

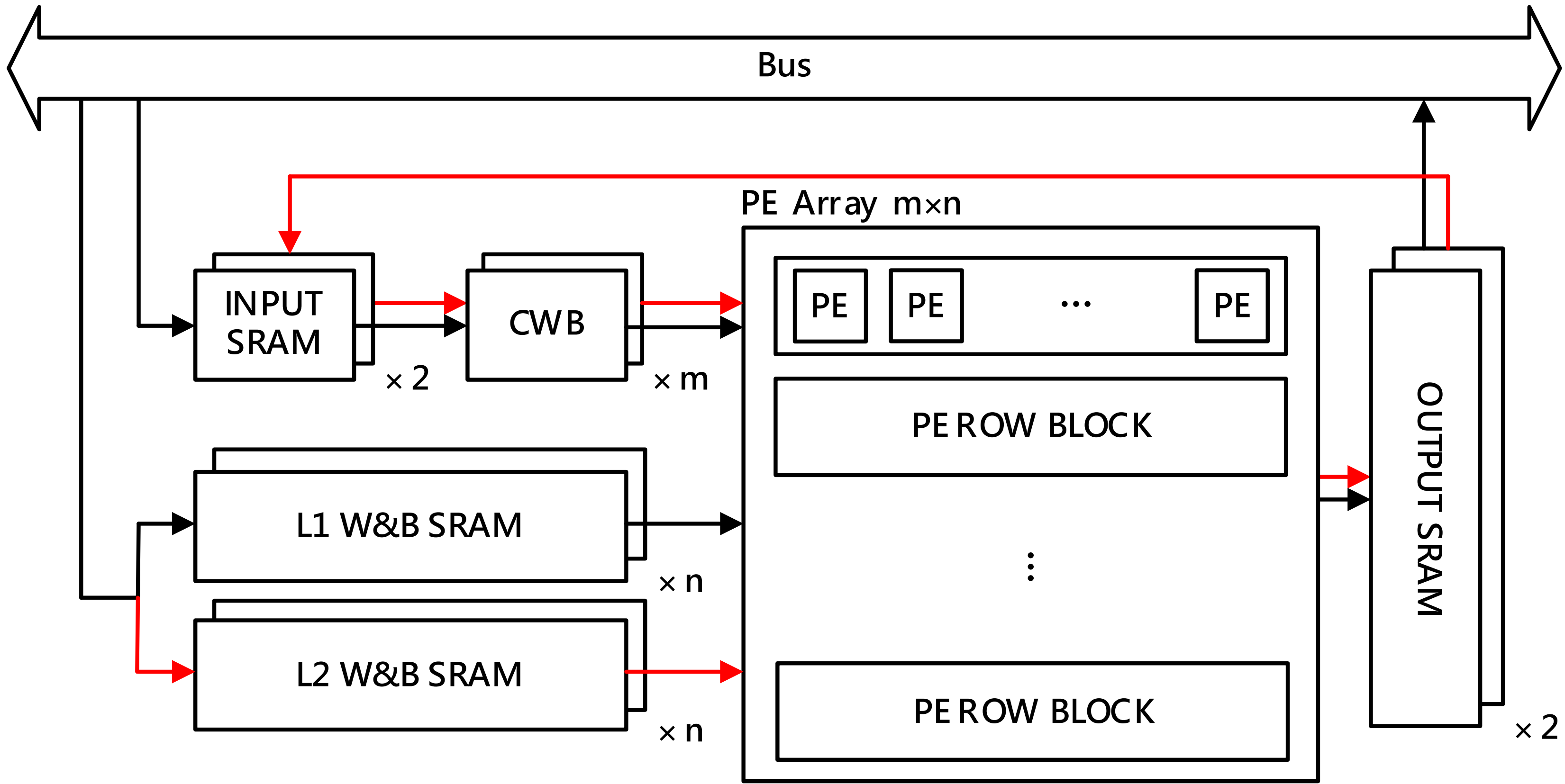
卷積神經網路加速器處理單元設計與實現

組員：林英傑、吳致廣、林穎杉、唐珣憶

摘要

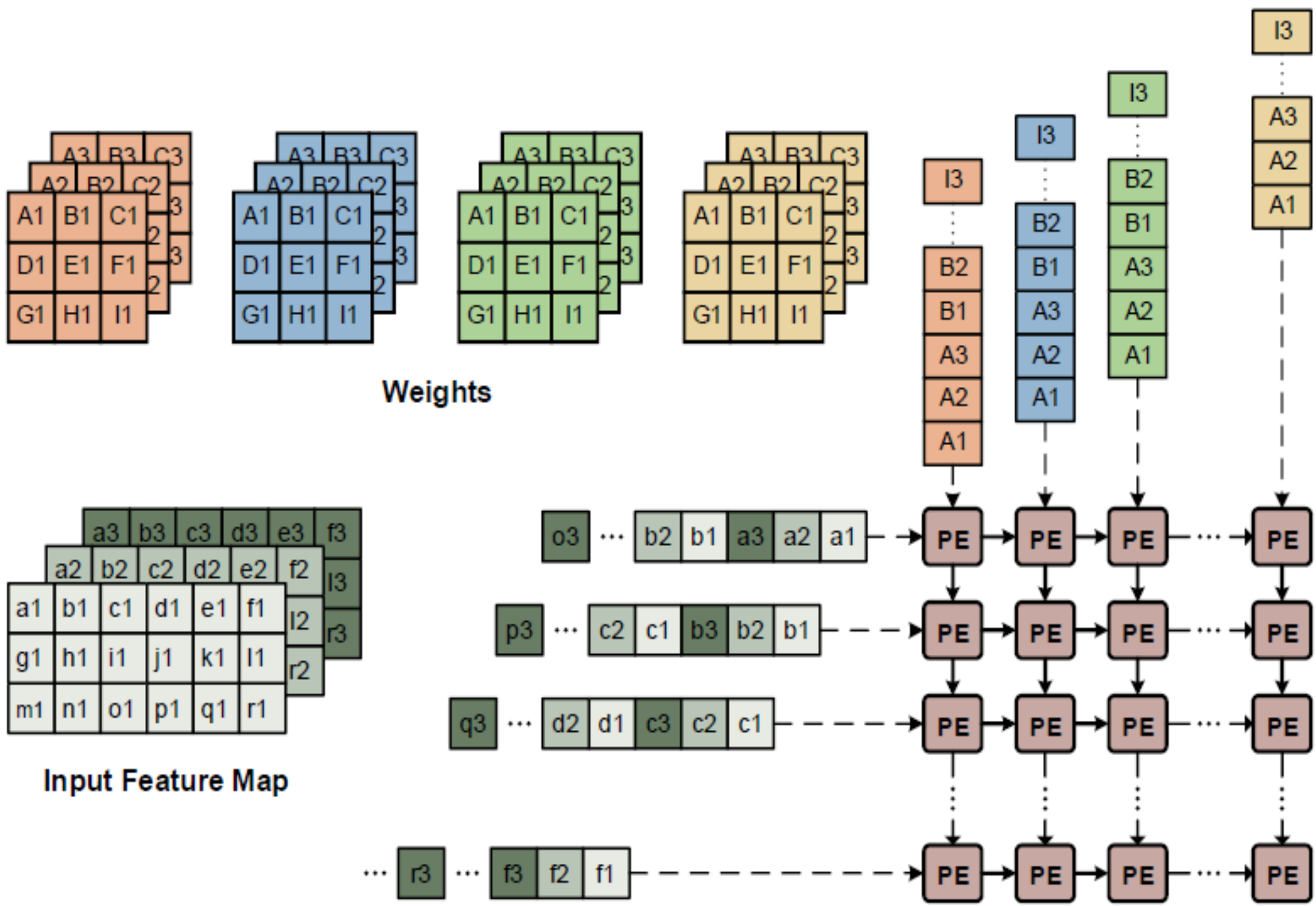
本研究提出了一種基於可重構脈動陣列的卷積神經網路（CNN）加速器架構。該架構將資料進行量化預處理並以特定通道格式排列後，輸入至脈動陣列進行卷積運算，並採用 Pipeline 技術以及高效的乘法器與設計。

