Trace comparison using a sequence alignment algorithm

Alef Farah and Lucas Mello Schnorr

Informatics Institute

Federal University of Rio Grande do Sul

Porto Alegre, Rio Grande do Sul, Brazil

.inf INSTITUTO DE INFORMÁTICA UFRGS

CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico

UFRGS UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Overview

- <u>Concepts</u>
- Motivation
- Challenges
- Strategy
- Results
- Suggestions

Concepts

- Trace file
  - Sequence of events
    - Function calls etc
  - Performance analysis
  - Timestaped
    - In depth analysis

```
12 0.000000000 rank0 STATE main
12 0.000003679 rank0 STATE MPI_Init
12 0.007094880 rank1 STATE main
12 0.007098212 rank1 STATE MPI_Init
14 0.256268134 rank1 STATE
12 0.256271425 rank1 STATE MPI_Comm_size
14 0.256273626 rank1 STATE
12 0.256274072 rank1 STATE MPI_Comm_rank
14 0.256274449 rank1 STATE
12 0.256274773 rank1 STATE MPI_Finalize
14 0.256275228 rank0 STATE
```

Overview

- Concepts
- <u>Motivation</u>
- Challenges
- Strategy
- Results
- Suggestions

Motivation

- Development → Comparison
  - Multiple versions and environments
- Trace file comparison
  - Manual comparison → inconvenient → automatic comparison

Overview

- Concepts
- Motivation
- <u>Challenges</u>
- Strategy
- Results
- Suggestions

Challenges

- Which events are comparable?
- Scalability
  - Execution times
  - Space
  - Visualization

Overview

- Concepts
- Motivation
- Challenges
- <u>Strategy</u>
- Results
- Suggestions

Strategy

- What events are comparable?
- Aligning sequence's similar regions
  - Pair-wise comparison
    - Textual identifiers
    - Only the order matters!
  - Used in bioinformatics
    - Linear space
    - Quadratic time → parallelism

Strategy

Alignment 1.0: `A L I N H A M E N T O`
Alignment 2.0: `A L I ~~N H A~~ M E N T O`

Pair-wise comparison without alignment:

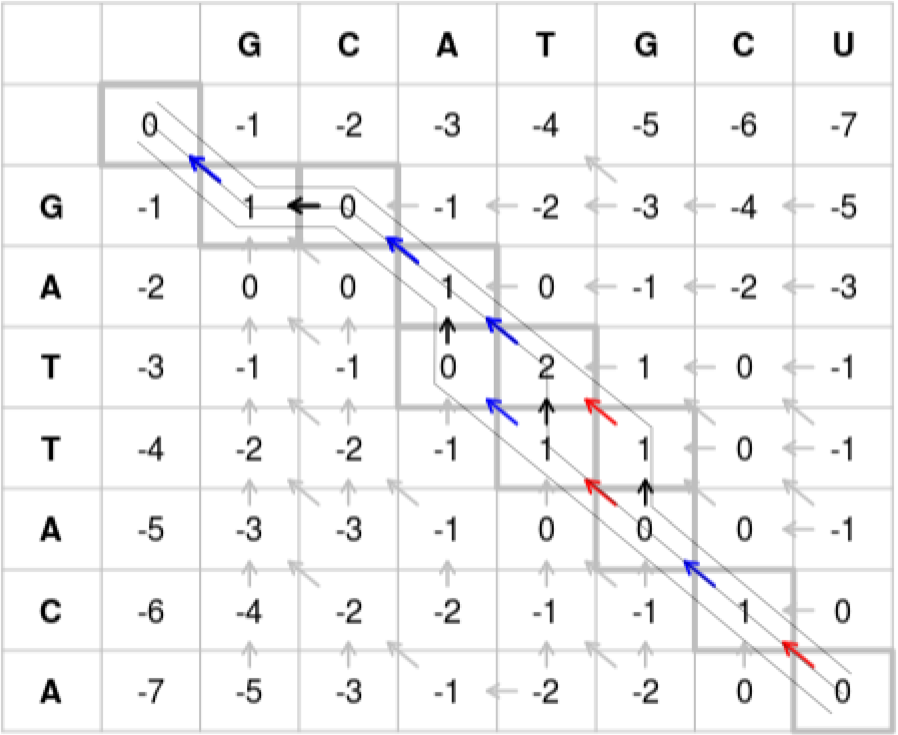
```
A L I N H A M E N T O
A L I M E N T O - - -
```

