

LEHRSTUHL FÜR RECHNERARCHITEKTUR UND PARALLELE SYSTEME
Praktikum RechnerarchitekturParallelle Berechnung der Mandelbrotmenge
Wintersemester 2018/19

Maximilian Frühauf

Tobias Klasuen

Florian Lercher

Niels Mündler

Inhaltsverzeichnis

1 Einleitung	3
1.1 Didaktische Ziele	3
1.2 Verwendung der Mandelbrotmenge	3
1.3 Darstellung der Mandelbrotmenge	4
1.4 MPI	5
1.5 Qualitätsanforderungen	5
1.6 Einschränkungen	5
2 Problemstellung und Motivation	6
3 Dokumentation der Implementierung	7
3.1 Übersicht	7
3.2 Installation der Anwendung	7
3.2.1 Lokales Backend	8
3.2.2 Backend auf HimMUC Cluster	8
3.2.3 Installation des Frontends	11
3.3 Erläuterung des Backends	12
3.4 Implementierung der Mandelbrotberechnung	12
3.4.1 Inkludierte Header und CMake Anweisungen	13
3.4.2 Mainfunktion und Initialisierung	13
3.5 Host Funktionalitäten	14
3.5.1 WebSocketverbindung	14
3.6 Lastbalancierung	19
3.6.1 Naive Strategie	19
3.6.2 Strategie mit Vorhersage	22
3.6.3 Erweiterung	24
3.7 Berechnung der Mandelbrotmenge	25
3.7.1 Berechnung mithilfe von SIMD	25
3.8 Erläuterung des Frontends	28
3.8.1 Kommunikation mit dem Backend	29
3.8.2 Darstellung der Regionsdaten	29
3.8.3 Visualisierung der Rechenzeiten	32
3.8.4 MISC	32
4 Ergebnisse / Evaluation	33

1 Einleitung

Die Leistung und Geschwindigkeit des individuellen Rechenkerns stagniert seit einigen Jahren. Moderne Computer erlangen einen Großteil ihrer erhöhten Rechenleistung seit einiger Zeit nur noch durch Parallelisierung. Diese sollte jedoch geschickt gestaltet werden, um unerwünschte Seiteneffekte wie Leerlauf zu vermeiden.

1.1 Didaktische Ziele

Das zugrundeliegende Problem ist, dass bei der Lastaufteilung einer unabhängigen Menge von Berechnungen in einem Cluster eine fixe Zuordnung von zu berechnenden Bereichen auf Rechenkerne erzeugt wird. Dauert die Bearbeitung eines Bereiches jedoch deutlich kürzer als diejenige anderer Abschnitte, so verbringt der reservierte Rechenkern die Zeit bis zum Abschluss der anderen Berechnungen ohne Arbeit und verbraucht Strom und Platz im Idle-Mode. Da die Kerne eines Clusters jedoch darauf ausgelegt sind, ständig zu arbeiten, sollte dieser Zustand vermieden werden, um Zeit, Kosten und Energie zu sparen.

Dies kann erreicht werden, indem die Einteilung der Rechenbereiche die vorraussichtliche Rechendauer berücksichtigt. Dazu werden rechenintensive Bereiche verkleinert und umgekehrt Bereiche mit geringerer Rechenlast vergrößert. Ziel sollte sein, dass alle Knoten für die Bearbeitung in etwa gleich lang brauchen, sodass die gegenseitige Wartezeit minimiert wird.

Dieses Projekt soll intuitiv vermitteln, dass bei der Aufteilung unabhängiger Berechnungen auf ein Cluster eine Abschätzung der benötigten Rechenlast die Gesamtrechendauer deutlich verringern kann. Außerdem soll ersichtlich sein, wie die verwendete Aufteilung bestimmt wird.

1.2 Verwendung der Mandelbrotmenge

Die Mandelbrotmenge ist eine Teilmenge der komplexen Zahlen. Um sie zu berechnen wendet man folgende Formel wiederholt auf jede komplexe Zahl c an:

$$z_{n+1} = z_n^2 + c, \quad z_0 = 0 \tag{1}$$

In der Mandelbrotmenge befinden sich alle c , für die der Betrag von z_n für beliebig große n endlich bleibt. Wenn der Betrag von z nach einer Iteration größer als 2 ist, so strebt z gegen unendlich, das zugehörige c liegt also nicht in der Menge. Sobald $|z_n| > 2$ kann die Berechnung daher abgebrochen werden.

Um nun für eine beliebige Zahl zu bestimmen, ob diese in der Mandelbrotmenge liegt, müssen theoretisch unendlich viele Rechenschritte durchgeführt werden. Zur computergestützten Bestimmung werden die Rechenschritte nach einer bestimmten Iteration abgebrochen die Zahl als in der Menge liegend betrachtet.

Es handelt sich also um eine Berechnung, die sehr zeitaufwändig ist, wobei die benötigte Zeit durch Erhöhen der Iterationszahl beliebig erhöht werden kann. Zusätzlich ist die Berechnung für jede einzelne komplexe Zahl unabhängig von jeder anderen Zahl.

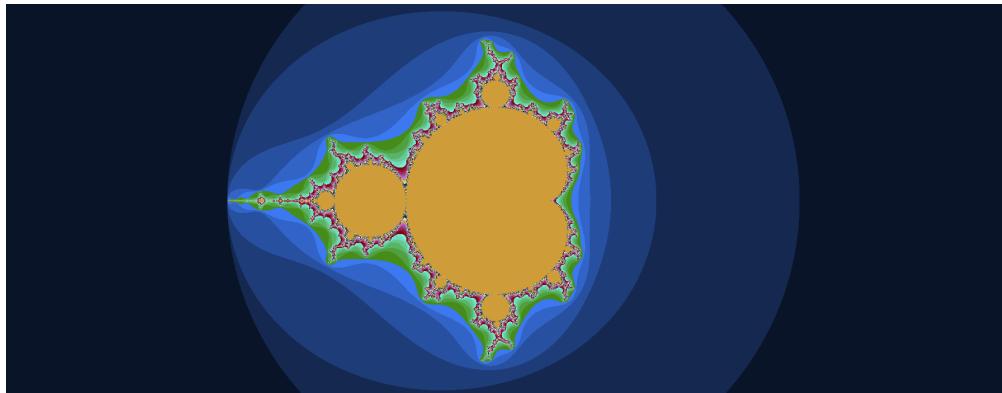


Abbildung 1: Die Mandelbrotmenge, visualisiert in einem Ausschnitt des komplexen Zahlenraumes.

Diese Eigenschaften ermöglichen es, zweierlei Dinge zu kontrollieren:

- Die Dauer der Berechnung
- Die Aufteilung der Berechnung auf unterschiedliche Rechenkerne

Somit kann gesichert werden, dass eine wahrnehmbare Zeit (100-200 ms) zur Berechnung benötigt wird. Zudem kann die Unterteilung des zu berechnenden Raumes frei gewählt werden, sodass für verschiedenste Aufteilungen die Gesamtrechenzeit visualisiert werden kann.

1.3 Darstellung der Mandelbrotmenge

Komplexe Zahlen lassen sich auch grafisch darstellen, indem man sie in ein Koordinatensystem einträgt. Dabei entspricht die x-Koordinate dem Realteil und die y-Koordinate dem Imaginärteil der Zahl. Für das Projekt wird ein Ausschnitt des Bildschirmes als zweidimensionale Darstellung des komplexen Raumes betrachtet und für jeden darin liegenden Punkt die Zugehörigkeit zur Mandelbrotmenge bestimmt. Dabei wird der Raum jedoch diskretisiert, indem jedem Pixel des Bildschirmes die komplexen Koordinaten c der linken oberen Ecke zugeordnet werden.

Die grafische Darstellung der Mandelbrotmenge wird durch Einfärbung des zu c gehörigen Pixels erhalten. Die Zahl der benötigten Iterationen bis zum Abbruch der Berechnung bestimmt dabei die Farbe, sodass alle Pixel innerhalb der Menge und alle Pixel außerhalb jeweils gleichfarbig sind.

Das entstehende Fraktal ist aufgrund seiner Form auch als "Apfelmännchen" bekannt (siehe Abbildung 1). Die Menge ist zusammenhängend, jedoch bilden sich an ihren Rändern viele kleine und sehr komplexe Formen, die visuell ansprechend sind. Es eignet sich daher gut, um optisch Interesse am Projekt zu wecken.

1.4 MPI

Das Message Passing Interface¹ ist eine weit verbreitete Spezifikation, für die Kommunikation zwischen unabhängigen Rechenkernen. Dadurch existieren viele gut funktionierende Umsetzungen in einer Vielzahl von Programmiersprachen. Für dieses Projekt wichtig ist, dass es echte Parallelisierung mit geringem Overhead ermöglicht. So können die einzelnen Berechnungen auf jeweils eigenen unabhängigen Rechenkernen laufen und die Art der Aufteilung erhält größtmögliche Bedeutung. Die Gestaltung von MPI erlaubt dabei beliebige Zuordnungen, von Kernen auf einem Prozessor bis hin zu unabhängigen Clusterknoten, die lediglich eine SSH-Verbindung besitzen.

1.5 Qualitätsanforderungen

Die Benutzeroberfläche soll so leicht und intuitiv wie möglich zu bedienen sein. Hierbei soll zudem darauf geachtet werden, dass alle Funktionen nur mit einer minimalen Anzahl an Mausklicks auszuführen sind und die Oberfläche nicht überladen wird.

Zudem soll das System robust gestaltet werden. Dies wird durch die Verwendung von Buttons, Listen, Drop-Down Menüs und Slider gewährleistet, die die Möglichkeit der Eingabe von ungültigen Werten verhindern.

Es sind keine Sicherheitsfeatures (Benutzerauthentisierung, Verschlüsselung) geplant, da keine sensiblen Daten verarbeitet werden und die Anwendung nicht uneingeschränkt über das Internet zugänglich ist.

1.6 Einschränkungen

Die Benutzeroberfläche soll in einem Webbrower lauffähig sein, sodass sie auf beliebigen Endgeräten zugänglich gemacht werden kann. Um die erwartete Performanzsteigerung an einem echten Beispiel zu demonstrieren, soll die Berechnung der Mandelbrotmenge parallel auf mehreren Raspberry Pi's² oder ähnlichen unabhängigen Kleincomputern oder Rechenkernen zum Einsatz kommen soll.

