

CS2323 Homework 4

CS18BTECH11001

October 19, 2019

This document is generated by L^AT_EX

Q1. The General formula for the time steps in which an array multiplication could be done using systolic array = $3n - 2$.

Q2.

```
for(rr=0; rr < R;rr=rr+B)
  for ( row=rr; row < min(R,rr+B); row++)
    for ( col =0; col < C; col++)
      for ( to=0; to < M; to++)
        for ( ti =0; ti < N; ti++)
          for ( i =0; i < K; i++)
            for ( j =0; j < K; j++)
              Output_fmaps[to][row][col] += Weights[to][ti][i][j] × Input_fmaps[ti][S×row+i][S×col+j]
```

Q3.

```
_device_ void addFunc1(int *a, int *b, int *c)
_global_ void addFunc2(int *a, int *b, int *c)
_host_ void random_ints(int* x, int size)
_host_ int main(void)
```

Q4. Location of array/variables to be stored

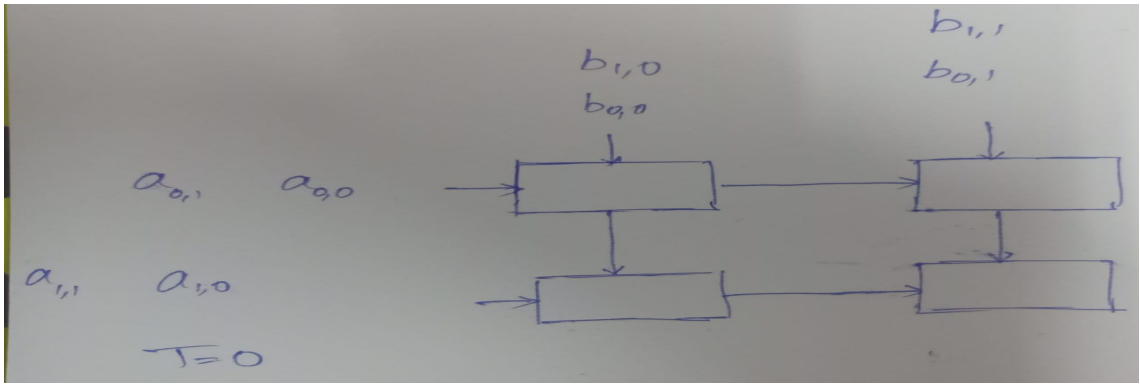
Variables	Location
<i>x_ dim</i>	local(register)
<i>y_ dim</i>	local(register)
<i>iteration</i>	local(register)
<i>pqr</i>	local memory
<i>ABC</i>	global memory
<i>maxValue</i>	global memory

Q5. a) The dimension of matrix = 16×16
Cache Size=128B

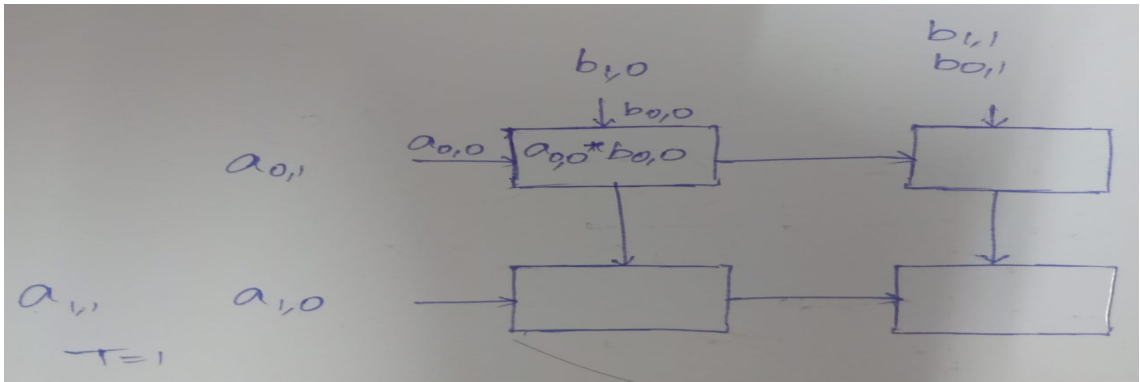
b) Unblocked Cache :
Total hits for unblocked cache = 192
Total misses for unblocked cache = 320
Misses coming from input matrix=64
Misses coming from output matrix=256
Blocked Cache :
Total hits for blocked cache = 384
Total misses for blocked cache = 128
Misses coming from input matrix=64
Misses coming from output matrix=64

Q6. Snapshot from $T=0$ to $T=4$:

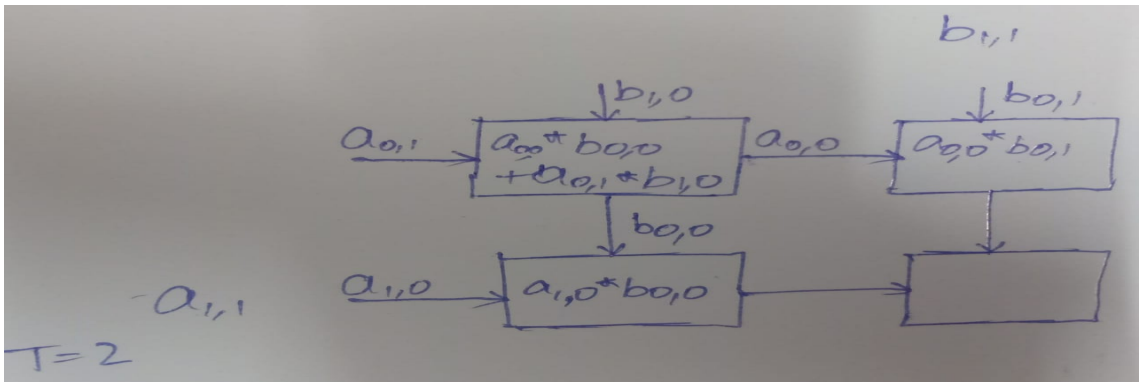
$T=0$:



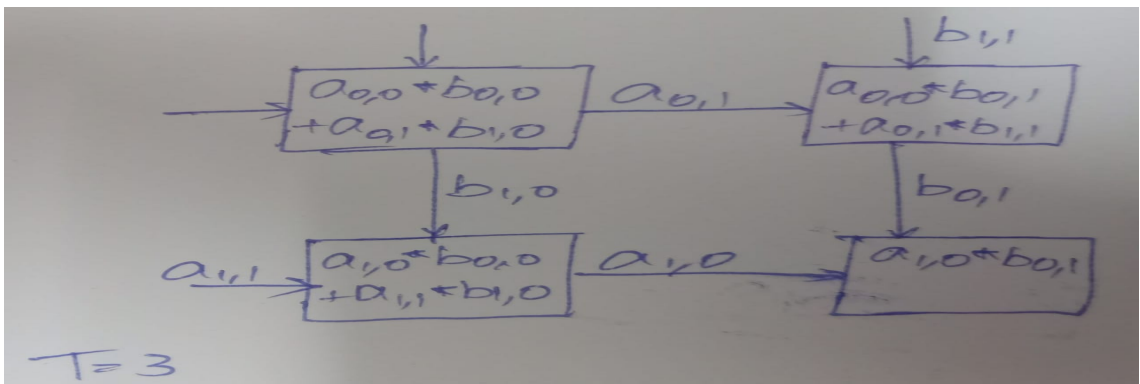
$T=1$:



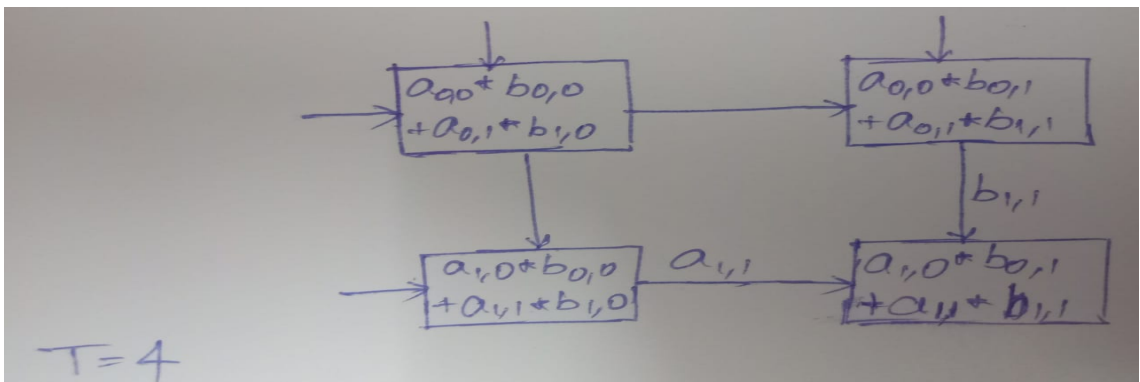
$T=2$:



$T=3$:



$T=4$:



- Q7. a) 4
b) 5.5
c) 5.6875
d) 5.6953125

Q8. *Instruction : v.ld vr1,20[r2]*
Semantics : vr1 $\leftarrow ([r2+20],[r2+24])$

Q9. a) $AI \text{ for case 1} = \frac{\text{Total no. of fp operations}}{\text{Total DRAM bytes}} = \frac{N^2}{N^2 * 3 * 8} = 1/24$

$AI \text{ for case 2} = \frac{\text{Total no. of fp operations}}{\text{Total DRAM bytes}} = \frac{N^2/4}{N^2 * 3 * 8} = 1/96$

b) $AI \text{ for case 1} = \frac{\text{Total no. of fp operations}}{\text{Total DRAM bytes}} = \frac{N^2}{N^2 * 3 * 8} = 1/24$

$AI \text{ for case 2} = \frac{\text{Total no. of fp operations}}{\text{Total DRAM bytes}} = \frac{N^2/4}{\frac{N^2}{4} * 3 * 8} = 1/24$

Q10. a) *No. of GOPS required to classify 1 image = 1.5*
No. of GOPS available at peak performance = $0.75 \times 66 \times 1000 = 49500$
No. of images that can be classified = $\frac{49500}{1.5} = 33000$

b) $AI \text{ for 8b fixed point versions} = \frac{1.5 \times 10^9}{50 \times 1024 \times 1024 \times 1024} = 28.61022 \text{ operations/B}$

$AI \text{ for binarized versions} = \frac{1.5 \times 10^9}{7.4 \times 1024 \times 1024 \times 1024} = 193.31236 \text{ operations/B}$

Q11. $\text{Arithmetic intensity required for achieving peak FLOP on using MCDRAM} = \frac{2199}{372} = 5.91129$

$\text{arithmetic intensity required for achieving peak FLOP on using DRAM} = \frac{2199}{77} = 28.55844$