Übung 5
Computational Physics III

Matthias Plock (552335) Paul Ledwon (561764)

14. Juni 2018

Inhaltsverzeichnis

1 Conjugate-Gradient auf der GPU

Die Funktionen aus den letzen Uebungen wurden so in das Grundgeruest von Uebung5.zip eingefuegt, dass nun der Conjugate-Gradient-Algorithmus parallelisiert auf der GPU ausgefuehrt werden kann. Fuer verschiedene Gittergroessen N wurden die Laufzeiten bei verschiedenen execution-configurations und 10 Durchlaufen ermittelt. Aus dem Durchschnitt der Laufzeiten für ein festes N wurde mit der schnellsten execution-configuration der Speedup berechnet. Die Ergebnisse sind in Abb. 1 dargestellt. Mit steigender Gittergroesse steigt auch der Speedup, auch wenn dieser erst ab $N = 512$ den Wert 1 uebersteigt.

Es wird vermutet, dass die GPU-Implementierung nicht optimal ist, da bei der Benutzung von nvprof auffiel, dass ein Grossteil der Laufzeit durch die Ausführung von den Unroll-Funktionen bei der Berechnungen der Skalarprodukte verursacht wird. Um ein Skalarprodukt zu berechnen, wird die Unroll-Funktion zweimal ausgefuehrt. Moeglicherweise waere es effizienter, nur eine Unroll-Funktion auszufuehren und dann das vergleichsweise kleine Array auf dem Host zu berechnen.

2 Amdahlsches Gesetz

Das Amdahlsche Gesetz besagt, dass fuer den Speedup $S_p(N) = \frac{T_s(N)}{T_p(N)}$ gilt

$$S_p(N) \leq \frac{1}{f}, \tag{1}$$

wobei f der Anteil des Problems ist, der seriell ausgefuehrt werden muss.

2.1 Beweis

Wir teilen die Laufzeit der seriellen Loesung des Problems auf in

$$T_s(N) = t_s + t_p^s = f \cdot T_s(N) + (1 - f) \cdot T_s(N). \tag{2}$$

Analog gilt dann fuer die Laufzeit des parallelen Problems

$$T_p(N) = t_s + t_p^p = f \cdot T_s(N) + t_p^p. \tag{3}$$

Hierbei ist t_s die Laufzeit, die benoetigt wird um den seriellen Anteil des Problems zu loesen. t_p^s und t_p^p sind die Laufzeiten, die das beste serielle bzw. parallelisierte Programm zur Loesung des parallelisierbaren Anteils des Problems benoetigen. Theoretisch kann t_p^p auf die Dauer einer Operation reduziert werden, fuer den Fall, dass man eine ausreichende Menge an Threads zur Verfuegung hat.

Setzt man (2) und (3) in die Definition des Speedups ein erhaelt man

$$\begin{aligned} S_p(N) &= \frac{f \cdot T_s(N) + (1-f) \cdot T_s(N)}{f \cdot T_s(N) + t_p^p} \leq \frac{f \cdot T_s(N) + (1-f) \cdot T_s(N)}{f \cdot T_s(N)} \\ &= \frac{1 + \frac{1-f}{f}}{1} = \frac{f + 1 - f}{f} = \frac{1}{f} \end{aligned}$$