

Baseliner: GPU benches

Baseliner :

- Reproducible benchmarking
- Cross-architecture
- Config-driven
- Unified framework
- Accurate & repeatable measurement

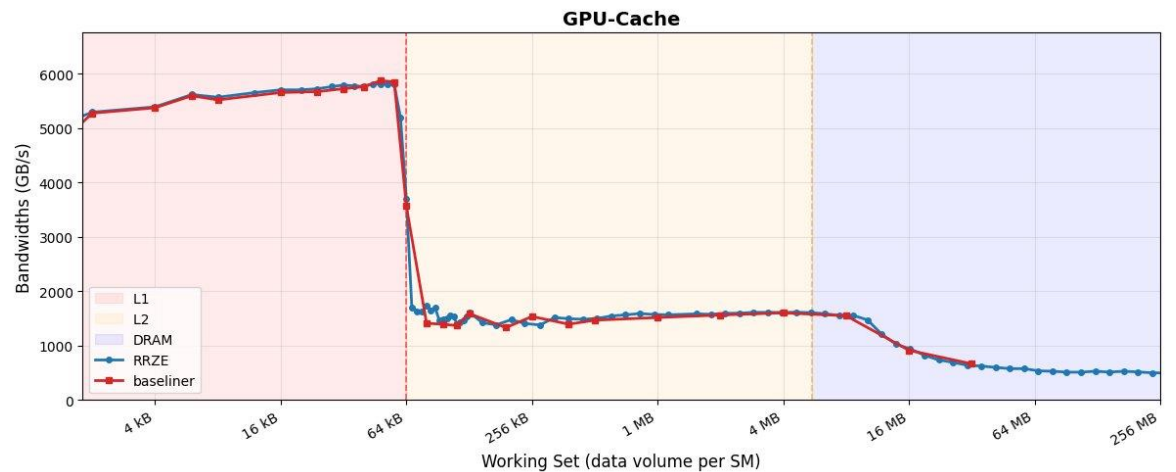
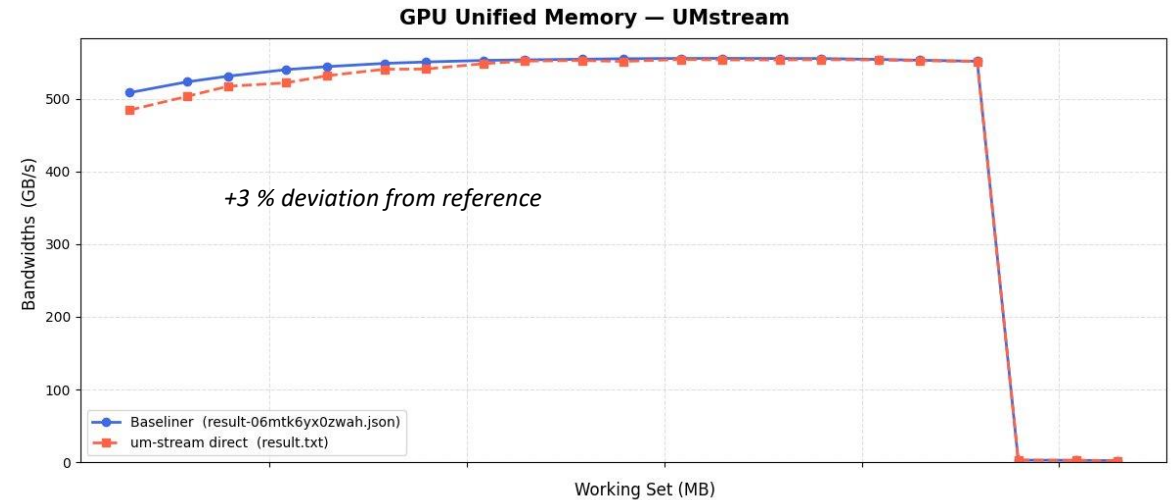
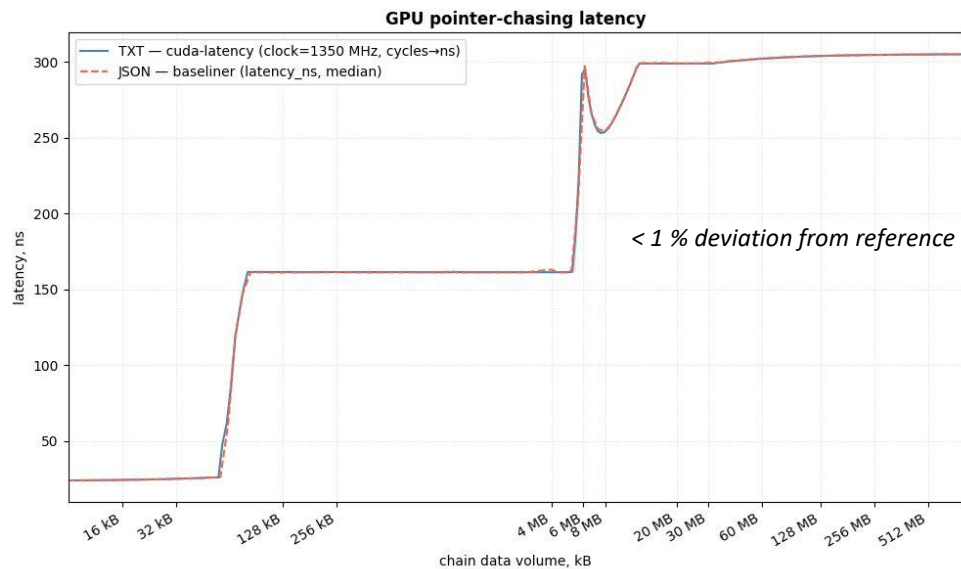
My part :

