



# Introduction to Paraver

Jesus Labarta, BSC

POP Webinar 22

July 28<sup>th</sup> 2021

EU H2020 Centre of Excellence (CoE)



Grant Agreement No 824080

1 December 2018 – 30 November 2021

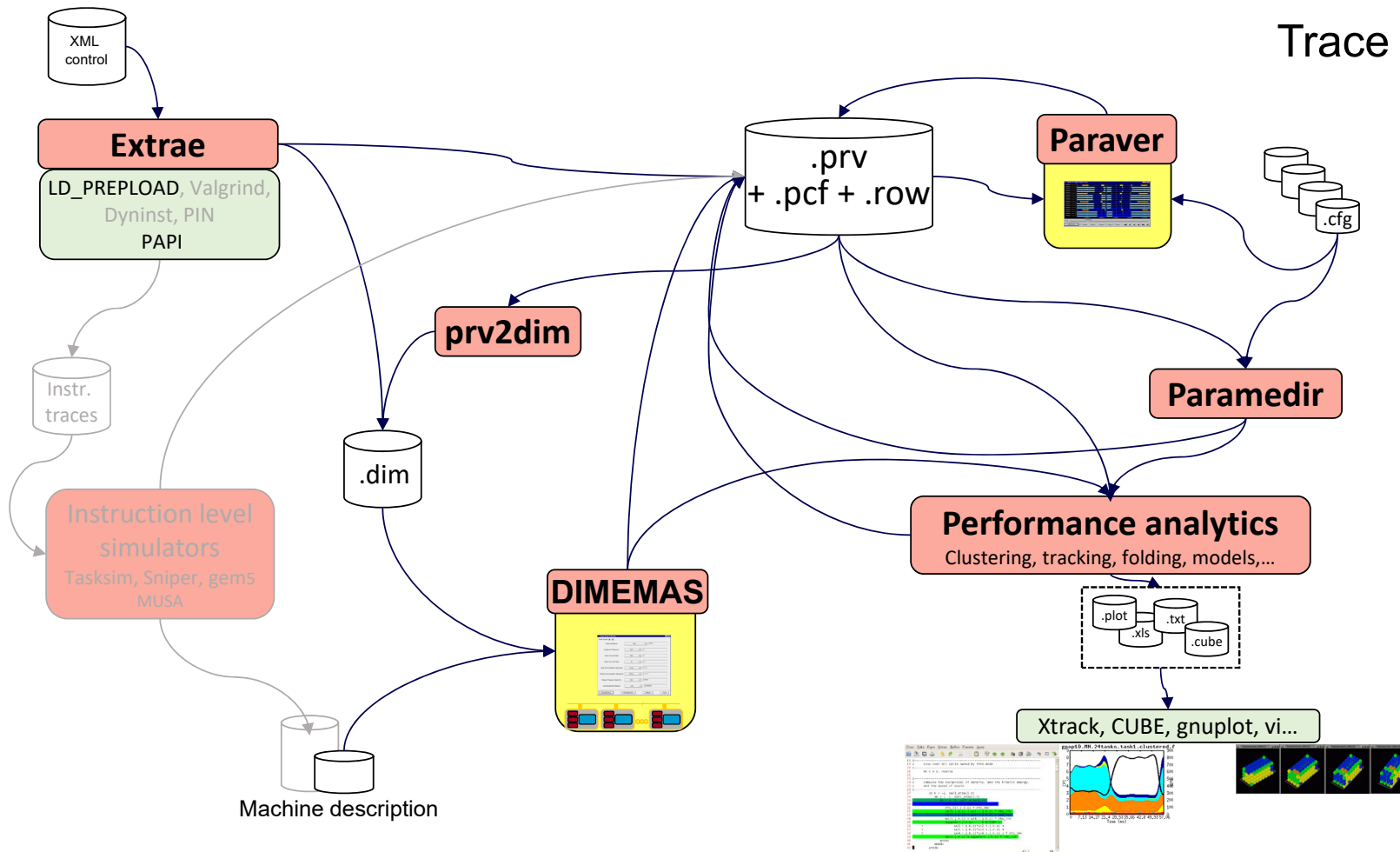
# Objective



- Introduce basic features of Paraver and the philosophy on how to use it
- Agenda:
  - BSC tools
  - Paraver trace format
  - Paraver fundamentals
    - Overall structure
    - Interpreting displays
    - Internals (generating displays)
  - Useful metrics
  - Links to more detailed material
  - Demo



