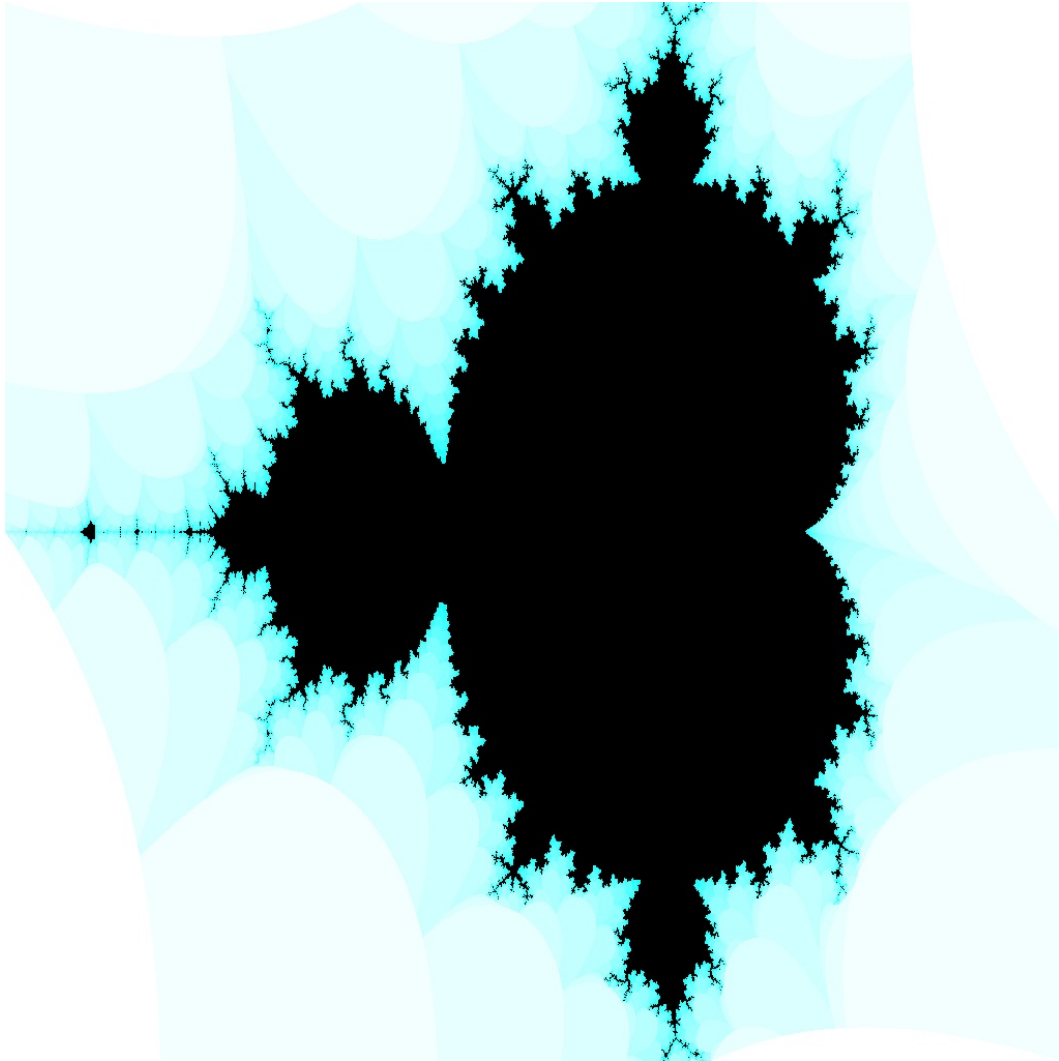


LEHRSTUHL FÜR RECHNERTECHNIK UND RECHNERORGANISATION
**Aspekte der systemnahen Programmierung
bei der Spieleentwicklung**

Projektaufgabe - 1: Mandelbrotmenge

Robin Ostner
Oliver Jung
Florian Sprang



Inhaltsverzeichnis

1	Problemstellung und Spezifikation	3
2	Lösungsalternativen	4
3	Dokumentation der Implementierung	5
3.1	Implementierung des C-Programms	5
3.2	Implementierung Single-Pixel-Calculation	6
3.3	Implementierung Mult-Pixel-Calculation	7
3.3.1	Konzept	7
3.3.2	Konkrete Umsetzung	8
3.4	Anwendung des Programmes	10
4	Ergebnisse / Fazit	13
4.1	Ergebnisse und Optimierung	13
4.2	Fazit	14

1 Problemstellung und Spezifikation

2 Lösungsalternativen

3 Dokumentation der Implementierung

3.1 Implementierung des C-Programms

Bevor wir zu der Umsetzung der Mandelbrotmenge kommen, wird zuerst die Ein/Ausgabe und Laufzeitanalyse in unserem C-Programm genauer erläutert. Hierfür werden zuerst die einzelnen Methode genauer analysiert und anschließend wird der Ablauf in der Main-Methode spezifiziert. Da wir zwei Assemblercodes für die Mandelbrotmenge implementiert haben, stehen die externen Verlinkungen zu diesen ganz oben. In diesen werden die durch die Angabe spezifizierten Übergabeparamter von xStart, xEnd, yStart, yEnd und Auflösung bei beiden Aufrufen eingehalten. Lediglich die Methodennamen weichen von der Vorgabe ab, aufgrund der unterschiedlichen Implementierungen.

Die Methode createBMP ist dafür verantwortlich unser zuvor berechnetes Array an Pixeln in eine Bitmapdatei zu speichern. Hier ist zu beachten, wie das Dateiformat genauer spezifiziert ist¹. Zunächst wird ein Fileheader erstellt und initialisiert, anschließend erstellen wir den Infoheader. Da in diesem die Dateigröße spezifiziert werden muss, berechnen wir die Größe des Bildes und addieren anschließend in einer separaten Variable noch die Größe der beiden Header dazu. Dies geschieht in zwei verschiedenen Variablen, da später die Größe des Bildes noch benötigt wird. Da wir immer nur ein Byte auf einmal schreiben wollen, legen wir unsere Header so an, dass wir den Integerwert der Bildgröße in vier Byte speichern. Das bedeutet, dass wir viermal die Größe schreiben müssen, jedoch immer geschiftet, damit die Bytes in der richtigen Reihenfolge im Header stehen. Das gleiche geschieht bei Höhe, Breite und Bildgröße, welche im Infoheader geschrieben werden. Anschließend benutzen wir die Standardmethoden von C um eine Datei zu öffnen, die beiden Header zu schreiben und anschließend das Bild anzuhängen. Wichtig ist hier den Dateinamen mit .bmp zu beenden, damit das Bild später auch als Bild interpretiert werden kann.

In der nächsten Methode secondes erzeugen wir einen Zeitstempel. Dieser wird als double für genauere Berechnungen zurück gegeben. Um den Zeitstempel zu erzeugen verwenden bereits vorimplementierte C Methoden, mit welchen wir eine Genauigkeit im Bereich von Nanosekunden erhalten.

Um sicherzustellen, dass wir keine falschen Übergaben in unserem Programm verwenden, wurde eine Methode dafür entwickelt um Eingaben zu überprüfen. Da ein Übergabeparameter zunächst ein String ist, versuchen wir bereits in diesem Fehler zu finden. Hierfür wird die Länge des Strings bestimmt und anschließend jeder einzelne Character darauf überprüft ob er eine Zahl ist. Da wir führende Nullen nicht zulassen, überprüfen wir den String auf solche.

