COMPUTER ORGANIZATION AND ARCHITECTURE
COA LAB ASSIGNMENT – 3
Name: Y.Krishna Vamshi
Roll No: 21CS01019

Q1. What is the runtime for each configuration?

Example: Plot showing the IPC on Y-axis and application name on X-axis, legend: different warp schedulers.

Plot of IPC vs Configurations:

Firstly I have calculated the values of IPC for all 6 configurations

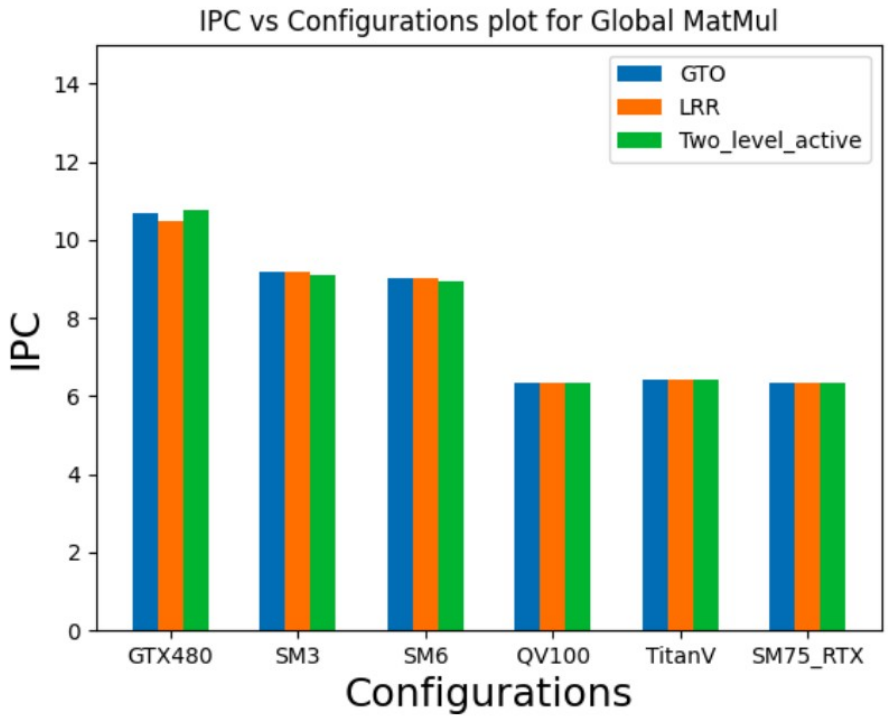
1. SM2_GTX480
2. SM3_KEPLER_TITAN
3. SM6_TitanX
4. SM7_QV100
5. SM7_TITANV
6. SM75_RTX2060

and for each configuration, I have run for each warp schedulers (GTO - Greedy-then-oldest ,LRR - Loose Round Robin, and TL - Two Level).

This process is repeated for both global and shared memory and plot the grap.

Plot:

1. Global Matrix Multiplication:



- The following screenshots contain data of instuctions per sec,cycles per sec and run time for each warp scheduler and for each of the 6 configurations for global memory implementation of matrix multiplication.

Warp Scheduler:

GTO:

SM2_GTX480:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 18 sec (18 sec)
gpgpu_simulation_rate = 132309 (inst/sec)
gpgpu_simulation_rate = 12370 (cycle/sec)
gpgpu_silicon_slowdown = 56588x
```

SM3_KEPLER_TITAN:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 27 sec (27 sec)
gpgpu_simulation_rate = 88206 (inst/sec)
gpgpu_simulation_rate = 9617 (cycle/sec)
```

SM6_TITANX:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 38 sec (38 sec)
gpgpu_simulation_rate = 62672 (inst/sec)
gpgpu_simulation_rate = 6943 (cycle/sec)
```

SM7_QV100:

```
gpgpu_simulation_time = 0 days, 0 hrs, 2 min, 37 sec (157 sec)
gpgpu_simulation_rate = 15169 (inst/sec)
gpgpu_simulation_rate = 2395 (cycle/sec)
```

SM7_TITANV:

```
gpgpu_simulation_time = 0 days, 0 hrs, 2 min, 6 sec (126 sec)
gpgpu_simulation_rate = 18901 (inst/sec)
gpgpu_simulation_rate = 2944 (cycle/sec)
```

SM75_RTX2060:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 57 sec (57 sec)
gpgpu_simulation_rate = 41781 (inst/sec)
gpgpu_simulation_rate = 6598 (cycle/sec)
```

LRR:

SM2_GTX480:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 18 sec (18 sec)
gpgpu_simulation_rate = 161749 (inst/sec)
gpgpu_simulation_rate = 15447 (cycle/sec)
gpgpu_silicon_slowdown = 45316x
```

SM3_KEPLER_TITAN:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 24 sec (24 sec)
gpgpu_simulation_rate = 121312 (inst/sec)
gpgpu_simulation_rate = 13208 (cycle/sec)
gpgpu_silicon_slowdown = 63370x
```

SM6_TITANX:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 35 sec (35 sec)
gpgpu_simulation_rate = 83185 (inst/sec)
gpgpu_simulation_rate = 9205 (cycle/sec)
gpgpu_silicon_slowdown = 153938x
```

SM7_QV100:

```
gpgpu_simulation_time = 0 days, 0 hrs, 2 min, 33 sec (153 sec)
gpgpu_simulation_rate = 19029 (inst/sec)
gpgpu_simulation_rate = 3002 (cycle/sec)
gpgpu_silicon_slowdown = 377081x
```

SM7_TITANV:

```
gpgpu_simulation_time = 0 days, 0 hrs, 1 min, 43 sec (103 sec)
gpgpu_simulation_rate = 23122 (inst/sec)
gpgpu_simulation_rate = 3603 (cycle/sec)
gpgpu_silicon_slowdown = 333055x
```

SM75_RTX2060:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 45 sec (45 sec)
gpgpu_simulation_rate = 52923 (inst/sec)
gpgpu_simulation_rate = 8357 (cycle/sec)
gpgpu_silicon_slowdown = 163336x
```

TWO_LEVEL_ACTIVE:

SM2_GTX480:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 18 sec (18 sec)
gpgpu_simulation_rate = 132309 (inst/sec)
gpgpu_simulation_rate = 12309 (cycle/sec)
gpgpu_silicon_slowdown = 56868x
```

SM3_KEPLER_TITAN:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 26 sec (26 sec)
gpgpu_simulation_rate = 91598 (inst/sec)
gpgpu_simulation_rate = 10054 (cycle/sec)
gpgpu_silicon_slowdown = 83250x
```

SM6_TITANX:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 38 sec (38 sec)
gpgpu_simulation_rate = 62672 (inst/sec)
gpgpu_simulation_rate = 7000 (cycle/sec)
gpgpu_silicon_slowdown = 202428x
```

SM7_QV100:

```
gpgpu_simulation_time = 0 days, 0 hrs, 2 min, 43 sec (163 sec)
gpgpu_simulation_rate = 14610 (inst/sec)
gpgpu_simulation_rate = 2310 (cycle/sec)
gpgpu_silicon_slowdown = 490043x
```

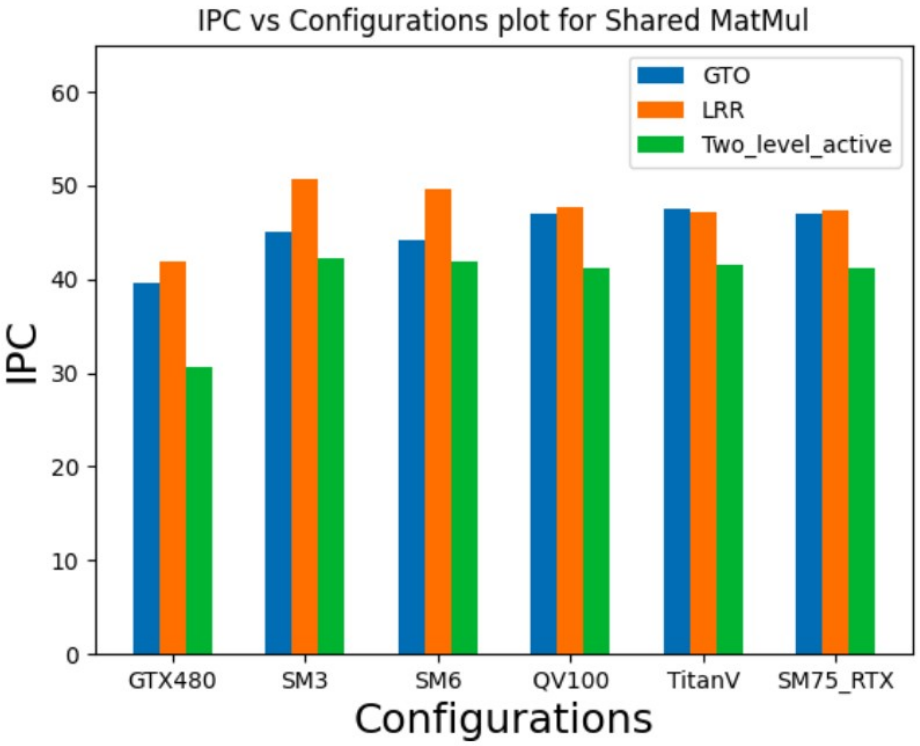
SM7_TITANV:

```
gpgpu_simulation_time = 0 days, 0 hrs, 2 min, 14 sec (134 sec)
gpgpu_simulation_rate = 17772 (inst/sec)
gpgpu_simulation_rate = 2773 (cycle/sec)
gpgpu_silicon_slowdown = 432744x
```

SM75_RTX2060:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 57 sec (57 sec)
gpgpu_simulation_rate = 41781 (inst/sec)
gpgpu_simulation_rate = 6597 (cycle/sec)
gpgpu_silicon_slowdown = 206912x
```

2. Shared Matrix Multiplication:



- The following screenshots contain data of instuctions per sec,cycles per sec and run time for each warp scheduler and for each of the 6 configurations for global memory implementation of matrix multiplication.

GTO:

SM2_GTX480:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 16 sec (16 sec)
gpgpu_simulation_rate = 424976 (inst/sec)
gpgpu_simulation_rate = 10738 (cycle/sec)
```

SM3_KEPLER_TITAN:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 18 sec (18 sec)
gpgpu_simulation_rate = 377756 (inst/sec)
gpgpu_simulation_rate = 8383 (cycle/sec)
```

SM6_TITANX:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 23 sec (23 sec)
gpgpu_simulation_rate = 295635 (inst/sec)
gpgpu_simulation_rate = 6701 (cycle/sec)
```

SM7_QV100:

```
gpgpu_simulation_time = 0 days, 0 hrs, 1 min, 3 sec (63 sec)
gpgpu_simulation_rate = 107930 (inst/sec)
gpgpu_simulation_rate = 2295 (cycle/sec)
```

SM7_TITANV:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 51 sec (51 sec)
gpgpu_simulation_rate = 133325 (inst/sec)
gpgpu_simulation_rate = 2805 (cycle/sec)
```

SM75_RTX2060:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 24 sec (24 sec)
gpgpu_simulation_rate = 283317 (inst/sec)
gpgpu_simulation_rate = 6024 (cycle/sec)
```

LRR:

SM2_GTX480:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 15 sec (15 sec)
gpgpu_simulation_rate = 561288 (inst/sec)
gpgpu_simulation_rate = 13395 (cycle/sec)
gpgpu_silicon_slowdown = 52258x
```

SM3_KEPLER_TITAN:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 15 sec (15 sec)
gpgpu_simulation_rate = 561288 (inst/sec)
gpgpu_simulation_rate = 11082 (cycle/sec)
gpgpu_silicon_slowdown = 75527x
```

SM6_TITANX:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 21 sec (21 sec)
gpgpu_simulation_rate = 400920 (inst/sec)
gpgpu_simulation_rate = 8088 (cycle/sec)
gpgpu_silicon_slowdown = 175197x
```

SM7_QV100:

```
gpgpu_simulation_time = 0 days, 0 hrs, 1 min, 2 sec (62 sec)
gpgpu_simulation_rate = 135795 (inst/sec)
gpgpu_simulation_rate = 2850 (cycle/sec)
gpgpu_silicon_slowdown = 397192x
```