Aligned sequences:

```
A L I N H A M E N T O
A L I - - - M E N T O
```

Strategy

$$D(i,j) = \begin{cases} j * \gamma & i = 0 \\ i * \gamma & j = 0 \\ max \begin{cases} D(i-1, j-1) + \theta(i,j) \\ D(i, j-1) + \gamma \\ D(i-1, j) + \gamma \end{cases} & i \neq 0, j \neq 0 \end{cases}$$



		G	C	A	T	G	C	U
	0	-1	-2	-3	-4	-5	-6	-7
G	-1	1	0	-1	-2	-3	-4	-5
A	-2	0	0	1	0	-1	-2	-3
T	-3	-1	-1	0	2	1	0	-1
T	-4	-2	-2	-1	1	1	0	-1
A	-5	-3	-3	-1	0	0	0	-1
C	-6	-4	-2	-2	-1	-1	1	0
A	-7	-5	-3	-1	-2	-2	0	0

Strategy

- Alignment result → Paje trace file
    - Only one run
    - Any visualization tool

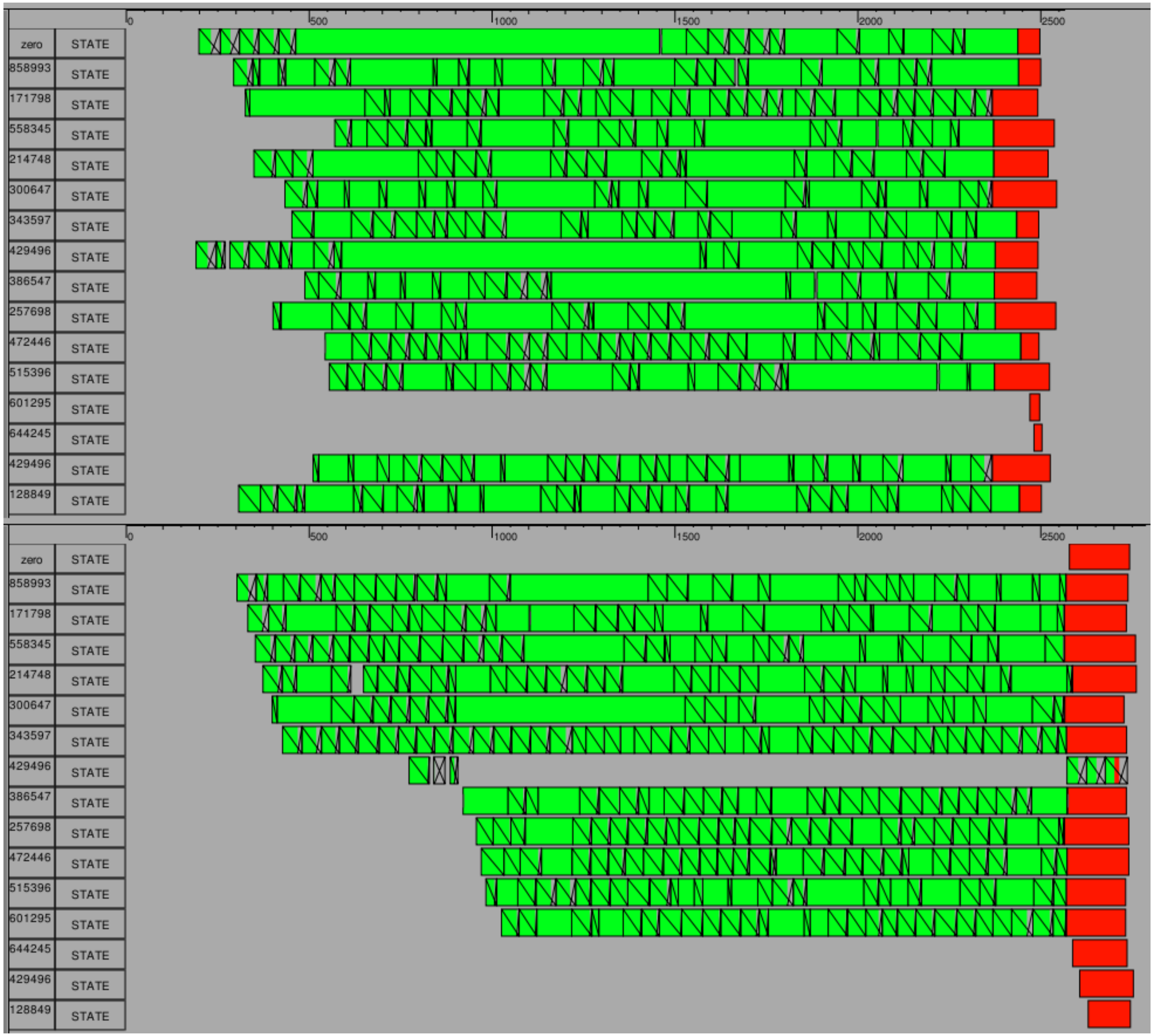
Strategy

- Scalability
  - Visualization

Overview

- Concepts
- Motivation
- Challenges
- Strategy
- <u>Results</u>
- Suggestions

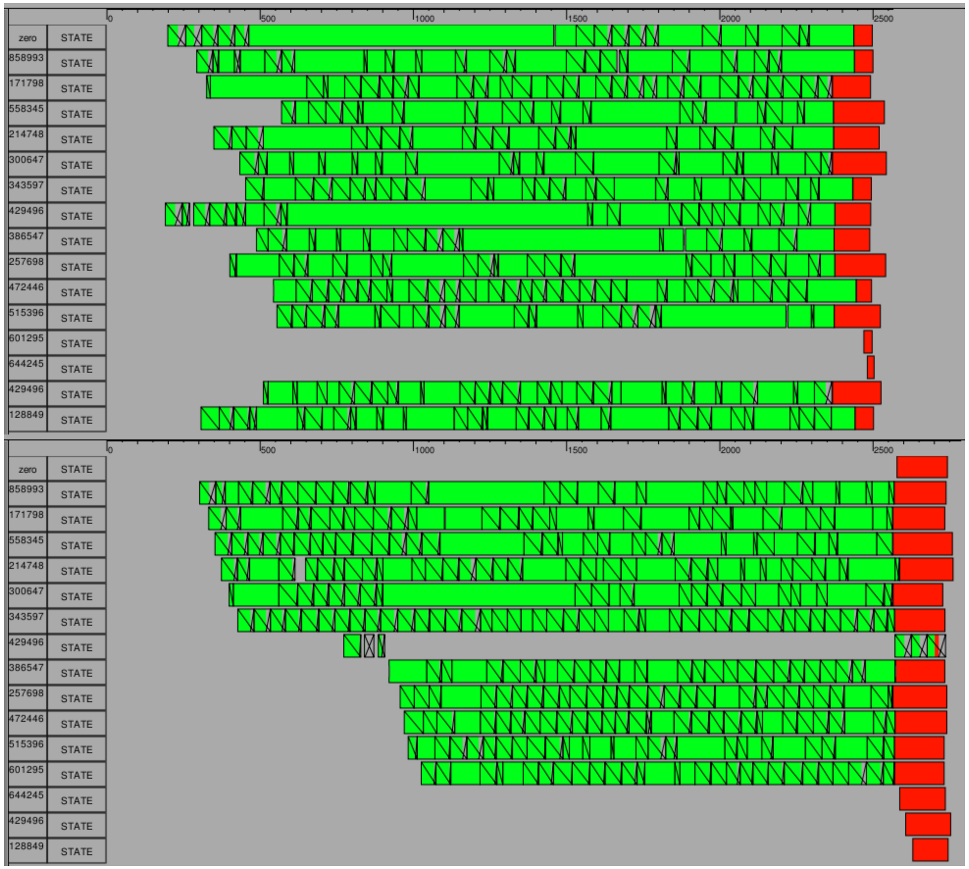
Results

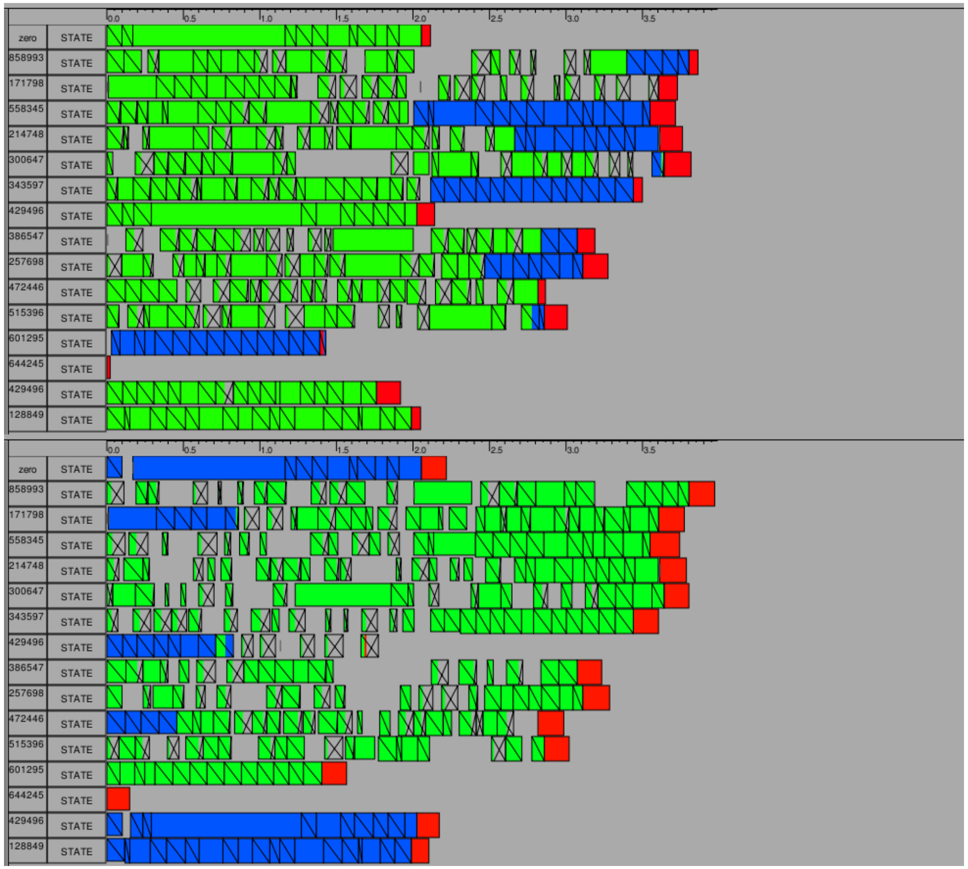

0
500
1000
1500
2000
2500
zero
858993
171798
558345
214748
300647
343597
429496
386547
257698
472446
515396
601295
644245
429496
128849
STATE