¹<https://www.mpi-forum.org/>

²Für ein Beispiel, siehe <https://www.raspberrypi.org/products/raspberry-pi-3-model-b/>

2 Problemstellung und Motivation

- Fachliche Spezifikation in Anhang
- NFRs in Einleitung schreiben

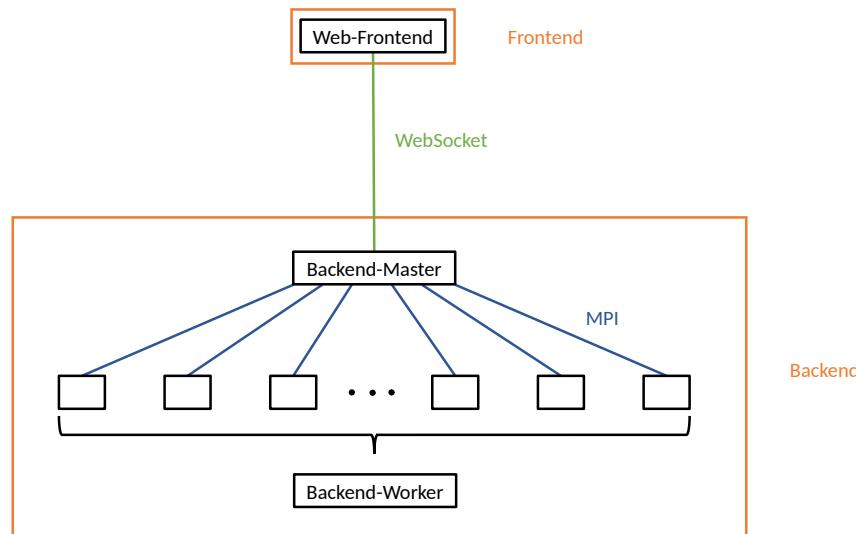


Abbildung 2: Architekturübersicht

3 Dokumentation der Implementierung

3.1 Übersicht

Das Problem fordert eine Unterteilung in drei wesentliche Bausteine, wie sie in Abbildung 2 zu sehen sind. Eine Benutzeroberfläche in einem Web Browser des Benutzers ("Frontend"), welches die Benutzerinteraktionen entgegennimmt und mit dem Backend kommuniziert. Ein Backend-Host übernimmt dazu die Kommunikationsfunktion, verwaltet die eingehenden Rechenaufträge und verteilt sie an die Backend-Worker. Das Ergebnis dieser Berechnung sendet der Host dann an das Frontend zurück. Die Backend-Worker nehmen die zugewiesenen Rechenaufträge entgegen und führen die eigentlichen Berechnungen verteilt aus.

3.2 Installation der Anwendung

Um das System zu installieren, muss das Repository mit git³ lokal geklont werden. Dabei werden die Quelldateien für das Front- sowie Backend heruntergeladen.

```
1 git clone https://gitlab.lrz.de/lrr-tum/students/eragp-mandelbrot.git
```

Quelltext 1: Klonen des Repositorys

³<https://git-scm.com/>

3.2.1 Lokales Backend

Eine lokale Installation des Backends zu Entwicklungszwecken ist durch einen Docker⁴ Container möglich. Dieser bietet eine ähnliche Umgebung zu der des Clusters und ermöglicht schnellere Feedbackzyklen.

```

1 # Systemabhängige Installation der Docker Anwendung
2 $ sudo apt install docker
3 cd backend/ && ./run_docker.sh
4 Starting the Build Process
5 ...
6 Host: Core 37 ready!
7 # ^C beendet das Backend und verbindet sich mit der shell des Containers
8 ^C[mpiexec@9cc2d5ac2cd1] Sending Ctrl-C to processes as requested
9 [mpiexec@9cc2d5ac2cd1] Press Ctrl-C again to force abort
10 # exit schliesst die shell des Containers
11 root@9cc2d5ac2cd1:~/eragp-mandelbrot/backend# exit

```

Quelltext 2: Starten der Entwicklungsumgebung des Backends

Das `run_docker.sh` Skript lädt das benötigte Basis Image, welches alle benötigten Bibliotheken bereits enthält, herunter und erstellt basierend darauf den Entwicklungscontainer. In diesen werden dann die aktuellen Quelldateien hinein kopiert und kompiliert, wonach das Backend mit Adresse `ws://localhost:9002` gestartet wird.

3.2.2 Backend auf HimMUC Cluster

[Der] HimMUC ist ein flexibler Cluster von ARM-Geräten, bestehend aus 40 Raspberry Pi 3 sowie 40 ODroid C2 Single-Board-Computers (SBC).⁵

Schnellstart Um das Programm auf dem HimMUC Cluster zu starten, wurde ein Python Skript erstellt, das alle notwendigen Schritte übernimmt. Es führt die Befehle aus Abschnitt 3.2.2 aus, es kann daher bei Problemen zur Fehlerbehebung herangezogen werden.

Stellen sie zunächst sicher, dass sie ein Konto mit Zugangsberechtigungen auf dem HimMUC Cluster besitzen. Um den eigenen Quellcode auf dem Cluster zu kompilieren muss für die korrekte Funktionsweise des Skriptes zudem ihr SSH-Key auf dem Cluster abgelegt sein⁶.

Außerdem sollten folgende Programme lokal installiert sein:

- rsync
- ssh
- python3 (3.5 oder neuer)

Starten sie anschließend aus dem Ordner `backend/` den Befehl aus Quelltext 3

⁴<https://www.docker.com/>

⁵<http://www.caps.in.tum.de/himmuc/>

⁶siehe `ssh-copy-id`

```
1 python3 himmuc/start_himmuc.py <Rechnerkennung> <Anzahl Prozesse> <Anzahl  
Rechenknoten>
```

Quelltext 3: Start der Entwicklungsumgebung auf dem HimMUC

```
1 $ eragp-mandelbrot/backend$ python3 himmuc/start_himmuc.py muendl 10 9  
2 Uploading backend... sending incremental file list  
3 backend/himmuc/start_backend.py  
4      3,897 100%    3.05MB/s   0:00:00 (xfr#1, to-chk=35/62)  
5 done  
6 Start mandelbrot with 1 host and 9 workers on 9 nodes... started  
     mandelbrot  
7 Search host node... srun: error: Could not find executable worker  
odr00 found  
8 Establish port 9002 forwarding to host node odr00:9002 ... established  
9 System running. WebSocket connection to backend is now available at  
ws://himmuc.caps.in.tum.de:9002  
10 Press enter (in doubt, twice) to stop Warning: Permanently added the  
ED25519 host key for IP address '10.42.0.54' to the list of known  
hosts.  
11 # Enter  
12  
13 Stopping port forwarding... stopped (-9)  
14 Stopping mandelbrot host and workers... stopped (-9)
```

Quelltext 4: Beispielausgabe bei Start der Entwicklungsumgebung auf dem HimMUC

Das Ergebnis wird ähnlich zu Quelltext 4 aussehen. Details zu weiteren Optionen des Skripts sind via --help verfügbar.

Detaillierter Start Um Mandelbrot manuell auf dem Host-System zu installieren, müssen zunächst die notwendigen Bibliotheken installiert werden. Eine Anleitung dazu findet sich in Abschnitt 3.2.2. Dies muss nur einmal ausgeführt werden, anschließend können die Programme wie in Abschnitt 3.2.2 beschrieben kompiliert werden. Ist dies nicht gewünscht oder erledigt muss das Backend lediglich noch wie in Abschnitt 3.2.2 beschrieben gestartet werden.

Lokale Installation der Bibliotheken Da hierbei davon ausgegangen wird, dass keine root-Rechte auf dem Server existieren, werden die Bibliotheken hier lokal in `~/eragp-mandelbrot` installiert. Achten Sie darauf, dass sie Schreibrechte auf dem Ordner haben und falls sie einen anderen Ordner verwenden wollen, ersetzen sie jedes Vorkommen des Pfades durch ihren Pfad (insbesondere in der Datei `CMakeLists.txt`). Die MPI-Bibliothek ist auf dem HimMUC Cluster bereits vorinstalliert und muss daher nicht mehr aufgesetzt werden.

```
1 mkdir ~/eragp-mandelbrot
```

Quelltext 5: Erstellen des Installationsordners

Aus der Bibliothek `boost` muss die Teilbibliothek `boost_system` lokal kompiliert

werden Dazu werden die Befehle aus Quelltext 6 ausgeführt, um die Version 1.67.0 herunterzuladen, zu entpacken und lokal zu installieren.

```

1 mkdir "~/.eragg-mandelbrot/install"
2 mkdir "~/.eragg-mandelbrot/local"
3 cd "~/.eragg-mandelbrot/install"
4 wget "https://dl.bintray.com/boostorg/release/1.67.0/source/boost_1_67_0.
      tar.bz2"
5 tar --bzip2 -xf boost_1_67_0.tar.bz2
6 cd boost_1_67_0
7 ./bootstrap.sh --prefix="$HOME/.eragg-mandelbrot/local/" --with-libraries
      =system
8 ./b2 install

```

Quelltext 6: Lokale Installation der Bibliothek boost.

Die 'Header-only' Libraries `websocketpp` und `rapidjson` müssen lediglich an einen fixen Ort kopiert werden. Dies erledigen die Befehle aus Quelltext 7.

```

1 mkdir "~/.eragg-mandelbrot/install"
2 cd "~/.eragg-mandelbrot/install"
3 # Installation von websocketpp
4 git clone --branch 0.7.0 https://github.com/zaphoyd/websocketpp.git
      websocketpp --depth 1
5 # Installation von rapidjson
6 git clone https://github.com/Tencent/rapidjson/

```

Quelltext 7: Lokale Installation der Bibliotheken `websocketpp` und `rapidjson`.

Kompilieren des Backends Stellen Sie zunächst sicher, dass auf dem Cluster die Quelldateien des Backendes (im Ordner `backend/`) liegen (zum Beispiel über `rsync` oder indem sie das Repository dort auch klonen) Zum Kompilieren des Backends sollte sich auf einen Raspberry Pi oder ODroid per ssh eingeloggt werden⁷

Auf dem Board, aus dem Ordner des Backendquellcodes müssen sie zum kompilieren des Backendes die Befehle aus Quelltext 8 ausführen.

```

1 # Erstellen und betreten eines build Ordners
2 mkdir build
3 cd build
4 # Aktivieren der MPI Bibliothek
5 module load mpi
6 # Kompilieren
7 cmake ..
8 make

```

Quelltext 8: Kompilieren des Backends

⁷Es existiert ein Entwicklerzugang zu einem geteilten Raspberry Pi über die Adresse `sshgate-gepasp.in.tum.de`. Dieser wird auch vom Python Skript genutzt

Ausführen des Backends Um das Backend auf dem HimMUC Cluster laufen zu lassen, muss sich zunächst darauf per ssh eingeloggt werden. Damit für das Frontend kein Unterschied dazwischen besteht, ob das Backend im Dockercontainer , oder auf einem externen Server ausgeführt wird, ist bei der ssh-Verbindung der Port 9002 des `himmuc.caps.in.tum.de`-Servers an den lokalen Port 9002 gebunden. So ist das Backend stets unter `localhost:9002` verfügbar. Der zugehörige Befehl zum Login lautet demnach:

```
1 ssh <rechnerkennung>@himmuc.caps.in.tum.de -L localhost:9002:localhost
:9002
```

Anschließend muss aus dem Ordner, in dem die ausführbaren Dateien liegen, für gewöhnlich also der `~/eragp-mandelbrot/build/` Ordner, folgender Befehl ausgeführt werden:

```
1 srun -p <odr|rpi> -n <number of workers+1> -N <number of nodes/raspis> -l
--multi-prog <path to eragp-mandelbrot/backend>/himmuc/run.conf &
2 ssh -L 0.0.0.0:9002:localhost:9002 -fN -M -S .tunnel.ssh <odr|rpi><host
number>
```

Dabei bestimmt `-n` die Anzahl der laufenden Prozesse (Also Hostprozess und Workerprozesse) und `-N` die Anzahl zu verwendender Rechenknoten. Damit anschließend noch alle Anfragen an den WebSocketserver auf dem Hostknoten weitergeleitet werden, muss noch der Port 9002 des `himmuc.in.caps.tum.de`-Servers an den Port 9002 des Rechenknotens gebunden werden, auf dem der Hostprozess läuft. Der korrekte Knoten ist dabei der Ausgabe des `srun`-Befehles zu entnehmen. Eine beispielhafte Ausgabe ist in Quelltext 9 zu sehen.

Stoppend des Backends Um das Backend wieder zu stoppen, müssen der ssh-Tunnel zur Verbindung der Ports und der `srun`-Prozess gestoppt werden. Letzterer lässt sich nach dem dämonisieren im vorigen Aufruf nur über die Prozess-ID finden. Diese zeigt das Tool `ps` an.

```
1 ssh -S .tunnel.ssh -O exit rpi<host number>
2 # To stop the node allocation
3 ps -eo comm,pid | grep srun
4 kill <srun pid>
```

3.2.3 Installation des Frontends

Das Frontend ist in TypeScript⁸ (erweiterung von JavaScript⁹) geschrieben und kann somit auf einem beliebigen Endgerät mit einem modernen Webbrowser ausgeführt werden. Um eine Version lokal zu starten, muss die Paketverwaltung `npm`¹⁰ installiert werden. Diese verwaltet alle für das Frontend benötigten Bibliotheken und installiert diese lokal.

⁸<https://www.typescriptlang.org/>

⁹<https://en.wikipedia.org/wiki/JavaScript>

¹⁰<https://www.npmjs.com/>

```

1 muendl@vmschulz8:~/eragp-mandelbrot/backend/build$ srun -N4 -n5 -l --
2   multi-prog ./himmuc/run.conf
3 srun: error: Could not find executable worker
4: Worker: 4 of 5 on node rpi06
4: Worker: 2 of 5 on node rpi04
5: Worker: 3 of 5 on node rpi05
6: Host: 0 of 5 on node rpi03
7: Host init 5
8: 1: Worker: 1 of 5 on node rpi03
9: 0: Core 1 ready!
10: 1: Worker 1 is ready to receive Data.
11: 2: Worker 2 is ready to receive Data.
12: 0: Listening for connections on to websocket server on 9002
13: 0: Core 2 ready!
14: 3: Worker 3 is ready to receive Data.
15: 0: Core 3 ready!
16: 4: Worker 4 is ready to receive Data.
17: 0: Core 4 ready!
18 muendl@vmschulz8:~/eragp-mandelbrot/backend/build$ ssh ssh -L
    0.0.0.0:9002:localhost:9002 -fN -M -S .tunnel.ssh rpi03

```

Quelltext 9: Beispielhafter Start des Backends. Hierbei ist der Knoten des Hostprozesses rpi03.

```

1 # Systemabhängige Installation der npm Paketverwaltung
2 sudo apt install npm
3 # Installiert benötigte Bibliotheken und startet WebServer
4 cd frontend/ && npm install ; npm start
5 ...
6 Version: webpack 4.25.1
7 Time: 7230ms
8 Built at: 12/28/2018 10:48:32 PM
9          Asset      Size  Chunks             Chunk Names
10        index.html  1.65 KiB          [emitted]
11 mandelbrot.js   11.7 MiB     main  [emitted]  main
12       style.css   519 KiB     main  [emitted]  main
13 Entrypoint main = style.css mandelbrot.js
14 ...

```

Quelltext 10: Starten des Frontends mit beispielhafter Ausgabe

Das Kommando `npm start` startet dabei einen lokalen WebServer, welcher eine kompilierte Version des Frontends unter der Adresse `http://localhost:3000` anbietet. Danach wird der Standardwebbrowser des Systems verwendet, um diese URL zu öffnen.

3.3 Erläuterung des Backends

3.4 Implementierung der Mandelbrotberechnung

Zur hardwarenahe Berechnung der Mandelbrotmenge wird ein sogenanntes Backend gestartet. Das in C++ programmierte Teilprojekt nimmt Rechenaufträge von einem Nutzer durch ein Frontend entgegen (auch ein solches wird bereitgestellt), zerlegt sie

und verteilt sie per MPI auf dedizierte Rechenknoten. Dazu besteht das Backend aus zwei ausführbaren Dateien, host und worker.

3.4.1 Inkludierte Header und CMake Anweisungen

Die zusammenstellung der ausführbaren Dateien wird in CMake definiert. Dabei unterscheiden sich diese lediglich in den eingebundenen Quelldateien: In die Datei host werden host.main.cpp und actors/Host.cpp eingebunden, während in worker worker.main.cpp und actors/Worker.cpp eingebunden werden.

Diese und alle weiteren Build-Vorgaben werden in der Datei CMakeLists.txt für cmake¹¹ in der hier beschriebenen Reihenfolge spezifiziert. Es sollte hierbei eine CMake-Version über 3.7.0 gewählt werden und die C++11 Standards¹² werden vorrausgesetzt. Zudem werden für das Projekt "Mandelbrot" werden alle Dateien im Order include eingebunden. In diesem Ordner liegen die Header-Dateien für alle projektinternen C++-Quelldateien. Anschließend werden alle C++-Quelldateien (Endung ".cpp") aus dem Ordner src in einer Liste gesammelt, mit Ausnahme jedoch der oben genannten, exklusiven Quelldateien. Die erzeugte Liste und die jeweils exklusiven Dateien werden dann den ausführbaren Dateien host und worker zugeordnet.

Um die verwendeten Bibliotheken verfügbar zu machen werden anschließend die Header der installierten MPI-Bibliothek sowie die Header der Bibliotheken rapidjson¹³, websocketpp¹⁴ und boost¹⁵. Diese werden respektive verwendet um JSON zu parsen und enkodieren, WebSocket-Verbindungen aufzubauen und darüber zu kommunizieren sowie um diese Bibliothek zu unterstützen. Da für die boost Bibliothek dabei Header nicht genügen und die systemweite Verfügbarkeit der kompilierten boost-Bibliothek nicht garantiert werden kann, wird die Teilbibliothek boost_system statisch in die ausführbaren Datei host eingebunden.

Zuletzt werden über Compilerflags alle Kompilierfehler und -warnungen aktiviert sowie die POSIX-Thread-Bibliothek eingebunden und spezielle Flags für die Websocket-library und MPI gesetzt.

3.4.2 Mainfunktion und Initialisierung

Zur Initialisierung der Prozesse muss zunächst die MPI-Umgebung aktiviert und abgerufen werden. Dies geschieht für beide Programme gleich, über die Initialisierungsfunktion in Quelltext 12. Sie erwartet lediglich eine Beschreibung des Prozesses für den Log und eine Initialisierungsfunktion, die erst zurückkehrt, wenn das Programm abgeschlossen ist und MPI beendet werden soll. Die Funktion muss als Parameter den Rang bzw. die Id des aktuellen MPI-Prozesses und die Anzahl der initialisierten Prozesse entgegen nehmen.

¹¹Ein Programm, welches die Erstellung von Makefiles vereinfacht in dem es sie automatisch an die Umgebung des Build-Systems anpasst. <https://cmake.org/>

¹²<https://isocpp.org/wiki/faq/cpp11>

¹³<http://rapidjson.org>

¹⁴<https://github.com/zaphoyd/websocketpp>

¹⁵<https://www.boost.org/>

Ein beispielhafter Aufruf ist in Quelltext 11 zu sehen. Damit wird MPI initialisiert und nach der erfolgreichen Initialisierung der eigentliche Host-Prozess über `Host::init` gestartet.

```

1 #include "init.h"
2 #include "Host.h"
3
4 int main(int argc, char *argv[])
{
5     return init(argc, argv, "Host", Host::init);
6 }
7 
```

Quelltext 11: Initialisierung des Host-Prozesses in host.main.cpp

3.5 Host Funktionalitäten

3.5.1 Websocketverbindung

Direkt nach der Initialisierung des Host-Programms in Quelltext 13, wird ein separater Thread gestartet, der über Websocket Anfragen zur Berechnung einer Region entgegen nimmt sowie ein Thread, der berechnete Regionen an den verbundenen Client übergibt. Die Methode `Host::start_server` initialisiert dabei lediglich den Websocketserver mit den Methoden zum Behandeln geöffneter und geschlossener Verbindungen und einer Methode um Nachrichten des Clients zu behandeln, `Host::handle_region_request`. Der zweite gestartete Thread führt die Methode `Host::send` aus. Zudem wird der Websocketserver mit `server.init_asio()` mit der Transport Policy "transport::asio" konfiguriert, sodass Multithreadzugriff auf Sende- und Empfangsmethoden problemlos möglich ist [2].

Host::handle_region_request Die Methode versucht, den Inhalt der empfangenen Nachricht als JSON zu dekodieren und entnimmt die für das struct `Region` notwendigen Werte unter gleichem Namen dem Objekt, das unter dem Schlüssel "region" in der geprästen Anfrage gespeichert ist. Außerdem erwartet es unter dem Schlüssel "balancer" einen String, der den zu wählenden Lastbalancierer bestimmt und unter dem Schlüssel "type" den String "regionRequest". Mögliche Zeichenketten hierfür sind in den Klassen der Balancierer unter `backend/src/balancer` in der globalen Variable `Klassenname::NAME` gespeichert. Ein Beispiel für einen Namen kann in Quelltext 14 gefunden werden:

Ein Beispiel für eine gültige Regionsanfrage ist in Quelltext 15 zu finden. Einerseits wird hierbei eine Region in komplexen Koordinaten beschrieben, wobei der obere linke Punkt (`maxImag, minReal`) und der rechte untere Punkt (`minImag, maxReal`) in der komplexen Ebene einen zu berechnenden Bereich aufspannen. Da die reelle Ebene jedoch beliebig genau aufgelöst werden kann, muss zudem noch die Anzahl an Pixeln pro Seite des Rechteckes definiert werden, `width` und `height`. Wie in Abbildung 3 zu sehen, ist zudem der horizontale Offset und vertikale Offset die linke obere Koordinate

```

15 int init(int argc, char **argv, const char* type, void (*initFunc) (int
16   world_rank, int world_size)) {
17
18   MPI_Init(&argc, &argv);
19   int world_rank;
20   MPI_Comm_rank(MPI_COMM_WORLD, &world_rank);
21   int world_size;
22   MPI_Comm_size(MPI_COMM_WORLD, &world_size);
23   // Retreive the processor name to check if host and
24   // worker share a node
25   char* proc_name = new char[MPI_MAX_PROCESSOR_NAME];
26   int proc_name_length;
27   MPI_Get_processor_name(proc_name, &proc_name_length);
28   std::cout << type << ":" << world_rank << " of " << world_size <<
29     " on node " << proc_name << std::endl;
30
31   if (world_size < 2) {
32     std::cerr << "Need at least 2 processes to run. Currently have "
33           << world_size << std::endl;
34     MPI_Finalize();
35     return -1; // return with error
36   }
37   initFunc(world_rank, world_size);
38   MPI_Finalize();
39
40   return 0;
}

```

Quelltext 12: Initialisierung der MPI-Prozesse in init.cpp

```

382     std::cerr << e.what() << "\n" << "Refresh websocket
383     connection.\n";
384
385     delete [] data.data;
386
387     std::cout << "Host Websocket-Result-Thread: Sending RegionData
388     from Worker " << workerInfo.rank << " done." << std::endl;
389
390   }
391
392 void Host::init(int world_rank, int world_size) {
393   MPI_Comm_set_errhandler(MPI_COMM_WORLD,MPI_ERRORS_RETURN); /* return
     info about errors */

```

Quelltext 13: Starten des Websocketservers bei der Initialisierung des Host-Programmes in Host.cpp

```
11 const std::string PredictionBalancer::NAME = "prediction";
```

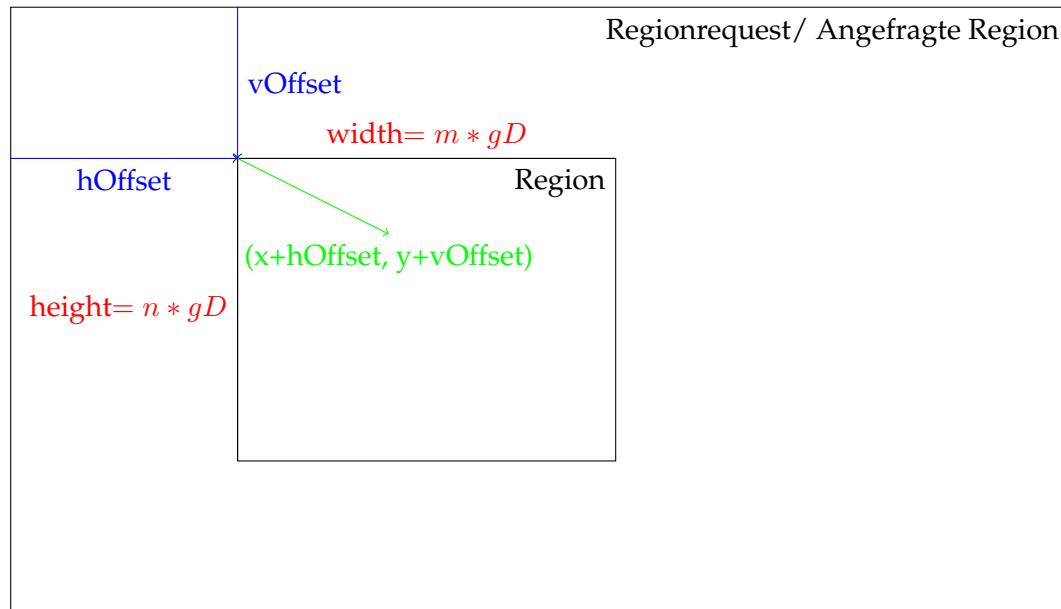
Quelltext 14: Namensdefinition des PredictionBalancers

```
1 {
2     "type": "regionRequest",
3     "region": {
4         "minReal": -0.251953125,
5         "maxImag": -0.8388671875,
6         "maxReal": -0.2216796875,
7         "minImag": -0.8505859375,
8         "width": 1984,
9         "height": 768,
10        "hOffset": 0,
11        "vOffset": 0,
12        "validation": 8,
13        "guaranteedDivisor": 64,
14        "maxIteration": 1019
15    },
16    "balancer": "naiveRecursive",
17    "fractal": "mandelbrot"
18}
```

Quelltext 15: Eine gültige Anfrage einer Region in JSON

der Region bezüglich der gesamten sichtbaren Anfrage in Pixeln (diese Werte gewinnen in den Regionsaufteilungen an Bedeutung). Der Wert validation ist technisch gesehen nicht mehr notwendig, wird aber mit dem Zoom-wert der Leafletkarte gefüllt um zu vermeiden dass Regionsdaten von zuvor berechneten Regionen falsch verwendet werden.

Nach dem Parsen der Nachricht wird die Region global festgelegt und beim korrekten Lastbalancierer eine Zerlegung der angefragten Region über Balancer::balanceLoad gestartet. Aus dem Ergebnis werden leere Regionen werden anschließend aussortiert und alle anderen in eine per mutex Thread-gesicherte Datenstruktur gelegt. Der MPI verwendende Thread, der diese Regionen anschließend an die Worker sendet wird über das Setzen des booleschen Wertes mpi_set_regions auf true darüber informiert, dass neue Regionen zum versenden zur Verfügung stehen. Dabei ist der Rang des Workers, der eine Region berechnen soll genau der Index der Region in der Datenstruktur sein. Ist ein Worker nicht verfügbar, so werden alle folgenden Ränge um eins erhöht (siehe Quelltext 16). Damit kann der WebSocketprozess unabhängig von der tatsächlichen Verteilung den Rang des berechnenden Worker-Prozesses bestimmen und in der Antwort an den Client zu der Aufteilung der Prozesse einfügen. Die Struktur und eine gültige Antwort des Websocketservers kann Quelltext 17 entnommen werden.



$gd = \text{guaranteedDivisor}, n, m \in \mathbb{N}^+, x < width, y < height$

Abbildung 3: Konzept der Koordinaten in den Regionsobjekten. Alle Koordinaten beziehen sich auf die Darstellungsebene und sind daher in Pixeln.

```

231     for (int i = 0 ; i < nodeCount ; i++) {
232         blocks[i].fractal = fractal_type;
233         websocket_request_to_mpi[i] = blocks[i];
234     }
235     mpi_send_regions = true;
236 }
237 std::cout << "Sending Region division" << std::endl;
238
239 // Determine which Worker gets which Region
240 int region_to_worker[nodeCount];

```

Quelltext 16: Algorithmus zur Zuordnung von Regionen auf Worker in Host.cpp

```
1  {
2      "type": "region",
3      "regionCount": 37,
4      "regions": [
5          {
6              "rank": 1,
7              "computationTime": 0,
8              "region": {
9                  "minReal": -0.251953125,
10                 "maxImag": -0.8408203125,
11                 "maxReal": -0.24609375,
12                 "minImag": -0.8427734375,
13                 "width": 384,
14                 "height": 128,
15                 "hOffset": 0,
16                 "vOffset": 0,
17                 "maxIteration": 1019,
18                 "validation": 8,
19                 "guaranteedDivisor": 64
20             }
21         },
22         {
23             "rank": 2,
24             "computationTime": 0,
25             "region": {
26                 "minReal": -0.251953125,
27                 .....
28             },
29             "rank": 37,
30             "computationTime": 0,
31             "region": {
32                 "minReal": -0.2275390625,
33                 "maxImag": -0.84765625,
34                 "maxReal": -0.2216796875,
35                 "minImag": -0.8486328125,
36                 "width": 384,
37                 "height": 64,
38                 "hOffset": 1600,
39                 "vOffset": 448,
40                 "maxIteration": 1019,
41                 "validation": 8,
42                 "guaranteedDivisor": 64
43             }
44         }
45     ]
46 }
```

Quelltext 17: Ausschnitt aus einer gültigen Antwort auf die Region aus Quelltext 15 in JSON

Host::send Während ein Thread das Empfangen von Nachrichten übernimmt, behandelt diese Methode von den Workern fertig berechnete Regionen. Er versendet in einer Dauerschleife Regionsdaten an den aktuell verbundenen Client. Nach jedem Durchlauf oder wenn der Thread per `notify` (oder auch durch mögliche andere Nebeneffekte) aufgeweckt wird, überprüft er auf das Vorhandensein neuer Regionen, wählt die erste aus und entfernt sie aus der Datenstruktur. Nachdem das Lock auf die geteilte Datenstruktur gelöst wurde, werden die Regionsdaten JSON codiert versendet. Ein Beispiel für versendete Regionsdaten kann Quelltext 18 entnommen werden. Dabei wird der Punkt (x, y) in der gesendeten Region (Punkt $(x+hOffset, y+vOffset)$ in der angefragten Region) im Datenarray an Index $i = x + y * width$ gespeichert. Ist keine Region in der Datenstruktur abgelegt wird per `wait` auf eine Änderung gewartet. Mit Hilfe der `condition_variable`¹⁶ `Host::mpi_to_websocket_result_available` nutzt sie die C++11 nativen mutex-Mechanismen um über das Vorhandensein neuer Regionen informiert zu werden.

3.6 Lastbalancierung

Um die Mandelbrotmenge effizient parallel zu berechnen, muss die Last gleichmäßig auf die Worker verteilt werden. Die Aufgabe der Lastbalancierung besteht darin zu einer gegebenen Region und einer Anzahl von Workern eine solche Unterteilung in sogenannte Teilregionen zu finden. Wichtig dabei ist, dass der garantierte Teiler von Höhe und Breite der Teilregionen dem der angeforderten Region entspricht, da es sonst im Frontend zu Schwierigkeiten bei der Darstellung kommt. Die Klassenstruktur der Lastbalancierer entspricht dem Strategy-Pattern. So kann der Balancierer zur Laufzeit leicht gewechselt werden und auch die Erweiterung des Projekts um eine weitere Strategie gestaltet sich einfach. Was dabei genau beachtet werden muss findet sich im Teil Erweiterung.

Damit die Unterschiede zwischen guter und schlechter Lastverteilung deutlich werden, wurden hier verschiedene Strategien zur Lastbalancierung implementiert.

3.6.1 Naive Strategie

Bei der naiven Strategie (zu finden in den Klassen `NaiveBalancer` und `RecursiveNaiveBalancer`) wird versucht den einzelnen Workern etwa gleich große Teilregionen zuzuweisen. Dies geschieht allerdings ohne Beachtung der eventuell unterschiedlichen Rechenzeiten innerhalb der Teilregionen. Die naive Strategie wurde hier in einer nicht-rekursiven und einer rekursiven Variante implementiert.

Zur nicht-rekursiven Aufteilung wird zuerst die Anzahl der zu erstellenden Spalten und Zeilen berechnet. Dazu wird der größte Teiler der Anzahl der Worker bestimmt. Dieser gibt die Anzahl der Spalten an, das Ergebnis der Division ist die Anzahl der

¹⁶https://en.cppreference.com/w/cpp/thread/condition_variable

```

1  {
2      "type": "regionData",
3      "workerInfo": {
4          "rank": 3,
5          "computationTime": 1120852,
6          "region": {
7              "minReal": 0.33984375,
8              "maxImag": -0.583984375,
9              "maxReal": 0.400390625,
10             "minImag": -0.5859375,
11             "width": 1984,
12             "height": 64,
13             "hOffset": 0,
14             "vOffset": 128,
15             "maxIteration": 1019,
16             "validation": 7,
17             "guaranteedDivisor": 64
18         }
19     },
20     "data": [22, 22, 22, 22, 22, 22, 22, 22, 22, 22, 22,
21     22, 22, 22, 22, 22, 22, 22, 22, 22, 22, 22, 22,
22, 22, 22, 22, 22, 22, 22, 22, 22, 22, 22, 22,
22, 22, 22, 22, 22, 22, 22, 22, 22, 22, 22, 22,
22, 22, 22, 22, 22, 22, 22, 22, 22, 22, 22, 22,
22, 22, 22, 23, 23, 23, 23, 23, 23, 23, 23, 23,
23, 23, 23, 23, 23, 23, 23, 23, 23, 23, 23, 23,
23, 23, 23, 23, 23, 23, 23, 23, 23, 23, 23, 23,
23, 23, 23, 23, 23, 23, 23, 23, 23, 23, 23, 23,
24, 24, 24, 24, 24, 24, 24, 24, 24, 24, 24, 24,
24, 24, 24, 24, 24, 24, ...]
21 ]

```

Quelltext 18: Ausschnitt aus den Daten einer versendeten Teilregion. Punkt (x,y) liegt in "data" an Index $i = x + y * width$.

Zeilen. Damit ist sichergestellt, dass die Region in die richtige Menge von Teilregionen unterteilt wird.

Als nächstes wird die Breite und Höhe der Teilregionen berechnet. Hierbei ist wichtig, dass der garantierte Teiler erhalten wird. Die Breite berechnet sich also durch:

$$\frac{\text{region.width}}{\text{region.guaranteedDivisor} * \text{nodeCount}} * \text{region.guaranteedDivisor}$$

Dabei ist `nodeCount` die Anzahl der Worker und `region` die zu unterteilende Region. Es ist zu beachten, dass es sich hier um eine Ganzzahldivision handelt, deren Rest angibt, wie viele Teilregionen um `region.guaranteedDivisor` breiter sind. Die Höhe berechnet sich analog.

Bevor die eigentliche Aufteilung beginnt werden noch die Deltas für Real- und Imaginärteil bestimmt. Diese geben an, wie breit bzw. hoch der Bereich der komplexen Ebene ist, den ein Pixel der Region überdeckt. Die Deltas können über Methoden der Klasse `Fractal` berechnet werden.

Zur Aufteilung wird nun mittels zweier verschachtelter Schleifen über die Zeilen und Spalten iteriert. Die benötigten Start- und Endpunkt der Teilregionen (also die linke obere Ecke und die rechte untere Ecke auf der komplexen Ebene) können nun mithilfe der Schleifenzähler und der Deltas bestimmt werden. Zusätzlich wird noch der vertikale und horizontale Offset der Teilregion vom Startpunkt der Eingaberegion abgespeichert. Diese Information ermöglicht es die Region im Frontend einfach anzuzeigen.

Falls die aktuell betrachtete Teilregionen zu den Breiteren und/oder Höheren (s.o.) gehört, müssen alle Werte entsprechend angepasst werden. Die letzte Teilregion einer Spalte bzw. Zeile wird immer so gewählt, dass sie auf jeden Fall mit dem Rand der Eingaberegion abschließt.

Die Teilregionen werden in einem Ergebnisarray gespeichert, welches dann zurückgegeben wird.

Die Grundidee der rekursiven Balancierung ist, die Region solange zu halbieren, bis man genug Teile für jeden Worker hat. Dies funktioniert sehr gut, wenn die Anzahl der Worker eine 2er-Potenz ist. Wenn das nicht der Fall ist, so müssen für einige Worker die Regionen öfters geteilt werden als für andere. Die Anzahl der Blätter des Rekursionsbaumes (hier ein Binärbaum) muss also der Anzahl der Worker entsprechen. Die Rekursionstiefe kann man wie folgt berechnen:

$$\text{recCounter} = \lfloor \log_2 \text{nodeCount} \rfloor + 1 \quad (2)$$

Und die Anzahl der Teilregionen auf der untersten Ebene des Rekursionsbaumes ergibt sich als:

$$\begin{aligned} \text{missing} &= \text{nodeCount} - 2^{\lfloor \log_2 \text{nodeCount} \rfloor} \\ \text{onLowestLevel} &= \text{missing} * 2 \end{aligned} \quad (3)$$

Auch hier ist `nodeCount` die Anzahl der Worker. Die Multiplikation mit 2 ist nötig, da bei der Unterteilung einer Region immer zwei neue entstehen. Es müssen also `missing` Regionen aufgeteilt werden müssen, um die richtige Anzahl an Teilregion zu erzeugen.

Um diese Werte einfach durch die Rekursionsebenen zu reichen wurde die Struktur BalancingContext definiert, welche zusätzlich noch die beiden Deltas, den Index in das Ergebnisarray und einen Zeiger auf das Array selbst abspeichert.

Aus `recCounter` und `onLowestLevel` kann auch die Abbruchbedingung der Rekursion gefolgt werden:

$$\text{recCounter} = 0 \vee (\text{recCounter} = 1 \wedge \text{resultIndex} \geq \text{onLowestLevel}) \equiv \text{true} \quad (4)$$

`resultIndex` ist hierbei der Index in das Ergebnisarray, beschreibt also die Anzahl der bereits erstellten Teilregionen.

Ist die Abbruchbedingung (4) erfüllt, so genügt es die übergebene Region in das Ergebnisarray einzutragen und `resultIndex` zu inkrementieren. Ansonsten muss die Region halbiert werden. Ist die Region vertikal oder horizontal nicht mehr teilbar (d.h. `region.width` bzw. `region.height` $\leq \text{region.guaranteedDivisor}$), so wird in die andere Richtung geteilt. Kann die Region in beide Richtungen geteilt werden, so wird abwechselnd vertikal und horizontal geteilt. Dies gewährleistet, dass der Balancierer in beide Richtungen aufteilt, sofern das möglich ist. Wenn die Region unteilbar ist, so muss eine leere Region erzeugt werden.

Die beiden Hälften berechnen sich wie bei der Aufteilung auf zwei Worker mit der nicht-rekursiven Strategie. Dann wird die Funktion für jede Hälfte rekursiv aufgerufen.

3.6.2 Strategie mit Vorhersage

Bei dieser Strategie (zu finden in den Klassen `PredictionBalancer` und `RecursivePredictionBalancer`) basiert die Aufteilung der Region auf einer Vorhersage über die Rechenzeit. Die Teilregionen werden so gewählt, dass sie, entsprechend der Vorhersage, etwa einen ähnlichen Rechenaufwand haben.

Die Vorhersage (`struct Prediction`) wird von der Klasse `Predictor` angestellt. Dazu wird die Region in einer sehr viel geringeren Auflösung berechnet. Die benötigte Anzahl an Iterationen wird jeweils pro Kachel (Breite und Höhe sind der garantierte Teiler) abgespeichert. So wird sichergestellt, dass der garantierte Teiler auch nach der Aufteilung noch gilt, da die Balancierer die Vorhersage Eintrag für Eintrag verarbeiten. Die Genauigkeit der Vorhersage kann über das Attribut `predictionAccuracy` gesteuert werden:

- $\text{predictionAccuracy} > 0$: $(\text{predictionAccuracy})^2$ Pixel werden pro Kachel berechnet. Die Summe der Iterationen für die einzelnen Pixel ergibt die Vorhersage für die Kachel.
- $\text{predictionAccuracy} < 0$: Für $(\text{predictionAccuracy})^2$ Kacheln wird ein Pixel in der Vorhersage berechnet. Es erhalten also mehrere Kacheln dieselbe Vorhersage.
- $\text{predictionAccuracy} = 0$: Unzulässig, es wird ein Null-Pointer zurückgegeben.

Es ist wichtig eine gute Balance zwischen Güte und Geschwindigkeit der Vorhersage zu finden. Zusätzlich beinhaltet die Vorhersage die Summen der benötigten Iterationen pro

Spalte und Zeile, sowie die Gesamtsumme. Auch die Deltas für Real- und Imaginärteil pro Kachel und die Anzahl der Zeilen und Spalten werden angegeben.

Auch die Strategie mit Vorhersage wurde in einer rekursiven und in einer nicht-rekursiven Variante implementiert.

Für die nicht-rekursive Variante wird zuerst die benötigte Anzahl an Zeilen und Spalten bestimmt. Dies geschieht genauso wie bei der naiven Strategie.

Die erzeugten Teilregionen sollen in etwa den gleichen Rechenaufwand haben. Dieser berechnet sich durch:

$$\text{desiredN} = \frac{nSum}{nodeCount} \quad (5)$$

Dabei ist `nSum` die Gesamtsumme der Vorhersage und `nodeCount` wieder die Anzahl der Worker.

Danach wird die Region erst in Spalten aufgeteilt und in einem zweiten Schritt wird dann die horizontale Unterteilung in Teilregionen vorgenommen.

Zur Aufteilung wird über die Spaltensummen in der Vorhersage iteriert. Diese werden aufaddiert und bilden so den Zähler `currentN`. Sobald $\text{currentN} \geq \text{desiredN}$ gilt oder für alle restlichen Spalten nur noch je ein Eintrag in den Spaltensummen vorhanden ist, wird eine Spalte abgeschlossen. Dazu werden `maxReal` und `width` aus den Zählern berechnet. Es ist wichtig `maxReal` immer neu aus `region.minReal` zu berechnen, anstatt nur das Delta aufzuaddieren, da sich sonst der Fehler, der bei Fließkommaaddition unvermeidbar ist, auch mit aufaddiert. `minReal` und `hOffset` stehen bereits in der Region `tmp`, die Werte wurden bei der Berechnung der vorhergehenden Spalte bereits gesetzt. Jetzt wird `tmp` für die nächste Spalte vorbereitet, das heißt `tmp.minReal` wird auf `tmp.maxReal` gesetzt und `tmp.hOffset` wird um `tmp.width` erhöht. Ersteres vermeidet das Entstehen von Lücken zuverlässig. Außerdem wird `desiredN` für die verbleibenden Spalten nach (5) neu berechnet und die Zähler werden zurückgesetzt. Bevor die Aufteilung der Spalte in Teilregionen startet, wird eine Kopie der Vorhersage erstellt, die nur die Werte für die aktuelle Spalte enthält. Das Aufteilen einer Spalte in Teilregion geschieht analog zum Aufteilen in Spalten.

Die letzte Spalte wird gesondert behandelt: Sie muss so gewählt werden, dass sie den gesamten Rest der Eingaberegion abdeckt, ansonsten können Lücken entstehen. Dies gilt genauso bei der Unterteilung der einzelnen Spalten.

Die rekursive Variante der Strategie mit Vorhersage verwendet dasselbe Rekursions-schema wie ihr naives Gegenstück. Die Werte in `BalancingContext` berechnen sich also wie in (2) und (3). Auch die Abbruchbedingung ist wie in (4). Die Entscheidung, ob horizontal oder vertikal geteilt werden soll, wird auch auf die gleiche Art und Weise gefällt.

Bei dieser Strategie werden die Regionen allerdings nicht einfach halbiert, sondern in zwei Teile aufgeteilt, die laut der Vorhersage ähnlich rechenintensiv sind. Dazu wird so vorgegangen, wie bei der nicht-rekursiven Variante für die Aufteilung auf zwei Worker. Es ist hierbei auch wichtig die Vorhersage so zu teilen, dass es für jede Hälfte eine Vorhersage gibt, die dann an den rekursiven Aufruf übergeben werden kann.

```

1 #include "Region.h"
2
3 class Balancer {
4     public:
5         virtual Region* balanceLoad(Region region, int nodeCount) = 0;
6         virtual ~Balancer();
7 };

```

Quelltext 19: Das gemeinsame Interface der Lastbalancierung

3.6.3 Erweiterung

Da die Lastbalancierung nach dem Strategy-Pattern realisiert ist, gestaltet sich die Erweiterung um eine neue Balancierungsstrategie recht einfach. Zuerst muss eine Unterklasse von Balancer (Quelltext 19) erstellt werden, um ein gemeinsames Interface zu erzwingen und somit die polymorphe Nutzung der neuen Klasse zu ermöglichen.

Danach wird die neue Strategie über die Methode `BalancerPolicy::chooseBalancer` verfügbar gemacht. Dazu muss die neue Strategie mithilfe der statischen Variablen `Klassenname::NAME` benannt werden. `BalancerPolicy` muss nun so erweitert werden, dass, bei Eingabe des vorher festgelegten Namens, ein neues Objekt der entsprechenden Klasse zurückgegeben wird. In `BalancerPolicy` kann auch die Genauigkeit der Vorhersage für die oben aufgeführten Strategien festgelegt werden.

Bedingungen an `Balancer::balanceLoad` Bei der Eingabe von `region` und `nodeCount` erfüllt ein korrekter Rückgabewert `subregions` die folgenden Bedingungen:

- `subregions` ist ein Zeiger auf ein Array mit `nodeCount` Elementen vom Typ `Region`
- Die Regionen in `subregions` sind eine Partitionierung von `region`, d.h. sie überschneiden sich nicht und ihre Vereinigung ergibt genau `region`
- Für jede Region `subregion` in `subregions` gilt:
 - `guaranteedDivisor`, `validation`, `maxIteration` und `fractal` sind in `region` und `subregion` gleich
 - `subregion.guaranteedDivisor` teilt `subregion.width` und `subregion.height` ohne Rest
 - `subregion.hOffset` und `subregion.vOffset` sind so gesetzt, dass sie den Abstand der oberen linken Ecke von `subregion` zur oberen linken Ecke von `region` angeben
 - Die Deltas für die komplexen Werte sind unverändert, d.h. die Größe des Bereiches der komplexen Ebene, der von einem Pixel abgedeckt wird, ist unverändert

Ob diese Bedingungen erfüllt sind kann mit dem Test in der Klasse `BalancerTest` überprüft werden.

3.7 Berechnung der Mandelbrotmenge

3.7.1 Berechnung mithilfe von SIMD

Um die Berechnungen intern noch zu beschleunigen, kann SIMD zur parallelen Bearbeitung mehrerer Punkte verwendet werden. Dazu ist es hilfreich, sich zunächst vor Augen zu führen, wie ein Vektor komplexer Koordinaten ohne SIMD verarbeitet würde. Dies ist in Quelltext 20 zu sehen. Die Berechnung der einzelnen Koordinaten bleibt gleich, nur die Abbruchbedingung wird auf alle bearbeiteten Koordinaten erweitert.

Es muss dabei solange weiter iteriert werden, bis für alle Komponenten die Berechnung abgebrochen werden darf. Hierbei ist es kein Problem, mit den abgebrochenen Punkten weiter zu rechnen, sofern die Iterationszahl nur hochgezählt wird solange das z_n der Koordinate Betragmäßig kleiner gleich 2 ist. Dies gilt, da alle $|z_{n+i}| > 2$ sofern $|z_n| > 2$ [1].

Damit kann die Berechnung relativ simpel via Arm NEON Compiler Intrinsics¹⁷ implementiert werden (siehe dazu Quelltext 21). Diese Intrinsics ermöglichen eine Verwendung der nativen SIMD-Befehle, wobei der Compiler sich um die Verwendung der SIMD-Register kümmert. Dadurch wird lesbarer Code ermöglicht, der sich stärker am zu implementierenden Algorithmus orientiert.

Zu den hier benötigten mathematischen Operationen (z.B. der Addition) wird hierbei das Compiler Intrinsic nach folgendem Schema erzeugt: "vopcq_fpr" mit dem Operationscode *opc* (z.B. *add* für Addition). "v" ist das allgemeine Prefix für Vektoroperationen und "q" bedeutet, dass doppelt so viele Register verwendet werden wie ohne "q". Damit werden alle verfügbaren SIMD-Register des ARMv8-A Prozessors des ODroids und Raspberry Pi 3 B+ herangezogen. Das Postfix *fpr* bestimmt, dass die Register als Gleitkommazahlen der Präzision *pr* bit (in diesem Fall 32 oder 64) interpretiert werden sollen. Damit werden in jeder Operation 4 mal 32 bit Gleitkommazahlen oder 2 mal 64 bit Gleitkommazahlen verrechnet.

Bei der Implementierung wurden Optimierungen mithilfe der NEON-nativen multiply-add (*mla*) und multiply-subtract (*mls*) Befehle vorgenommen. Zudem kann der parallele Vergleich zweier Vektoren (*clt*) ausgenutzt werden, wobei als Ergebnis jedoch nicht 1 und 0 ausgegeben werden, sondern alle Bits der Ergebnisvektorkomponente auf 1 gesetzt werden sofern die Bedingung erfüllt ist und sonst auf 0. Um im weiteren Verlauf effizient die Vektorkomponenten aufzuaddieren (*addv*), sodass die Summe der Anzahl nicht abgebrochener Berechnungen entspricht, werden daher alle Komponenten des Ergebnisvektors des Vergleiches mit 1 verundet. Das Ergebnis des Vergleiches und der Verundung ist ein Vektor vorzeichenloser 32 oder 64 bit Ganzzahlen, weshalb das Postfix für diese Operationen *u32* oder *u64* lautet.

```

8 void MandelbrotVect::calculateFractal(precision_t* cReal, precision_t*
9   cImaginary, unsigned short int maxIteration, int vectorLength,
10  unsigned short int* dest) {
11    if(vectorLength <= 0){
12      throw std::invalid_argument("vectorLength may not be less than 1.
13      ");
14    }
15    std::fill_n(dest, vectorLength, 0);
16    precision_t* zReal = new precision_t[vectorLength];
17    precision_t* zImaginary = new precision_t[vectorLength];
18    precision_t* nextZReal = new precision_t[vectorLength];
19    precision_t* nextZImaginary = new precision_t[vectorLength];
20    // Factor that is added on iteration count
21    // is 0 or 1 and determines whether that point is still being
22    // computed
23    bool* factor = new bool[vectorLength];
24    std::fill_n(factor, vectorLength, 1);
25    // Integer storing number of Z components with absolute value
26    // below two => continue computation
27    unsigned int lessThanTwo = vectorLength; // as we begin with ZReal/
28    ZImag as 0
29    int i = 0;
30    while (i < maxIteration && lessThanTwo > 0){
31      lessThanTwo = 0;
32      for(int k = 0; k < vectorLength; k++){
33        // Compute next step in iteration
34        nextZReal[k] = (zReal[k] * zReal[k] - zImaginary[k] *
35          zImaginary[k]) + cReal[k];
36        nextZImaginary[k] = 2 * (zReal[k] * zImaginary[k]) +
37          cImaginary[k];
38        zReal[k] = nextZReal[k];
39        zImaginary[k] = nextZImaginary[k];
40        // Determine whether to stop
41        factor[k] = (zReal[k] * zReal[k] + zImaginary[k] * zImaginary
42          [k] < 4.0) ? 1 : 0;
43        // sum => if any number is still less than two, we need to
44        // continue
45        lessThanTwo += factor[k];
46        // increase number of iterations if this number wasnt aborted
47        yet
48        dest[k] += factor[k];
49      }
50    }
51    delete[] zReal;
52    delete[] zImaginary;
53    delete[] nextZReal;
54    delete[] nextZImaginary;
55    delete[] factor;
56  }

```

Quelltext 20: Bearbeitung eines Vektors komplexer Koordinaten in C++

```

29 // Load casted values from array to SIMD vector
30 float32x4_t cReal = vdupq_n_f32(0); // = vld1q_f32(cRealArray); if
   casting weren't necessary this would work
31 cReal = vsetq_lane_f32((float32_t) cRealArray[0], cReal, 0);
32 cReal = vsetq_lane_f32((float32_t) cRealArray[1], cReal, 1);
33 cReal = vsetq_lane_f32((float32_t) cRealArray[2], cReal, 2);
34 cReal = vsetq_lane_f32((float32_t) cRealArray[3], cReal, 3);
35 float32x4_t cImaginary = vdupq_n_f32(0);
36 cImaginary = vsetq_lane_f32((float32_t) cImaginaryArray[0],
   cImaginary, 0);
37 cImaginary = vsetq_lane_f32((float32_t) cImaginaryArray[1],
   cImaginary, 1);
38 cImaginary = vsetq_lane_f32((float32_t) cImaginaryArray[2],
   cImaginary, 2);
39 cImaginary = vsetq_lane_f32((float32_t) cImaginaryArray[3],
   cImaginary, 3);
40 // The z values
41 float32x4_t zReal = vdupq_n_f32(0);
42 float32x4_t zImaginary = vdupq_n_f32(0);
43 // Helper variables
44 float32x4_t two = vdupq_n_f32(2);
45 float32x4_t four = vdupq_n_f32(4);
46 uint32x4_t one = vdupq_n_u32(1);
47 // result iterations
48 uint32x4_t n = vdupq_n_u32(0);
49 // vector with 1 if absolute value of component is less than two
50 uint32x4_t absLesserThanTwo = vdupq_n_u32(1);
51 int i = 0;
52 // addv => sum all elements of the vector
53 while(i < maxIteration && vaddvq_u32(absLesserThanTwo) > 0){
54     // mls a b c -> a - b*c
55     float32x4_t nextZReal = vaddq_f32(vmlsq_f32(vmulq_f32(zReal,
      zReal), zImaginary, zImaginary), cReal);
56     // mla a b c -> a + b*c
57     float32x4_t nextZImaginary = vmlaq_f32(cImaginary, two, vmulq_f32
      (zReal, zImaginary));
58     zReal = nextZReal;
59     zImaginary = nextZImaginary;
60     // Square of the absolute value -> determine when to stop
61     float32x4_t absSquare = vmlaq_f32(vmulq_f32(zReal, zReal),
      zImaginary, zImaginary);
62     // If square of the absolute is less than 4, abs<2 holds -> 1
       else 0
63     absLesserThanTwo = vandq_u32(vcsltq_f32(absSquare, four), one);
64     // if any value is 1 in the vector (abs<2) then dont break
65     n = vaddq_u32(n, absLesserThanTwo);
66     i++;
67 }
68 // write n to dest
69 dest[0] = vgetq_lane_u32(n, 0);
70 dest[1] = vgetq_lane_u32(n, 1);
71 dest[2] = vgetq_lane_u32(n, 2);
72 dest[3] = vgetq_lane_u32(n, 3);

```

Quelltext 21: Parallelisierte Bearbeitung eines Vektors komplexer Koordinaten in 32 bit Gleitkommapräzision in C++ mit Arm NEON Compiler Intrinsics für SIMD

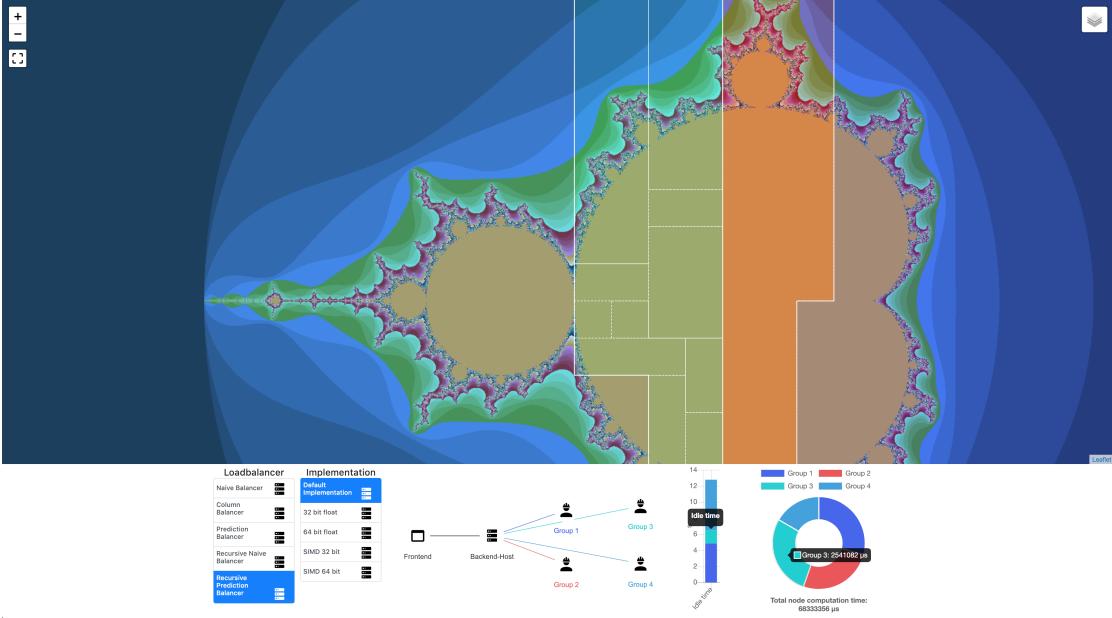


Abbildung 4: Benutzeroberfläche der Mandelbrot Anwendung

3.8 Erläuterung des Frontends

Um das Frontend zu implementieren, wurde sich für die Programmiersprache TypeScript¹⁸ (Erweiterung von JavaScript um Typisierung) entschieden. Da diese zu JavaScript kompiliert, werden die Vorteile von JavaScript, wie Ausführung im Webbrowser des Benutzers und eine vielzahl verfügbarer Bibliotheken, mit den Vorteilen einer typisierten Programmiersprache vereint.

Für die graphische Benutzeroberfläche wurde ebenfalls TypeScript mit dem React Framework¹⁹ verwendet, welches es ermöglicht, graphische Komponenten nativ in TypeScript zu erstellen und verändern. Es wurde für jede Komponente eine eigene TypeScript Klasse zu erstellt, welche dann den betreffenden das betreffende Verhalten und dessen Darstellung enthält.

Um das Fraktal darzustellen wird die Leaflet²⁰ Bibliothek verwendet. Diese, für Onlinekarten konzipierte, Bibliothek stellt den Bereich der komplexen Ebene, auf der die Mandelbrotmenge liegt dar. Dabei wird der momentan sichtbare Ausschnitt der Menge mit Hilfe der WebSockets Verbindung (siehe Unterunterabschnitt 3.8.1) an das Backend versendet und vom ausgewählten Lastbalancierer in Teilregionen unterteilt (siehe Unterabschnitt 3.6). Jeder der vom Backend berechneten Teilregionen wird, sobald diese empfangen wurden, im Frontend angezeigt und die Komponenten zu Visualisierung

¹⁷ Details im Abschnitt ‘Compiler Intrinsics’ unter <https://developer.arm.com/technologies/neon>

¹⁸ <https://www.typescriptlang.org/>

¹⁹ <https://reactjs.org/>

²⁰ <https://leafletjs.com/>

der Rechenzeit aktualisiert.

3.8.1 Kommunikation mit dem Backend

Zur Kommunikation mit dem Backend wird im Frontend ein Objekt der Klasse `WebSocketClient` erzeugt. Die zur Verbindung verwendeten Codestücke liegen dabei alle in dem Ordner `connection`, die genannte Klasse wird in `WSClient.ts` definiert.

Die dort definierte Klasse `WebSocketClient` abstrahiert von dem JavaScript-nativen `Websocketinterface` `WebSocket`²¹. Bei der Initialisierung baut das erzeugte Objekt eine Verbindung zu der lokalen Adresse `ws://localhost:9002` auf²². Dort sollte das Backend bereit sein, eine Websocketverbindung anzunehmen.

Zudem bietet die Klasse die Methode `sendRequest`, welche ein übergebenes Objekt JSON-kodiert und versendet.

Wichtig für den Rest des Frontends sind die Methoden `registerRegion` und `registerRegionData`. An diese Methoden können Callbacks übergeben werden, die aufgerufen werden, wenn das Frontend über die Websocketverbindung respektive ein Region-Objekt oder ein RegionData-Objekt empfängt. Diese sind respektive die Aufteilung einer angefragten Region (siehe Quelltext 17) oder die berechneten Iterationswerte einer Region (siehe Quelltext 18). Die übergebenen Callbacks erhalten als Parameter respektive die vorgruppierte Aufteilung als ein Array von Workergruppen (`RegionGroup`) oder das JSON-dekodierte `RegionData`-Objekt.

Die Definitionen der zugehörigen Objekt-Interfaces finden sich in den Dateien `RegionGroup.ts` und `ExchangeTypes.ts`.

Als weitere Hilfsfunktion wird in `RegionRequest.ts` `request` bereitgestellt. Diese extrahiert aus der übergebenen Sicht auf die Mandelbrotmenge, die in der Leaflet-Karte gespeichert ist, die Parameter zum Anfragen einer Region. Dazu werden mithilfe des aktuellen Zooms der linke obere und rechte untere Punkt des Sichtbereiches in dem Leaflet-internen Koordinatensystem auf entsprechende Punkte in der komplexen Ebene projiziert. Da zum Erzeugen des passenden Objektes auch der gewünschte Lastbalancierer und der Fraktaltyp notwendig sind, werden diese als weitere Parameter übergeben. Die Funktion gibt direkt ein Objekt zurück, dass das Interface `RegionRequest` erfüllt.

3.8.2 Darstellung der Regionsdaten

Die für die Darstellung der Mandelbrotmenge verwendete Bibliothek (leaflet) hat die Einschränkung, dass nur Quadrate einer vordefinierten Größe (`Tiles`) angezeigt werden können. Ebenfalls verwendet diese ein eigenes Koordinatensystem, unter welchem jede angezeigte Tile eindeutig mit dem Tripel $(x, y, zoom)$ identifiziert wird. Somit ist eine Übersetzung zwischen den Daten des Backends und den von leaflet erwarteten nötig.

²¹<https://developer.mozilla.org/en-US/docs/Web/API/WebSocket>

²² Faktisch wird eine Verbindung geöffnet, geschlossen und erneut geöffnet. Dies ist durch ein ungelöstes Problem beim Verbindungsauftbau bedingt, das dafür sorgt, dass ein Verbindungsauftbau erst bei der zweiten erzeugten Websocketverbindung fehlerfrei gelingt.

MatrixView.ts, RegionOfInterest.ts Die Klasse MatrixView.ts implementiert die Umsetzung einer vom Backend versendeten Region zu den von leaflet erwarteten Tiles. Dafür registrieren sich alle sichtbaren Tiles durch die Methode registerTile(point, draw) mit dem Callback draw, welcher ausgeführt wird, sobald die anzuzeigenden Daten für die entsprechende Tile verfügbar sind. Diese Iterationswerte des Backends werden dabei der draw Funktion als Parameter in Form eines RegionOfInterest Objekts übergeben.

Ein RegionOfInterest Objekt implementiert wiederum die Übersetzung von lokalen (x, y) Pixel-Werten einer Tile zu Indizes in die vom Backend gesendeten Regionsdaten (siehe Quelltext 18). Dabei gibt die get(x, y) Methode, für einen Tile (x, y) Pixel-Wert die benötigte Iterationsanzahl zurück (siehe Quelltext 22).

```

39   public get(x: number, y: number): number {
40     if (x > this.ROIWidth || y > this.ROIHeight) {
41       return -1;
42     }
43     let realX = this.topLeft.x + x;
44     let realY = this.topLeft.y + y;
45     return this.data[realY * this.width + realX];
46   }

```

Quelltext 22: Umsetzung von Pixel-Werten der Tiles zu Indizes in Regionsdaten

TileDisplay.tsx Die Klasse TileDisplay.tsx verwendet die leaflet Bibliothek direkt, um die Regionsdaten darzustellen. Dabei wird dieser die WebSocket Verbindung in Form eines WebSocketClient, sowie der vom Nutzer gewählte Balancer, die Implementierung und Gruppierung durch Observable Klassen (siehe Abschnitt 3.8.4) übergeben. Ebenfalls wird der anzuzeigende Ausschnitt der Mandelbrotmenge übergeben.

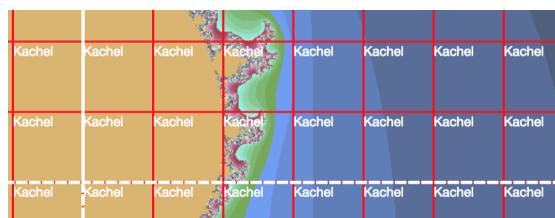


Abbildung 5: Relation von Backend Regionen zu leaflet Tiles

Da das Backend die berechneten Teilbereiche der Mandelbrotmenge als Regionen zurück gibt, dessen Höhe und Breite Vielfache der leaflet Tile-Größe sind, wird eine Region dem Benutzer durch mehrere Tiles dargestellt. Dieses Verhältnis wird in Abbildung 5 dargestellt. Dabei ist beispielhaft eine Region des Backends weiß eingezeichnet, alle leaflet Tiles sind rot umrandet angegeben. Eine detailliertere Version dieser Anzeige kann durch die Auswahl des DebugLayers jederzeit im Frontend eingeblendet werden.

Dargestellt werden die Tiles in einer eigenen Ebene der leaflet Karte, dem MandelbrotLayer.

Dieser erstellt für den sichtbaren Bereich²³ alle benötigten Tiles. Für jede der Tiles wird ein HTML5 `canvas`²⁴ Objekt erstellt, welches es ermöglicht, für jeden Pixel einen (r, g, b) Farbwert zu definierten und anzuzeigen. Wobei die Farbwerte aus den berechneten Iterationswerten des Backends mit einem Shader (siehe Abschnitt 3.8.2) ermittelt werden. Die Iterationswerte können wiederum mit einem `MatrixView` Objekt aus den Regionsdaten des Backends gelesen werden.

Weiterhin wird in `TileDisplay` der `WorkerLayer` (siehe Abschnitt 3.8.2), welcher die Regionsaufteilung des Backends visualisiert, hinzugefügt und callbacks auf die Observables definiert, welche die momentan sichtbare Region erneut vom Backend anfordern, falls sich einer der ausgewählten Lastbalancierer, oder die gewählte Implementierung ändert.

Shader.ts Berechnet für einen Iterationswert der Mandelbrotmenge ein Tripel (r, g, b) von Farbwerten. Implementiert ist eine simple Funktion, welche den Iterationswert für jeden Farbkanal mit einer Konstante multipliziert und auf den zulässigen ganzzahligen Wertebereich $[0, 255]$ abbildet (siehe Quelltext 23).

```

2  public static default(n: number, maxIteration: number): number[] {
3    const r = (n * 10) % 256;
4    const g = (n * 20) % 256;
5    const b = (n * 40) % 256;
6    return [r, g, b];
7 }
```

Quelltext 23: Berechnung der (r, g, b) Farbwerte

Project.ts Da die leaflet Bibliothek Tiles verwendet, um die Regionen des Backends anzuzeigen, liegen diese Tiles in einem neuen Koordinatensystem, welches jeder Tile auf einer Zoomstufe ein Triple $(tileX, tileY, zoom)$ zuordnet. Weiterhin besitzt leaflet ein internes Koordinatensystem (CRS²⁵), welches jeden Punkt auf der Karte durch das Paar $(latitude, longitude)$ identifiziert. Projekt.ts enthält Funktionen, um zwischen diesen 3 Koordinatensystemen (Tile-Koordinaten, leaflet-Koordinaten und Koordinaten der komplexen Ebene) zu konvertieren.

Regionsgruppierung (RegionGroup.ts)

WorkerLayer.ts Diese Klasse visualisiert die Regionsaufteilung des Lastbalancierers als Overlay, welche über den Iterationswerten dem Benutzer angezeigt wird. Dafür wird eine in leaflet bestehende GeoJSON API verwendet, mit welcher es möglich ist, beliebige

²³Da die Fenstergröße des sichtbaren Bereichs kein Vielfaches der Tilegröße sein muss, können Tiles erzeugt werden, welche teilweise außerhalb des sichtbaren Bereichs liegen

²⁴https://developer.mozilla.org/kab/docs/Web/API/Canvas_API

²⁵https://en.wikipedia.org/wiki/Spatial_reference_system

Polygone auf den bestehenden Kartendaten anzuzeigen. Die Eckpunkte für diese Polygone für jede RegionGroup mit Hilfe der Funktionen aus Project von Koordinaten der komplexen Ebene, welche im Backend verwendet werden, zu leaflet Koordinaten umgerechnet.

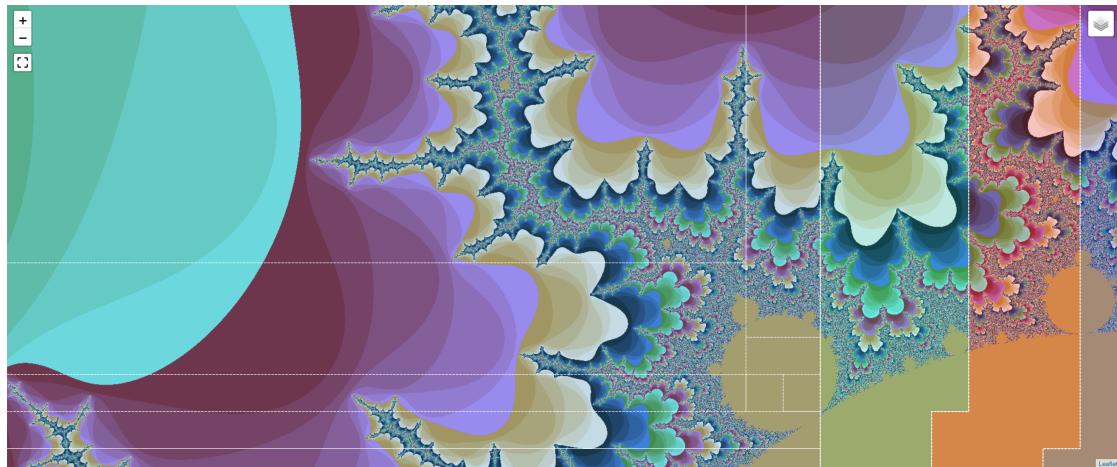


Abbildung 6: Gruppierung einer Aufteilung des Recursive PredictionBalancers mit 37 Workern

Da es wie in Abschnitt 3.8.2 beschrieben zu einer Gruppierung kommt, falls die Anzahl der Worker im Backend zu groß ist, werden ebenfalls alle Untergruppen einer Gruppe angezeigt (siehe Abbildung 6), falls der Benutzer mit der Maus über eine der dargestellten Gruppierungen geht.

3.8.3 Visualisierung der Rechenzeiten

3.8.4 MISC

Observable.ts

4 Ergebnisse / Evaluation

- Skalierbarkeitsgraph
- Wie gut ist SIMD / OpenMP / MPI / Mischformen?
- Frontend Overhead messen

5 Zusammenfassung

- Zusammenfassung
- Ausblick

Abbildungsverzeichnis

1	Die Mandelbrotmenge, visualisiert in einem Ausschnitt des komplexen Zahlenraumes.	4
2	Architekturübersicht	7
3	Konzept der Koordinaten in den Regionsobjekten. Alle Koordinaten beziehen sich auf die Darstellungsebene und sind daher in Pixeln.	17
4	Benutzeroberfläche der Mandelbrot Anwendung	28
5	Relation von Backend Regionen zu leaflet Tiles	30
6	Gruppierung einer Aufteilung des Recursive PredictionBalancers mit 37 Workern	32

Tabellenverzeichnis

Literatur

- [1] mrf (<https://math.stackexchange.com/users/19440/mrf>). Why is the bailout value of the mandelbrot set 2? Mathematics Stack Exchange. URL:<https://math.stackexchange.com/q/424331> (version: 2013-06-24).
 - [2] zaphoyd (<https://www.zaphoyd.com/>). Websocket++ user manual. Zaphoyd.com. URL:<https://www.zaphoyd.com/websocketpp/manual/reference/thread-safety> (version: 2013-03-02).
-