

Parallel and distributed computing with Julia

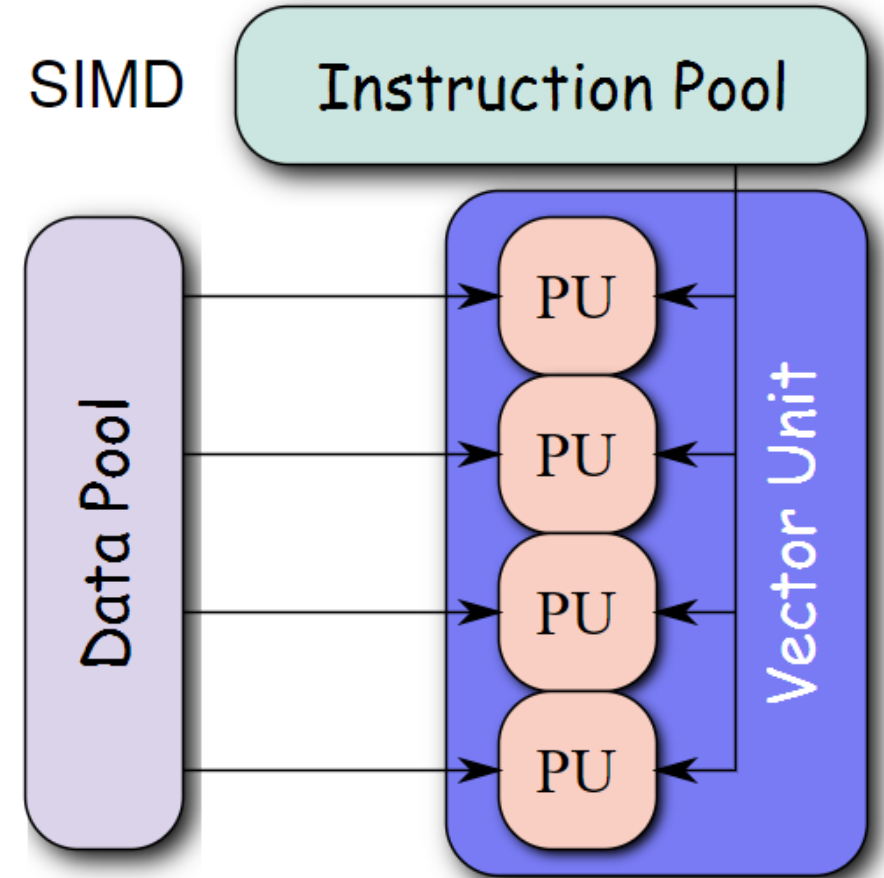
Bogumił Kamiński & Przemysław Szufel

Parallelization options in programming languages

- Single instruction, multiple data (SIMD)
- Green-threads
- Multi-threading
 - Language
 - Libraries
- Multi-processing
 - single machine
 - distributed (cluster)
 - distributed (cluster) via external tools
- GPU/TPU computations

SIMD

- Single instruction, multiple data (SIMD) describes computers with multiple processing elements that perform the same operation on multiple data points simultaneously. Such machines exploit data level parallelism, but not concurrency: there are simultaneous (parallel) computations, but only a single process (instruction) at a given moment.



Source: <https://en.wikipedia.org/wiki/SIMD>

Green threading

- In computer programming, green threads are threads that are scheduled by a runtime library or virtual machine (VM) instead of natively by the underlying operating system. Green threads emulate multithreaded environments without relying on any native OS capabilities, and they are managed in user space instead of kernel space, enabling them to work in environments that do not have native thread support.