Nach der Überprüfung auf die korrekte Anzahl an Übergabeparametern, werden zu Beginn in der Main-Methode fünf Variablen für die Übergabeparameter erstellt und mit Null initialisiert. Anschließend erzeugen wir sechs Zeitstempel und setzen zunächst den ersten. Vier der Zeiten dienen dazu die Gesamtlufzeit zu messen und die Laufzeit für das Berechnen der Menge selber. Bevor wir die zuvor erstellten Variablen für die Übergabeparamter mittels einer C-Funktion in Integer umwandeln, werden diese mit der

¹https://en.wikipedia.org/wiki/BMP_file_format

oben beschriebenen Funktion auf ihre Gültigkeit überprüft. Sollte hierbei ein Fehler auftreten, so beendet sich das Programm und gibt eine Fehlermeldung aus. Anschließend werden die Übergabeparameter noch auf ihre Gültigkeit gemäß unserer nachfolgenden Anforderung hin überprüft. Hier gilt, dass ein Startwert immer kleiner sein muss als sein korrespondierender Endwert und die Auflösung nicht negativ oder folgende Definition nicht verletzen darf: $Auflösung = 2^n$ für $n > 2$. Wurden alle Tests bis hierhin bestanden, wird der Benutzer dazu aufgefordert eine der beiden Berechnungsarten zu wählen. Die verbrauchte Zeit bei der Eingabeaufforderung wird gemessen und am Ende von der Gesamtlaufzeit abgezogen, damit das Ergebnis nicht verfälscht wird. Anschließend erstellen wir das Array und übergeben es an die gewünschte Umsetzung der Mandelbrotmenge. Vor der Übergabe wird ein Zeitstempel für den Beginn der Berechnung gesetzt und nach beenden dieser eine zweite Zeit gemessen. Im Anschluss wird die Laufzeit des ausgewählten Algorithmus auf der Konsole ausgegeben. Jetzt wird das berechnete Bildarray nur noch an die Speicherfunktion übergeben, wo es wie oben erwähnt abgespeichert wird. Bevor wir das Programm verlassen, wird die Gesamtlaufzeit noch berechnet und ebenfalls auf der Konsole ausgegeben.

3.2 Implementierung Single-Pixel-Calculation

3.3 Implementierung Mult-Pixel-Calculation

3.3.1 Konzept

Unser zweites Konzept basiert auf dem Prinzip eine dauerhafte Berechnung aufrecht zu erhalten. Wie bereits zuvor erwähnt fällt ein Pixel unter zwei Bedingungen aus der Berechnung heraus. Erstens wenn von einem Pixel eine maximale Anzahl an Schleifendurchläufen mitgemacht wurden, oder zweitens wenn gilt:

$$(\text{Imaginärteil} + \text{Realteil})^2 > \text{Schwellwert}$$

Diese dauerhafte Berechnung wird nur dann unterbrochen wenn ein Pixel eines seiner Abbruchkriterien erfüllt. Tritt dieser Fall ein besitzt jeder Pixel eine eigene Funktion. Diese speichert das Ergebnis des Pixels in das Bildarray und tauscht mittels der festgelegten Register für Konstanten die notwendigen neuen Werte (Positionswerte, Speicheradresse) mit den alten aus. Außerdem wird auch noch sein Laufwert zurückgesetzt, da wir diesen benutzen um den Grad der Konvergenz zu ermitteln und die korrespondierenden Pixel blau zu färben. Des Weiteren werden alle Werte für den neuen Pixel berechnet und gespeichert. Hierbei war die Schwierigkeit einen Zeilenumbruch in den Pixelreihen zu erkennen und die dazugehörigen Laufwerte zu aktualisieren. Mittels Abbildung 1 soll das gerade beschriebene Konzept noch einmal vereinfacht dargestellt werden. Im weiteren Verlauf wird nun auf die konkrete Implementierung unseres Konzeptes eingegangen.

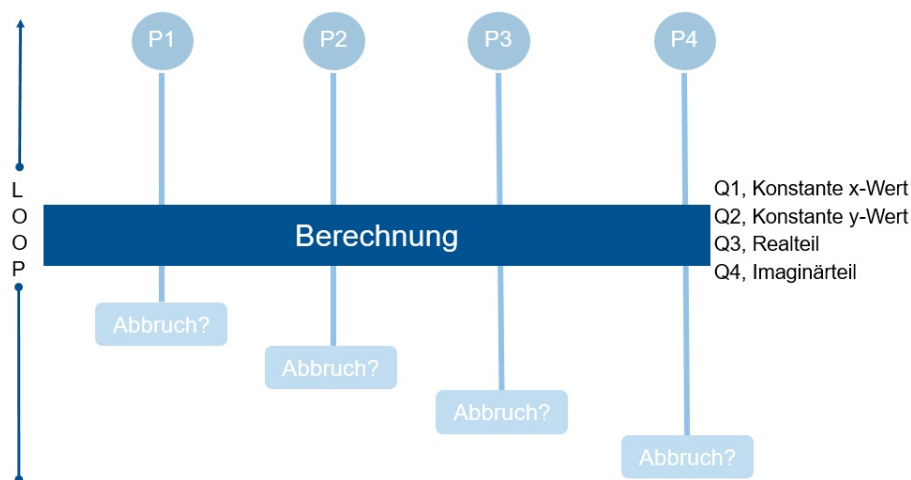


Abbildung 1: Grafische Darstellung des Berechnungsprinzips:
Durchlaufen der Berechnungsschleife von Pixel P1-P4,
nach jeder Iteration werden die Pixel sequentiell auf Abbruch überprüft
Bei einem Abbruch werden sie mit dem nächsten Pixel ausgetauscht

3.3.2 Konkrete Umsetzung

Bei unserer Umsetzung der Mandelbrotmenge benutzen wir fast alle Ressourcen, die uns der Prozessor zu Verfügung stellt. Konkret bedeutet das, dass wir alle zulässigen allgemeinen Register benötigen. Deshalb werden alle Register auf dem Stack gesichert. Das Link-Register wird auch abgesichert, damit ein Rücksprungsmöglichkeit in das C-Programm erhalten bleibt. Da wir sechs Übergabewerte haben und nur die ersten vier automatisch in R0-R3 geladen werden, werden die letzten beiden Werte vom Stack ausgelesen. Der Anfang der Multi-Pixel-Calculation ist größtenteils identisch mit dem der Single-Pixel-Calculation. Zu Beginn berechnen wir pro Achse die Strecke zwischen den Start- und Endwerten. Mittels der Auflösung wird anschließend der Unterschied zwischen zwei Pixeln in der FPU berechnet. Hierbei ist zu beachten, dass wir von der Auflösung zuvor eine Kopie erstellt welche um den Wert eins verringert wurde, da ansonsten immer das letzte Pixel einer Reihe nicht auf dem Endpunkt liegt sondern einen Abstand davor. Das Register q0 wurde anschließend als Register für Variablen festgelegt, welche während der dauerhaften Berechnung benötigt werden. Die freien allgemein Register werden nun für die Hauptberechnung initialisiert. Da zu Beginn festgelegt wurde, dass die Größe des Bildes immer $Auflösung^2$ ist, wird in Register zwei diese gespeichert. Register zwei ist also der Schleifenlaufwert, welcher verwendet wird um festzustellen, wann alle Pixel berechnet sind. Die weiteren Register drei bis sechs enthalten die Anzahl der bereits durchlaufenen Iterationen. Die Maximale Anzahl an erlaubten Iterationen wird in dem darauffolgendem Register R7 gespeichert. Somit ist es später möglich mit diesen Registern das bereits oben erwähnte Abbruchkriterium der maximalen Iterationsdurchläufe zu realisieren. Die Register R8-R11 werden nun mit der für den Pixel entsprechenden Speicheradresse initialisiert. Hierbei ist zu beachten, dass R0 immer den Wert enthält für den nächsten Pixel nicht für den letzten. Es werden vier separate Speicheradressen deshalb verwendet, da während der Berechnung die Pixel nicht immer gleich lang brauchen und somit Pixel4 vor Pixel1 fertig sein kann. Würde man eine einzelne Speicheradresse dafür verwenden, würde die Reihenfolge der berechneten Pixel im Bildarray nicht mehr stimmen. Um das Problem zu lösen müsste Pixel4 bei schnellerer Berechnungszeit auf die Berechnung von Pixel1 warten, was viel Zeit kosten kann. Um die Vorbereitungen für die Hauptberechnung abzuschließen werden die beiden float-Register s30 und s31 initialisiert. In s30 steht der aktuelle x-Wert, also der Wert der zuletzt verwendet wurde und in s31 wurde der Schwellwert geschrieben. Die Tabelle 1 und Tabelle 2 zeigen eine Übersicht über die allgemein und floating-point Register während der Berechnung.

Damit die Hauptschleife zu Beginn keinen Leerdurchlauf durchführt, werden vor Beginn der Schleife die entsprechenden Register (siehe: Abbildung 1) mit den ersten vier Pixelkoordinaten befüllt. Damit es zu keinen Fehlern in der maximalen Anzahl an zu berechnenden Pixeln kommt, werden bereits nach der Berechnung von R2 diese vier Pixel abgezogen. Demnach stehen jetzt in q1 und q2 die x-Werte bzw. y-Werte. Somit bilden q1, q2 unser Konstante. Da die erste Iteration der Schleife als Ergebniss die Konstante selber ist kann man diese überspringen, indem man in die Register q3 (Realteil von z_i) und q4 (Imaginärteil von z_i) direkt in die Konstante kopiert. Das Überspringen

der ersten Iteration wird auch bei einem Pixelwechsel umgesetzt. Dadurch lässt sich pro Pixel eine Iteration einsparen, was sich später positiv auf die Laufzeit auswirkt.

Zu diesem Zeitpunkt wird nun die Schleife gestartet. Um die binomische Formel von $(z_i)^2 = (a * i + b)^2 = b^2 - a^2 + 2 * a * b$ zu realisieren wird q5 als Zwischenspeicher von $2 * a * b$ verwendet. Wir berechnen $2 * a * b$ zuerst, da wir dadurch nur ein Register als Zwischenspeicher benötigen. Somit können die Register die Real- und Imaginärteil beinhalten anschließend quadriert werden. Gemäß der Formel werden nun die neuen Real- und Imaginärteile berechnet. Damit ist die eigentliche Berechnung von einer Iteration zu Ende.

Der weitere Teil an parallelen Schritten innerhalb von NEON dient dazu das zweite Abbruchkriterium zu testen. In q5 werden nun Real und Imaginärteil aufsummiert und anschließend das Quadrat der Summe in q6 berechnet. Hier wird das Quadrat verwendet, da somit nicht auf Divergenz von negative Zahlen getestet werden muss. Der Schwellwert wurde zuvor so ausgewählt, das dieser bereits das Quadratische des eigentlichen Schwellwertes ist. Im weiteren Verlauf erfolgen nun die sequentiellen Überprüfungen des Pixels gemäß der Abbruchbedingungen. Da die Überprüfung des einzelnen Pixel vom Aufbau her gleich ist, wird hier nur auf die Überprüfung eines Pixels eingegangen. Die sequentielle Überprüfung aller vier Pixel ist allerdings notwendig, da in den gespeicherten Variablen in r0, q0 und q7 immer nur die Koordinaten für das nächste Pixel stehen. Als erstes erfolgt die Überprüfung auf den Schwellwert, da dieser Fall häufiger eintritt und sich somit erneut ein paar Rechenoperationen einsparen lassen. Ist diese Überprüfung erfolgreich, springen wir in die Funktion `earlybailout` des jeweiligen Pixels. In dieser wird aufgrund der in R8-R11 gespeicherten Adressen die ersten beiden Bytes des Pixels mit dem *maximalenFarbwert* = 255, für weiß initialisiert. Um nun den Grad der Divergenz zu erhalten wenden wir folgende Formel an:

$$\text{zuspeichernderFarbwert} = \text{maxFarbwert} - (12 * \text{Anzahl der Schleifendurchläufe})$$

Die Zwölf ergibt sich aus der Ganzzahldivision des maximalen Farbwerts durch die maximale Anzahl an Schleifendurchläufen. Das Ergebnis wird anschließend an die Stelle des dritten Byte gespeichert. Sollte jedoch die andere Abbruchbedingung eintreten, auf welche im Anschluss geprüft wird, so werden an die gespeicherte Speicheradresse der minimale Farbwert null gespeichert, was in einen schwarzen Pixel resultiert. Dies geschieht in der `maxBailout` Funktion des jeweiligen Pixels.

