COA LAB

ASSIGNMENT-4

21/08/2023

TEAM MEMBERS

- SHLOK KR. SHAW(21CS02008)
- JATIN YADAV(21CS02007)
- ANIKET ROY(21CS01061)
- AJINKYA DESHMUKH(21CS01029)
- GIRISH JAIN(21CS01016)

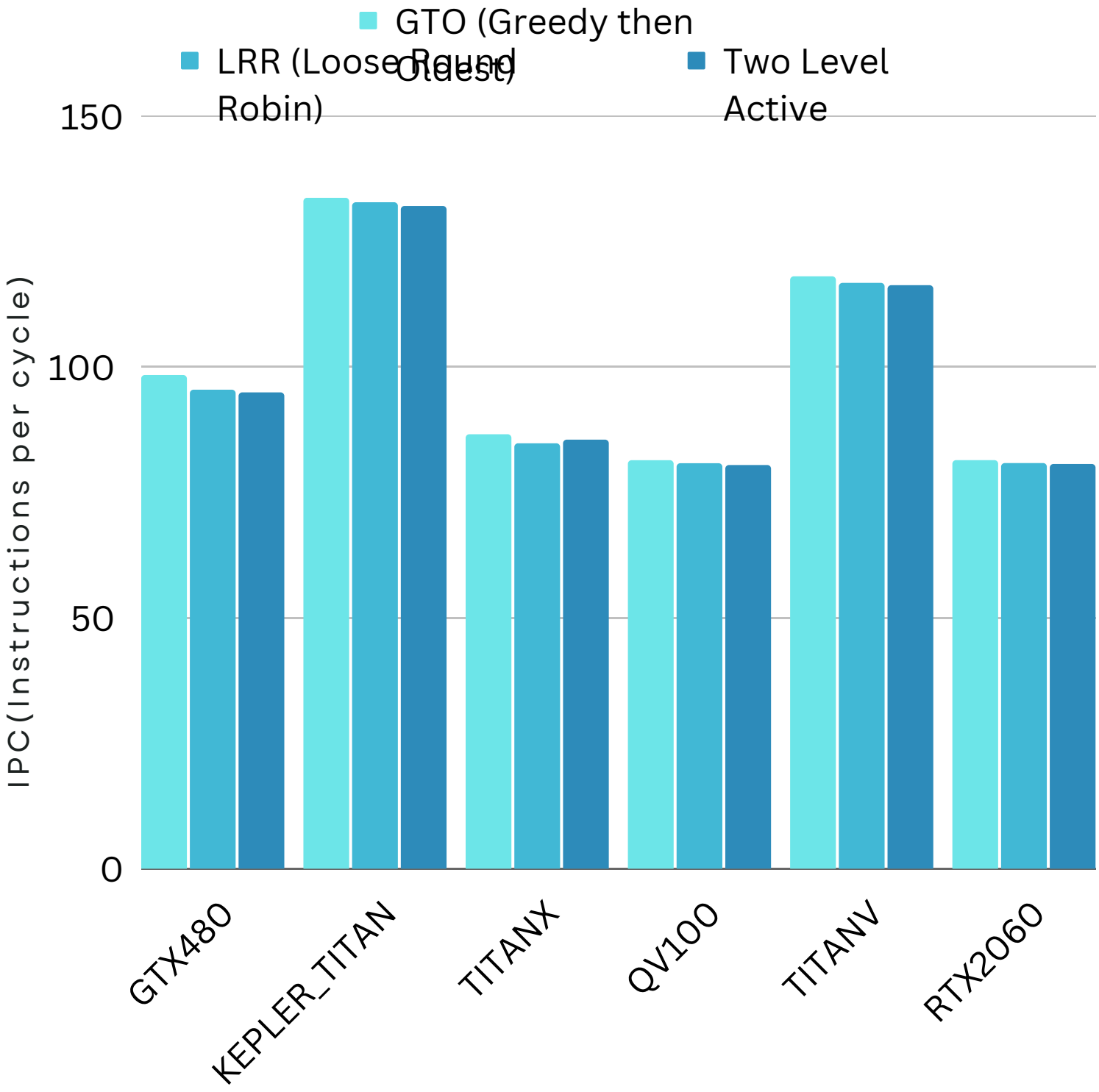
NOTE
Benchmarking application used: **pathfinder**
Inputs: **1000 100 20**

RUNTIMES

Scheduler	GTO (Greedy then Oldest)	LRR (Loose Round Robin)	Two Level Active
GTX480	7 secs	7 secs	7 secs
KEPLER_TITAN	8 secs	8 secs	8 secs
TITANX	13 secs	13 secs	12 secs
QV100	29 secs	30 secs	30 secs
TITANV	19 secs	20 secs	20 secs
RTX2060	14 secs	14 secs	14 secs

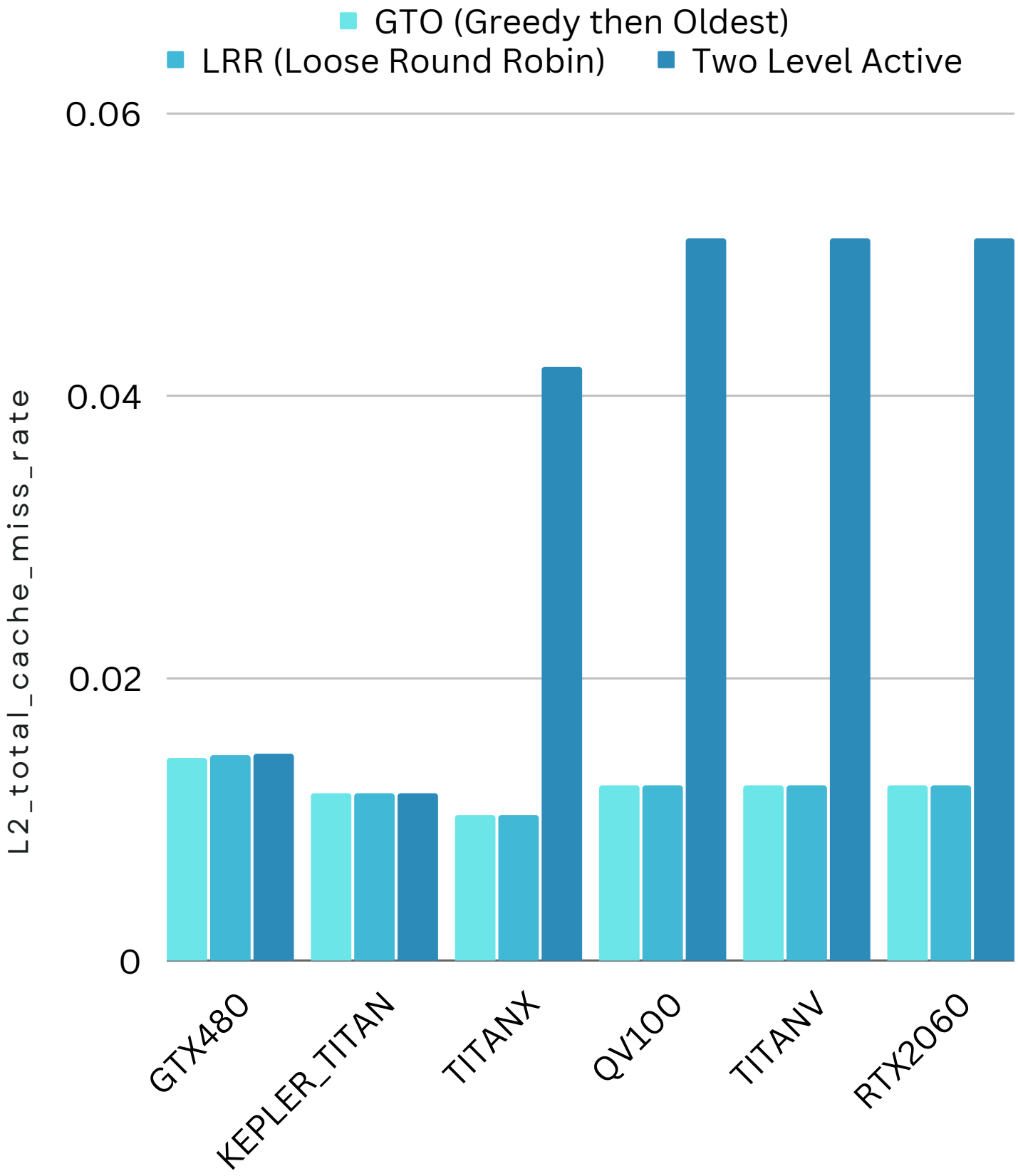
CONCLUSION: Out of these, GTX480 config is the fastest and QV100, the slowest.
All the schedulers seem to perform almost equally.