# BSC – tools framework



Trace handling & display  
Simulators  
Analytics

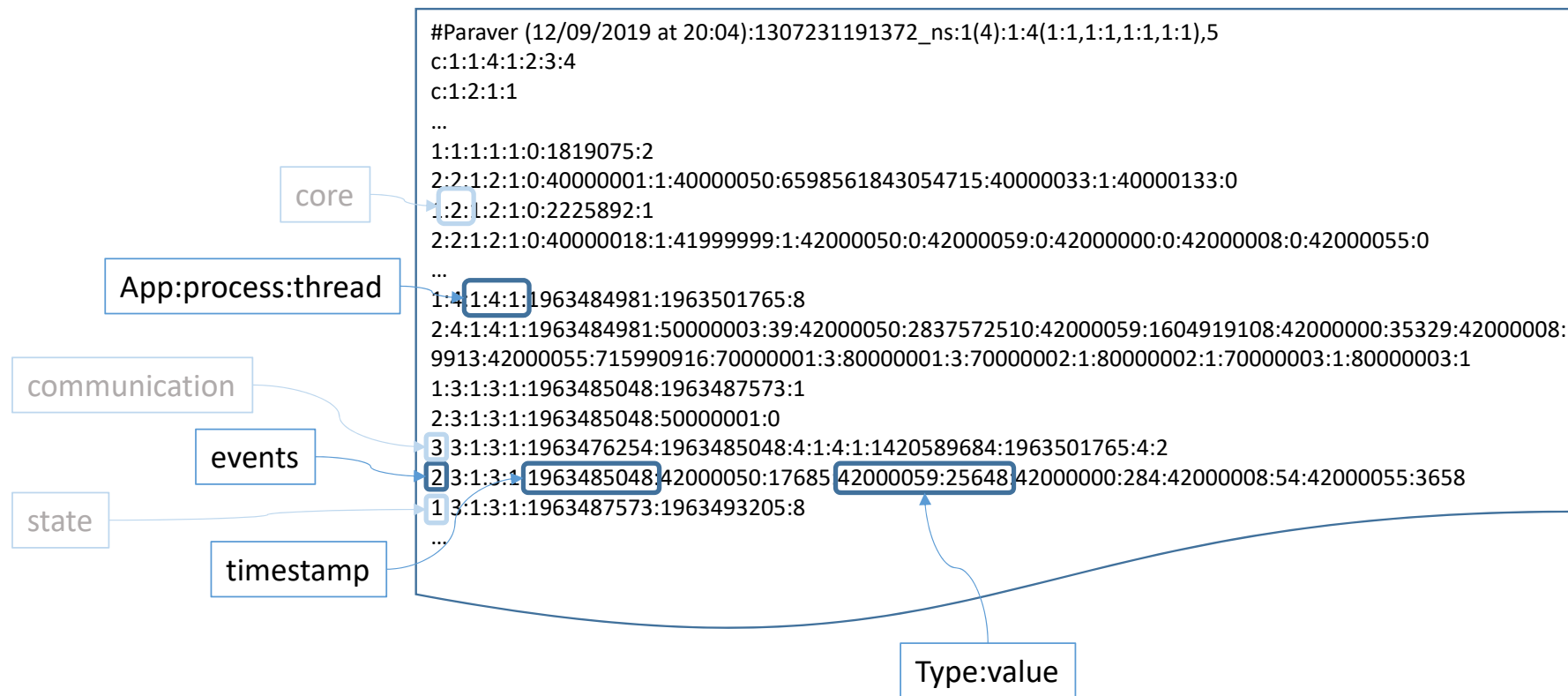
<http://www.bsc.es/paraver>



# Paraver trace format



- .prv



# Paraver trace format



- .pcf

```
...
EVENT_TYPE
9 50000001 MPI Point-to-point
VALUES
4 MPI_Irecv
3 MPI_Isend
0 Outside MPI
...
EVENT_TYPE
7 42000050 PAPI_TOT_INS [Instr completed]
7 42000059 PAPI_TOT_CYC [Total cycles]
7 42000002 PAPI_L2_DCM [L2D cache misses]
...
EVENT_TYPE
0 70000001 Caller at level 1
0 70000002 Caller at level 2
0 70000003 Caller at level 3
VALUES
0 End
1 Unresolved
2 _NOT_Found
3 MPI_Test
...
```



# What is Paraver

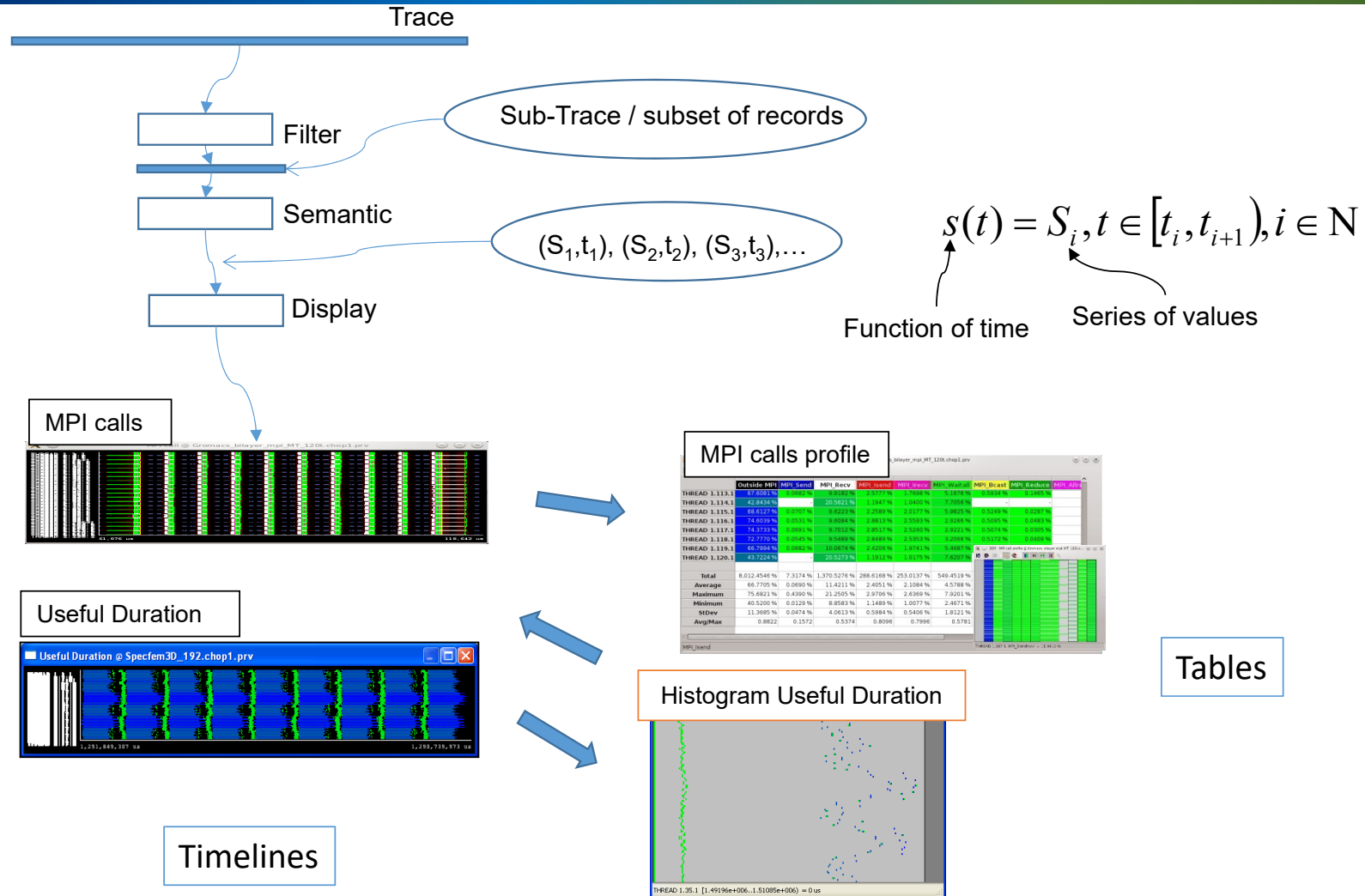
---



- A browser ...
- ...to manipulate (visualize, filter, cut, combine, ...) ....
- ... sequences of time-stamped events (traces) ...
- ... with a multispectral philosophy ...
- ... and a mathematical foundation ...
- ... agnostic of trace encoding semantics ...
- ... that happens to be mainly used for **performance analysis**



# Paraver mathematical foundation



# Views and .cfgs



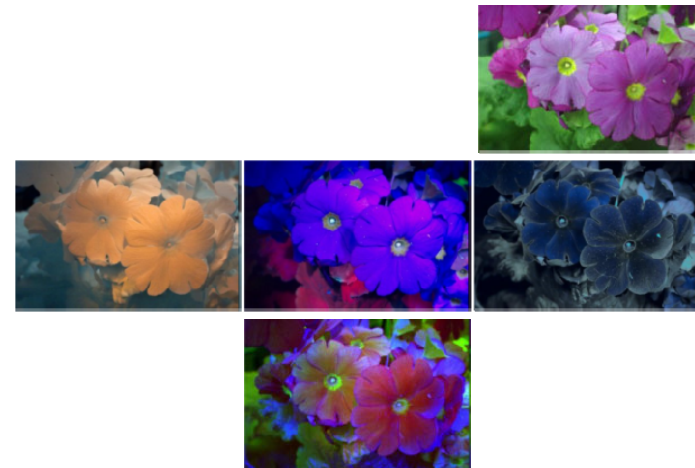
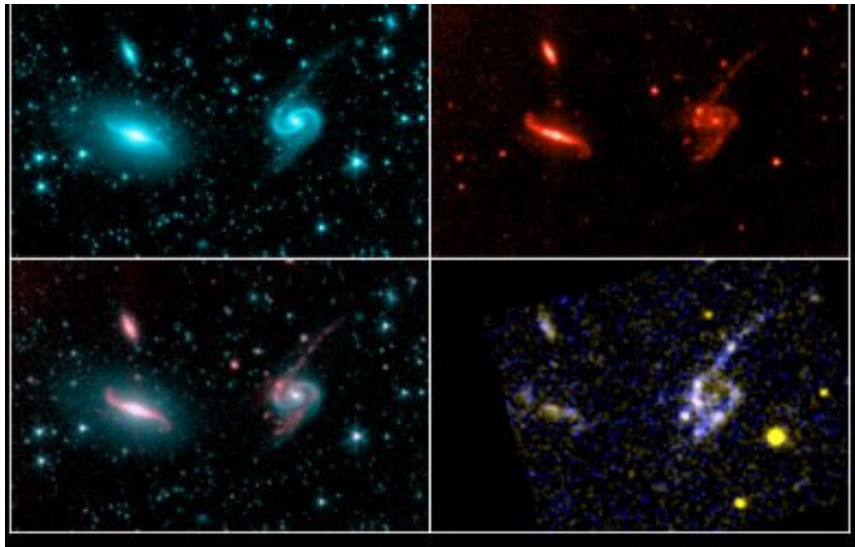
- For each view/window, the specification of how to compute the display from raw data is set through the main Paraver GUI
- Captured in configuration files (.cfg): ~DSL
  - Saved and restored
    - By instrumentation implementer / user
    - Capture semantics/insight, checkpoint, cooperative analysis, bug report, ...
  - Distribution repository and direct hints to relevant cfgs
  - Modes:
    - Basic: simple direct access to change relevant parameters
    - Full: complete access to all window setup parameters
- View/window names
  - ~ describe metric
  - Semantic consistency not enforced by tool
    - Rely on provider
    - May be checked by analyst



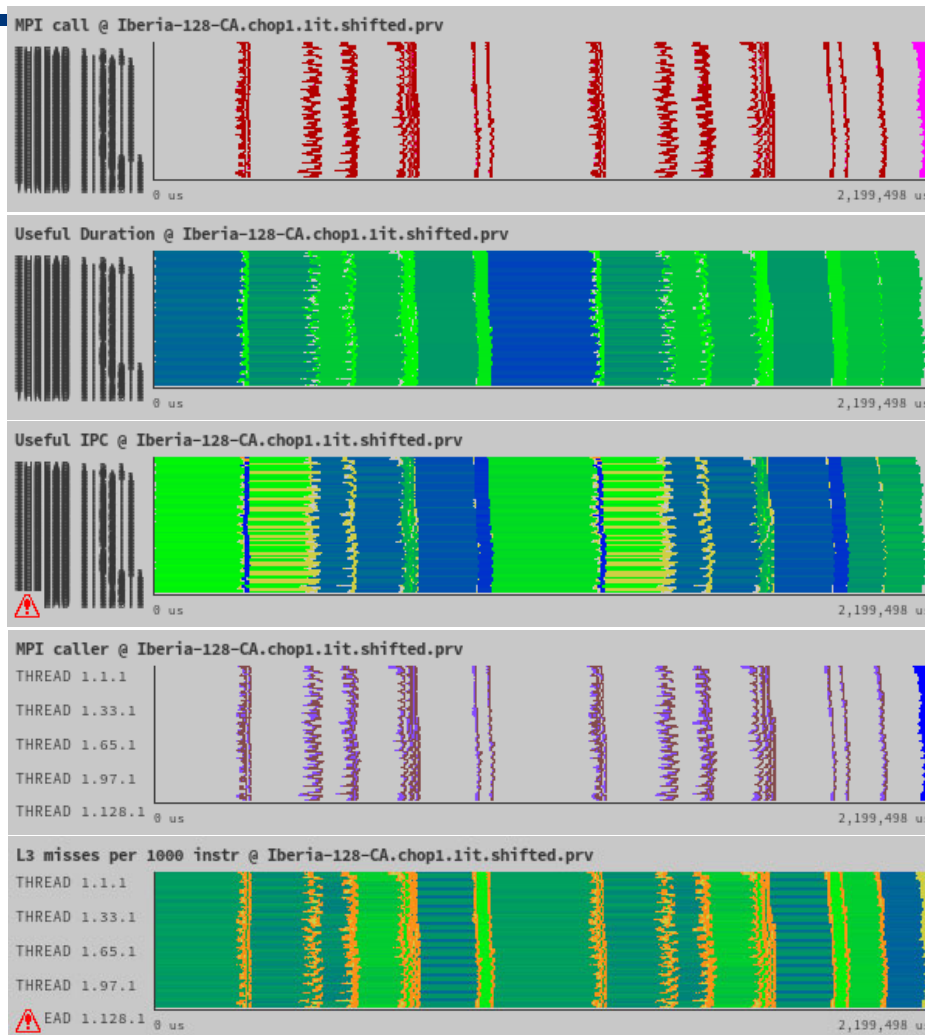
# Multispectral imaging



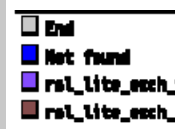
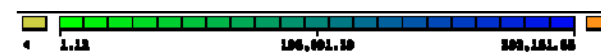
- Different looks at one reality
  - Different spectral bands (light sources and filters)
- Highlight different aspects
  - Can combine into false colored but highly informative images



# Multispectral timelines



Iberia WRF



|     |           |           |          |               |               |
|-----|-----------|-----------|----------|---------------|---------------|
| End | MPI_Isend | MPI_Irecv | MPI_Wait | MPI_Allreduce | MPI_Comm_rank |
|-----|-----------|-----------|----------|---------------|---------------|

zDP - MPI call profile @ Iberia-128-CA.chop1.iit.shifted.prv

| 2DP - MPI call profile @ Iberia-128-CA.chopi.iit.shifted.prv |           |           |               |
|--------------------------------------------------------------|-----------|-----------|---------------|
|                                                              | MPI_Isend | MPI_Wait  | MPI_Allreduce |
|                                                              | 26,780.15 | 26,780.15 | 1,776         |
|                                                              | 26,872.10 | 26,872.10 | 1,776         |
|                                                              | 26,872.10 | 26,872.10 | 1,776         |
|                                                              | 26,872.10 | 26,872.10 | 1,776         |
|                                                              | 26,872.10 | 26,872.10 | 1,776         |
|                                                              | 26,872.10 | 26,872.10 | 1,776         |
|                                                              | 26,872.10 | 26,872.10 | 1,776         |
|                                                              | 26,872.10 | 26,872.10 | 1,776         |
|                                                              | 27,128.89 | 27,128.89 | 1,776         |
|                                                              | 27,337.09 | 27,337.09 | 1,776         |
|                                                              | 27,753.48 | 27,753.48 | 1,776         |
|                                                              | 27,753.48 | 27,753.48 | 1,776         |
|                                                              | 27,753.48 | 27,753.48 | 1,776         |
|                                                              | 27,753.48 | 27,753.48 | 1,776         |
|                                                              | 27,753.48 | 27,753.48 | 1,776         |
|                                                              | 27,753.48 | 27,753.48 | 1,776         |
|                                                              | 27,802.07 | 27,802.07 | 1,776         |
|                                                              | 27,337.09 | 27,337.09 | 1,776         |
|                                                              | 27,753.48 | 27,753.48 | 1,776         |
|                                                              | 27,753.48 | 27,753.48 | 1,776         |
|                                                              | 27,753.48 | 27,753.48 | 1,776         |



# Multispectral histograms



- Histograms

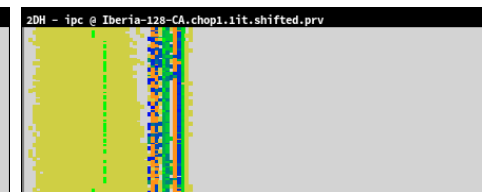
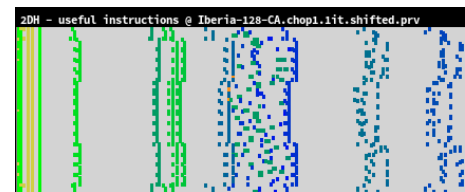
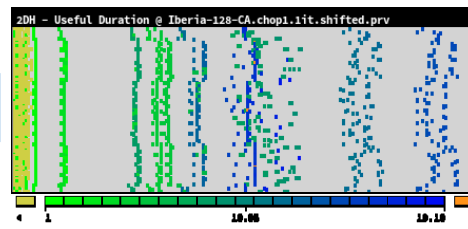
Useful duration

Useful instructions

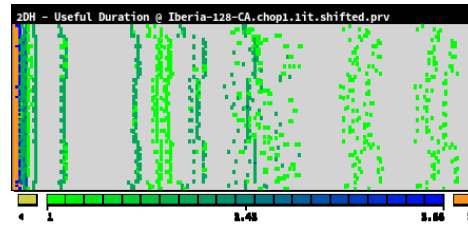
IPC

Iberia WRF

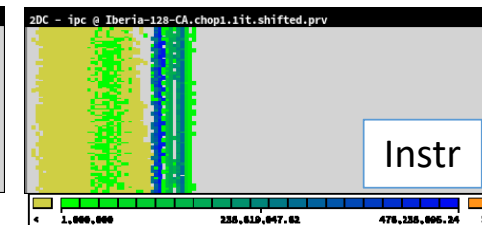
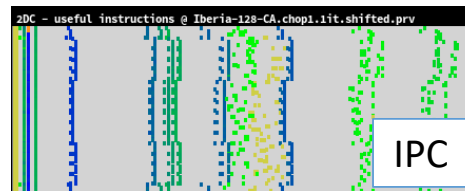
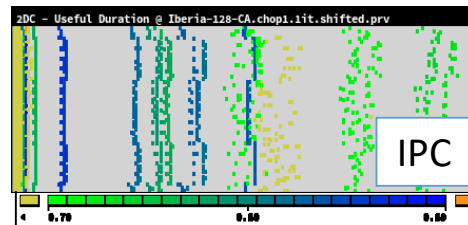
% time



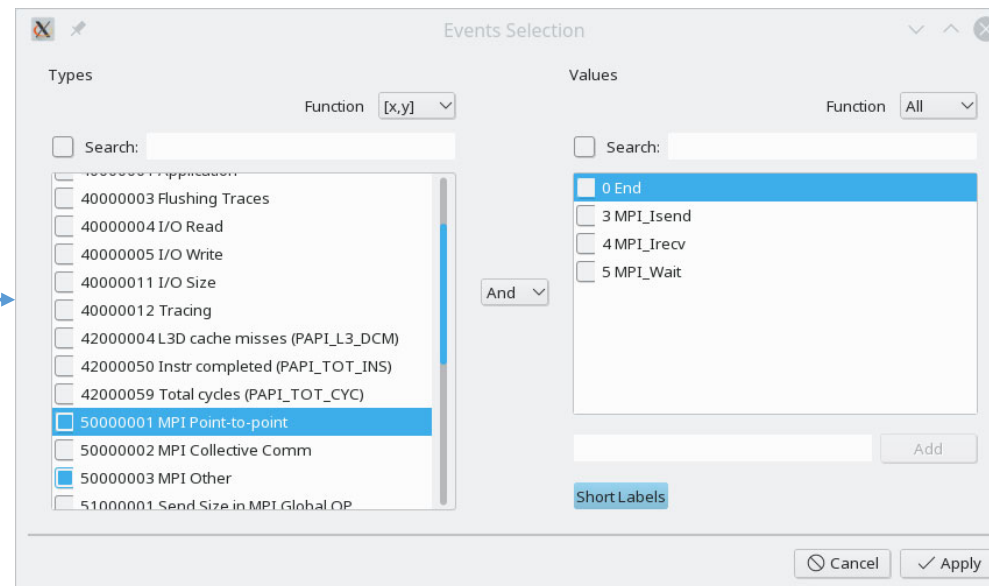
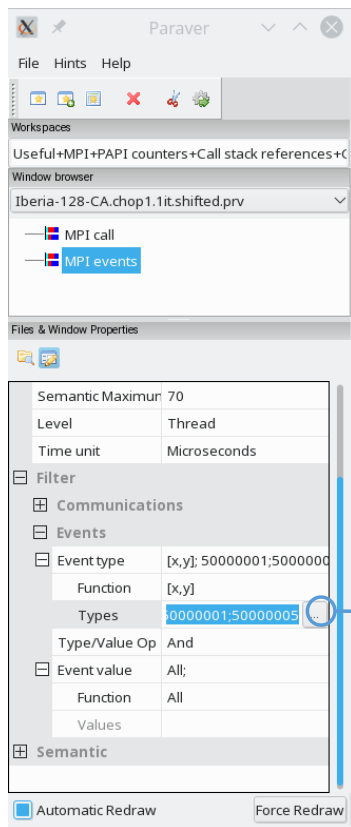
# bursts



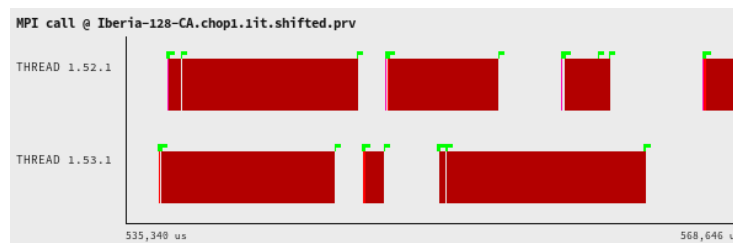
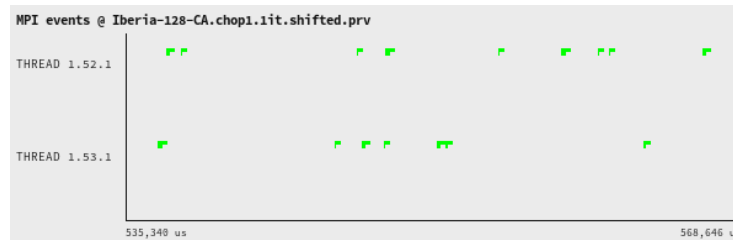
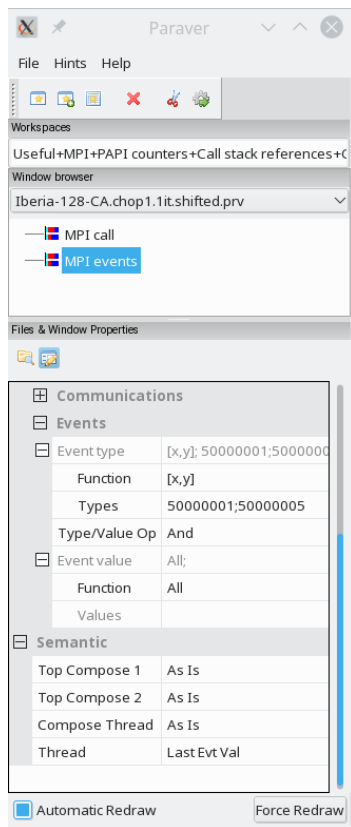
- Correlations



# From events to functions of time



# From events to functions of time



# From events to functions of time



Paraver

File Hints Help

Workspaces

Useful+MPI+PAPI counters+Call stack references+C

Window browser

Iberia-128-CA.chop1.1it.shifted.prv

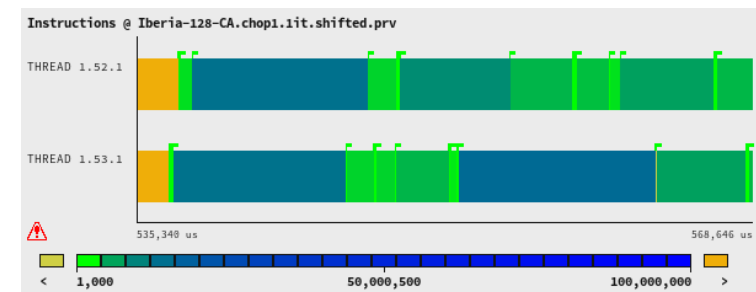
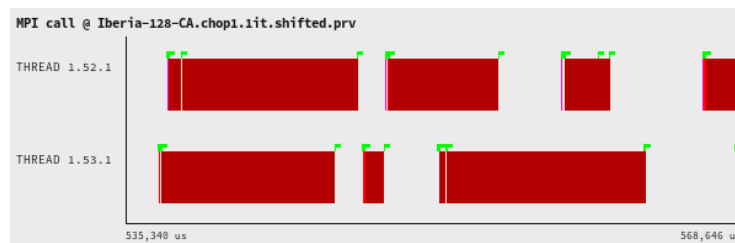
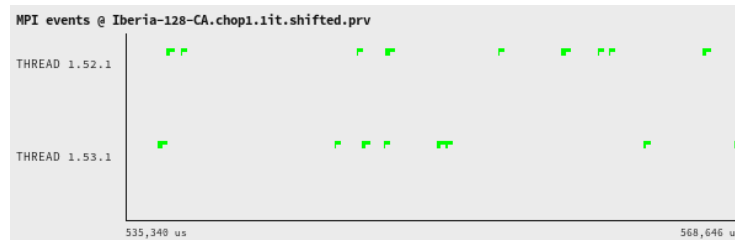
- MPI call
- MPI call
- Instruction Events

Files & Window Properties

Filter

- Communications
- Events
  - Event type = 42000050
    - Function =
    - Types 42000050
    - Type/Value Op And
  - Event value All
    - Function All
    - Values
- Semantic
  - Top Compose 1 As Is
  - Top Compose 2 As Is
  - Compose Thread As Is
  - Thread Next Evt Val

☒ Automatic Redraw Force Redraw



# Functions of time



## Semantic module

### From Events to functions of time

- Last event value  $S(t) = V(E_{V_{i-1}})$
- Next event value  $S(t) = V(E_{V_{i+1}})$
- Average Next Event Value  $S(t) = \frac{V(E_{V_{i+1}})}{T(E_{V_{i+1}}) - T(E_{V_i})}$
- Interval btw. Events  $S(t) = T(E_{V_{i+1}}) - T(E_{V_i})$



## Semantic module

### From communication records to functions of time

- Send Bytes  $S(t) = \sum_j Sx(C_j), j \mid (T_2(C_j) < t) \wedge (T_2(C_j) > t) \wedge (Dir(C_j) \rightarrow send)$
- Send Bandwidth  $S(t) = \sum_j \frac{Sx(C_j)}{T_2(C_j) - T_1(C_j)}, j \mid (T_2(C_j) < t) \wedge (T_2(C_j) > t) \wedge (Dir(C_j) \rightarrow send)$
- Mgs in transit  $S(t) = \sum_j sign(C_j), j \mid (T_2(C_j) < t) \wedge (T_2(C_j) > t) \wedge (Dir(C_j) \rightarrow send)$
- Recv. Bandwidth  $S(t) = \sum_j \frac{Sx(C_j)}{T_2(C_j) - T_1(C_j)}, j \mid (T_2(C_j) < t) \wedge (T_2(C_j) > t) \wedge (Dir(C_j) \rightarrow recv)$
- Recv. Negative Msgs  $S(t) = \sum_j sign(C_j), j \mid (T_2(C_j) < t) \wedge (T_2(C_j) > t) \wedge (Dir(C_j) \rightarrow recv)$
- Comm. Partner  $S(t) = Partner(C_j), j \mid (T_2(C_j) < t) \wedge (T_2(C_j) > t)$
- Bytes btw. Events  $S(t) = \sum_j Sx(C_j), j \mid T_2(C_j) \in [T(E_{V_i}), T(E_{V_{i+1}})) \vee T_2(C_j) \in [T(E_{V_i}), T(E_{V_{i+1}}))$



## Composition

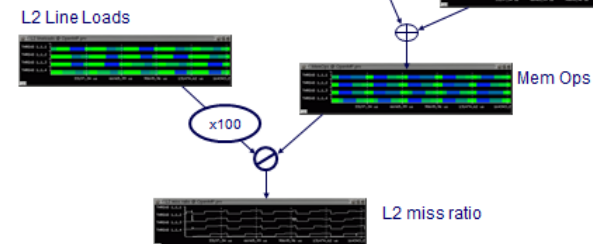
- $S'(t) = f(S(t))$   $S' = f \circ S$
- Sign  $S'(t) = sign(S(t))$
- 1-sign  $S'(t) = 1 - sign(S(t))$
- Select range  $S'(t) = S(t) \in [a, b] ? S(t) : 0$
- Sign \* Is equal  $S'(t) = sign(S(t) = a ? S(t) : 0)$
- Delta  $S'(t) = S_i + 1 - S_i$
- Stacked value



## Semantic module

### Derived windows

- Point wise operation
  - $S = \alpha * S^* <op> \beta * S^*$
  - $<op> : +, -, *, /, \dots$



# Piecewise constant functions of time



- Events certainly not uniformly distributed in time !
  - Piecewise constant functions, but not uniform length of constant interval
- Extrae tries to use same timestamps for several events if they are fired by the same physical action ... but not all events are necessarily generated at every timestamp (eg. instrumentation and sampling events)
  - Different functions of time may not change at the same points
- Still, in all cases you can derive/compose any pair of functions of time, still generating a piecewise constant function of time
  - Of course, with potentially different intervals



# Rendering

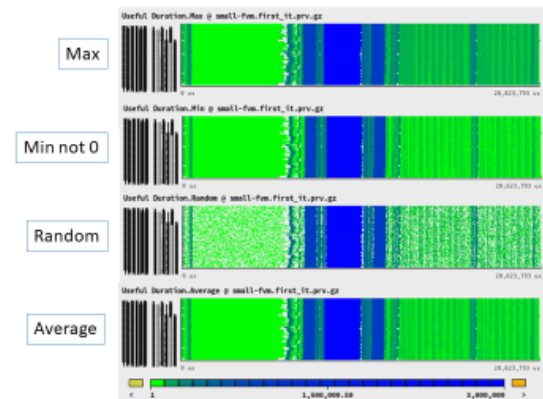
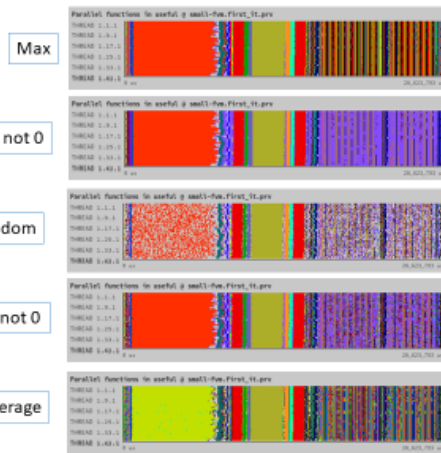


## Rendering

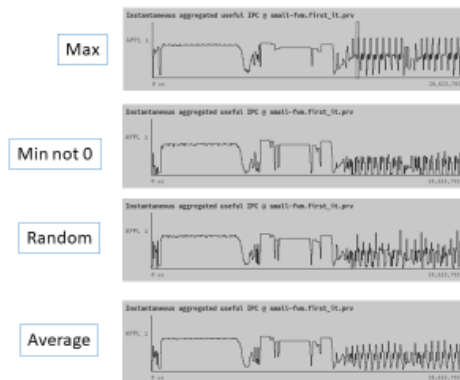
## Rendering

### • Categorical metrics

### • Numerical value



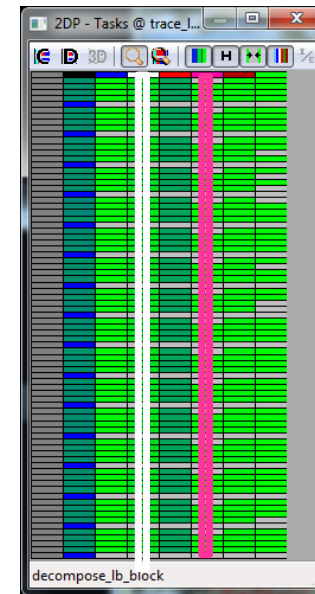
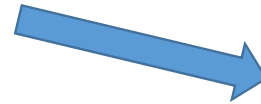
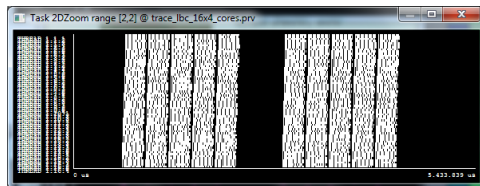
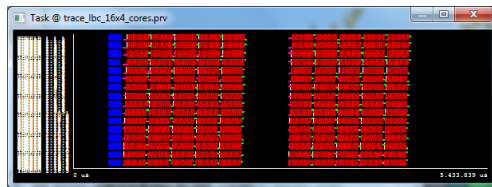
### • Numerical value



# From tables to timelines



- Where in the timeline do the values in certain table columns appear?
- ie. want to see the time distribution of a given routine?



Only showing when a  
given value happens



# The approach



- The “scientific method”

- The engine

- Curiosity, interest

- Tools

- “Just” instruments to help validate/reject

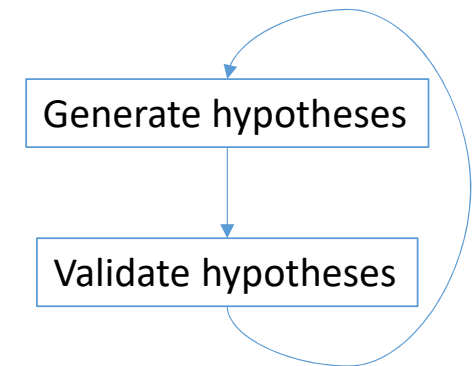
- **Do I have the data in the trace to validate/reject?**

- **Do I have configuration files exposing the metrics to validate/reject ?**

- **If not, can I generate the timelines/tables exposing the metrics to validate/reject**

- An “infinite loop”:

- there is always a next question
    - Decide to Exit: fatigue, honesty



# “useful” cfgs



- Per thread:
  - Useful duration
  - MPI calls, MPI call duration, message sizes, effective MPI call bandwidth
  - OpenMP outlined functions
  - Hardware counters: Instructions, IPC, frequency, cache miss ratios
  - Link to source: caller / caller line at different call stack levels
  - Sanity checks: flushing, preempted time
- Aggregated:
  - Instantaneous parallelism
  - Average instantaneous granularity (useful duration)
  - Aggregated useful IPC
  - Aggregated effective network bandwidth, memory bandwidth.

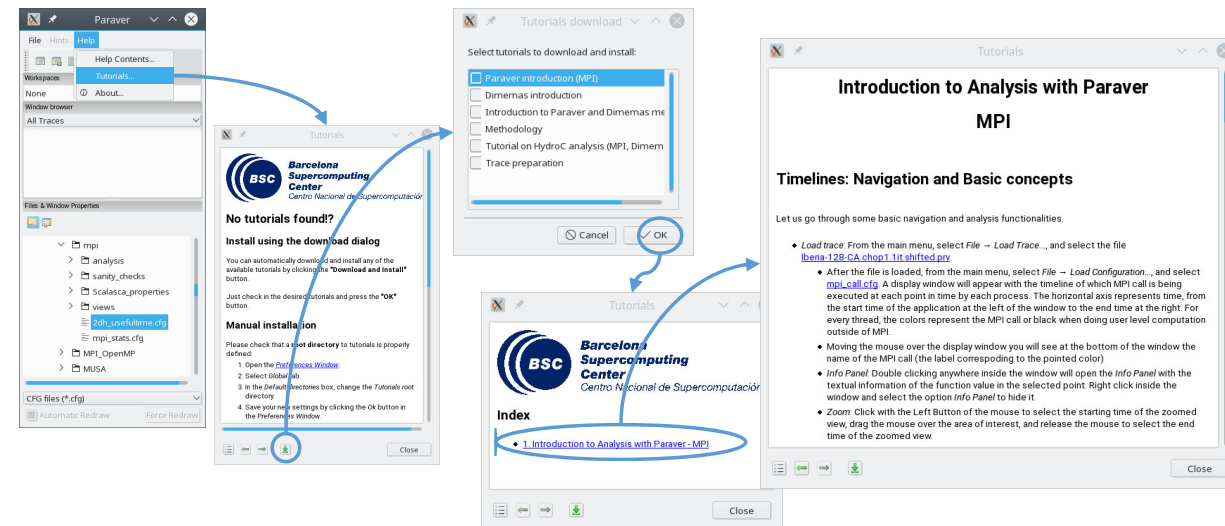


# Further resources



- BSC Tools

- More tutorial slides
  - Tools and methodology
  - Introduction & details
- Download (<https://tools.bsc.es/>)
  - Tools
  - Paraver hands on Tutorials:
    - Launch paraver
    - set preferences Tutorials root directory
    - Download and follow tutorial



- POP material

- Webinars
  - [Understand the Performance of your Application with just Three Numbers](#)
  - [Impact of Sequential Performance on Parallel Codes](#)
  - [Identifying Performance Bottlenecks in Hybrid MPI + OpenMP Software](#)
- Online training (<https://pop-coe.eu/further-information/online-training>)
  - Using Paraver
  - Advanced POP online Training on Paraver





# Performance Optimisation and Productivity

A Centre of Excellence in HPC

Contact:

 <https://www.pop-coe.eu>

 [pop@bsc.es](mailto:pop@bsc.es)

 [@POP\\_HPC](https://twitter.com/POP_HPC)

 [youtube.com/POPHPC](https://youtube.com/POPHPC)

