Benchmarking-Single Core Matrix Multiplication

Virtual Machine (Google-Colab) Specifications :

1. CPU (Processor) and Cache :

Component	Specification	Notes
CPU Model	AMD EPYC 7B12	Clearly states the processor model.
Architecture	x86_64	Standard 64-bit architecture.
Total CPU Cores	2	Your performance (GFLOPS) is limited by these two virtual cores.
Virtualization Type	KVM (full)	Confirms this is a virtualized environment.
L1d / L1i Cache	32 KiB	Relevant for the *Tiny/L1-friendly* case (64 × 64 × 64).
L2 Cache	512 KiB	Relevant for the *Small/L2-friendly* case (256 × 256 × 256).
L3 Cache	16 MiB	Relevant for the *Medium/L3-friendly* case (512 × 512 × 512) and *Large* cases.

2. RAM :

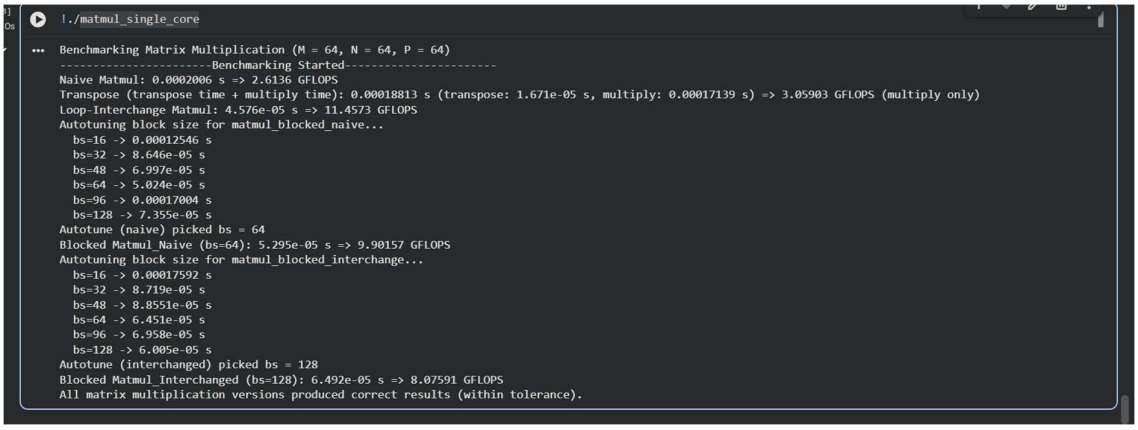
Component	Specification
Total System RAM	13.0 GB (13,286,956 kB)
Free RAM	8.57 GB (8,988,776 kB)
Swap Total	0 kb

3. Operating System(OS) :

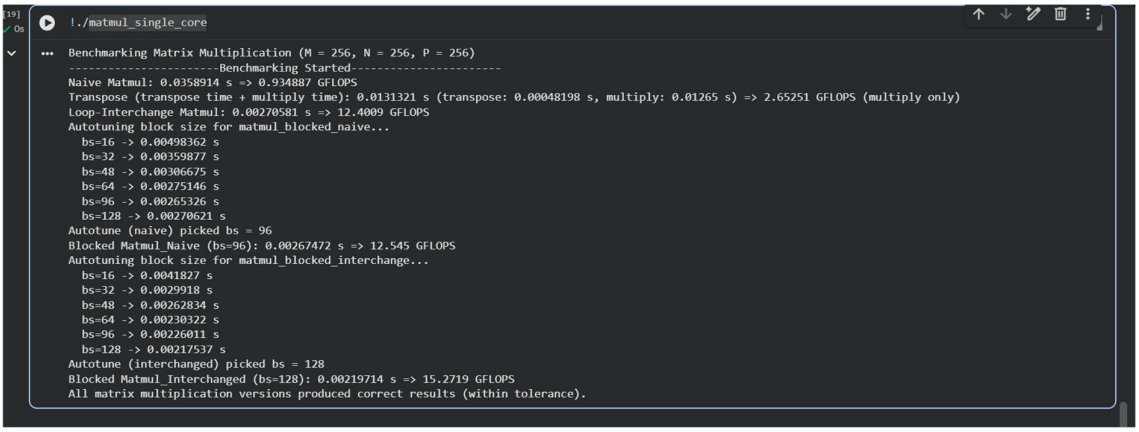
Component	Specification
Operating System (OS)	Ubuntu 22.04.4 LTS
Code Name	Jammy Jellyfish

Benchmarking - Different Cases :

