

Übungsblatt 10, für den 15.1.2026

Aufgaben:

1. **Optimieren** der am Anfang des Semesters implementierten Matrix-matrix Multiplikation.
2. Geschwindigkeits- und Skalierungsvergleich mit erster Version – Details:
 - Zu Messende Größen: 512^2 , 1024^2 , 2048^2 , 2000^2
 - Zu Messende Datentypen: float, double
 - Auf NV und AMD Plattformen

Berechnen von **GFLOP/s** Metrik für jeden dieser Fälle!

Ziel:

Minimum Ziel: $\geq 20\%$ der theoretischen Maximalperformance der jeweiligen HW, für mindestens eine Problemgröße.

Maxima: AMD float: 14.85 TFLOPS; AMD double: 464 GFLOPS
 NV float: 7.46 TFLOPS; NV double: 233 GFLOPS

Hinweise:

- Local memory und/oder Register als Zwischenspeicher für Tiling ist ein guter Ansatz. Auch umstrukturieren der Daten ist denkbar, muss aber in die Messungen einbezogen werden.
- Tuning der work group dimensions, buffer sizes, und anderer Parameter ist sicher zielführend.
- Online-recherche ist absolut kein Problem, aber jede spezialisierte Optimierung muss im Detail erklärt werden können.
- Alle Zeitmessungen sind mit OpenCL events durchzuführen.
- Zu Messen: Zeit von Inputdaten am GPU -> fertiger Output am GPU.
D.h. Anfang des Ersten involvierten Kernels bis Ende des Letzten.
(mehr als 1 Kernel könnte verwendet werden, z.B. für Preprocessing)

Abgabe:

- Per Email an peter.thoman@uibk.ac.at
Betreff: “[gpu-computing-2025] [UE10] NACHNAME1, NACHNAME2, NACHNAME3”
Vor (!) VU-Beginn
- 1 **Abgabe pro Gruppe** – Gruppenmitglieder in der Email auflisten

Format:

Archiv (.tar, .tar.gz, .zip, ...) mit einem Folder.

Folder enthält source + makefile.

Letzteres muss out of the box auf ifi-cluster.uibk.ac.at funktionieren.

Messdaten, schriftliche Antworten etc. als .txt, .md, .pdf und/oder .csv.