- Port the RRZE* benchmarks to the Baseliner format
- German academic center of expertise in high-performance computing (HPC)*
- Compare the performance of the ported benchmarks

Port the RRZE benchmarks to the Baseline format 2

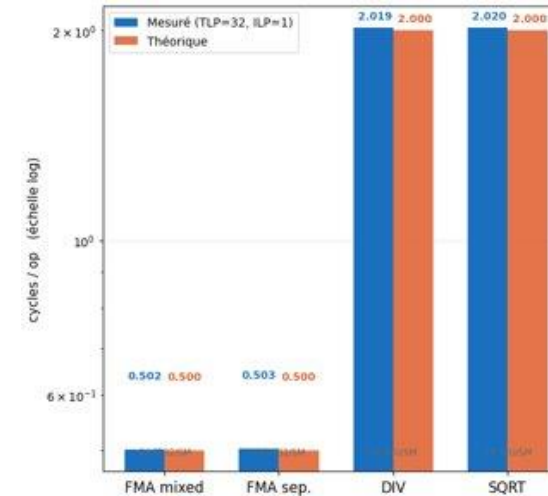
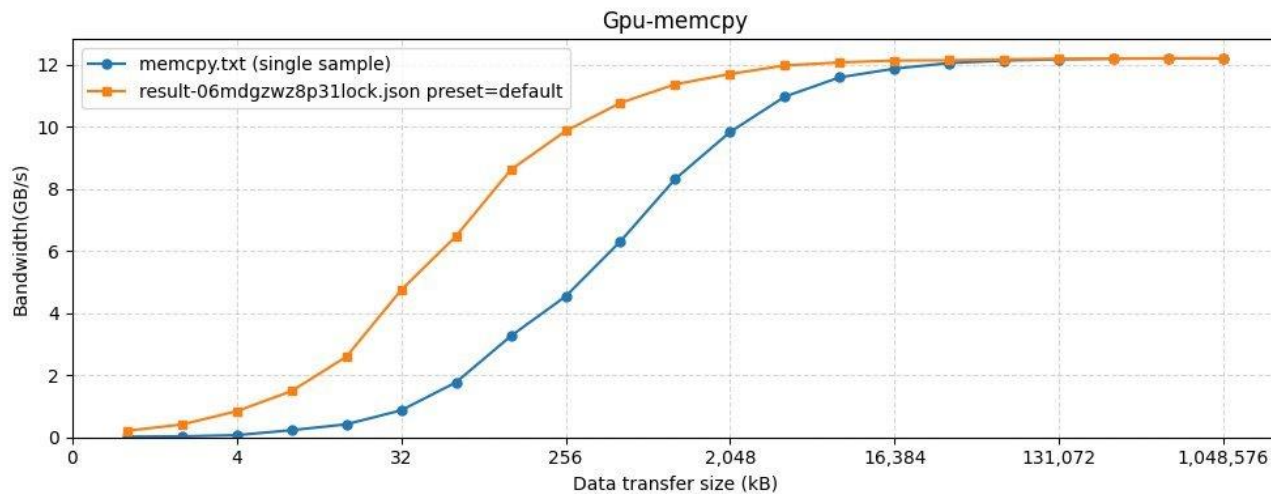
5 distinct benchmarks :

- Memory access latency
- Unified Memory bandwidth
- L1 & L2 cache bandwidth
- Host-to-device transfer rate
- FMA / DIV / SQRT throughput

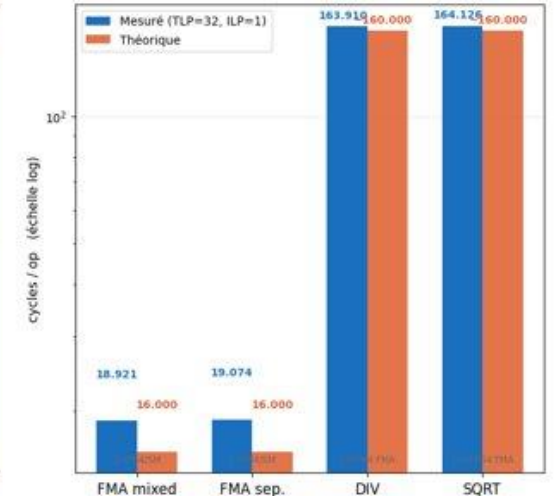


Port the RRZE benchmarks to the Baseline format 3

For small transfers, the RRZE measurement is dominated by launch overhead and is slightly less precise than baseliner, hence the large initial gap for small packets. As the transfer size grows, this fixed cost becomes negligible and both converge to the same bandwidth.