https://en.wikipedia.org/wiki/Green_threads

Comparison of parallelism types

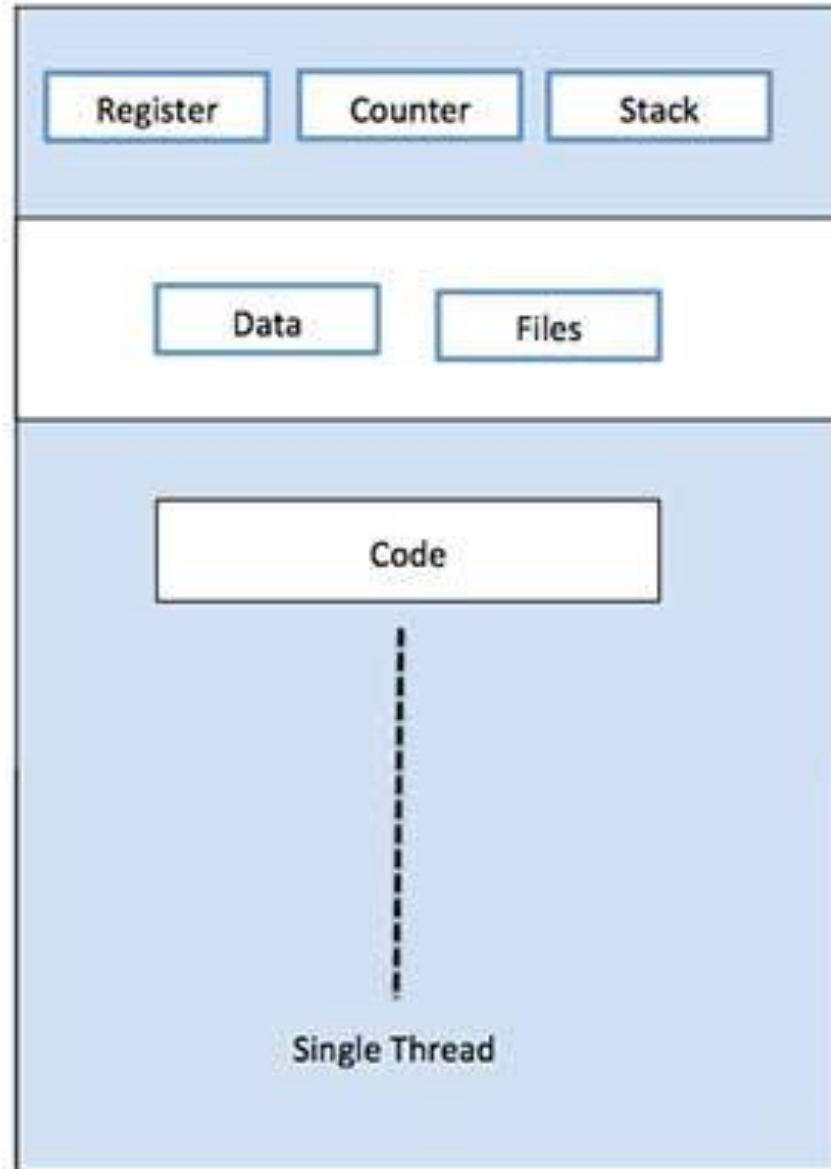
Threading

- Single process (cheap)
- Shared memory
- Number of threads running simultaneously limited by number of processors
- Possible issues with locking and false sharing

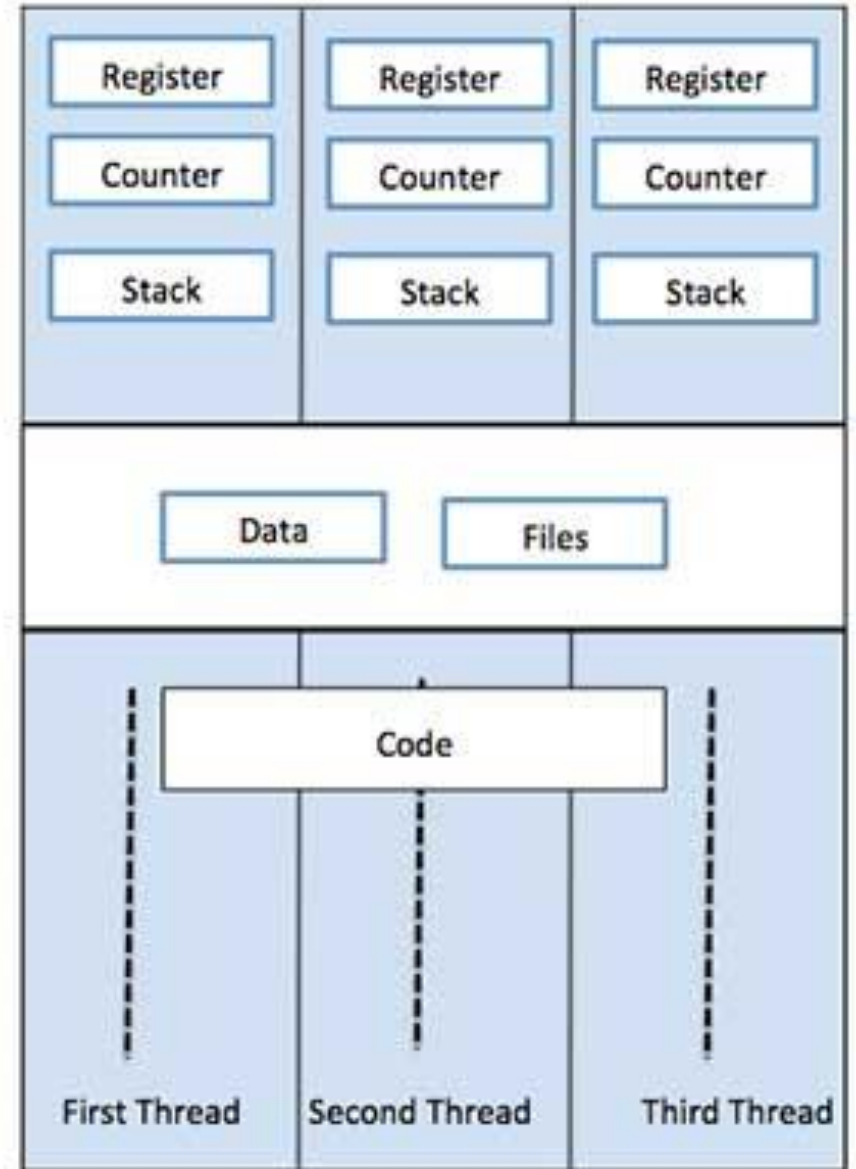
Multiprocessing

- Multiple processes
- Separate memory
- Number of processes running simultaneously limited by cluster size
- Possible issues if inter-process communication is needed

Threading



Single Process P with single thread



Single Process P with three threads

Sending data across cluster nodes

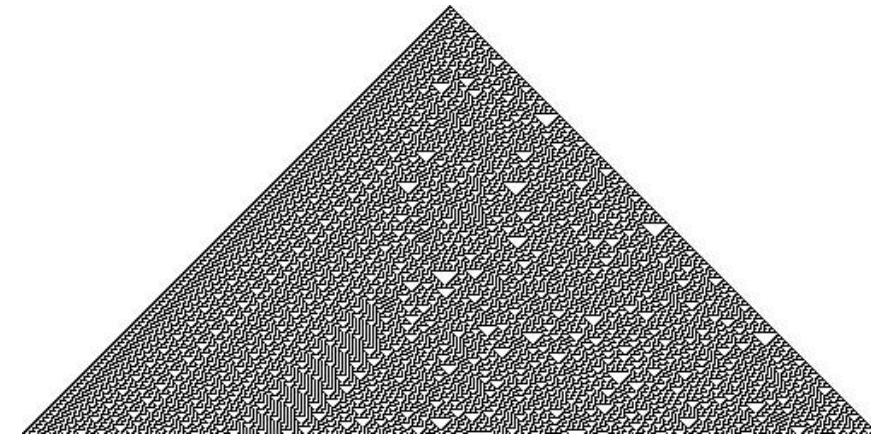
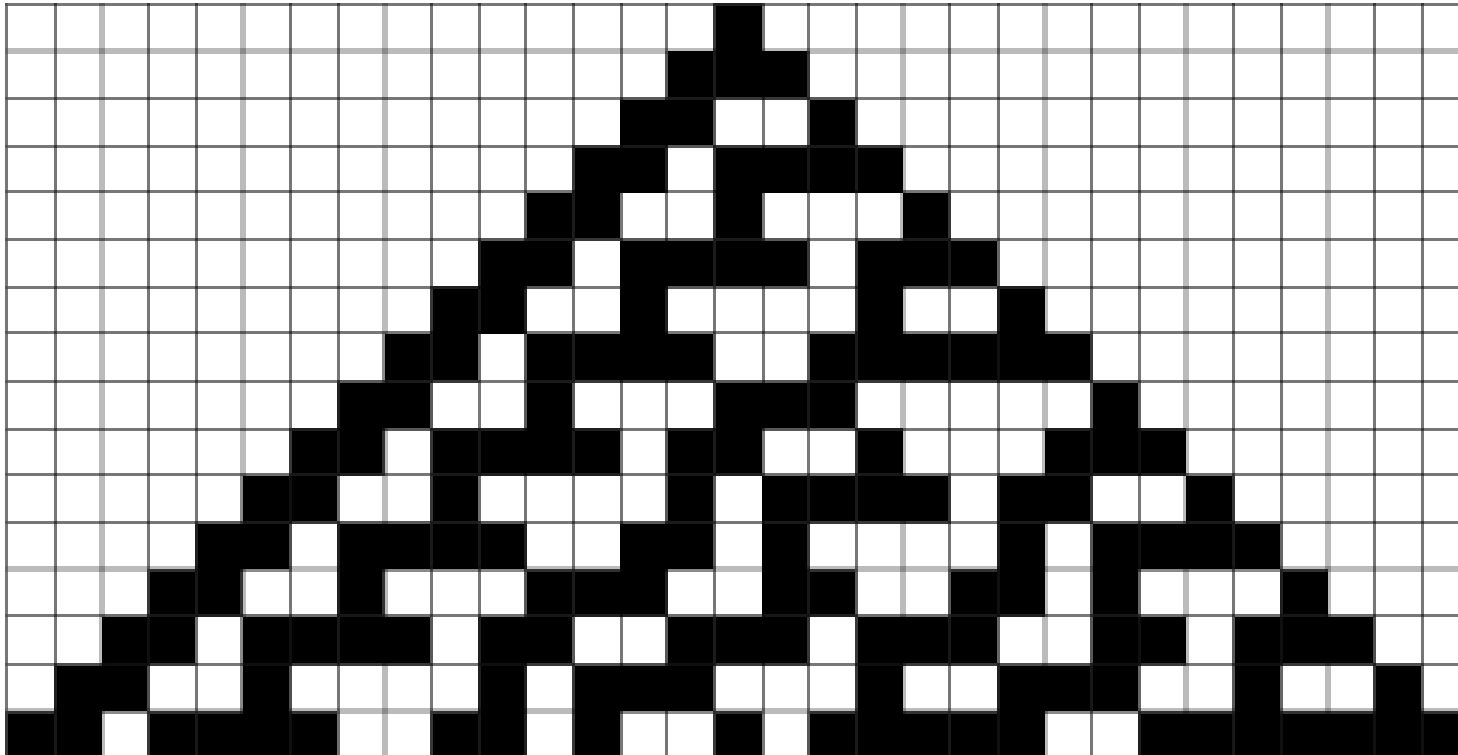
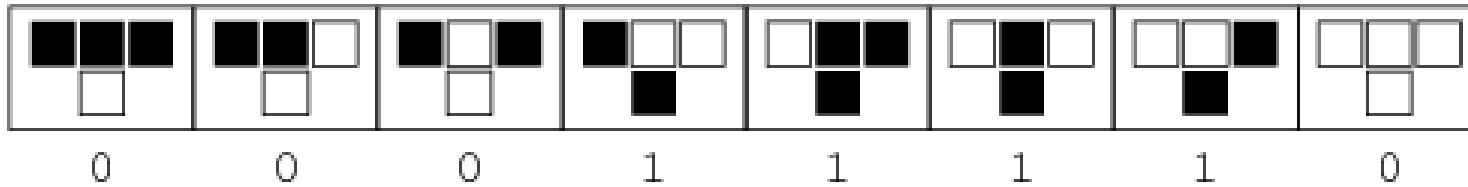
`@everywhere` using `ParallelDataTransfer`

```
sendto(workerid, vara = vara)
```

```
sendto([workerid1, workerid2], varb = varb)
```

Cellular automaton

rule 30

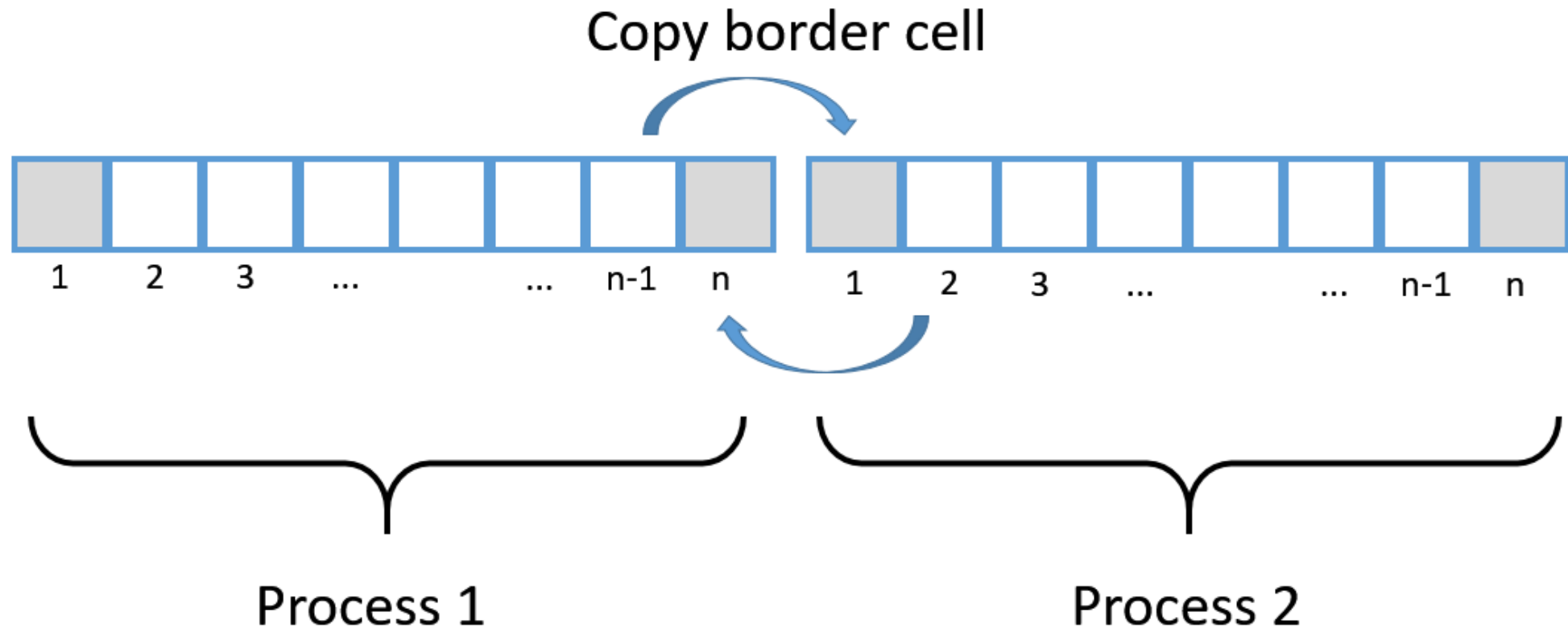


Source:

<http://mathworld.wolfram.com/Rule30.html>

Distributed cellular automaton

- Distributing data among worker processes



Julia cluster specification file and running distributed clusters (example file)

```
$ more machinefile_julia
```

```
4*ubuntu@172.31.10.229
```

```
4*ubuntu@172.31.11.44
```

```
4*ubuntu@172.31.0.243
```

```
4*ubuntu@172.31.13.134
```

```
4*ubuntu@172.31.14.219
```

```
$ julia --machinefile machinefile_julia program.jl
```

```
# REQUIRES PASSWORDLESS SSH TO BE CONFIGURED!
```

Julia set

We have a function

$$f(z) = z^2 - \frac{1}{2} + \frac{3}{4}i$$

If we start from some complex value z_0 and iterate

$$z_{n+1} = f(z_n)$$

How long does it take for this sequence to make $|z_k| > 2$?