硬體架構圖



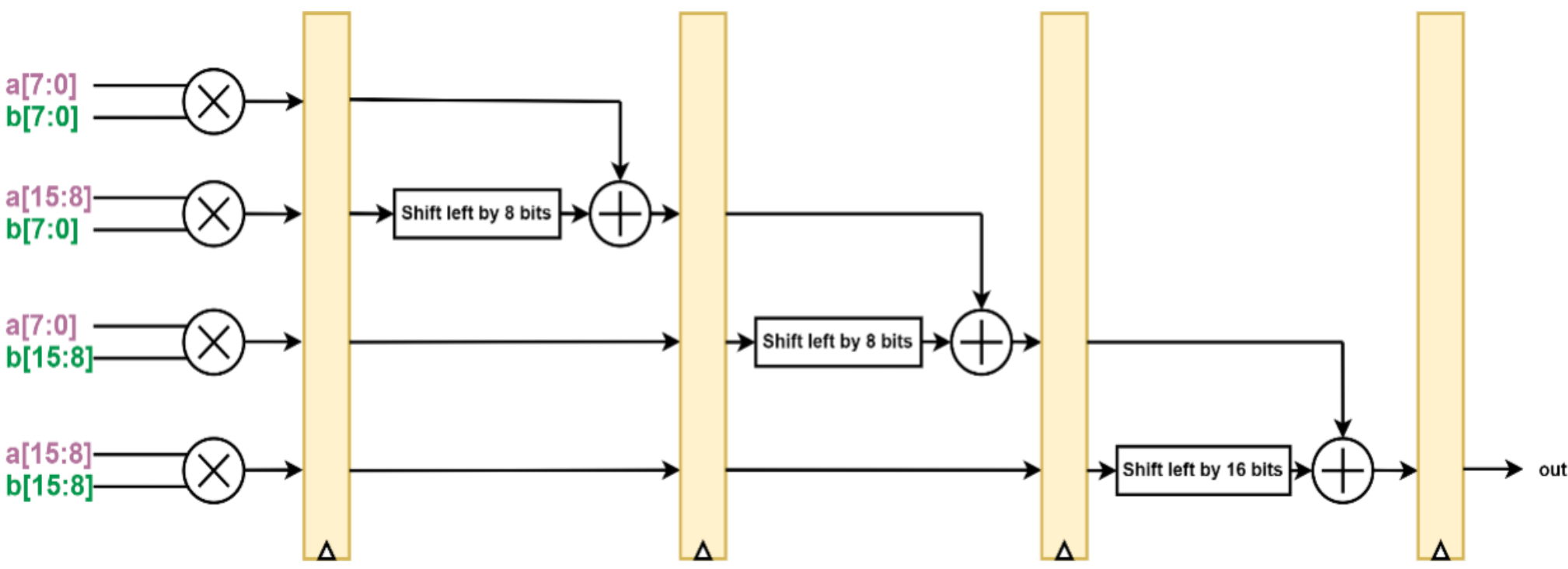
可重構脈動陣列

每個計算單元接收來自相鄰計算單元的數據，並對其執行卷積操作。Systolic array 的並行性和高效性可以顯著提高卷積神經網路的計算速度和效率。



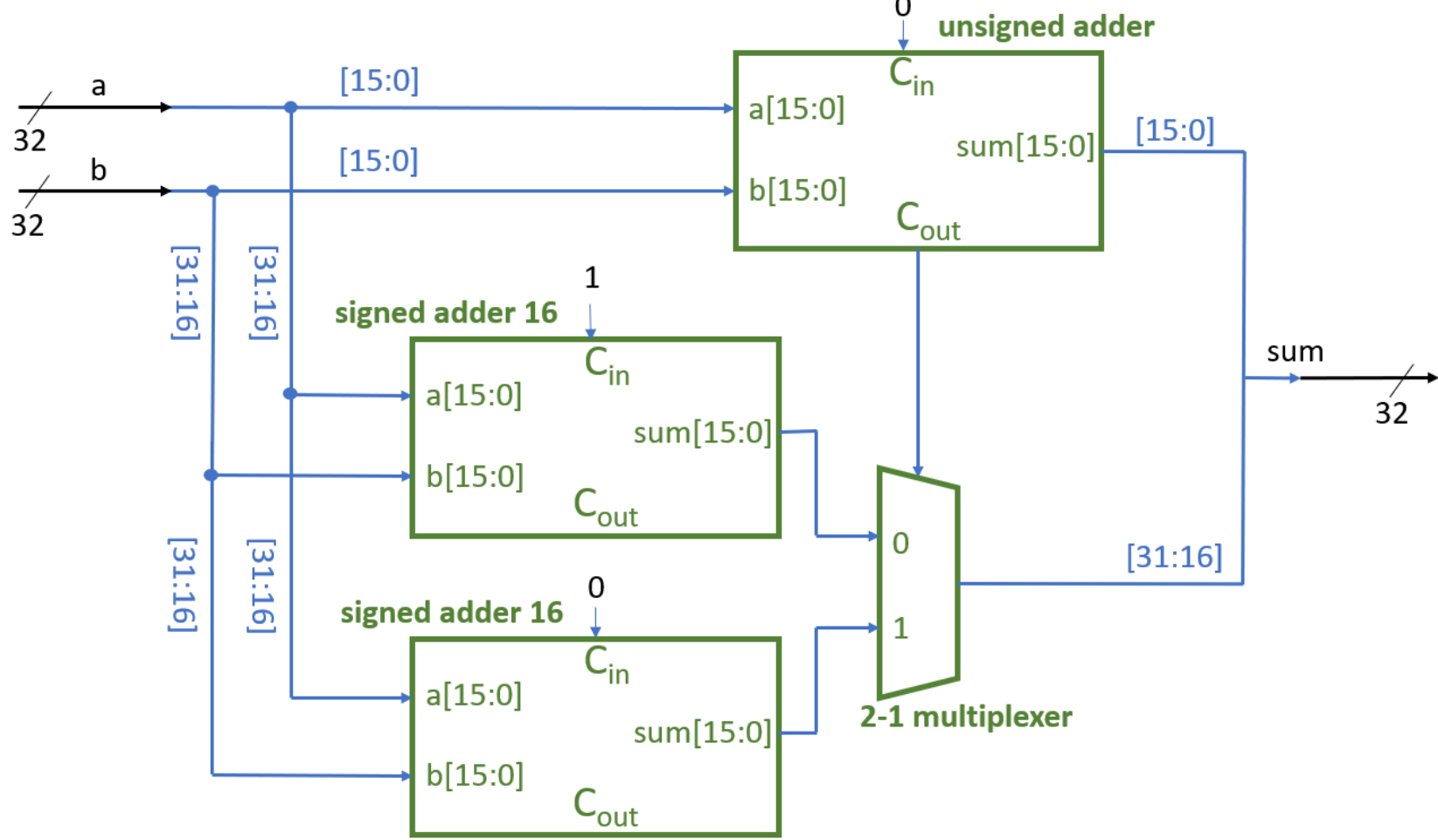
乘法器

將乘法運算拆分為較小位寬的子運算，且結合流水線設計。



加法器

32bits的數相加拆成兩個16bits相加，再利用與多工器去選擇是否有進位。



卷積神經網路加速器處理單元設計與實現

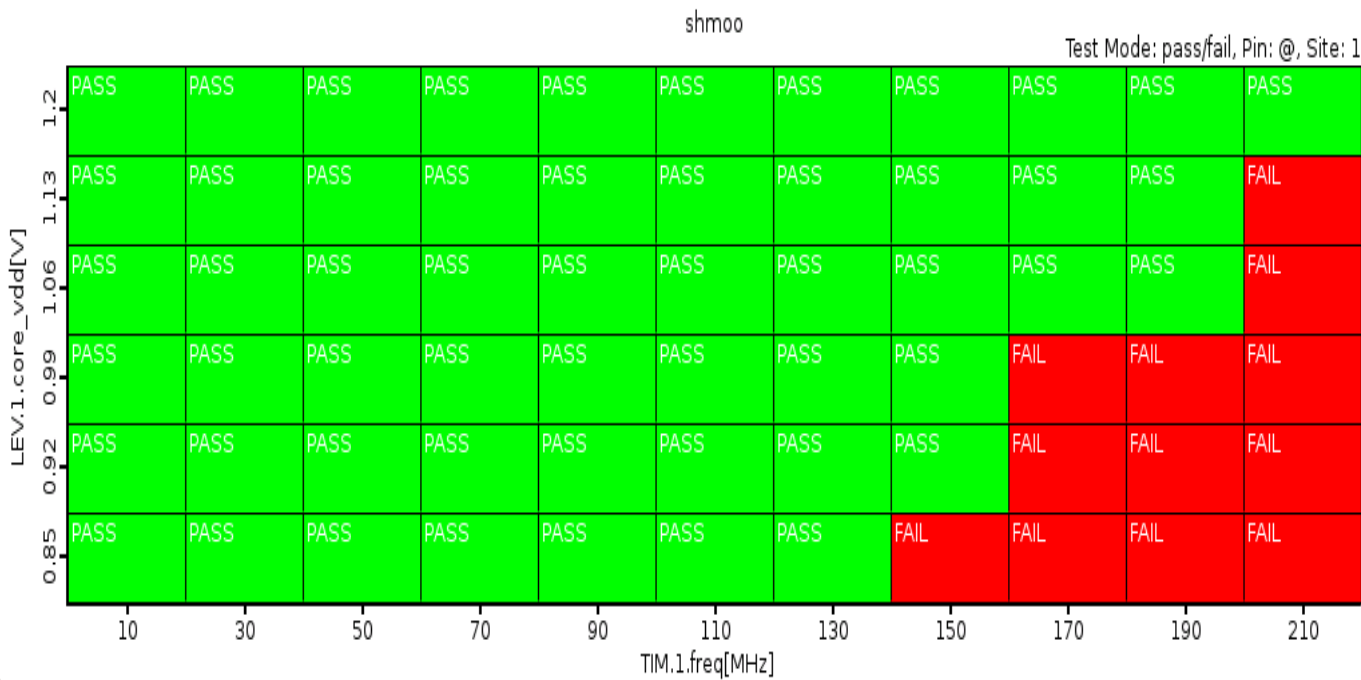
組員：林英傑、吳致廣、林穎杉、唐瑚憶

實驗結果

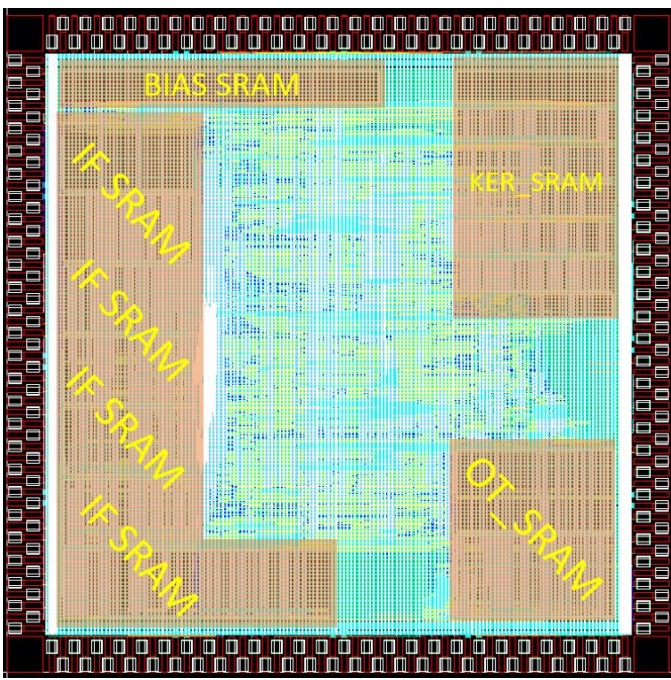
晶片已於2024年8月下線完成，並使用國家晶片中心(TSRI)所提供的ADVANTEST V93000 PS1600 自動測試系統來驗證晶片量測的Power與效能。在IO 3.3V, Core 1.2V可達200MHz, standby power 在IO 3.3V, Core 1V為38.76mW, operation power 為100.73mW。

	standby current(uA)	operation current(uA)
Core 1V	20062.391	81038.71
IO 3.3V	5666.568	5966.572
	standby power(mW)	operation power(mW)
	38.76	100.73