Fma efficiency = 99,6%
Div efficiency = 99,05%
Sqrt efficiency = 99,01%

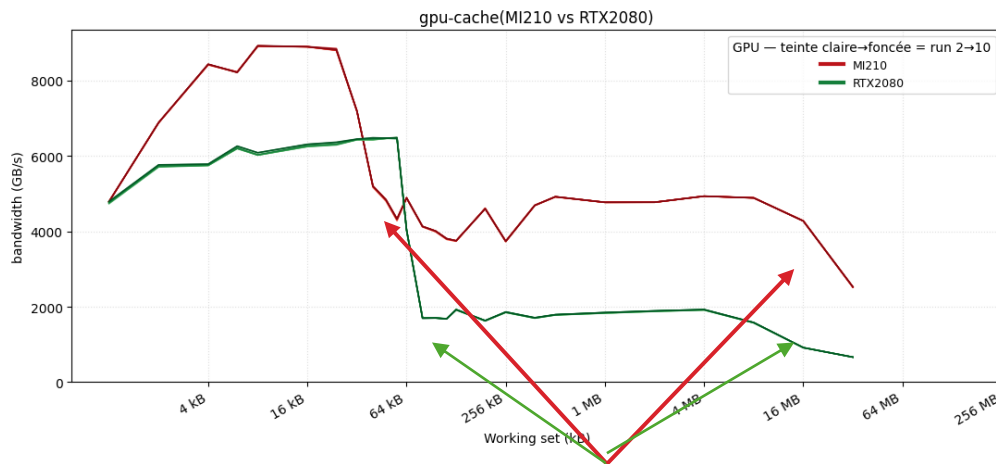


Fma efficiency = 84,5%
Div efficiency = 97,61%
Sqrt efficiency = 97,48%

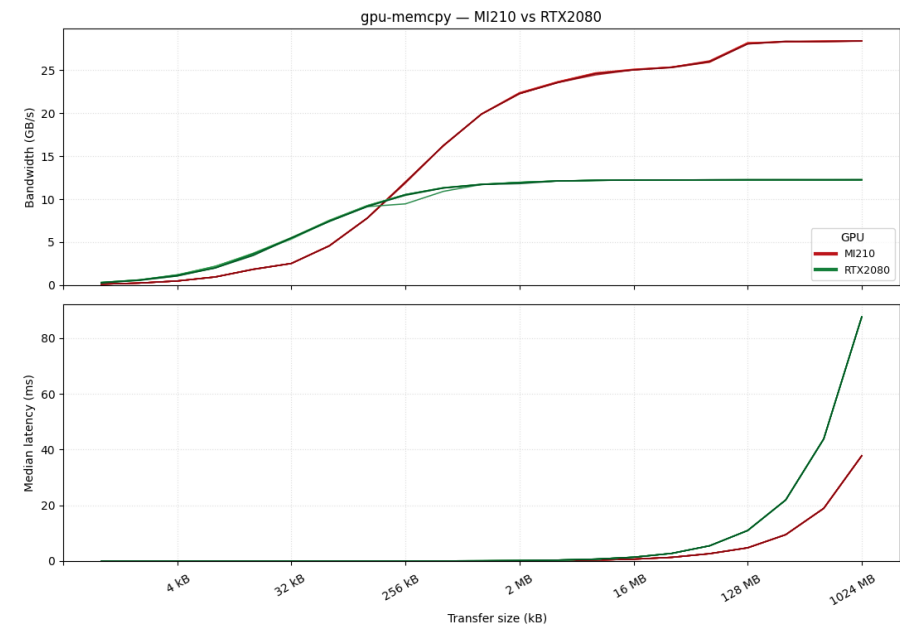
Compare the performance of the ported benchmarks 4

Devices under test :

- **MI210** — AMD, datacenter/HPC, CDNA 2 (2022), HBM2e · *HIP*
- **RTX 2080 Ti** — NVIDIA, public consumer GPU, Turing (2018), GDDR6 · *CUDA*



