Universität Regensburg

Wissenschaftliches Rechnen auf Grafikkarten

Übungsblatt

Thomas Karl

20. Januar 2020

Inhaltsverzeichnis

1 Bandbreitentest

Kopieren Sie ein einzelnes Array der Größe n in Bytes in den globalen Speicher. Messen Sie die Zeit t. Plotten Sie für verschiedene n die Zeiten und fitten sie linear die Funktion $t(n) = \frac{1}{b}n + l$ an. Dabei ist b die Bandbreite von PCIe (unidirektional) und l die Latenz.

2 Axpy

Parallelisieren Sie für zwei Vektoren $\vec{x}$ und $\vec{y}$ die Zuweisung

$$\vec{y} \leftarrow a \cdot \vec{x} + \vec{y} \tag{1}$$

auf der Grafikkarte. Schreiben Sie drei verschiedene Kernel für single-, double- und half-precision. Variieren Sie die Dimension der Vektoren und protokollieren Sie die Laufzeit der drei Versionen. Bestimmen Sie durch lineares Fitten den Speedup von half- und float- gegenüber double-precision.

3 Matrixaddition auf Cluster

Parallelisieren Sie eine einfache Addition zweier großer $n \times n$ Matrizen $C = A + B$ unter Verwendung mehrdimensionaler Blöcke. Benutzen Sie das Beispiel *cluster.cu* um das Problem

auf verschiedene GPUs aufzuteilen. Überprüfen Sie das Gesamtresultat.

4 Bubble Sort

Implementieren Sie den *Bubble Sort Algorithmus*[1] sequentiell. Teilen Sie die innere Schleife auf: Die erste behandelt nur jedes zweite Element, die zweite die restlichen. Parallelisieren Sie diese nun unabhängigen Schleifen.

5 Concurrency

Messen Sie die Bandbreite wie in Augabe 1 für den bidirektionalen Fall. Kopieren Sie ein Array der Größe s auf die Grafikkarte. Benutzen Sie zwei asynchrone Kopierfunktionen um gleichzeitig dieses Array zu lesen und ein anderes der selben Größe zu schreiben. Dazu müssen diese Funktionen einem anderen Stream zugeordnet werden. Messen Sie die Zeit t für beide Operationen und fitten Sie wie in Aufgabe 1 mit $n = 2s$.

6 Streams

mehrere Additionen

Führen Sie mehrere Vektoradditionen auf dem selben Device aus. Maximieren Sie den Durchsatz, indem Sie jede Addition einem anderen Stream zuordnen und damit Concurrency ausnutzen.

eine Addition

Führen Sie eine Vektoradditionen auf dem selben Device aus. Maximieren Sie den Durchsatz, indem Sie die Arrays auf gleich große Subarrays aufteilen und jede Addition einem anderen Stream zuordnen. Führen Sie ihre Anwendung mit *Nvidia nsight* aus und überprüfen Sie mit dem Profiler den Überlapp.

[1] https://de.wikipedia.org/wiki/Bubblesort