Wenn ein Pixel nun fertig berechnet und seine Werte in das Bildarray gespeichert wurden, muss er mit dem nächsten zu berechnenden Pixel ausgetauscht werden. Das Austauschen erfolgt in der jeweiligen `Reset` Funktion. In dieser wird zunächst der Laufwert auf -1 gesetzt. Der Laufwert wird nicht auf null gesetzt, da wir in der Hauptschleife erst nach allen Überprüfungen die Laufwerte inkrementieren und somit bei der ersten Iteration des neuen Pixels mit dem Laufwert 0 beginnt. Der nächste Schritt ist die Überprüfung auf einen Zeilenum sprung. Da uns die Länge einer Zeile bekannt ist, passiert dann ein Zeilenumbruch wenn gilt:

$$\text{restlicheAnzahl an Pixeln} \% \text{Auflösung} = 0$$

Da die vorhandene ARM-Architektur keine passende Modulo Operation bietet und auch keine Division von Ganzzahlen unterstützt wird die standardmässig vorimplementierte Funktion des gcc-Compilers für zwei unsigned Integer Werte verwendet. Mittels der Multiply and Subtract Instruktion (MLS) kann man jetzt den Rest einer Division berechnen. Ist das oben genannte Kriterium erfüllt, so findet ein Zeilenumbruch statt. Bei diesem werden der aktuelle y-Wert um einen Schritt erhöht und der aktuelle x-Wert auf $StartX - xSchritt$ gesetzt. Das Setzen des x-Wertes ergibt sich daraus, das wir im weitem Vorgehen des Initialisierens des neuen Pixels, den x-Wert standardmäßig um einen Schritt erhöhen. So ist gewährleistet, dass der neue Pixel bei xStart beginnt. Am Schluss wird nun die neue Konstante gespeichert und aus den bereits vorher erwähnten Gründen die Konstante auch in Real- und Imaginärteil kopiert. Auch der neue zu speichernde Adressbereich wird aktualisiert.

Sobald eines der beiden Abbruchkriterien eintritt wird auf das nächste Kriterium hin nicht mehr überprüft. Sonst kann der noch nicht berechnete neue Pixel direkt gespeichert werden. Nachdem alle Pixel überprüft wurden, erhöhen wir deren Laufwerte um eins und verifizieren ob wir noch weitere Berechnungen durchführen müssen. Der gerade geschilderte Ablauf passiert solange bis wir alle Pixel berechnet haben. Wurden alle benötigten Berechnungen durchgeführt so stellen wir den Ausgangszustand wieder her und springen zurück in das C-Programm.

3.4 Anwendung des Programmes

Um jetzt die bereits detailliert beschriebene Implementierung der Mandelbrotmenge zu verwenden, muss man das Programm mittels des Makefile compilieren. Der Programmaufruf erfolgt dann mittels der Konsole mit folgenden Übergabeparametern:

xStart: Integerwert der den Anfangswert auf der x-Achse repräsentiert
xEnd: Integerwert der den Endwert auf der x-Achse repräsentiert
yStart: Integerwert der den Anfangswert auf der y-Achse repräsentiert
yEnd: Integerwert der den Endwert auf der y-Achse repräsentiert
Auflösung: Integerwert für den gelten muss $Auflösung = 2^n$ für $n > 2$