晶片量測結果

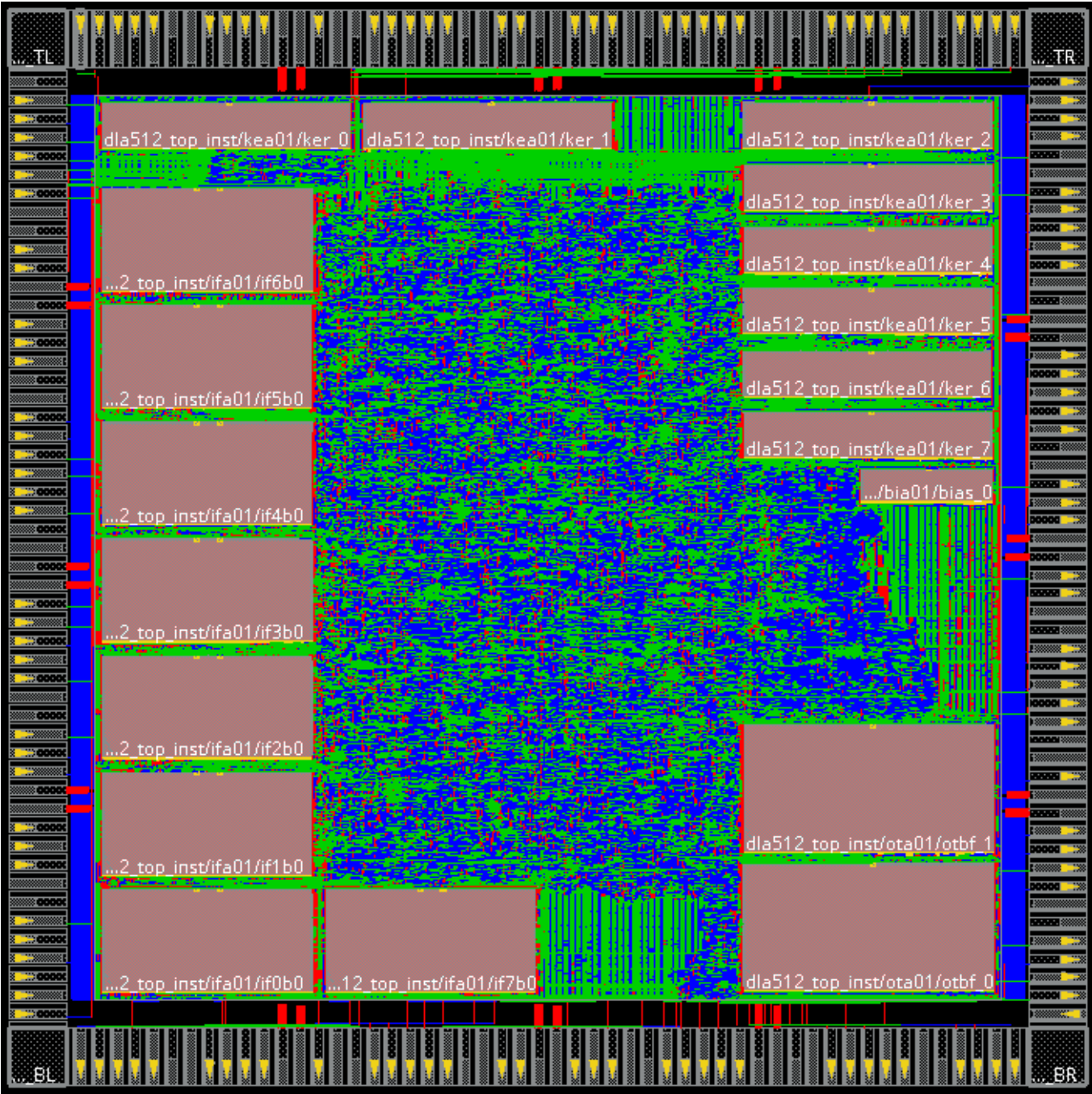


Shmoo Plot



晶片布局圖

CHIP & SPEC



Technology	TSMC 90nm
Clock Rate	200MHz
Chip Size	3.521x3.505mm ²
Core Area	2.948x2.943 mm ²
SRAM	194KB
IO power	23組
Core power	12組
Input	69
Output	67

Chip Implement 比較表

	[12]	[21]	[22]	[23]	[24]	This work
Technology	TSMC 65nm	TSMC 65nm	TSMC 65nm	TSMC 40nm	TSMC 40nm	TSMC 90nm
Core Area	3.5 x 3.5	2 x 2.5	3.2 x 3.3	-	73.81	3.61 x 3.61
Precision	Fixed16	Fixed16	Fixed16	Fixed16	8bit	Unit 8
CNN Net	AlexNet	LeNet-5	AlexNet	AlexNet	VGG-16	VGG-16
Number of PEs	168	144	64	216	2048	512
Clock Rate (Mhz)	250	500	250	454	200	200
Throughput (Gops)	67.2	152	32	132.8	522.35	204
Energy Efficiency (Gops/W)	241.72	434	209.15	-	296.79	2025
Area(MGE)	1.852	1.3	1.596	1.783	47.53	2.7
Area Efficiency (Gops/MGE)	57.1	116.9	20.0	74.5	11.0	750

a The area is shown in terms of the size of the million NAND2 gate
b Synthesised result