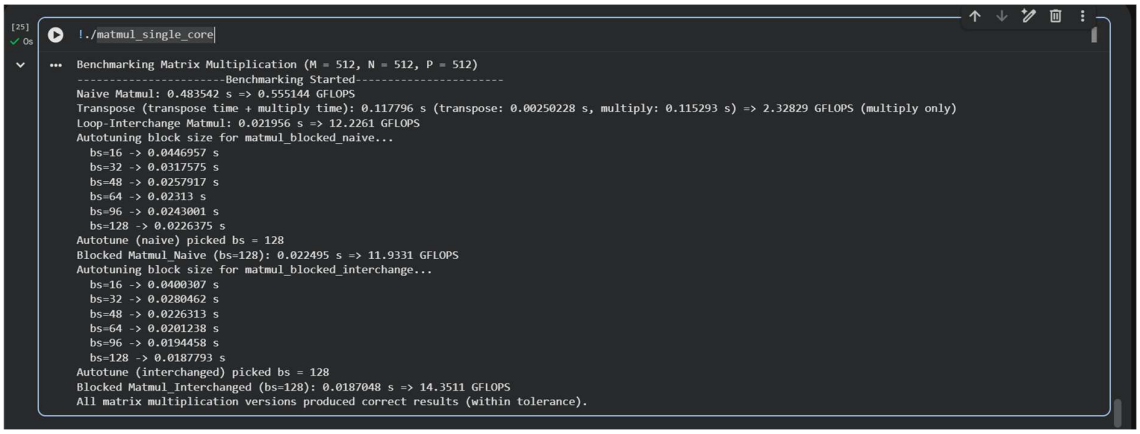
1. **Tiny / L1-friendly** — 64 × 64 × 64
   Purpose: sanity check, correctness, fastest small-run behavior (fits in L1).



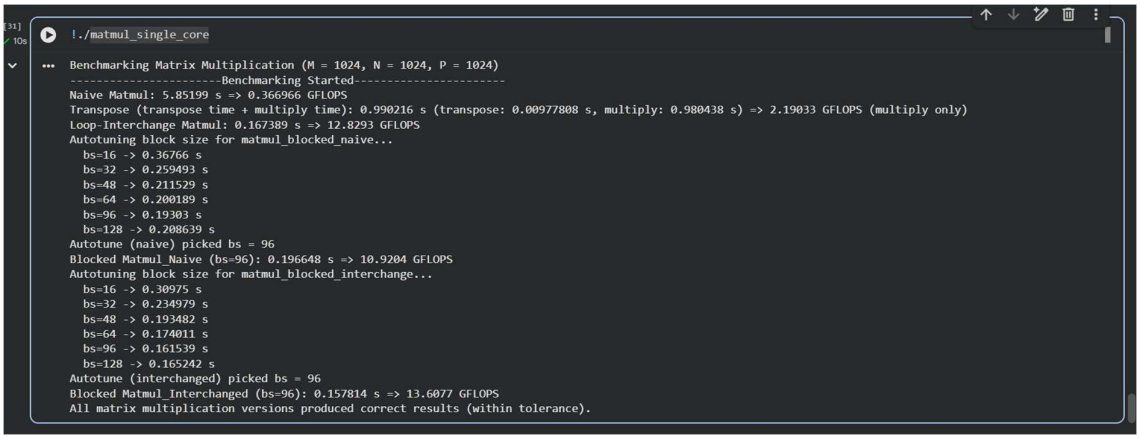
2. **Small / L2-friendly** — 256 × 256 × 256
   Purpose: checks cache-working-set beyond L1 (tests loop ordering benefits).



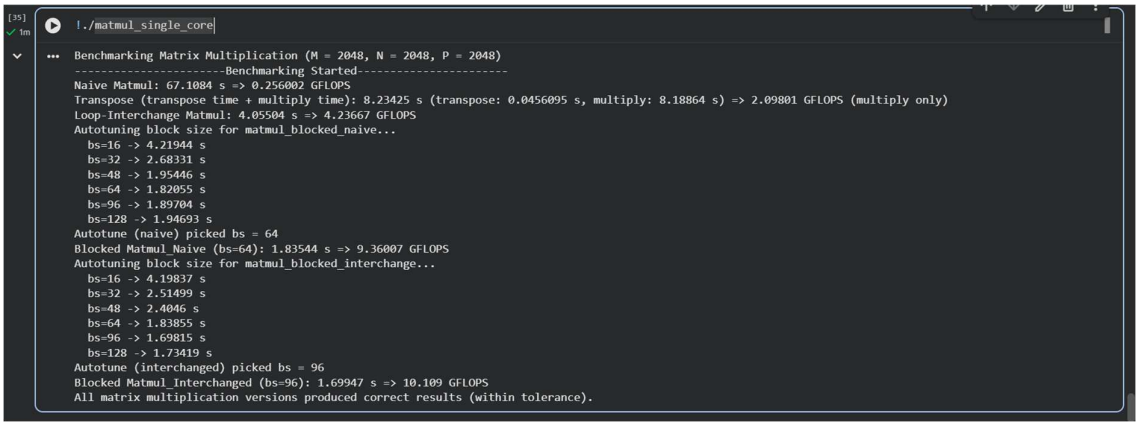
3. **Medium / L3-friendly** — 512 × 512 × 512
   Purpose: typical in-memory compute case; shows blocked gains.



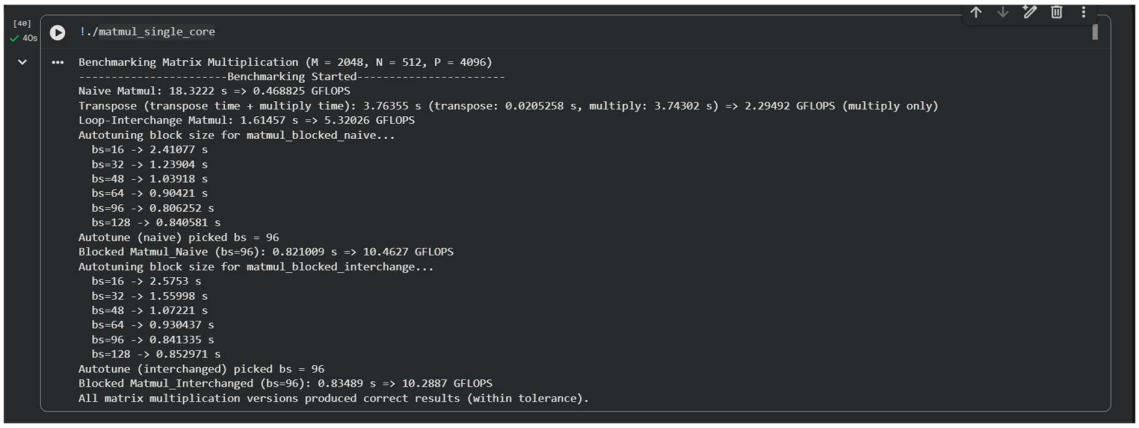
4. **Large / Memory-pressure** — 1024 × 1024 × 1024
   Purpose: approach L3 / main-memory limits on many machines — shows memory-bound behavior.



5. **Very large / stress test** — 2048 × 2048 × 2048
   Purpose: heavy compute & memory; use for final best-known-tweak timings (might take long).



6. **Rectangular / real-world pattern** — 2048 × 512 × 4096 (A: 2048×512, B: 512×4096 → C: 2048×4096)
   Purpose: tests non-square workloads (skinny K, wide N) — exposes different memory/compute balance.



Output Table - 6 Cases-Time

Method ↓ / Case →	64×64 ×64	256×256 ×256	512×512 ×512	1024×1024 ×1024	2048×2048 ×2048	2048×512 ×4096
Naive	200.6 µs	0.0358914 s	0.483542 s	5.85199 s	67.1084 s	18.3222 s
Transpose	188.13 µs	0.0131321 s	0.117796 s	0.990216 s	8.23425 s	3.76355 s
Loop-Interchange	4.576e-05 s	0.00270581 s	0.021956 s	0.167389 s	4.05504 s	1.61457 s
Blocked - Naive	bs=64 5.295e-05 s	bs=96 0.00267472 s	bs=128 0.022495 s	bs=96 0.196648 s	bs=64 1.83544 s	bs=96 0.821009
Blocked – Interchange	bs=128 6.492e-05 s	bs=128 0.00219714 s	bs=128 0.0187048	bs=96 0.157814 s	bs=96 1.69947 s	bs=96 0.83489 s

Output Table - 6 Cases - GFLOP

Method ↓ / Case →	64×64 ×64	256×256 ×256	512×512 ×512	1024×1024 ×1024	2048×2048 ×2048	2048×512 ×4096
Naive	2.6136 GFLOP	0.934887 GFLOP	0.555144 GFLOP	0.366966 GFLOP	0.256002 GFLOP	0.468825 GFLOP
Transpose	3.05903 GFLOP	2.65251 GFLOP	2.32829 GFLOP	0.990216 GFLOP	2.09801 GFLOP	2.29492 GFLOP
Loop-Interchange	11.4573 GFLOP	12.4009 GFLOP	12.2261 GFLOP	12.8293 GFLOP	4.23667 GFLOP	5.32026 GFLOP
Blocked - Naive	9.90157 GFLOP	12.545 GFLOP	11.9331 GFLOP	10.9284 GFLOP	9.36007 GFLOP	10.4627 GFLOP
Blocked – Interchange	8.07591 GFLOP	15.2719 GFLOP	14.3511 GFLOP	13.6077 GFLOP	10.109 GFLOP	10.2887 GFLOP