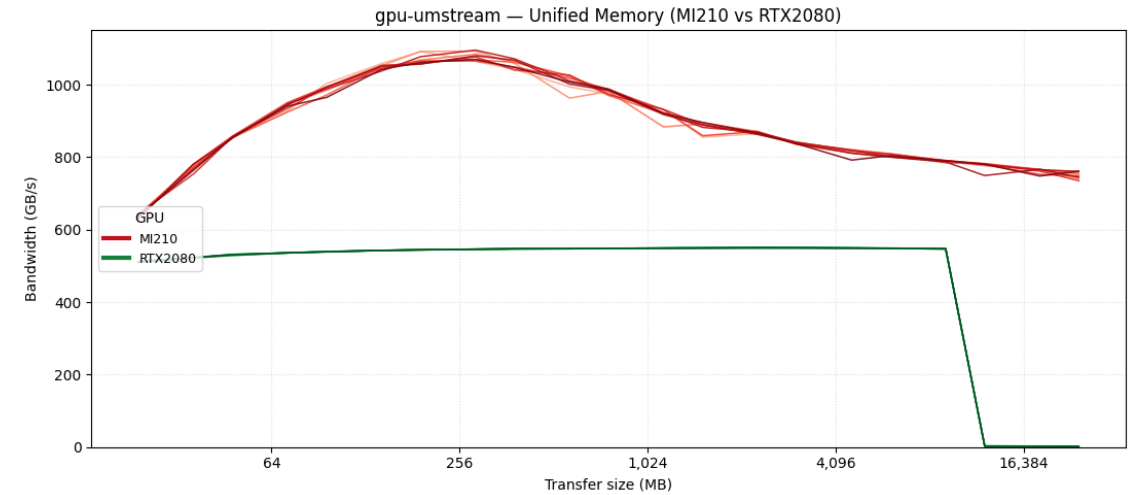
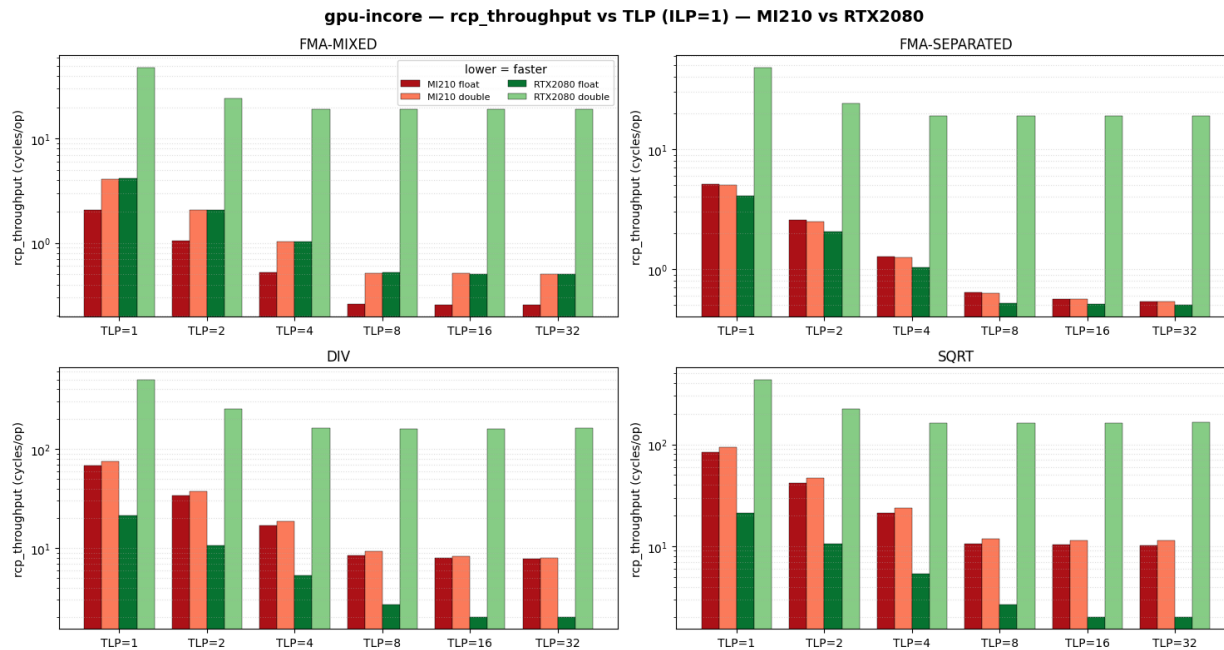
Each drop pinpoints a cache boundary (L1, then L2) — a clear sign the benchmark exposes the cache architecture correctly.



The MI210 is outperforming as attented the RTX2080ti

Compare the performance of the ported benchmarks 5

Not comparable (different architectures), yet consistent with each card's design.



HBM2e (MI210) 64Go vs GDDR6 (RTX) 16Go

Not the same numbers of FP32/64

Next steps

- **More benchmarks** : port the remaining RRZE benchmarks to Baseline.
- **More runs on current hardware** : expand testing on the GPUs already available
- **Porting on Intel GPU**: Once a SYCL / oneAPI backend is available in Baseline,
- **Automated plotting** : generate the comparison graphs automatically from the JSON results.