SM7_TITANV:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 43 sec (43 sec)
gpgpu_simulation_rate = 158130 (inst/sec)
gpgpu_simulation_rate = 3355 (cycle/sec)
gpgpu_silicon_slowdown = 357675x
```

SM75_RTX2060:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 19 sec (19 sec)
gpgpu_simulation_rate = 357874 (inst/sec)
gpgpu_simulation_rate = 7559 (cycle/sec)
gpgpu_silicon_slowdown = 180579x
```

TWO_LEVEL_ACTIVE:

SM2_GTX480:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 20 sec (20 sec)
gpgpu_simulation_rate = 339980 (inst/sec)
gpgpu_simulation_rate = 11128 (cycle/sec)
gpgpu_silicon_slowdown = 62904x
```

SM3_KEPLER_TITAN:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 18 sec (18 sec)
gpgpu_simulation_rate = 377756 (inst/sec)
gpgpu_simulation_rate = 8933 (cycle/sec)
gpgpu_silicon_slowdown = 93697x
```

SM6_TITANX:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 27 sec (27 sec)
gpgpu_simulation_rate = 251837 (inst/sec)
gpgpu_simulation_rate = 6009 (cycle/sec)
gpgpu_silicon_slowdown = 235812x
```

SM7_QV100:

```
gpgpu_simulation_time = 0 days, 0 hrs, 1 min, 15 sec (75 sec)
gpgpu_simulation_rate = 90661 (inst/sec)
gpgpu_simulation_rate = 2198 (cycle/sec)
gpgpu_silicon_slowdown = 515013x
```

SM7_TITANV:

```
gpgpu_simulation_time = 0 days, 0 hrs, 0 min, 27 sec (27 sec)
gpgpu_simulation_rate = 251837 (inst/sec)
gpgpu_simulation_rate = 6111 (cycle/sec)
gpgpu_silicon_slowdown = 223367x
```

SM75_RTX2060:

```
gpgpu_simulation_time = 0 days, 0 hrs, 1 min, 2 sec (62 sec)
gpgpu_simulation_rate = 109671 (inst/sec)
gpgpu_simulation_rate = 2638 (cycle/sec)
gpgpu_silicon_slowdown = 454890x
```

2. Do you notice any correlation between the L1D, L2 and DRAM Statistics in the output/log file of gpgpusim for both the implementations (shared memory vs global memory). Mention your observations with supporting data.

Obseved correlation between L1D, L2 and DRAM is:

1. L1D:

- Access time for memory is fastest for L1D cache.
- Miss rate for L1D in case of global is less than that of shared memory.
- Total cache access for L1D are also more for global than shared memory.
- This can be understood because in case of shared memory, data is majorly accessed from L2 Cache and in case of global memory implementation it is majorly accessed from L1cache.

2. L2 :

- Access time for L2 is in between that of L1 and DRAM.
- Miss rate for L2 in case of global is more than that of shared memory.
- Accesses for L2 is greater for shared than global memory.
- This can be understood because in case of global more is already found from L1 cache.
- In case of Shared memory implementation, data is stored in shared memory and that is also why hit rate is higher.

3. DRAM:

- Total accesses for dram in both global and shared memory is zero as per outputs.
- As accesses and hit rate in both global and shared memory implementation is very high.
- Thus high accesses and low miss rates could be resulted in low access from DRAM.

Supporting Data:

Shared:

1. L1D :

```
L1D_cache:
	L1D_cache_core[0]: Access = 3700, Miss = 1999, Miss_rate = 0.540, Pending_hits = 834, Reservation_fails = 0
	L1D_cache_core[1]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[2]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[3]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[4]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[5]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[6]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[7]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[8]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[9]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[10]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[11]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[12]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[13]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[14]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_total_cache_accesses = 3700
	L1D_total_cache_misses = 1999
	L1D_total_cache_miss_rate = 0.5403
	L1D_total_cache_pending_hits = 834
	L1D_total_cache_reservation_fails = 0
	L1D_cache_data_port_util = 0.004
	L1D_cache_fill_port_util = 0.010
```

2. DRAM :