IPC VS GPU CONFIG PLOTS WITH DIFFERENT WARP SCHEDULERS

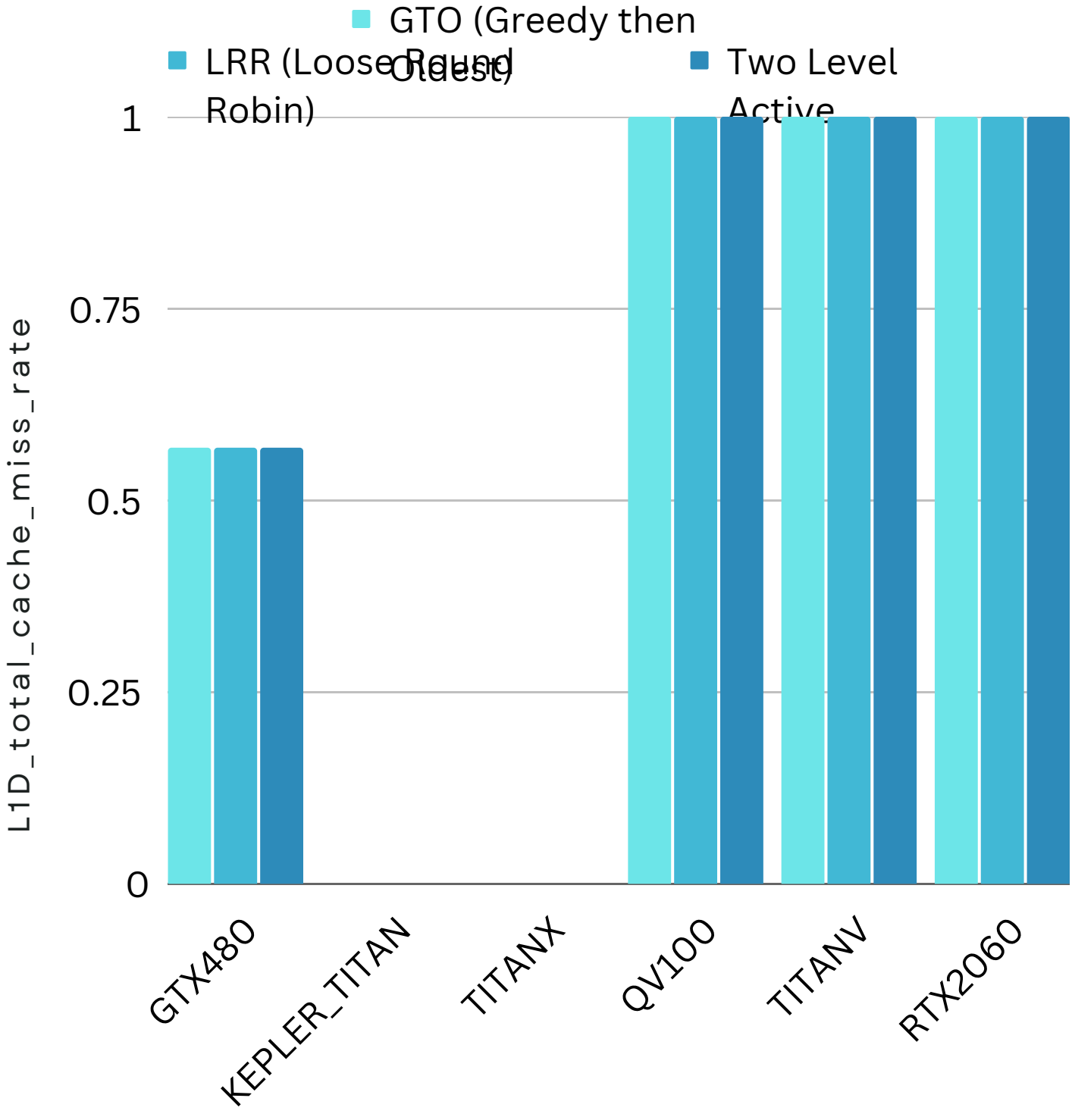


CONCLUSION: KELPER_TITAN with GTO scheduler seems to provide the maximum IPC.

L2_TOTAL_CACHE_MISS_RATE



L1D_TOTAL_CACHE_MISS_RATE



CONCLUSION: GTO scheduler has the highest cache hit rate, hence the lowest cache miss rate because it doesn't re-schedule a warp till its finished due to which data corresponding to the warp remains in the cache leading to higher hit-rates.

CATEGORIZATION OF GPU CONFIGS WRT L1D AND L2 CACHE HIT RATES

For L1D Cache:
With respect to L1D cache hit rates, the KELPER_TITAN and TITANX configs had the best performance with a hit-rate of 1, followed by GTX480 with a hit-rate of 0.432.

For L2 Cache:
With respect to L2 cache hit rates, the KELPER_TITAN config had the best performance with a hit-rate of 0.9882, followed by GTX480 with a hit-rate of 0.9857.

One more curious observation was made for TITANX, QV100, TITANV, RTX2060 configs with the two-level-active scheduler. The miss-rate for these are notably higher as compared to other schedulers

HIT RATES VS L1D CACHE SIZE

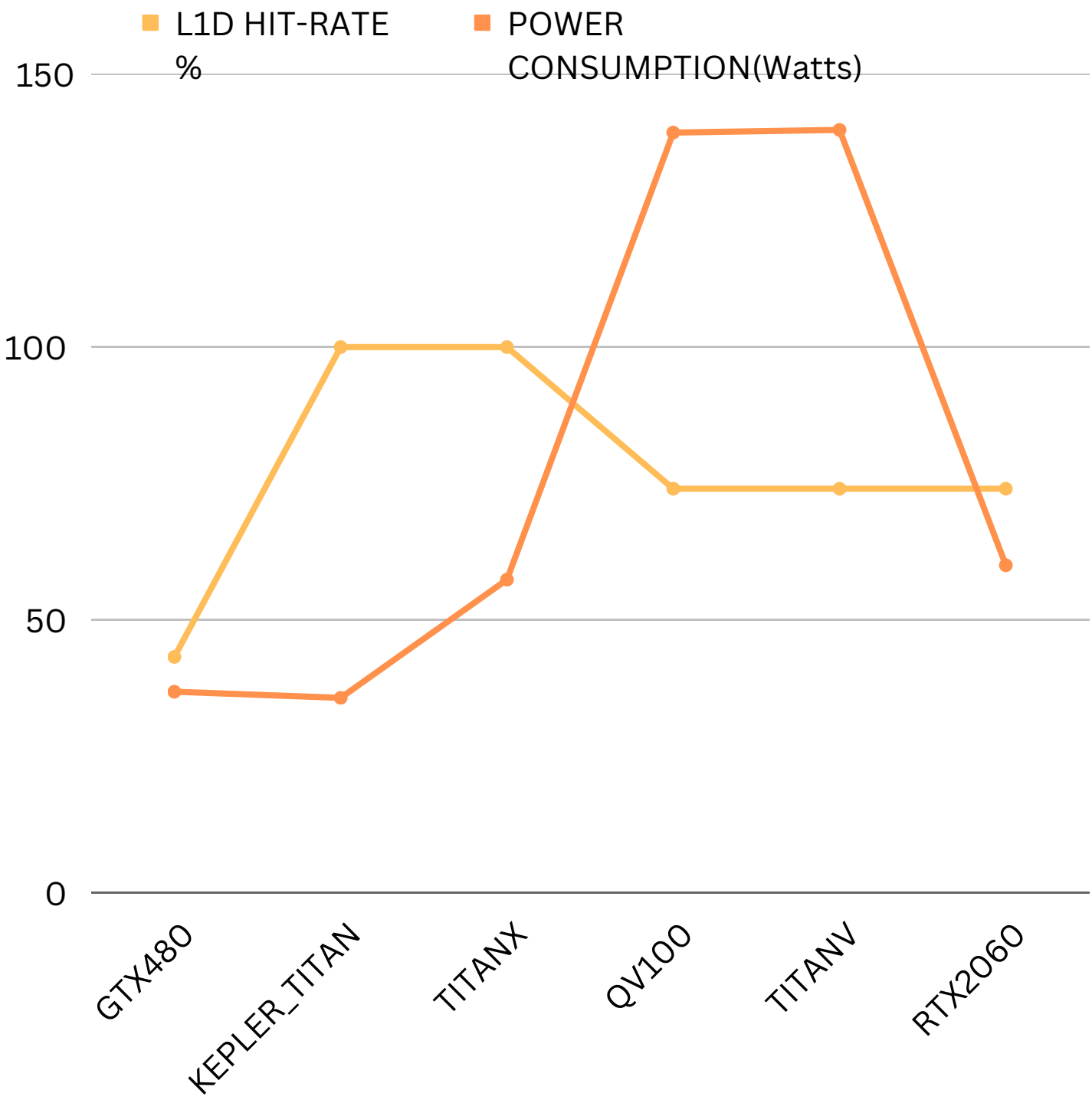
	L1D		L2	
	32 KB	8 MB	32 KB	8 MB
GTX480	0.1689	0.7852	0.8567	0.5047
KEPLER_TITAN	1	1	0.9882	0.9882
TITANX	1	1	0.9897	0.9897
QV100	0.5204	0.5204	1	1
TITANV	0.5182	0.5182	0.9785	0.9785
RTX2060	0.4265	0.5886	1	1

Conclusion:
Increasing the L1D cache size generally improves the L1D cache hit rate, leading to better performance by reducing cache misses. The impact on L2 cache hit rates varies among different GPU architectures, with some GPUs maintaining high L2 cache hit rates even with larger L1D caches.

POWER ANALYSIS

Configuration	Warp Scheduler	Execution Units Avg Power	DRAM Avg Power	Register Files Avg Power	Total Avg Power	% Execution	% DRAM	% Register Files
SM2_GTX480	gto	23.84169181	0.00173819	14.28387	38.1273	62.53181266	0.004558911856	22.84256507
	lrr	23.95666134	0.00166866	13.80647	37.7648	63.43648408	0.004418559081	21.76424214
	two_level	23.75106181	0.00173819	14.1196	37.8724	62.71337916	0.004589595589	22.51449402
SM3_KEPLER_TITAN	gto	26.5822	0	11.2415	37.8237	70.27921647	0	15.99548282
	lrr	17.27633	0	20.45497	37.7313	45.78779422	0	44.67341209
	two_level	17.29839	0	20.44701	37.7454	45.82913415	0	44.61574582
SM6_TITANX	gto	36.88296	0	20.49314	57.3761	64.28279371	0	31.87966611
	lrr	45.51155	0	11.70325	57.2148	79.54506526	0	14.71272914
	two_level	45.58968	0	11.67702	57.2667	79.60940651	0	14.66788978
SM7_QV100	gto	127.48635	0	11.82165	139.308	91.5140193	0	12.91785684
	lrr	128.06606	0	11.18494	139.251	91.96778479	0	12.16180212
	two_level	128.10281	0	11.09719	139.2	92.02788075	0	12.05850869
SM7_TITANV	gto	123.73838	0	15.97062	139.709	88.56865342	0	18.03191014
	lrr	123.69409	0	15.96291	139.657	88.56991773	0	18.02294776
	two_level	123.31502	0	16.33798	139.653	88.30101752	0	18.50259539
SM75_RTX2060	gto	48.75084	0	11.24356	59.9944	81.25898417	0	13.83669771
	lrr	48.75395	0	11.16325	59.9172	81.36887238	0	13.71931265
	two_level	48.8833	0	11.0211	59.9044	81.60218615	0	13.50588816

CONTINUED...



CONCLUSION: There seems to be a negative correlation between L1D cache hit rates and power consumption. This might be because of higher number of accesses from the L1D cache which consumes less power as compared to higher level caches since its located closer to the processing units.

THANK YOU!