Results

- Alignment
    - Same test case

Results

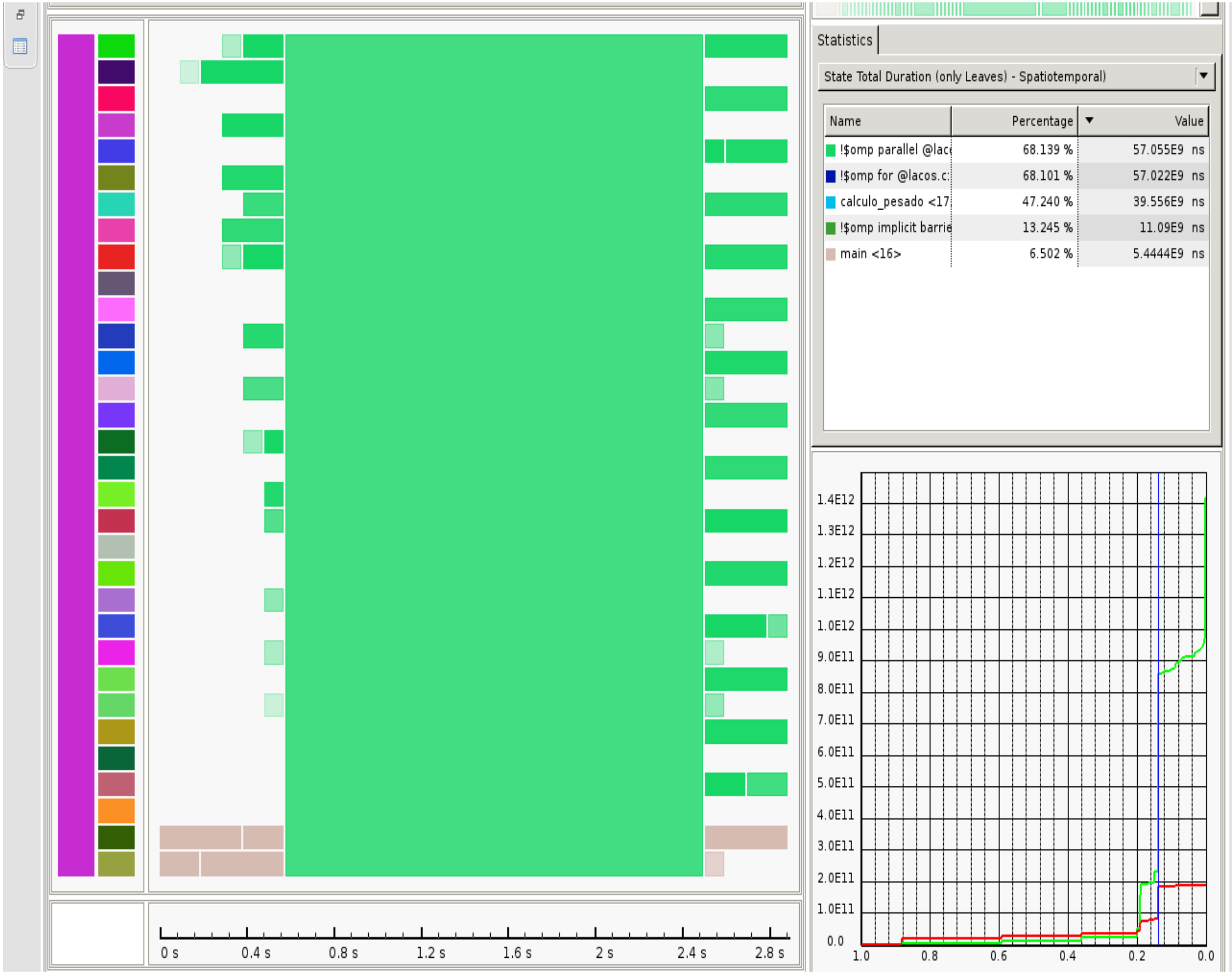

0
500
1000
1500
2000
2500
zero
858993
171798
558345
214748
300647
343597
429496
386547
257698
472446
515396
601295
644245
429496
128849
STATE


0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
zero
858993
171798
558345
214748
300647
343597
429496
386547
257698
472446
515396
601295
644245
429496
128849
STATE

