



Proseminar „Rechnerarchitektur“

Thema:

Leistungsanalyse von GPU-beschleunigten Programmen

Grafikbeschleuniger sind ein fester Bestandteil im Hochleistungsrechnen. Die Anwendungsbereiche umfassen u.a. die Beschleunigung von komplexen wissenschaftlichen Simulationen und des maschinelle Lernens. Um GPUs möglichst effizient zu nutzen, müssen entsprechende Anwendungen als Basis für eine systematische Optimierung zuerst analysiert werden.

Inhaltliche Schwerpunkte und Hinweise:

- Grundlegende Analysemethoden und -techniken (top-down, bottom-up, Sampling, Instrumentierung, Profile, Traces)
- vorhandene Werkzeuge und Schnittstellen für die Analyse von GPU-Steuerungsschnittstellen und GPU-Aktivitäten
- Beispiel(e) für typischen Performance-Probleme in GPU-Anwendungen und ihre Erkennung
- Möglichkeiten zur Analyse von Deep-Learning-Anwendungen

Literatur:

- Keren Zhou and John Mellor-Crummey. 2019. A tool for performance analysis of GPU-accelerated applications. In Proceedings of the 2019 IEEE/ACM International Symposium on Code Generation and Optimization (CGO 2019). IEEE Press, 282. online: <https://dl.acm.org/doi/10.5555/3314872.3314914>
- Allen D. Malony, Scott Biersdorff, Sameer Shende, Heike Jagode, Stanimire Tomov, Guido Juckeland, Robert Dietrich, Duncan Poole, and Christopher Lamb. 2011. Parallel Performance Measurement of Heterogeneous Parallel Systems with GPUs. In Proceedings of the 2011 International Conference on Parallel Processing (ICPP '11). IEEE Computer Society, USA, 176–185. DOI:<https://doi.org/10.1109/ICPP.2011.71>
- NVIDIA, CUPTI Documentation, online: https://docs.nvidia.com/cupti/Cupti/r_overview.html

Betreuender Hochschullehrer:

Prof. Dr. Wolfgang E. Nagel

Betreuer:

Dr.-Ing. Dietrich, Robert

robert.dietrich@tu-dresden.de

Raum: WIL A106