```
number of total memory accesses made:
dram[0]:         0         0         0         0         0         0         0         0         0         0         0
         0         0
dram[1]:         0         0         0         0         0         0         0         0         0         0         0
         0         0
dram[2]:         0         0         0         0         0         0         0         0         0         0         0
         0         0
dram[3]:         0         0         0         0         0         0         0         0         0         0         0
         0         0
dram[4]:         0         0         0         0         0         0         0         0         0         0         0
         0         0
dram[5]:         0         0         0         0         0         0         0         0         0         0         0
         0         0
total accesses: 0
```

3. L2:

```
L2_cache_bank[0]: Access = 674, Miss = 10, Miss_rate = 0.015, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[1]: Access = 668, Miss = 0, Miss_rate = 0.000, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[2]: Access = 676, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[3]: Access = 660, Miss = 0, Miss_rate = 0.000, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[4]: Access = 668, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[5]: Access = 664, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[6]: Access = 672, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[7]: Access = 664, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[8]: Access = 668, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[9]: Access = 664, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[10]: Access = 668, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[11]: Access = 668, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_total_cache_accesses = 8014
L2_total_cache_misses = 82
L2_total_cache_miss_rate = 0.0102
L2_total_cache_pending_hits = 0
L2_total_cache_reservation_fails = 0
```

Global:

1. L1D:

```
L1D_cache:
	L1D_cache_core[0]: Access = 62736, Miss = 2001, Miss_rate = 0.032, Pending_hits = 6862, Reservation_fails = 108492
	L1D_cache_core[1]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[2]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[3]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[4]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[5]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[6]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[7]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[8]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[9]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[10]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[11]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[12]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[13]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_cache_core[14]: Access = 0, Miss = 0, Miss_rate = -nan, Pending_hits = 0, Reservation_fails = 0
	L1D_total_cache_accesses = 62736
	L1D_total_cache_misses = 2001
	L1D_total_cache_miss_rate = 0.0319
	L1D_total_cache_pending_hits = 6862
	L1D_total_cache_reservation_fails = 108492
	L1D_cache_data_port_util = 0.194
	L1D_cache_fill_port_util = 0.007
```

2. DRAM :

```
number of total memory accesses made:
dram[0]:         0        0        0        0        0        0        0        0        0        0        0        0        0        0
         0        0
dram[1]:         0        0        0        0        0        0        0        0        0        0        0        0        0        0
         0        0
dram[2]:         0        0        0        0        0        0        0        0        0        0        0        0        0        0
         0        0
dram[3]:         0        0        0        0        0        0        0        0        0        0        0        0        0        0
         0        0
dram[4]:         0        0        0        0        0        0        0        0        0        0        0        0        0        0
         0        0
dram[5]:         0        0        0        0        0        0        0        0        0        0        0        0        0        0
         0        0
total accesses: 0
min_bank_accesses = 0!
min_chip_accesses = 0!
```

3. L2:

```
========= L2 cache stats =========
L2_cache_bank[0]: Access = 670, Miss = 10, Miss_rate = 0.015, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[1]: Access = 672, Miss = 0, Miss_rate = 0.000, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[2]: Access = 672, Miss = 0, Miss_rate = 0.000, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[3]: Access = 660, Miss = 0, Miss_rate = 0.000, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[4]: Access = 668, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[5]: Access = 672, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[6]: Access = 668, Miss = 4, Miss_rate = 0.006, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[7]: Access = 660, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[8]: Access = 664, Miss = 0, Miss_rate = 0.000, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[9]: Access = 660, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[10]: Access = 656, Miss = 0, Miss_rate = 0.000, Pending_hits = 0, Reservation_fails = 0
L2_cache_bank[11]: Access = 672, Miss = 8, Miss_rate = 0.012, Pending_hits = 0, Reservation_fails = 0
L2_total_cache_accesses = 7994
L2_total_cache_misses = 54
L2_total_cache_miss_rate = 0.0068
L2_total_cache_pending_hits = 0
L2_total_cache_reservation_fails = 0
```