Ein konkreter Aufruf wäre `./main -2 1 -1 1 1024`. Für die jeweiligen Start- bzw Endwerte muss gelten, dass $Start < End$. Wird bei der Auflösung für n ein Wert größer als 12 gewählt, so wird das Bild berechnet jedoch steigt die Zeit, die dafür benötigt wird, sehr schnell an. Während man das Programm laufen lässt wird man aufgefordert sich zwischen der Single-Pixel-Calculation oder der parallelisierte Version zu entscheiden. Am Ende erhält man ein Datei des .bmp Formates. Diese kann man sich jetzt mit dem bevorzugten Programm für Bilder anschauen. Sollten für das Programm fehlerhafte Aufrufe durchgeführt werden oder während der Ausführung so beendet sich das Programm und muss neu aufgerufen werden.

R0	Speicheradresse zum nächsten Pixel
R1	Auflösung
R2	Anzahl der Restlichen zu berechnenden Pixel
R3	Durchlaufene Iterationen Pixel 1
R4	Durchlaufene Iterationen Pixel 2
R5	Durchlaufene Iterationen Pixel 3
R6	Durchlaufene Iterationen Pixel 4
R7	Maximale Anzahl an erlaubten Iterationen
R8	Speicheradresse von Pixel 1
R9	Speicheradresse von Pixel 2
R10	Speicheradresse von Pixel 3
R11	Speicheradresse von Pixel 4
R12	allzweck Register zum berechnen oder zur Farbwertspeicherung

Tabelle 1: Einteilung der allzweck Register zur Berechnungszeit

q0	s0	Der Startwert auf der x-Achse
	s1	Abstand zwischen zwei Pixeln entlang der x-Achse
	s2	Aktuelle y-Position
	s3	Abstand zwischen zwei Pixeln entlang der y-Achse
q1	s4	Pixel 1 x-Position / Konstante Realteil
	s5	Pixel 2 x-Position / Konstante Realteil
	s6	Pixel 3 x-Position / Konstante Realteil
	s7	Pixel 4 x-Position / Konstante Realteil
q2	s8	Pixel 1 y-Position / Konstante Realteil
	s9	Pixel 2 y-Position / Konstante Realteil
	s10	Pixel 3 y-Position / Konstante Realteil
	s11	Pixel 4 y-Position / Konstante Realteil
q3	s12	Pixel 1 z_i Realteil
	s13	Pixel 2 z_i Realteil
	s14	Pixel 3 z_i Realteil
	s15	Pixel 4 z_i Realteil
q4	s16	Pixel 1 z_i Imaginarteil
	s17	Pixel 2 z_i Imaginarteil
	s18	Pixel 3 z_i Imaginarteil
	s19	Pixel 4 z_i Imaginarteil
q5	s20	Berechnung Pixel 1: $2 * a * b$ / Realteil+Imaginärteil
	s21	Berechnung Pixel 2: $2 * a * b$ / Realteil+Imaginärteil
	s22	Berechnung Pixel 3: $2 * a * b$ / Realteil+Imaginärteil
	s23	Berechnung Pixel 4: $2 * a * b$ / Realteil+Imaginärteil
q6	s24	Pixel 1: $(\text{Imaginärteil} + \text{Realteil})^2$ / Abbruchtest gegen Schwellwert
	s25	Pixel 2: $(\text{Imaginärteil} + \text{Realteil})^2$ / Abbruchtest gegen Schwellwert
	s26	Pixel 3: $(\text{Imaginärteil} + \text{Realteil})^2$ / Abbruchtest gegen Schwellwert
	s27	Pixel 4: $(\text{Imaginärteil} + \text{Realteil})^2$ / Abbruchtest gegen Schwellwert
q7	s28	unbenutzt
	s29	unbenutzt
	s30	Aktuelle x-Position
	s31	Schwellwert

