MULTIDIMENSIONAL DATA AGGREGATION AND VISUALIZATION FOR HUGE EXECUTION TRACE ANALYSIS

Visu 2014, 7th October 2014

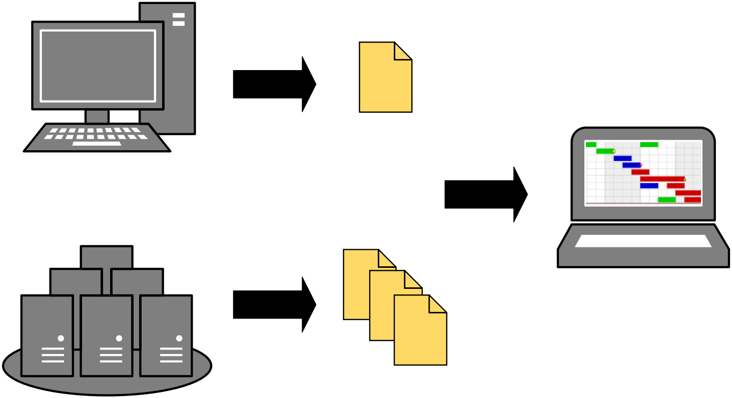
Damien Dosimont [1 2], Robin Lamarche-Perrin [3], Youenn Corre [1 2], Lucas M. Schnorr [4], Guillaume Huard [2 1], Jean-Marc Vincent [2 1]

[1] Inria,
first.last@inria.fr,
[2] Univ. Grenoble Alpes, LIG, CNRS, F-38000 Grenoble, France
first.last@imag.fr
[3] MPI for Mathematics in the Sciences, 04103 Leipzig, Germany
robin.lamarche-perrin@mis.mpg.de
[4] Informatics Institute, UFRGS, Porto Alegre
schnorr@inf.ufrgs.br

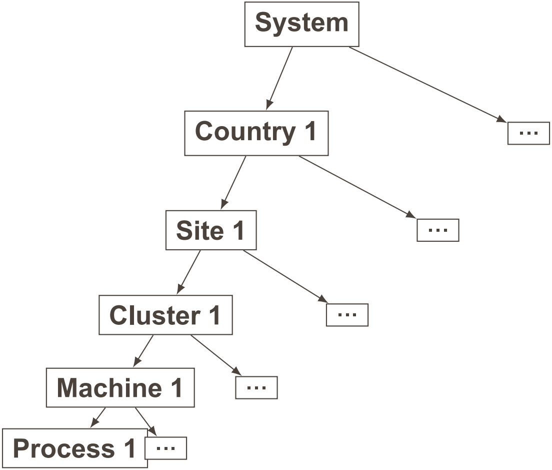
TRACE VISUALIZATION PROBLEMATIC

PARALLEL AND DISTRIBUTED SYSTEM ANALYSIS

TRACE-BASED ANALYSIS



STRUCTURE: EXAMPLE OF HW/SW HIERARCHY

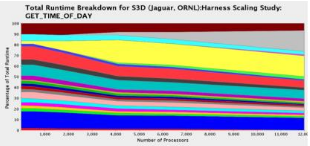


TIMESTAMPED EVENTS THAT OCCUR DURING EXECUTION

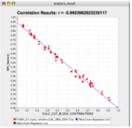
- **Punctual events**
  - Synchronization
  - Flag
  - Function call, return
- **States**
  - Function (call $\rightarrow$ return)
  - CPU state (idle, running)
- **Variables**
  - CPU load
  - Memory use
  - HW/SW Counters
- **Links**
  - Communication
  - Context switch

EXAMPLE OF CLASSIC ANALYSES

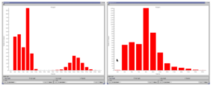
Global Analysis



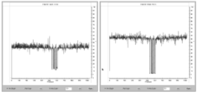
Execution Comparison



Correlations

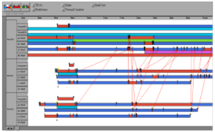


Outliers

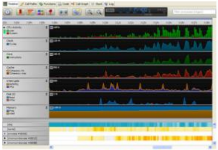


Workload

Run behavior

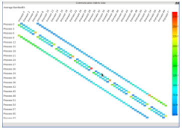


Causality relations

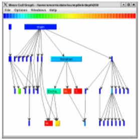


Resource usage

Structure

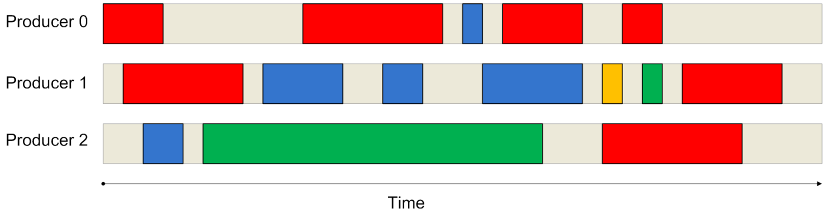


Communications



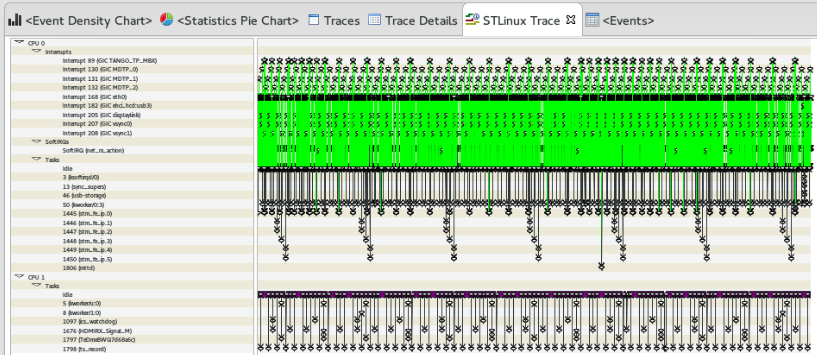
Call graphs

SPACE-TIME REPRESENTATIONS PROBLEMATIC



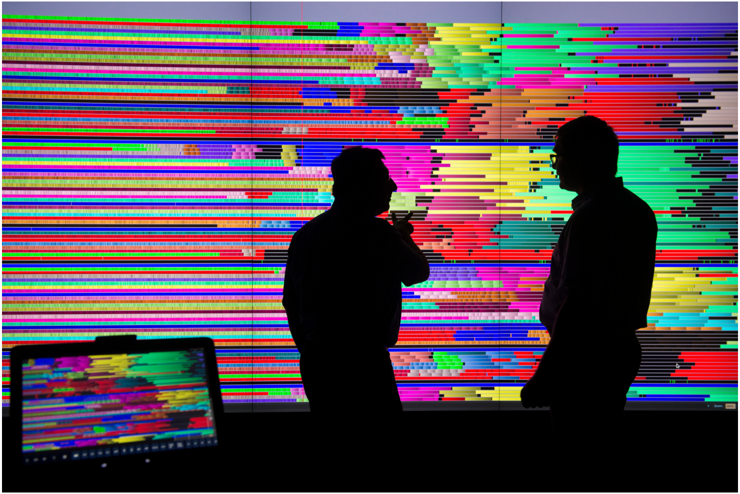
- Structure can be composed of millions of resources
- Trace can contain billions of events (up to TB)

LIMITED SCREEN SIZE ISSUES



COMPUTATION - RENDERING - INTERACTIVITY ISSUES

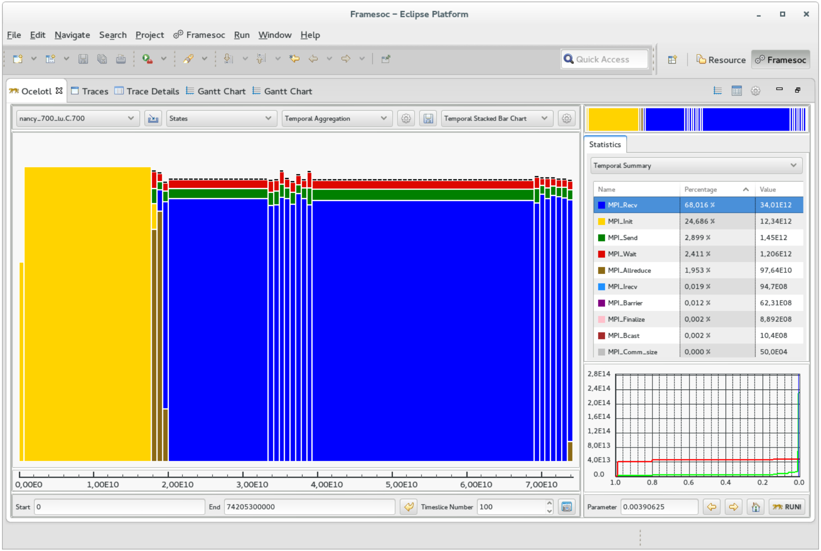
ANALYST CAPABILITY LIMITS



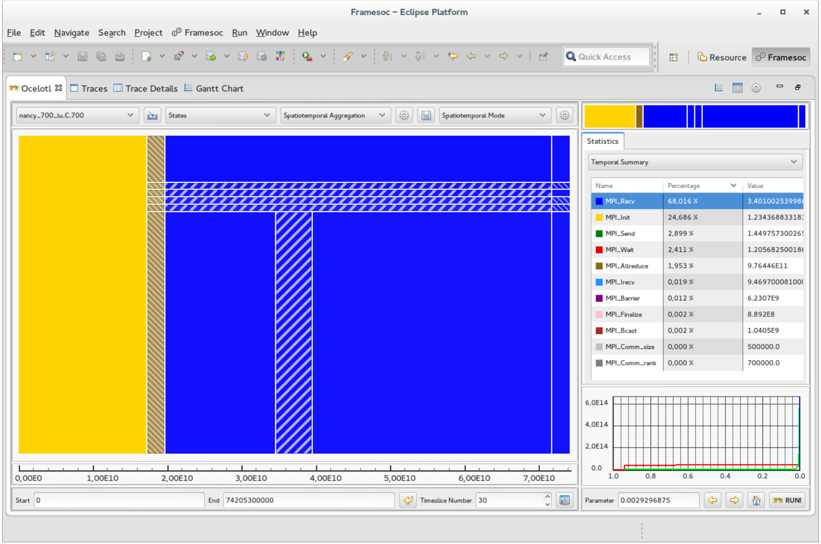
OUR PROPOSAL: METHODOLOGY TO BUILD OVERVIEWS

- Overviews generated using **data and visual aggregation**
  - Temporal
  - Spatiotemporal
- Showing **meaningful information** (phases, perturbations)
- Enabling to adjust dynamically the **level of details**
- **Interaction**:
  - Zoom
  - Filtering
  - Synchronized statistics
  - Switch to other representations

TEMPORAL ANALYSIS WITH OCELOTL TOOL

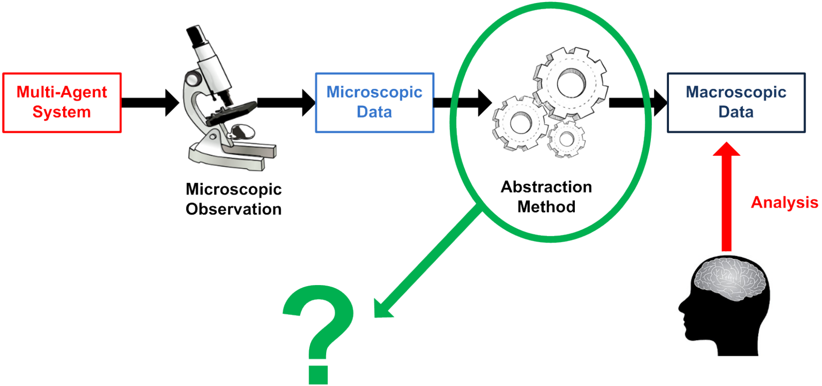


SPATIOTEMPORAL ANALYSIS WITH OCELOTL TOOL

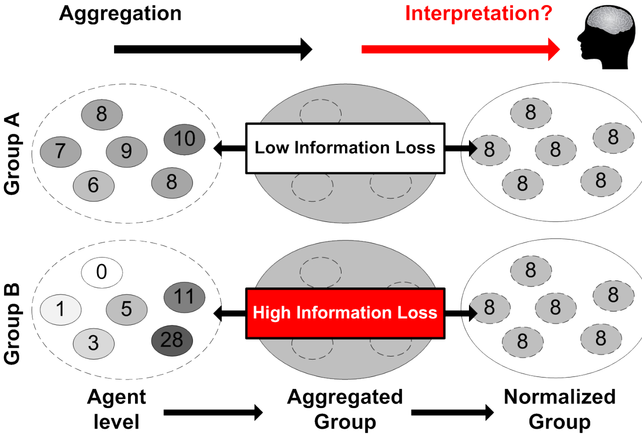


DATA AGGREGATION METHODOLOGY

ADAPTING AN AGGREGATION METHODOLOGY

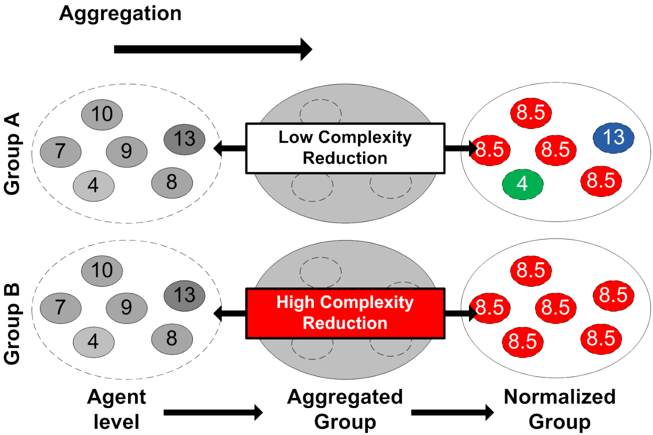


INFORMATION LOSS: KL DIVERGENCE



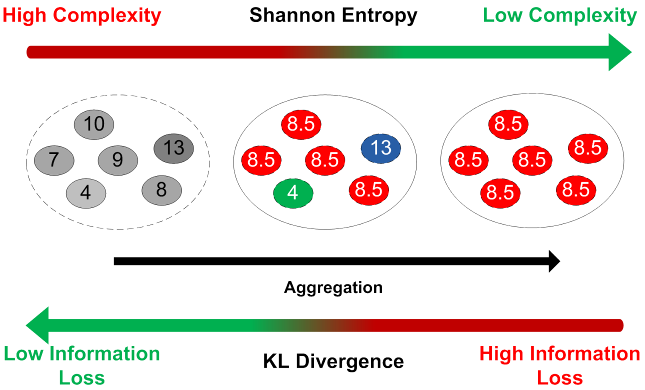
$$\mathrm{loss}_E = \sum_{e \in E} \rho_e \log_2 \left(\frac{\rho_e}{\rho_E} \right)$$

COMPLEXITY REDUCTION: SHANNON ENTROPY



$$\mathrm{gain}_E = \rho_E \log_2 \rho_E - \sum_{e \in E} \rho_e \log_2 \rho_e$$

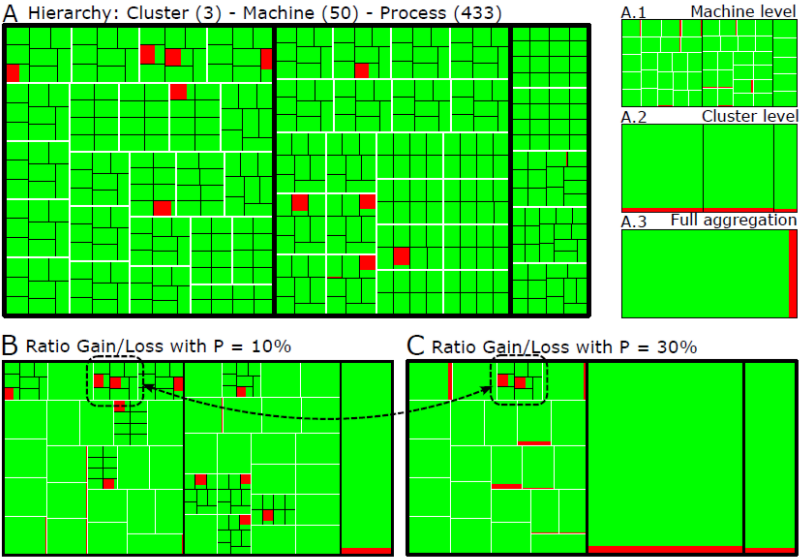
TRADE-OFF: PIC



$$\mathrm{pIC}_E = p\,\mathrm{gain}_E - (1-p)\,\mathrm{loss}_E \qquad \mathrm{pIC}_{\mathcal{P}} = \sum_{E \in \mathcal{P}} \mathrm{pIC}_E$$

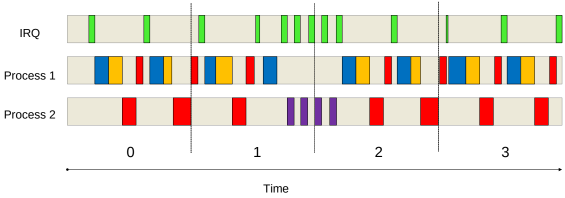
- For a given p: choose $\mathcal{P}$ with the highest pIC
- Aggregate in priority most homogeneous values

VIVA: SPATIAL AGGREGATION (SCHNORR & LP)



TEMPORAL OVERVIEW

GENERATE A TRACE MICROSCOPIC MODEL

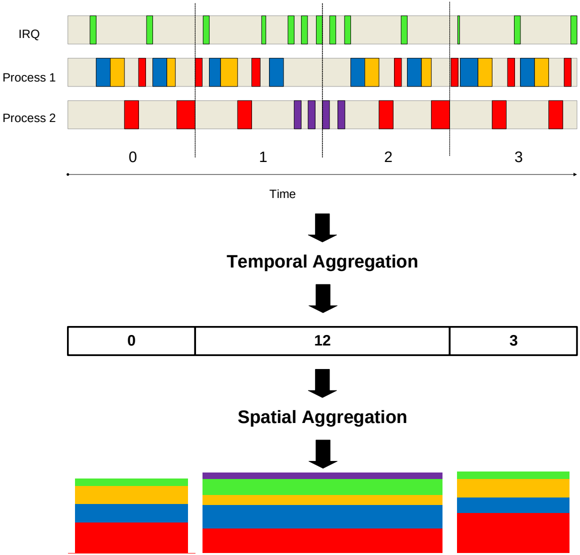


IRQ	0	0	0	0
Process 1	1	2.1	1	3
Process 2	4.1	2	4.1	4

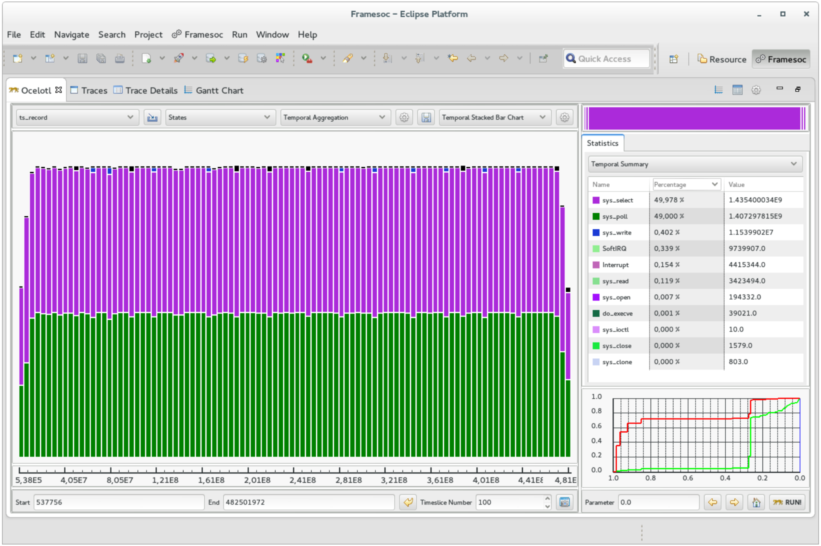
IRQ	2	4.9	3	2.4
Process 1	0	0	0	0
Process 2	0	0	0	0

And so on...

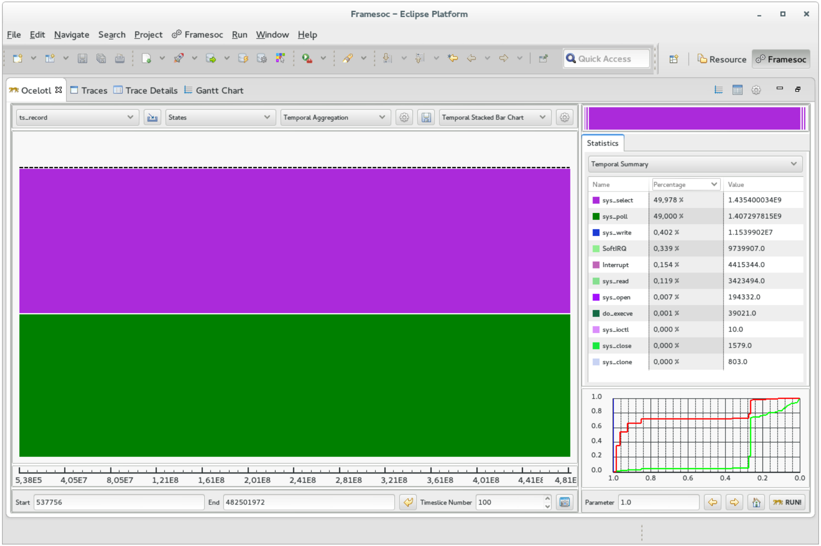
TEMPORAL AGGREGATION AND VISUALIZATION



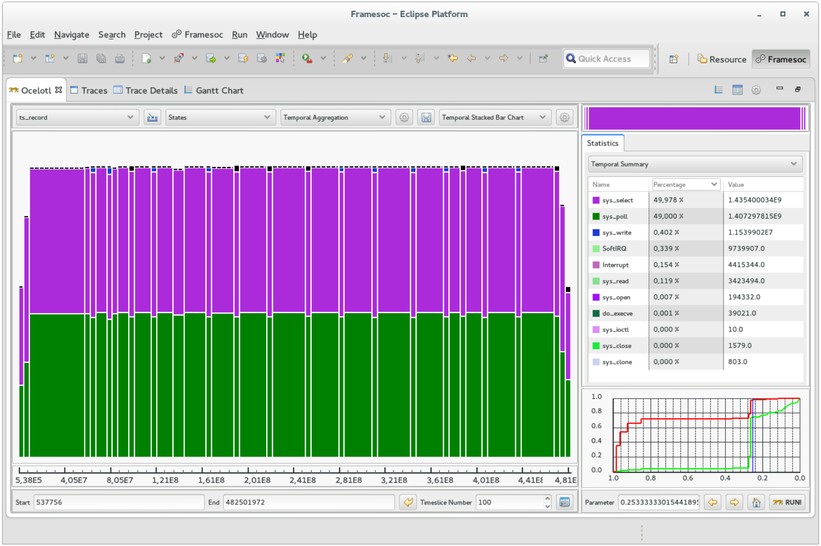
MINIMUM INFORMATION LOSS: P=0



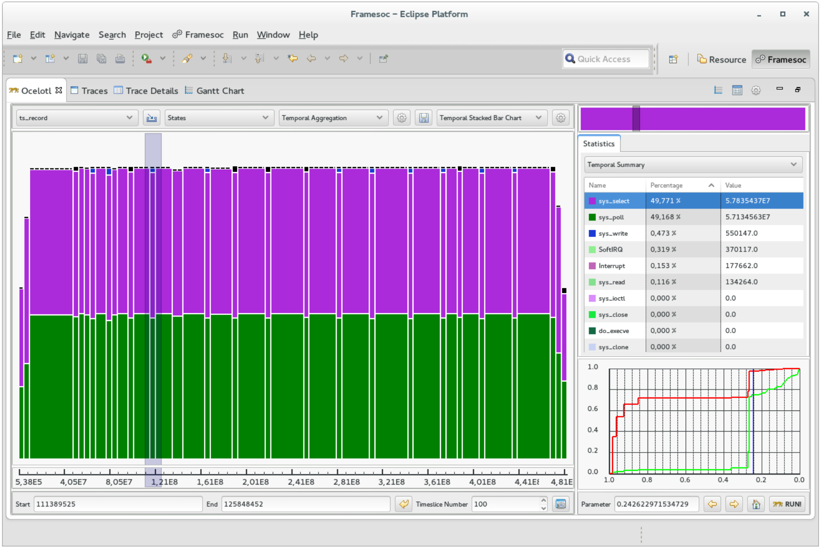
MAXIMUM COMPLEXITY REDUCTION: P=1



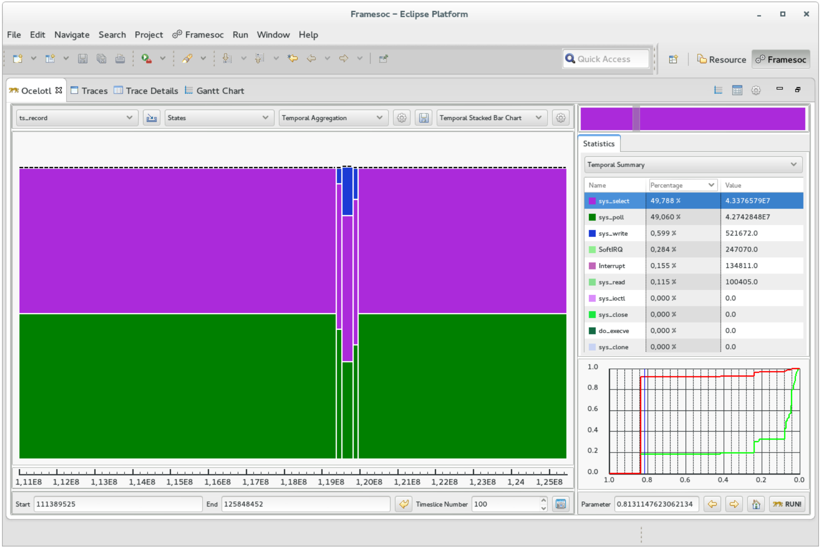
INTERESTING TRADE-OFF



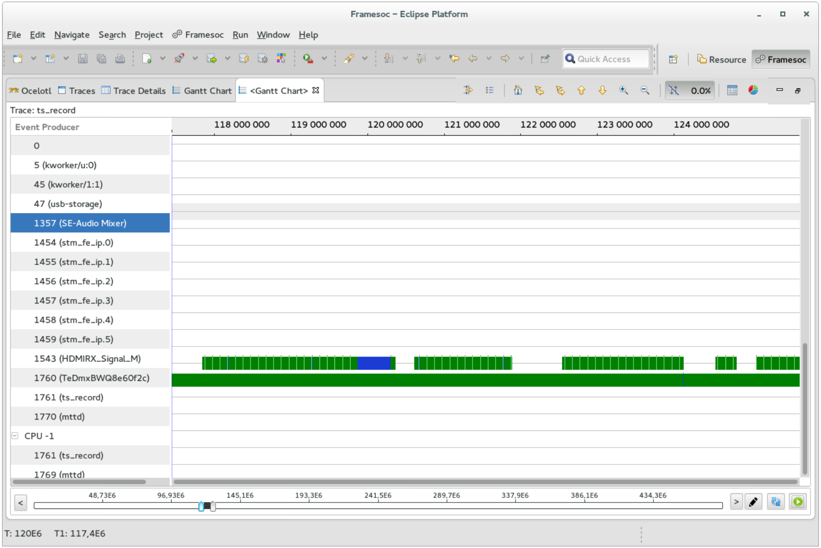
SELECT A TEMPORAL AREA



ZOOM

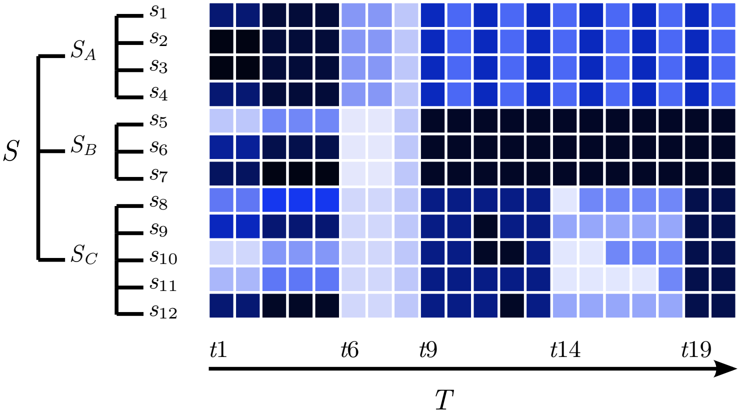

Framesoc – Eclipse Platform
File Edit Navigate Search Project Framesoc Run Window Help
Quick Access
Resource
Framesoc
Ocelotl
Traces
Trace Details
Gantt Chart
ts_record
States
Temporal Aggregation
Temporal Stacked Bar Chart
1,11E8 1,12E8 1,13E8 1,14E8 1,15E8 1,16E8 1,17E8 1,18E8 1,19E8 1,20E8 1,21E8 1,22E8 1,23E8 1,24 1,25E8
Start 111389525
End 125848452
Timeslice Number 100
Statistics
Temporal Summary
Name Percentage Value
sys_select 49,788 % 4.3376579E7
sys_poll 49,060 % 4.2742848E7
sys_write 0,599 % 521672.0
SoftIRQ 0,284 % 247070.0
Interrupt 0,155 % 134811.0
sys_read 0,115 % 100405.0
sys_ioctl 0,000 % 0.0
sys_close 0,000 % 0.0
do_execve 0,000 % 0.0
sys_clone 0,000 % 0.0
1.0 0.8 0.6 0.4 0.2 0.0
1.0 0.8 0.6 0.4 0.2 0.0
Parameter 0.8131147623062134
RUN!

MORE DETAILS WITH SPACE-TIME REPRESENTATION



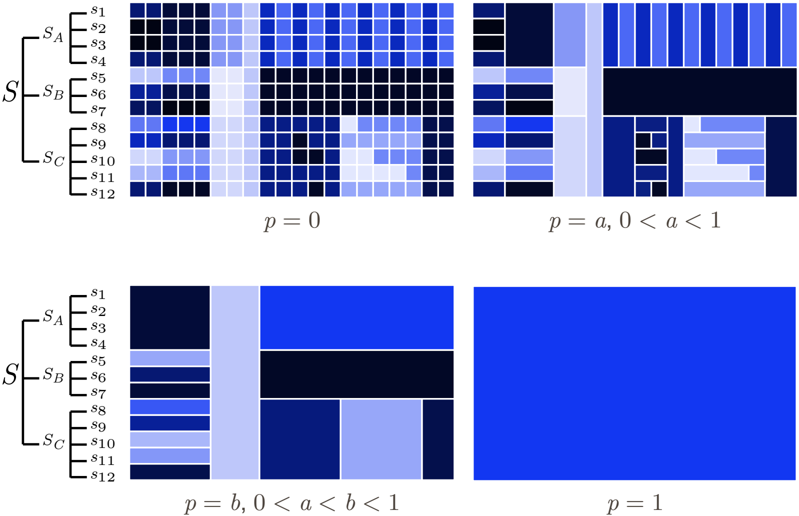
SPATIOTEMPORAL OVERVIEW

GENERATE A TRACE MICROSCOPIC MODEL

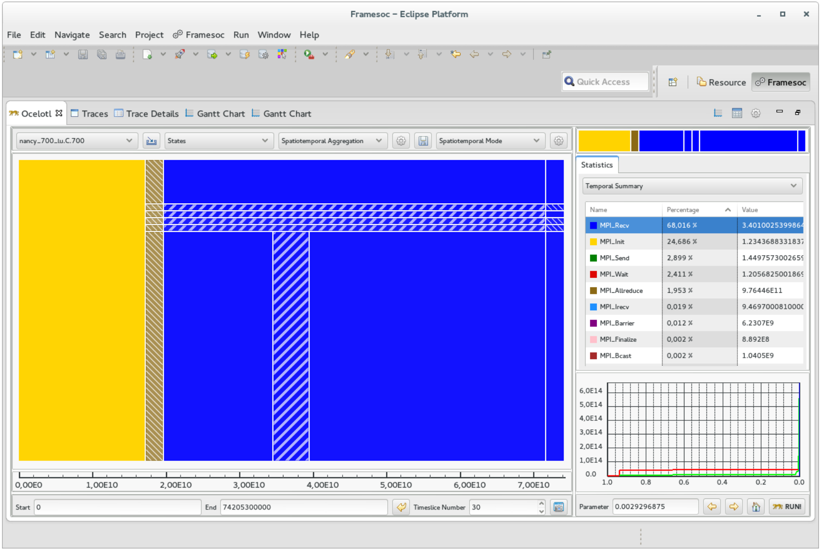


$$|X| = 2,\ \rho_x(s,t) = d_x(s,t)/d(t) \in [0,1],\ \rho_1(s,t) = 1 - \rho_2(s,t)$$

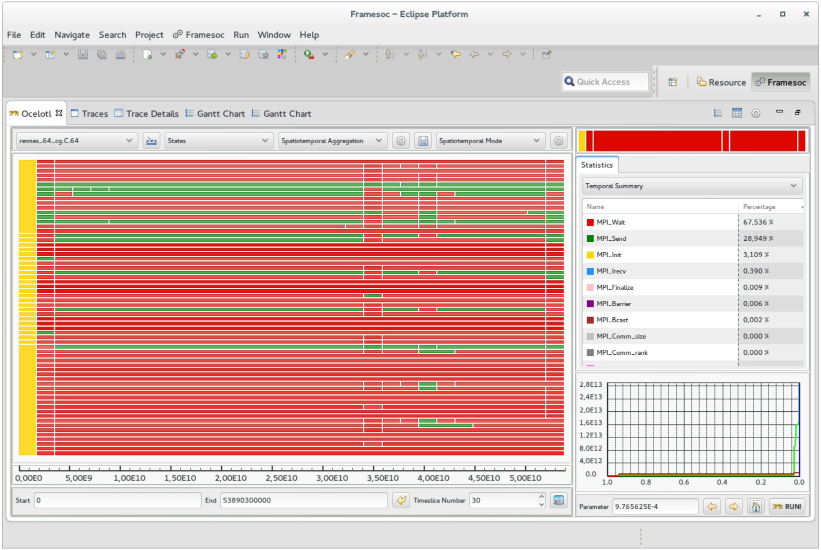
AGGREGATE THE MICROSCOPIC MODEL



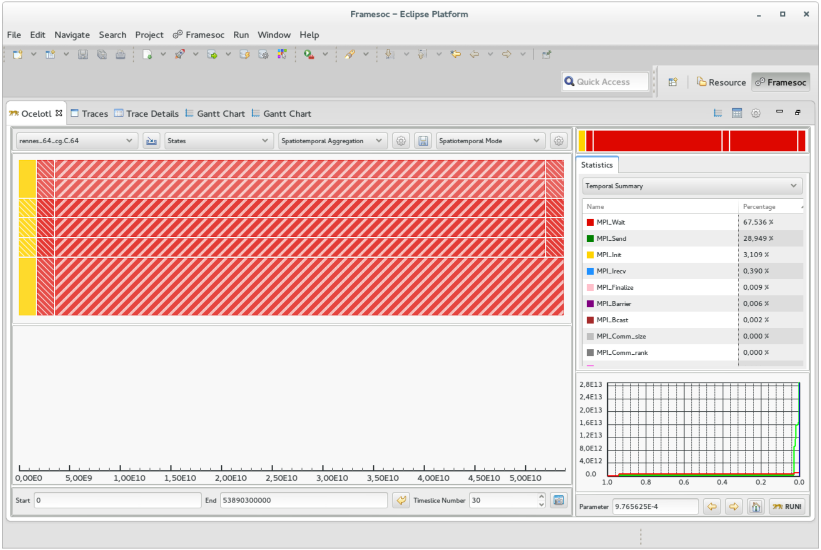
INTERESTING TRADE-OFF



NO VISUAL AGGREGATION



VISUAL AGGREGATION



CONCLUSION

CONCLUSION

- **Visualizations** based on **data and visual aggregation**
  - Solves screen, computing and analyst capability **limitations**
  - Gives **meaningful information** about homogeneity (phases, perturbations)
- **Implementation**:
  - **Interaction** (zoom, switch to other tools)
  - **Performance** 5 min for a 12 GB trace (220 millions of events)
- **Globally**:
  - HPC tools mainly designed by the HPC community, to fit with its need
  - Collaboration with the InfoVis community would help to improve our visualization techniques
  - On the other side, our traces represent interesting use cases of huge quantities of data

LINKS

Ocelotl:

http://dosimont.github.io/ocelotl/

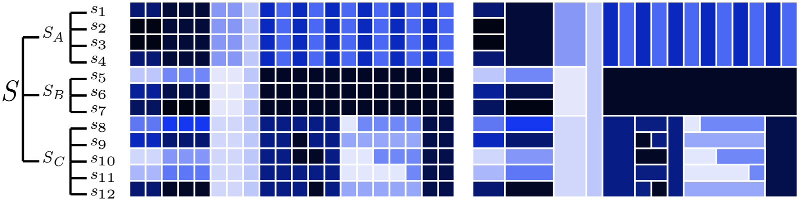
THANK YOU FOR YOUR ATTENTION

DATA AGGREGATION METHODOLOGY

- A1. Choose a **model** and a **metric**
- A2. Choose on **which dimension(s)** aggregate
- A3. Define the **operands**
- A4. **Constrain** the aggregation : $\rightarrow$ partitions $\mathcal{P}$ allowed
- A5. Define the **operator**
- A6. Define the **trigger -** the aggregation condition
- A7. Build the **algorithm** satisfying A1-A6

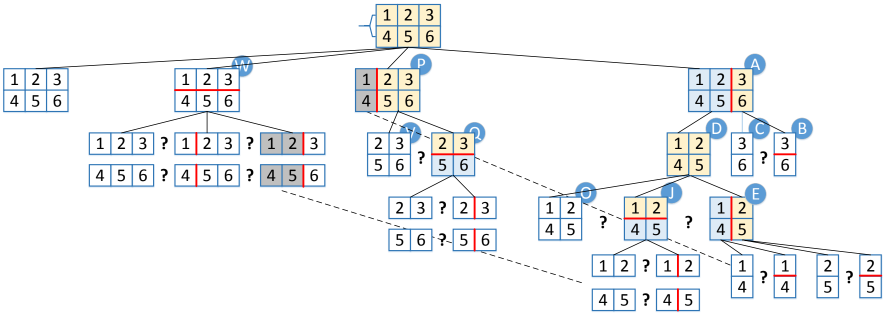
A2-A5

- A2. We aggregate simultaneously on T and S
- A3. Operands: $(s, t) \in S \times T$
- A4. Constraint: $\mathcal{A}(S \times T) = \mathcal{H}(S) \times \mathcal{I}(T)$
  Aggregation result is a partition $\mathcal{P}(S \times T) \in \boldsymbol{\mathcal{A}}(S \times T)$
- A5. Operator: $+$
- A6. Trigger: maximize pIC of the partition $\mathcal{P}(S \times T)$



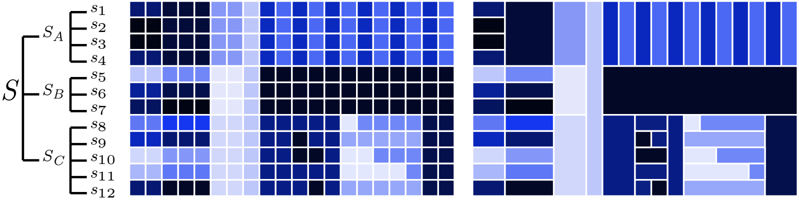
BEST CUT ALGORITHM

- Compute the partition with the highest pIC :
  - Cut an area : time, space (or no cut)
  - Best cut: the partition $\mathcal{P}$ where $\sum_{E \in \mathcal{P}} \mathrm{pIC}_E$ is max
  - Recursively cut and evaluate the partitions of $E_1, E_2 \in \mathcal{P}$
  - Useless recomputation is avoided



A6. TRIGGER THE AGGREGATION

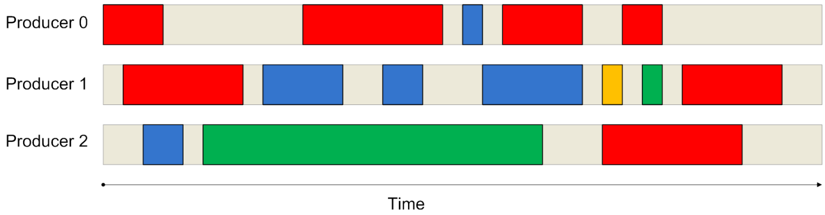
- Quantification of data reduction and information loss
  - aggregate the homogeneous areas
  - preserve the microscopic information of the heterogeneous areas
- Each $(S_k, T_{(i,j)}) \in \mathcal{A}(S \times T)$ has an associated gain and loss
- gain and loss of a partition $\mathcal{P}(S \times T)$ is the sum of gain and loss of its content $(S_k, T_{(i,j)}) \in \mathcal{P}(S \times T)$



ELMQVIST-FEKETE CRITERIA

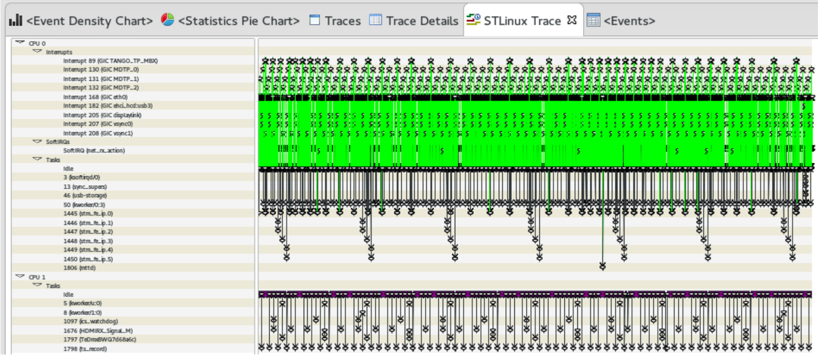
- **Shneiderman** : **overview**, zoom and filter, then get details on demand
- **Elmqvist & Fekete**: guidelines to design an **overview** visualization based on hierarchical aggregation
  - G1. Entity Budget
  - G2. Visual Summary
  - G3. Visual Simplicity
  - G4. *Discriminability*
  - G5. Fidelity
  - G6. *Interpretability*

VISUALIZATIONS NOT FULFILLING THESE CRITERIA (1)



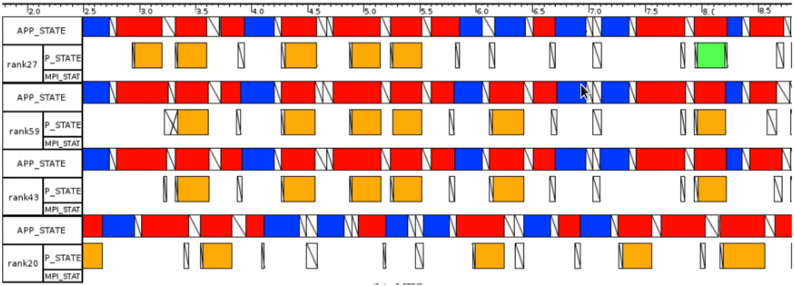
Example of Gantt chart - space-time diagram

VISUALIZATIONS NOT FULFILLING THESE CRITERIA (2)



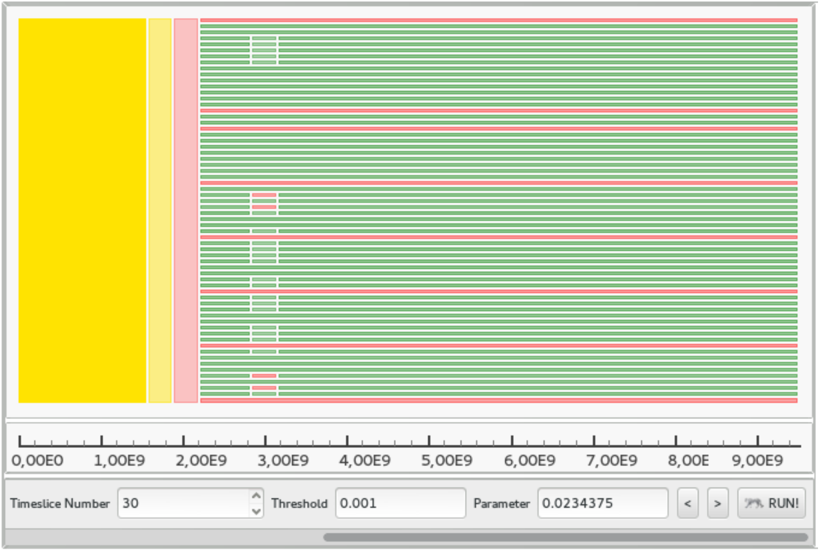
KPTrace: $\overline{\text{G1}}$ (time), $\overline{\text{G2}}$, $\overline{\text{G4}}$, $\overline{\text{G5}}$

