CSE530 Project 2

Name- Niramay Vaidya
PSU ID- 939687597
Date- 12/09/21
Time- 01:00 AM

Table of Contents

1. VGG19 Topology Configuration
2. Systolic Array Configuration
3. Results
    1. AlphaGoZero
        1. Compute Cycles
        2. Overall Utilization %
    2. FasterRCNN
        1. Compute Cycles
        2. Overall Utilization %
    3. Resnet50
        1. Compute Cycles
        2. Overall Utilization %
    4. VGG19
        1. Compute Cycles
        2. Overall Utilization %
4. Explanation

1. VGG19 Topology Configuration-

Layer information like input dimensions, filter dimensions, number of channels, filters, and strides was obtained from the VGG19 paper (mentions that the whole net uses very small 3x3 receptive fields which are convolved with the input at every pixel i.e. with stride 1) and the following references-
https://arxiv.org/pdf/1512.03385.pdf (Resnet paper which gives information about the input dimensions for every layer for VGG19 in one of the diagrams)
https://www.researchgate.net/figure/Details-on-the-VGG19-architecture-For-each-layer-number-of-filters-parameters-and_tbl1_314237915 (Gives information on the number of filters per layer)

VGG19.csv

```
Layer name, IFMAP Height, IFMAP Width, Filter Height, Filter Width, Channels, Num Filter, Strides,

conv1_1, 224, 224, 3, 3, 3, 64, 1,
conv1_2, 224, 224, 3, 3, 64, 64, 1,
conv2_1, 112, 112, 3, 3, 128, 128, 1,
conv2_2, 112, 112, 3, 3, 128, 128, 1,
conv3_1, 56, 56, 3, 3, 256, 256, 1,
conv3_2, 56, 56, 3, 3, 256, 256, 1,
conv3_3, 56, 56, 3, 3, 256, 256, 1,
conv3_4, 56, 56, 3, 3, 256, 256, 1,
conv4_1, 28, 28, 3, 3, 512, 512, 1,
conv4_2, 28, 28, 3, 3, 512, 512, 1,
conv4_3, 28, 28, 3, 3, 512, 512, 1,
conv4_4, 28, 28, 3, 3, 512, 512, 1,
conv5_1, 14, 14, 3, 3, 512, 512, 1,
conv5_2, 14, 14, 3, 3, 512, 512, 1,
conv5_3, 14, 14, 3, 3, 512, 512, 1,
conv5_4, 14, 14, 3, 3, 512, 512, 1,
fc6, 1, 1, 1, 1, 512, 4096, 1,
fc7, 1, 1, 1, 1, 512, 4096, 1,
fc8, 1, 1, 1, 1, 512, 1000, 1,
```

NORMAL VGG19.csv utf-8[unix] 4% 1/21 ln : 1

2. Systolic Array Configuration-

scale_256_2048.cfg

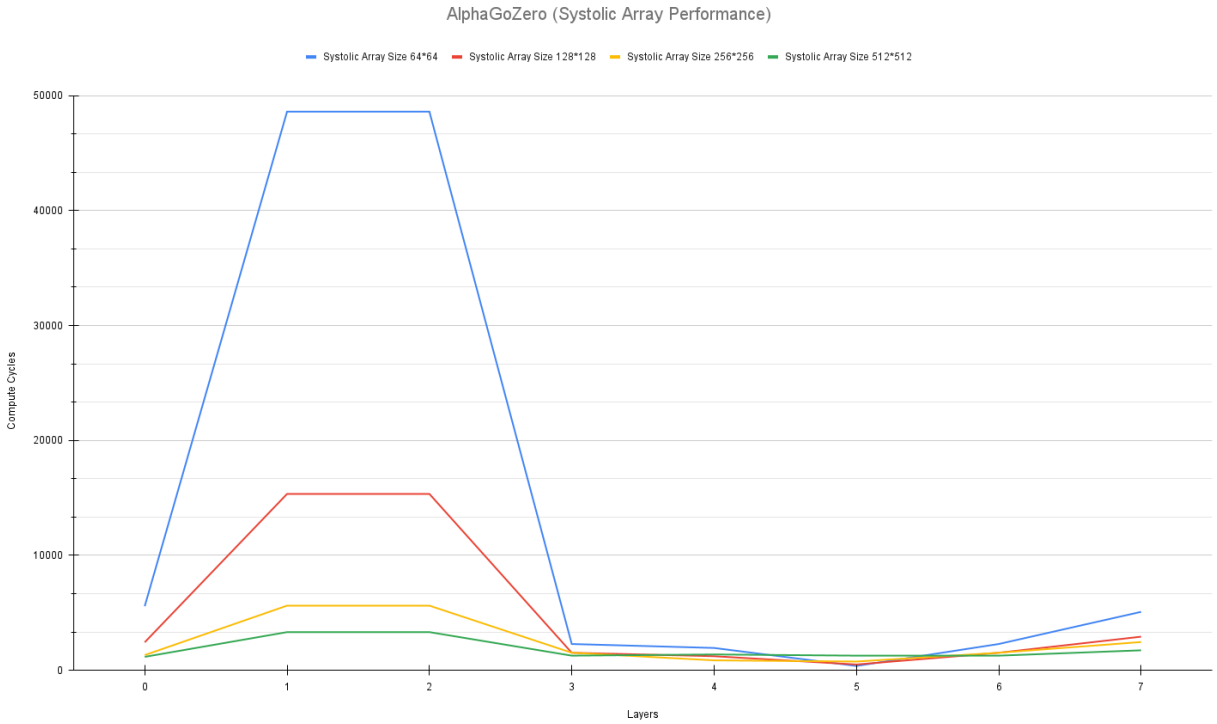
```
[general]
run_name = scale_example_run_256_2048_os

[architecture_presets]
ArrayHeight:    256
ArrayWidth:     256
IfmapSramSzkB:   2048
FilterSramSzkB:  2048
OfmapSramSzkB:   2048
IfmapOffset:    0
FilterOffset:   10000000
OfmapOffset:    20000000
Bandwidth : 10
Dataflow : os
MemoryBanks:   1

[run_presets]
InterfaceBandwidth: CALC
```

NORMAL main <le_256_2048.cfg cfg utf-8[unix] 5% 1/18 ln : 1 [18]tra…

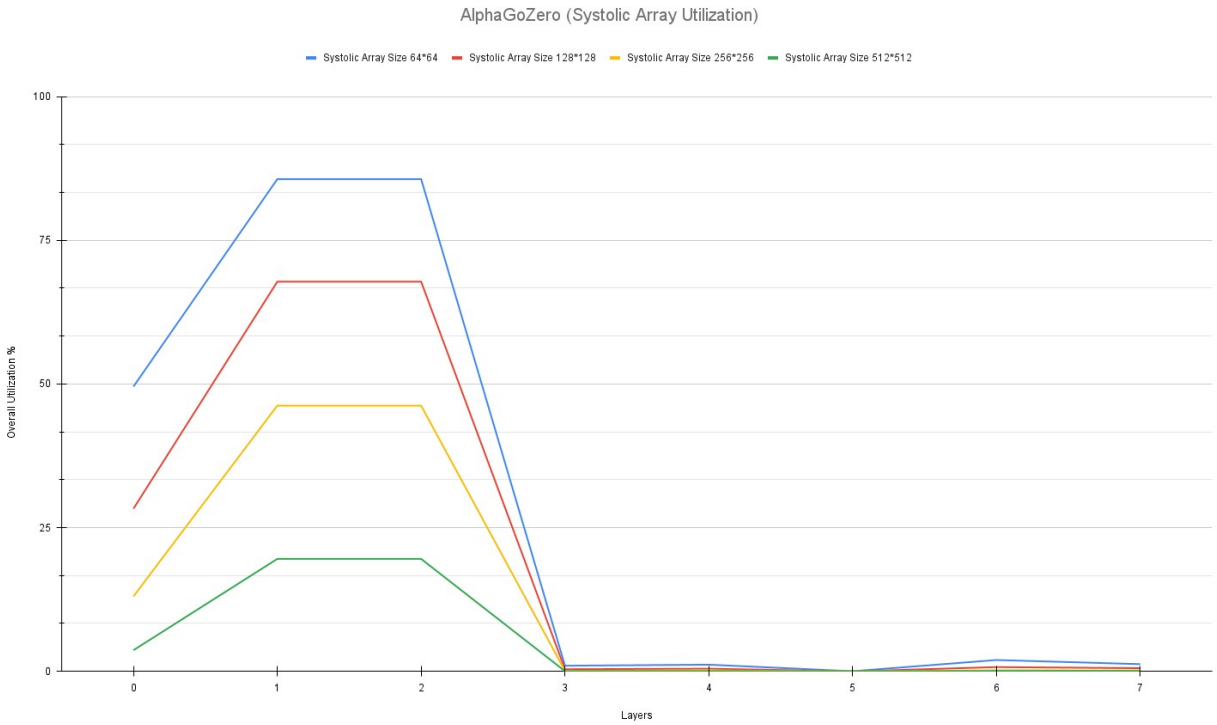
3. Results-

1. AlphaGoZero

1. Compute Cycles-

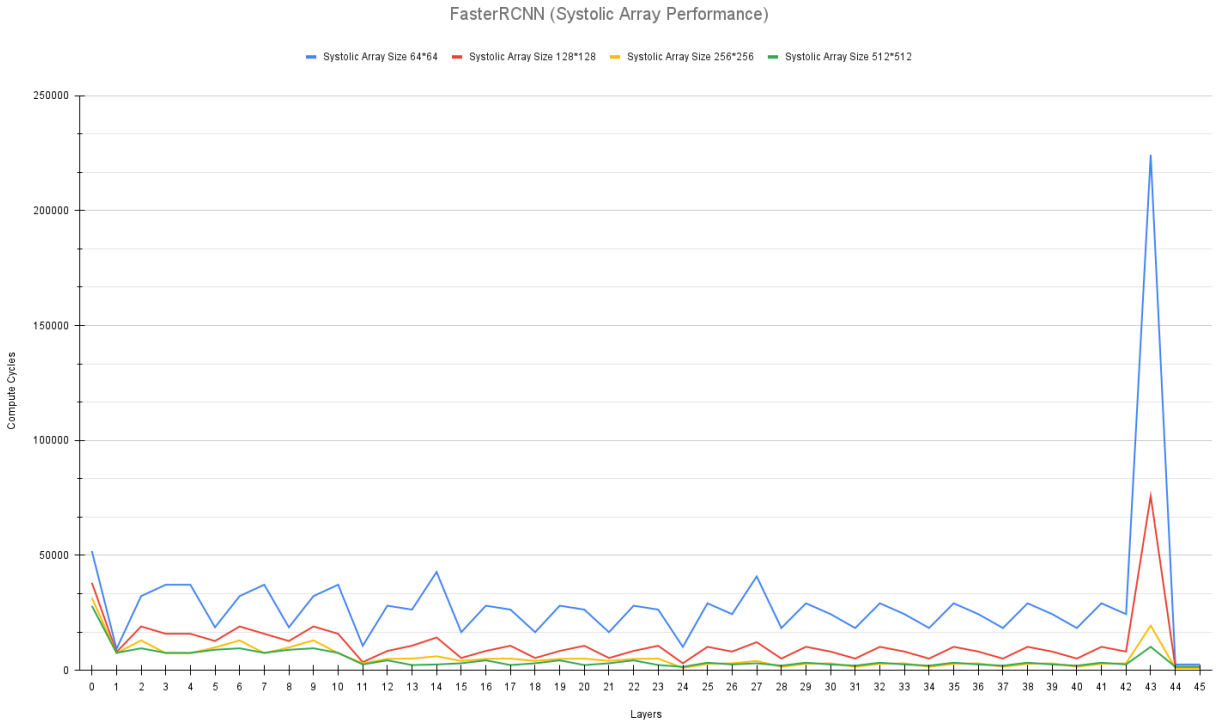


2. Overall Utilization %-

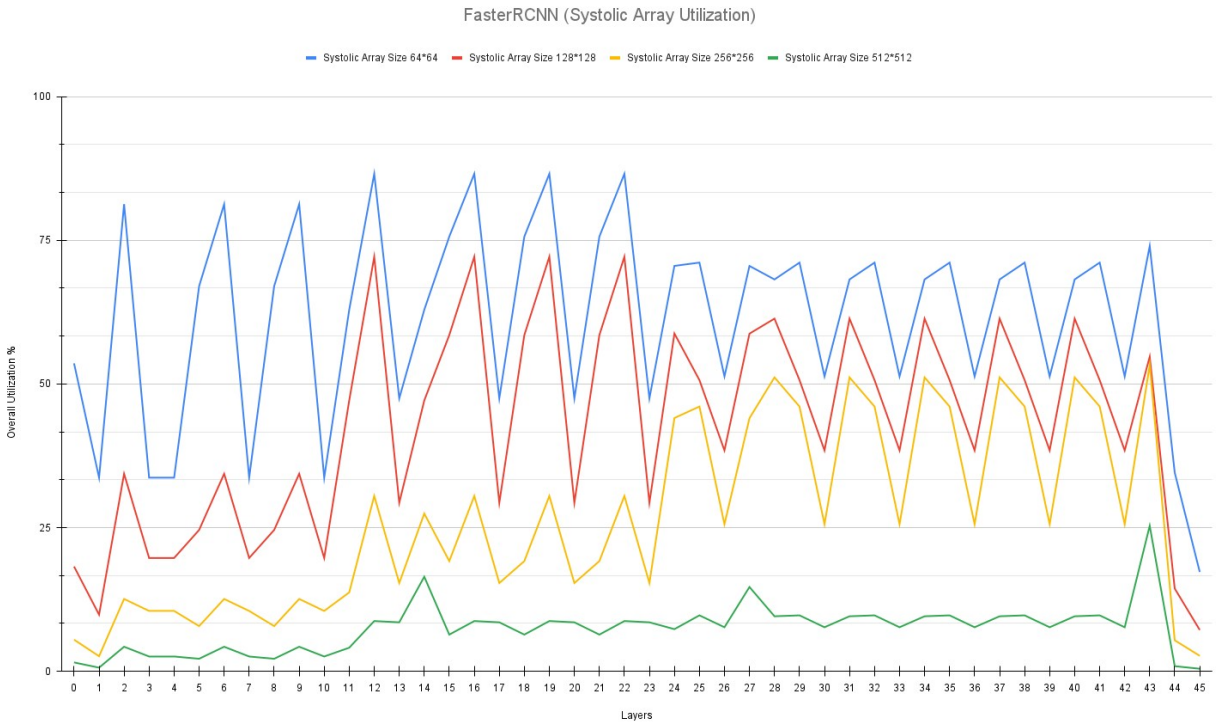


2. FasterRCNN

1. Compute Cycles-

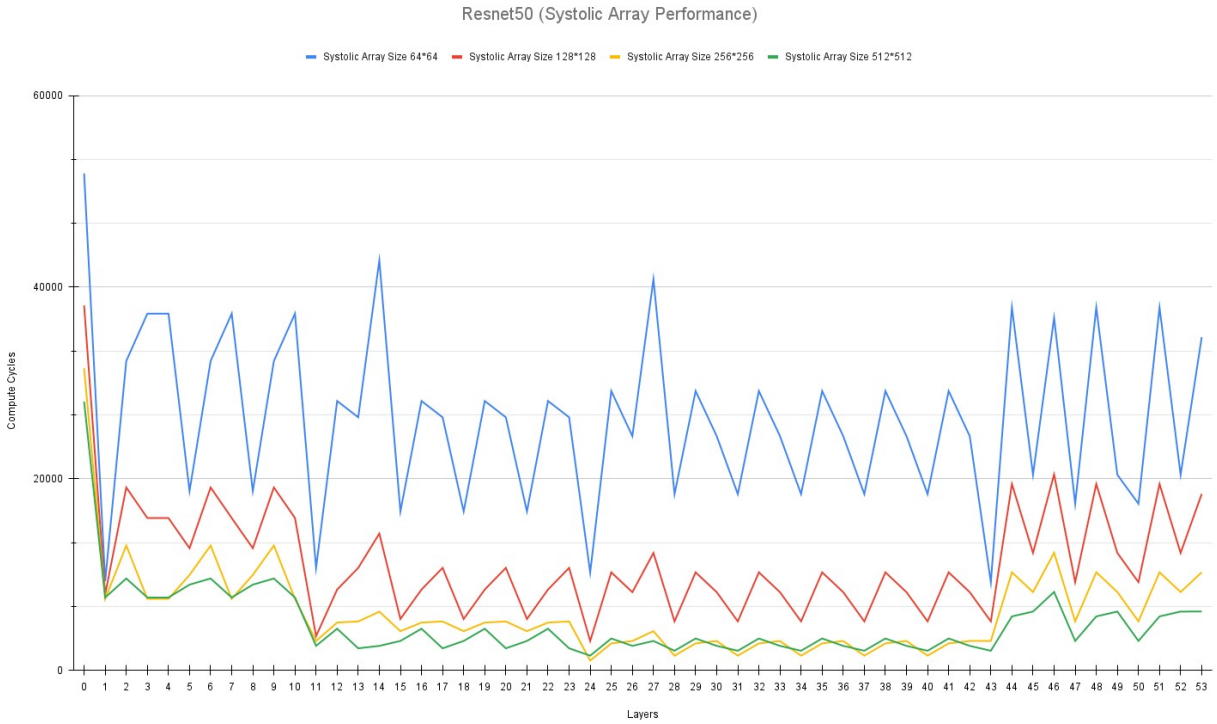


2. Overall Utilization %-

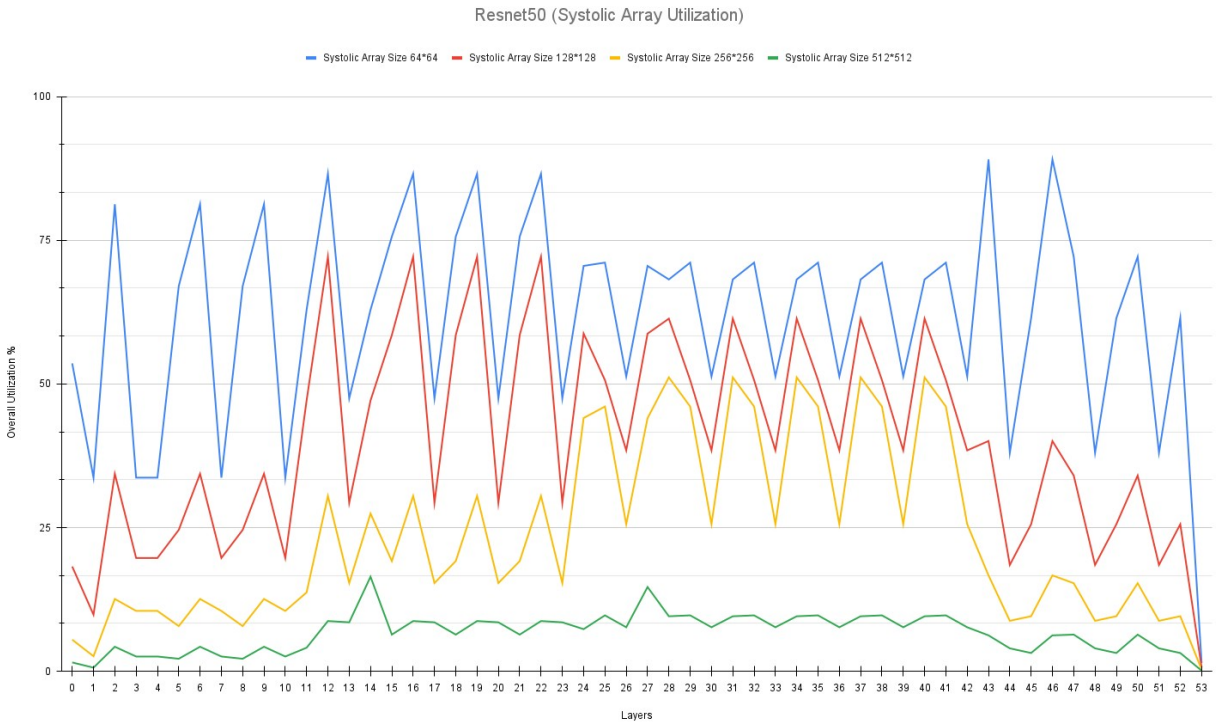


3. Resnet50

1. Compute Cycles-

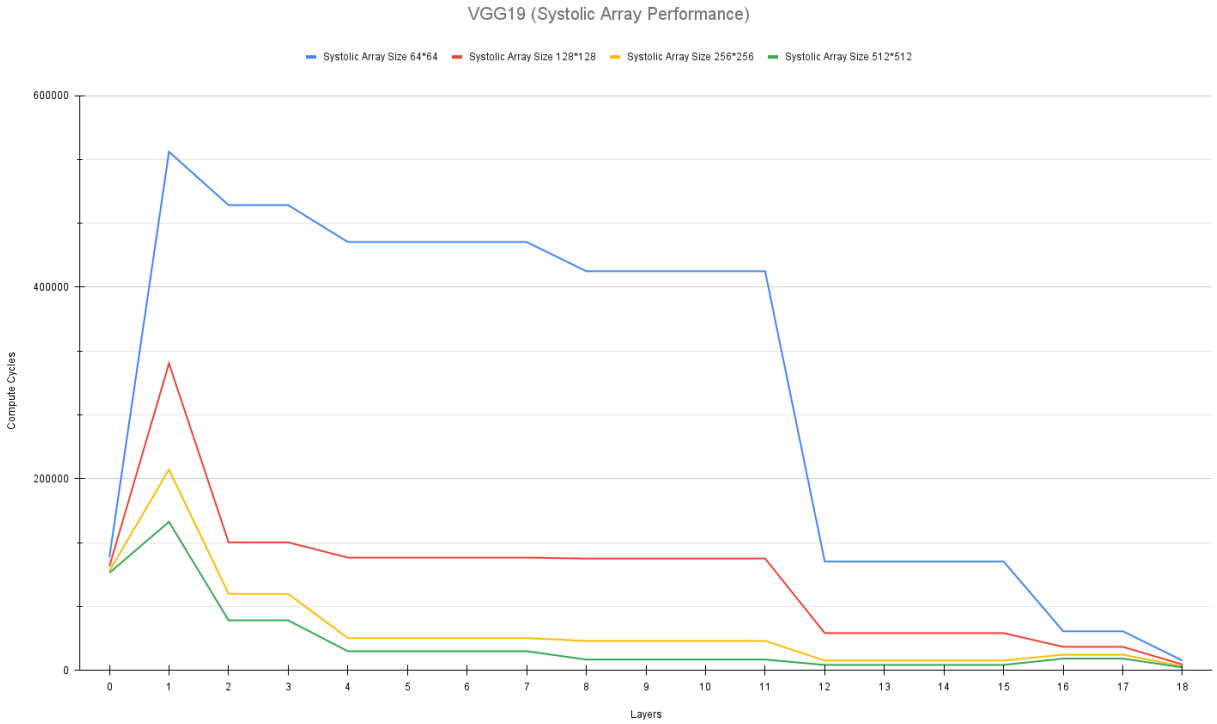


2. Overall Utilization %-

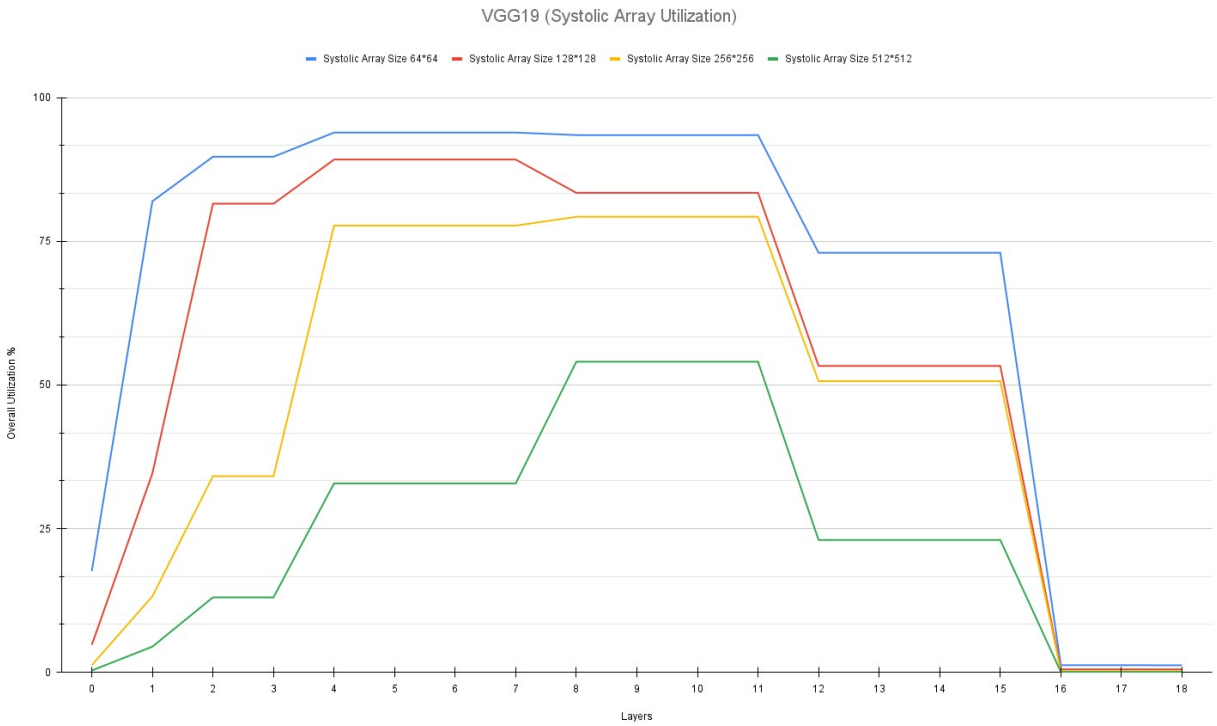


4. VGG19

1. Compute Cycles-



2. Overall Utilization %-



4. Explanation-

On the basis of the obtained results, the systolic array configuration having PE array dimensions 256*256 with the 3 SRAM (Ifmap, Filter and Ofmap) sizes all equal to 2048 KB seems to be the optimal configuration for all the 4 models/topologies. This is because with a further increase to 512*512, the improvement in the compute cycles is minimal but with this PE dimension, for most of the models (3 out of 4 i.e. AlphaGoZero, FasterRCNN and Resnet50), the overall utilization % is observed to be less than 10%, which indicates over-design of the system (under utilization of the available hardware). Also, increasing and decreasing the SRAM sizes to 4096 KB and 1024 KB has no effect on the obtained results (i.e. they come out to be the exact same for all the models), which is why the baseline config of 2048 KB has been considered.