練
  3. TCB未提代替乘法以C++方式硬體實現