VISUALIZATIONS NOT FULFILLING THESE CRITERIA (2)

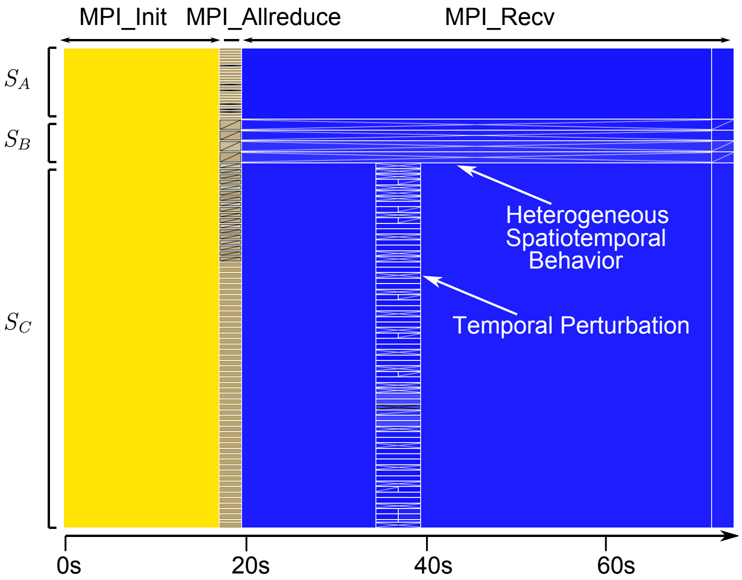


Pajé: $\overline{\text{G1}}$ (space), $\overline{\text{G2}}$

CG CLASS C, 64 PROCESSES ON G5K RENNES



LU CLASS C, 700 PROCESSES ON G5K NANCY



PERFORMANCES (SPATIOTEMPORAL)

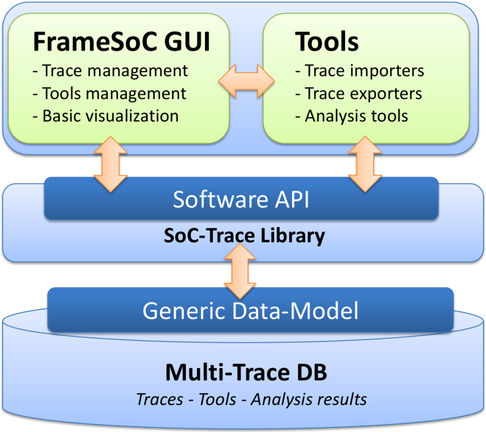
	Case A	Case B	Case C	Case D
Application	CG, class C	CG, class C	LU, class C	LU, class B
Processes	64	512	700	900
Site	Rennes	Grenoble	Nancy	Rennes
Clusters (nodes)	parapide(8)	adonis(9), edel(24), genepi(31)	graphene(26), graphite(4), griffon(67)	paradent(38), parapide(21), parapluie(18)
Event number	3,838,144	49,149,440	218,457,456	177,376,729
Trace size	136.9 MB	1.8 GB	8.3 GB	6.7 GB

	Ocelotl computation times (30 time slices)			
Trace reading + Microscopic description	5 s	31 s	222 s	174 s
Aggregation	$<$1s	$<$1s	2s	2s

OCELOTL TOOL

- Implementation of the overview techniques
- Generic architecture. Add:
  - Your own **aggregation operator** (dimensions, metric)
  - Your own **visualization**
- Persistent caches to avoid long recomputations
- Integrated in **Framesoc**:
  - Trace and tools management
  - **Fast** trace reading (DB queries)
  - **Interaction** with other analysis tools
  - Also enable to **add you own tools**

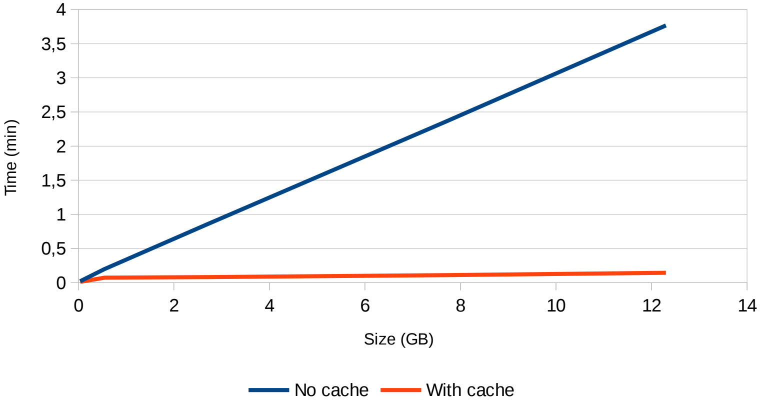
FRAMESOC



- Trace format compatibility : Pajé (Akypuera: tool to convert from OTF2, Tau), LTTng, KPTrace

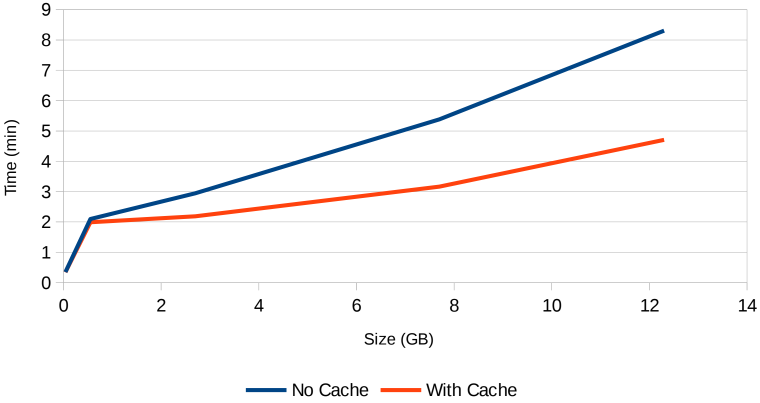
PERFORMANCE: TEMPORAL ANALYSIS

Total analysis time as a function of trace size (100 time slices)



PERFORMANCE: TEMPORAL ANALYSIS

Total analysis time as a function of trace size (1000 time slices)



PERFORMANCE: SPATIOTEMPORAL ANALYSIS

Total analysis time as a function of trace size (30 time slices)