Tabelle 2: Einteilung der NEON/FPU-Register zur Berechnungszeit

4 Ergebnisse / Fazit

4.1 Ergebnisse und Optimierung

Nachdem wir unsere parallele Implementierung fertiggestellt hatten, beginnen wir damit unsere beiden Programme zu vergleichen. Mit der implementierten Laufzeitanalyse wurden für alle zulässigen Auflösungen zwischen 8 und 8192 die Laufzeit von beiden Berechnungsarten zehn mal gemessen. Anschließend wurde der Mittelwert aus den zehn erhaltenen Zeiten gebildet und miteinander verglichen. In Abbildung 2 wurde die Laufzeit der Single-Pixel-Calculation durch die der parallelen Berechnung geteilt um somit einen Faktor zu erhalten um wie viel schneller die Multi-Pixel-Calculation ist. Zunächst nahmen wir an, dass die einfache Berechnung in den niedrigen Auflösungsbereichen schneller als die parallele sei. Jedoch zeigt sich auch in diesen Bereich, dass die parallele Berechnung schneller ist. Allerdings erwartet man, dass sie circa drei mal so schnell sei, da anstelle von einem Pixel vier Pixel auf einmal bearbeitet werden. Diese Diskrepanz lässt sich dadurch erklären, dass die NEON-Einheit für die parallele Berechnung zusätzliche Schritte unternehmen muss um die Berechnungen parallel durchzuführen. Die durchschnittliche Halbierung der Laufzeit ist in etwa das, was am Ende zu erwarten war.

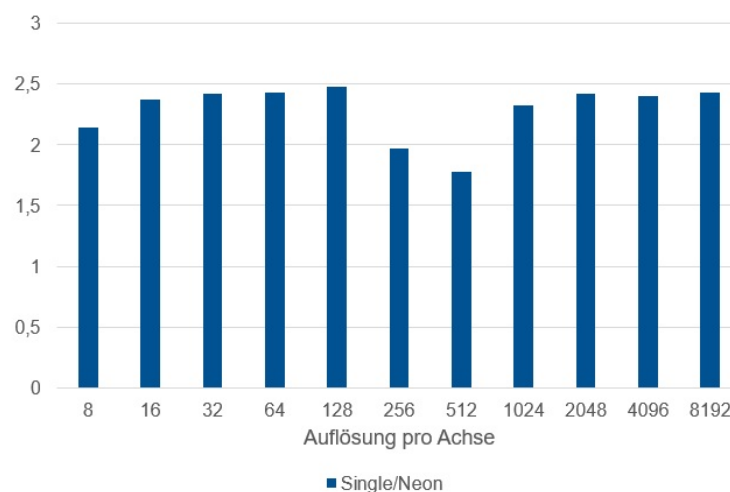


Abbildung 2: Vergleich von paralleler Berechnung zur einfachen Berechnung

Da wir bei unserer Implementierung eher den Ansatz verfolgt haben die Laufzeit zu minimieren sind wir nicht Speichereffizient, was die Anzahl an Register angeht. So finden sich in dieser Hinsicht einige Möglichkeiten der Optimierung. Als erster Ansatzpunkt wäre das Vereinigen der Pixelabbruchfunktionen. Hier ist es sicherlich möglich auf Kosten von Laufzeit eine einheitliche Funktion zu schreiben. Außerdem könnte man auch Speicheradressen auf den Stack auslagern und diese nur bei Bedarf laden und wieder absichern. In Bezug auf Speichereffizienz im Hauptspeicher, benutzen wir

das Minimum was unserer Meinung nach möglich ist, in Bezug auf die Speicherung einer bitmap Datei im Arbeitsspeicher. An dieser Stelle haben wir die Aufgabenstellung so interpretiert, dass die Ausgabefunktion nur das tatsächliche Array ohne weitere Berechnungen schreibt. Ist dies nicht der Fall lässt sich der Speicher auf $1/3$ des ursprünglich benötigten Speichers verkleinern. Da nun nur noch gespeichert werden muss ob das Pixel konvergiert oder nicht. Beim Abspeichern des Bildes in C müsste man dann nur noch die Pixelfarben entsprechend dem Eintrag berechnen und die drei Bytes in die Bitmapdatei schreiben.

4.2 Fazit