# **Protokoll 1: Serielle Optimierung der Matrix Multiplikation auf CPUs**

Gruppe 9: Marcel Beyer, Martin Carnein, Franz Lehmann

## Aufgabe 1:

Berechnung der Peak Performance mit folgender Formel:

Laut Intel Webseite besitzt der *Intel Xeon Platinum 8470 Prozessor* zwei AVX-512 FMA-Einheiten, sodass sich daraus 64 Instruktionen pro Zyklus ergeben. Die Basisfrequenz pro Kern ist 2.00 GHz.

Damit ergibts sich folgende Lösung für einen Kern:

## Aufgabe 2:

* Schleifenvertauschung
  + man hat mehrer Schleifen miteinander verschachtelt bspw. eine for-Schleife innerhalb einer anderen, um zum Beispiel durch eine Matrix zu iterieren
  + je nachdem wie man durch die Matrix iteriert, erst durch alle Spalten oder durch alle Zeilen, kann es vorkommen, dass Zugriffe schlechter oder besser sind
  + so wäre es besser, wenn man so iteriert, dass man eine ganze Cache Zeile komplett durchläuft und nicht immer nur ein Element aus einer Cache Zeile nimmt und dann die Cache Zeile verdrängt, weil man das Element aus der nächsten Cache Zeile braucht
* Loop unrolling – Schleifen entrollen
  + wenn man davon ausgeht, dass man eine Schleife hat, in der man auf ein Array zugreift und mit diesem Berechnungen ausführt, wird für jede Berechnung ein Sprung und ein Vergleich benötigt => wenig Arbeit, viele Befehle
  + mit Loop Unrolling würde man in einem Schritt auf mehrere Stellen im Array zugreifen und darauf Berechnungen ausführen und dann in der Schleife direkt mehrere Schritte weiter gehen, somit müssen weniger Befehle für die gleiche Arbeit dekodiert werden (es wird aber auch mehr Befehlscache gebraucht)
* Blocking/Tiling
  + Bei großen Daten(mengen) kann es vorkommen, dass die Daten nicht alle auf einmal in den Cache passen und werden nur einmal genutzt bevor die verdrängt werden
  + Daten werden in Blöcke aufgeteilt und die Berechnungen werden auf den kleineren Blöcken ausgeführt, welche komplett in den Cache passen, sodass Elemente wiedergenutzt werden können
* Wiedernutzung von Werten statt Neuberechnung
  + Wenn Werte im Programm mehrfach berechnet werden, ist es sinnvoll diese bei der ersten Berechnung zwischenzuspeichern
  + damit braucht man zwar ein wenig mehr Speicher jedoch kann man sich damit Rechenleistung und ggf. auch Zeit sparen

## Aufgabe 3:

## Aufgabe 4:

## Aufgabe 5:

## Aufgabe 6:

## Aufgabe 7:

## Aufgabe 8: