

PROJECT REPORT

FOUNDATIONS OF HIGH PERFORMANCE COMPUTING 2022-2023

Guido Cera

Table of Contents

- 0. [Introduction](#)
 - 0.1 [Game](#)
 - 0.2 [Orfeo](#)
- 1. [Methodology](#)
 - 1.1 [Static Evolution](#)
 - 1.2 [Ordered Evolution](#)
 - 1.3 [Multithreading](#)
 - 1.4 [Multiprocessing](#)
 - 1.5 [Board generation](#)
 - 1.6 [Read and write image](#)
 - 1.7 [Time tracking](#)
 - 1.8 [Experimentation Environment](#)
- 2. [Implementation](#)
 - 2.1 [Code management](#)
 - 2.2 [Cells states](#)
 - 2.3 [Static Evolution](#)
 - 2.4 [Board generation](#)
 - 2.5 [check_neighbours](#)
 - 2.6 [Multiprocessing](#)
- 3. [Results & Discussion](#)
 - 3.1 [OpenMP scalability](#)
 - 3.2 [Strong MPI scalability](#)
 - 3.3 [Weak MPI scalability](#)
- 4. [Conclusions](#)
 - 4.1 [Simpler splitting of the grid](#)

0. Introduction

The project consists of a scaling study on a personal and highly parallel implementation of the so called Conway's Game of Life.

0.1 Game

Game of Life is played on a grid, where each cell can be either dead or alive. Starting with a grid in a state - with some cells alive and some dead - the grid "evolves", meaning each cell keeps or changes its state depending on the state of the neighbouring cells. The specific implementation for the evolution rules used in this project is presented in the assignment description (github.com) and varies slightly from the classical rules of Conway's Game of Life. The grid is virtually infinite, the outer cells of the grid use the cells at the opposite

site as neighbours; imagine that the grid multiplies and each copy sits adjacent to the original in the same orientation. In a computer the cells have to be updated one by one and this offers multiple possibilities in how the cells are navigated and when they are updated; the two methods I implemented are called **ordered evolution** and **static evolution**. The former is perhaps the simplest both conceptually and in code, it is a loop over all cells in order and each is updated right away, before passing on to the next cell. The latter method holds the update of the cells after all cells have been checked and their next state have been decided. While in the ordered case the neighbours of a cell will be half in the old state and half in the new one, for the static evolution all neighbours will be in the same state at the time of determining the new state of a cell.

Interesting fact, the classic implementation of Conway's Game of Life is Turing complete. ([Wikipedia](#))

0.2 Orfeo

This project is being developed to run specifically on the high performance computer Orfeo located in Area Science Park, Trieste ([areasciencepark.it](#)). This is important for the parallelization of the code, which has to be built using the resources available on Orfeo ([ofeo-doc.areasciencepark.it](#)). Thus the code will contain an hybridization of MPI ([open-mpi.org](#)) - for multiprocessing - and OpenMP ([openmp.org](#)) - for multithreading.

1. Methodology

From an abstract point of view the core of the project - the Game of Life - is pretty straightforward since it comprises of a matrix, some nested for loops and few simple checks to update the cells. I approached the project implementing this core structure first and then dwelling in the parallelization; and even if I have experience in multi-processing and multi-threading it felt like going from swimming in a pool to being thrown in the ocean during a storm. The most interesting choice of algorithm I encountered at this stage was approaching the static evolution [1.1]. Now, I do not want to give the impression that no optimization can be done, on the contrary, there are a great number of articles and book's chapters that analyze the problem and make use of clever details of computer architecture or theory on sparse matrices and data structures in general to shrink dramatically the number of operations needed to play the game; here are some links to provide an overview of the optimization landscape for Game of Life.

- ([stackoverflow.com](#)) Post with some good answers and links to useful sources;
- ([docplayer.net](#)) Make use of theory on **sparse matrices**;
- ([jaregoy.com](#)) Michael Abrash's Graphics Programming Black Book, chapter 17 spends 33 pages to optimize Game of Life;
- ([Wikipedia](#)) **Hashlife** is an algorithm that makes use of hash tables and reaches impressive speeds, with the drawback of occupying large amounts of memory
- ([dotat.at](#)) An implementation of Game of Life which uses **bit-wise operations** and **lookup tables**.

I want to stress the many possibilities of optimization since the philosophy of coding for an high performance computer is all about that, but one of the focuses of this course is parallelization, in particular through MPI and OpenMP, and the objective of this project is to show the scalability of multithreading and multiprocessing. In light of all this I decided to focus on implementing a good parallelization and I left a simple representation of the game underneath; which might actually help showing the power of multiprocessing when scaling.

The following subsections explore the choices made in specific situations or show the methodological approach to specific subproblems.

1.1 Static Evolution

For this approach to evolution I need first to check all the cells and determine their next state and then, as a separate step, I can update the grid. This way the current state of the table is the only one used to update it. The problem resides in the computed next state, where do I put it while I check the following cells if I cannot use the grid itself? I need a way to store the next state until all cells have been checked and their next state computed. The way I store the next state needs also to be easily accessible and fast to read because as soon as I am done writing all next states there I will read them and transfer them to the actual grid.

I saw two possible solutions to this problem, each with their own positives and negatives; one being to use middle-states to mark cells to modify and the other being to store two separate lists with the coordinates of cells to kill and to revive. Follows an explanation and then an analysis of the two approaches in terms of memory needed, computational load and ease of use.

- **Middle-states.** With middle-states I mean adding states that a cell can be in other than "alive" and "dead". In particular I add the states "will die" to mark a cell that is alive but will have to die at the next evolution step, and "will live" to mark a cell that is dead but will live at the next evolution step. This approach consists of marking the cells directly on the grid, without any additional data structure, but in a way that still leaves visible the current state of the grid so that checking the following cells will not be influenced by the mark. The drawback of this approach presents itself at the moment of updating the grid for the next state, here I have to change all cells in a middle-state to the state they are meant to have. There are no shortcuts to accomplish this, an additional full scan of the grid is necessary to find the middle-states and write the corresponding final state.
- **Separate Lists.** Two separate dynamic lists store coordinates, one of the cells to revive and the other of the cells to kill. This approach uses additional memory but minimizes the effort to update the grid afterwards, since I only have to loop through the lists and access only the cells that need modifying and no other.

From initial experiments I carried out I formed some expectations on the behaviour of this game.

Unfortunately I did not save the precise numbers but my takeaway is that a significant number of cells change state each evolution step if we start from a random initial state [2.4]. Translated to the two solutions above, this means that the lists will occupy a significant amount of memory each step, and if I update the memory allocated for each list doubling it when it is full I could have to allocate up to four times the size of the grid at each evolution step. Moreover middle-states are much easier to handle than dynamic memory. All this is why I opted for the middle-states solution instead of using separate lists. Further implementation details can be found in the next section [2.3].

1.2 Ordered Evolution

The ordered evolution is simple in its implementation and does everything in a couple of nested loops; but this same simplicity is also reason reason of impairment for parallelization, and consequently scalability performance. Not all parallelization is impossible, multiprocessing works and is as effective as for the static evolution, only multithreading cannot bring any benefit and the reason is in the ordered nature of the approach and the way OpenMP parallelizes the work; let's analyze both in order. The ordered evolution is intrinsically serial because when one cell gets updated its state will affect the other cells coming after that are around it. OpenMP handles multithreading of for loops assigning iterations to the threads, the assignment can be static or dynamic but in both cases a thread with its assigned iterations has to wait for all previous iterations to be done before he can execute its own. In the end, the whole computation is serial even if it is

carried on multiple threads. For this reason I did not implement multithreading for this evolution. MPI implementation is the same for both evolutions and it is discussed in section [1.4].

1.3 Multithreading

I choose to focus my multithreading effort only on the heaviest part of the code, which is the evolution of the board itself. All other components that are necessary but only function as a frame to the main part, even if they could exploit multithreading, I choose to leave single threaded. The reason for my choice was code clarity but most importantly it should not affect overall performance because all initializations and the set up for multiprocessing [1.4] is orders of magnitude faster than the evolution of the board. To cite one of the sources I linked above, Michael Abrash in his book *Graphics Programming Black Book* says this.

The first rule of optimization is: Only optimize where it matters.

Where I did apply multithreading I tried to be strict with the data handling, all outside variables used inside the openMP area are explicitly marked and even the division policy I fixed the one which makes the most sense.

1.4 Multiprocessing

As I anticipated above there is a little set up to do for MPI to work properly and efficiently. Main reason being the need for explicit communication between the processes which for Game of Life, at least as I implemented it, is very much essential; but let's start from the beginning and explain why there is this need. My reasoning is pretty simple, I have a grid that I need to split up in blocks to hand off one per process. So this I did, the grid gets divided in rectangles and each gets sent to a process. The rectangles are all of the same size, so that each process has the same workload. Here comes the communication problem, the outer cells in a block have some of the neighbours in other blocks and thus in other processes, where they can no longer be accessed. Now I need to carefully propagate all outer cells of all blocks to the right corresponding blocks; top row, right column, bottom row and left column go respectively to top block, right block, bottom block and left block, but the four cells in the corners also need to be sent to their respective block in diagonal of the origin block. All this propagation has to be done at each evolution step before updating the grid and before checking the neighbours.

Last thing to consider for inter-process communication is the need to save a snapshot of the grid during the evolution. When this necessity arises all blocks are gathered in the root process, the snap is saved and then the blocks are sent back each to the same process it came from, before finally proceeding with the evolution.

There are many operations that need to be computed by a single process, I picked the root process to execute them. Picking the root process specifically ensures that the program will run even if launched with only one process.

1.5 Board generation

The assignment requires a mode where a grid gets generated given its dimensions but there is no indication as what the state of this grid should be, so here is what I did. The simplest solution would be to return an image of the requested size with always the same initial state. This would be possible since the minimum size of the grid is 100x100, the requirement would be for the state to be contained in this limit, and all the extra space could be all dead cells. What I did instead is generating a random state each time. Generated the image I iterate through it and with a probability of 0.2 a cell gets marked as alive.

Generating a new grid is one of the tasks my program executes with only one process, but I implemented multithreading since in initialization mode creating a new grid is all the program does. Multithreading brings a problem with the random number generation that I discuss in section [2.4].

To create a different seed for each grid instead of going all fancy with the `time.h` library I implemented a little seed generator using the string containing the name to give the final file. This works on the assumptions that the name of the image would contain some information about the size of the grid and there is no need to generate different grids of the same size. Since I am the only user of this code I can assure that these assumptions are met.

1.6 Read and write image

I will not comment on any function in the file `read_write_pgm_image.c` since I did not personally write it. I just used the functions in there, hopefully the way they were intended to.

1.7 Time tracking

I used three different timers to gather data on the execution. The first, most general, timer I call "whole evolution timer" because it starts after the initial set up of reading the external parameters, reading the image and initializing MPI and just before launching the evolution. This timer ends just after the evolution. A second timer is all inside the evolution function but still times the whole evolution and is outside all evolution loops. Compared to the first timer it excludes the division of the grid into blocks [1.4]. The third and final timer is actually a measure of average time; the average time to propagate the outer cells of the blocks between processes. I decided to implement the last one because I was worried that it would take a significant portion of the execution; turns out I was wrong as you can see in the results section [3].

1.8 Experimentation Environment

At the time I executed my program and got the tests results Orfeo was pretty busy all the time with, I guess, other students doing their own projects. I had a little hope I could run more tests than the strictly necessary but in the end I did not manage to. I ran all tests on EPYC nodes, I only touched THIN nodes for the second exercise. More over I only used OpenMPI and never IntelMPI. I do not have much to say about this because these were not choices I made but rather constraints I found myself in.

2. Implementation

2.1 Code management

- How is the code managed, what is in which files, why is it.

2.2 Cells states

- The matrix that stores grid, and also all auxiliary variables and arrays that store cells, are of type `unsigned char`. Firstable `char` is the type that occupies the least memory in C at 1 Byte, which is plenty to store the numbers from 0 to 255 that the grid uses. Secondable since, as just noted, the numbers to be stored are all positive using `unsigned` provides an hardcoded guard against negative values; this way the code better resembles the logic behind it.

2.3 Static Evolution

- 128,127: values to mark cells that will change

2.4 Board generation

- `random_board()`, multithreading problem with `rand()` and the seed. <https://pvs-studio.com/en/blog/posts/0012/>

2.5 `check_neighbours`

- `check_neighbours()`

2.6 Multiprocessing

- not bundling together the sends and receives of the propagation for clarity

2.7 Multithreading

Since I choose a simple approach to the Game of Life all the heavy computation is inside for loops. Since each iteration carries out almost the same amount of computation (check neighbours, if the state has to be changed, change it), there is no need to dynamically assing the iterations to the threads, I can save on the overhead and assign statically. The computational differences between iterations

3. Results & Discussion

for all tests:

- commands of allocated resources
- commands for running experiment
- data plots
- explanation

3.1 OpenMP scalability

```
salloc -N1 -p EPYC -n128 --time=1:0:0
export OMP_PLACES=cores
srun -n1 --cpus-per-task=64 mpirun --map-by socket gol.x -r -f
start_1000px.pgm -n 1000 -e 1 -s 1000
```

```
size,steps,evolution,"processor number","thread number","total time","evolution time","average propagation
time" 1000,1000,1,1,1,72635616,72632154,7 1000,1000,1,1,2,36536630,36534854,7
1000,1000,1,1,4,18318647,18317337,7 1000,1000,1,1,8,9218737,9217823,15
1000,1000,1,1,16,4652385,4651514,19 1000,1000,1,1,32,2367646,2366288,21
1000,1000,1,1,64,1302963,1300866,24
```

3.2 Strong MPI scalability

3.3 Weak MPI scalability

4. Conclusions

4.1 Simpler splitting of the grid

The simplest method for splitting the grid would be to assign to each process groups of n rows or columns, depending on which is major, where n is the result of the size of the matrix divided by the number of processes. This approach has the same number of cells to propagate as the one implemented but is simpler because they are all grouped in only two rows. There would be no need for external columns propagation.