Results

- Parenthesis
- Can aggregation make it better?
  - Between the traces, that is

Results

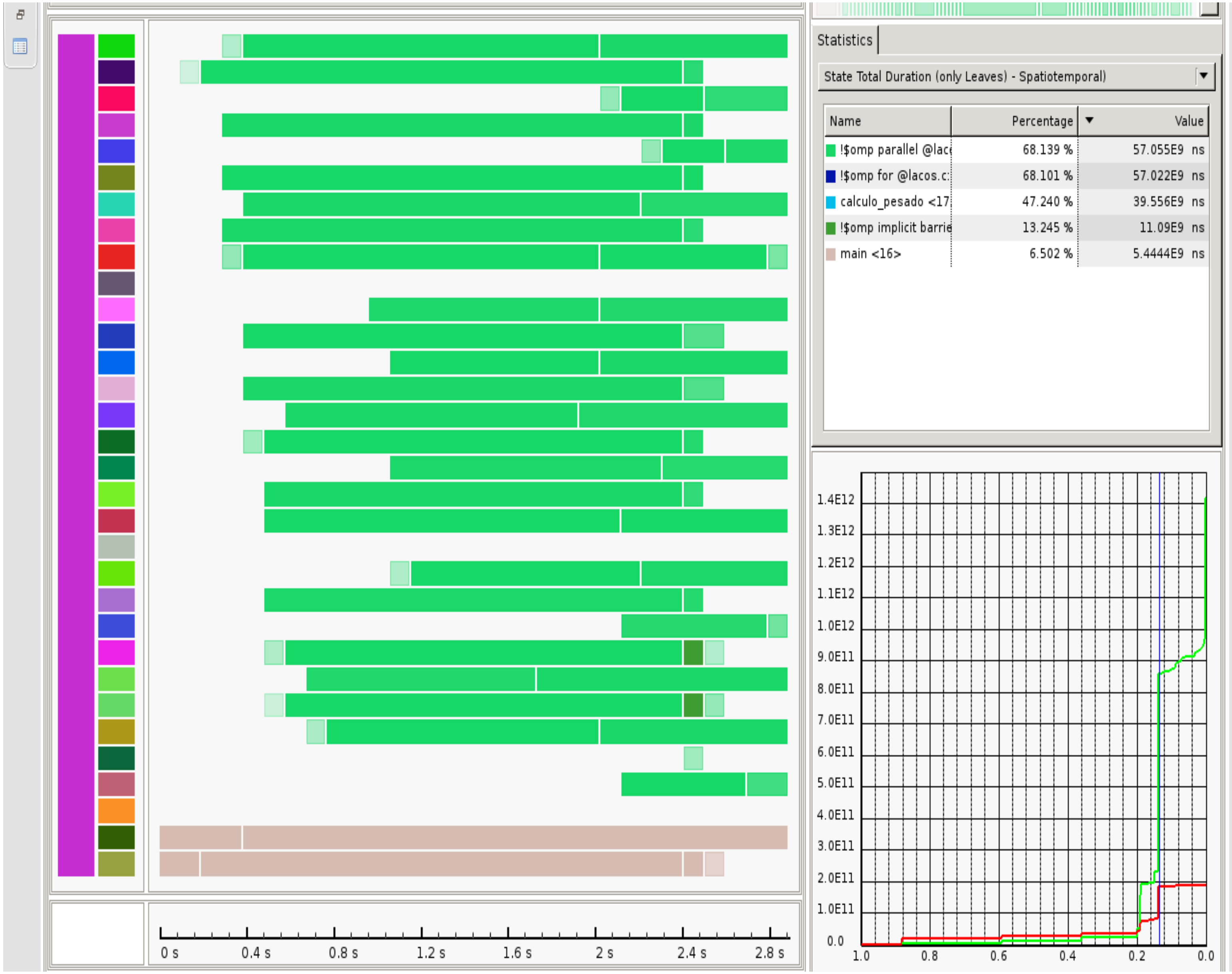


Statistics

State Total Duration (only Leaves) - Spatiotemporal)

Name	Percentage	Value
🟩 !$omp parallel @laci	68.139 %	57.055E9 ns
🟦 !$omp for @lacos.c:	68.101 %	57.022E9 ns
🟦 calculo_pesado <17	47.240 %	39.556E9 ns
🟩 !$omp implicit barrie	13.245 %	11.09E9 ns
🟫 main <16>	6.502 %	5.4444E9 ns

Results



State Total Duration (only Leaves) - Spatiotemporal)

Name	Percentage	Value
🟩 !$omp parallel @lac	68.139 %	57.055E9 ns
🟦 !$omp for @lacos.c:	68.101 %	57.022E9 ns
🟦 calculo_pesado <17	47.240 %	39.556E9 ns
🟩 !$omp implicit barri	13.245 %	11.09E9 ns
🟫 main <16>	6.502 %	5.4444E9 ns

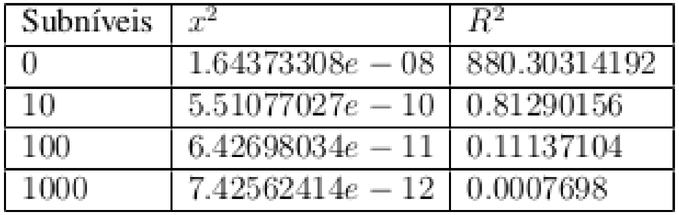
Results

- MPI Mandelbrot
  - 961,621 events
  - 60 MB

Worst	Mean	Best
2.57 seconds	2.54 seconds	2.44 seconds

- Intel® MPI Benchmarks 4.0 Update 2)
  - 34,611,196 events
  - 2.4 GB

Worst	Mean	Best
3.31 minutes	3.30 minutes	3.29 minutes



Subníveis	x^2	R^2
0	$1.64373308e-08$	880.30314192
10	$5.51077027e-10$	0.81290156
100	$6.42698034e-11$	0.11137104
1000	$7.42562414e-12$	0.0007698

Results

- Can the events be automatically compared?
  - Yes.
- Can a meaningful visualization of the comparison be provided?
  - In some cases, no. But not because of scalability issues.
- Can it be done in reasonable time and space?
  - It seems so → more tests!

Overview

- Concepts
- Motivation
- Challenges
- Strategy
- Results
- <u>Suggestions</u>

Suggestions

- Revalidation
  - Large scale
- Different alignment types
  - Semi-global
- Aggregation AND alignment?
- GPU?
- Free Software

Thank you. Questions?

PajeNG + PajeAligner:

https://github.com/a442/pajeng

PajeNG:

https://github.com/schnorr/pajeng

Contact:

afarah@inf.ufrgs.br

schnorr@inf.ufrgs.br

Acknowledgments:

.inf
INSTITUTO
DE INFORMÁTICA
UFRGS

CNPq
Conselho Nacional de Desenvolvimento
Científico e Tecnológico

UFRGS
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL