



Barcelona Supercomputing Center Centro Nacional de Supercomputación





# Performance analysis: Paraver hands-on using the HARMONIE model

Xavier Yepes-Arbós Mario C. Acosta Miguel Castrillo

ESiWACE Summer School on Effective HPC for Climate and Weather 2021



# Introduction

- In this session we will show the basics of Paraver to analyze a trace taking into account different aspects:
  - Load balance
  - Communication
  - Computation
- Using Paraver we will have both a global **qualitative** perception of the application and detailed **quantitative** information.
- The trace was obtained from a small execution of the **HARMONIE** weather forecast model, ran on the **ECMWF's** Cray XC40 system.
- It used **285 MPI processes** and without OpenMP.







# What is a trace made of?

- A Paraver **trace** is made up from **three files**: \*.prv, \*.pcf and \*.row
- For this **hands-on** we have:
  - harmonie\_285MPI\_noopenmp.pcf: Trace data (performance events).
  - harmonie\_285MPI\_noopenmp.pcf: Types of performance events in the trace.
  - harmonie\_285MPI\_noopenmp.row: ids of MPI processes and OMP threads.
- In particular, if you search for EVENT\_TYPE in the pcf file, you will see what the trace contains exactly. For example:

| EVE | NT_TYPE                     |
|-----|-----------------------------|
| 9   | 50000001 MPI Point-to-point |
| VAL | UES                         |
| 33  | MPI_Bsend                   |
| 4   | MPI_Irecv                   |
| 3   | MPI_Isend                   |
| 2   | MPI_Recv                    |
| 1   | MPI_Send                    |
| 5   | MPI_Wait                    |
| 59  | MPI_Waitany                 |
| 0   | Outside MPI                 |







#### **First steps**

• Navigate into the Paraver directory:

cd ~/perf-analysis/wxparaver-4.8.2-Linux\_x86\_64/bin

• **Open Paraver** by executing:

wxparaver&

• A Paraver window such as this one should pop up:













#### **Cut & filter**

In case of a trace is very big (> 1 GB), it might be **cut** and/or **filtered** to make it

manageable with Paraver Cut & Filter Traces Input harmonie 285MPI noopenmp.prv Browse Output Browse Load the processed trace Run application with the processed trace Cut/Filter Parameters Configuration file Browse Execution chain Paraver 2.- Filter Save... î 3.- Software Counters File Hints Help 1 X Cutter Filter Software Counters da Trace Limits Begin O Cut by time Cut by time % End Tasks Select Region.. All Window All Trace Trace Options Use original time Remove first state Don't break states Remove last state Keep boundary events Output Trace Maximum trace size 0 Ĵ MB Cancel·la Aplica

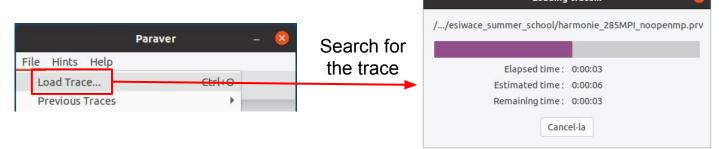






# (Un)load a trace

"File > Load Trace..." and search for the harmonie\_285MPI\_noopenmp.prv file. Loading trace...



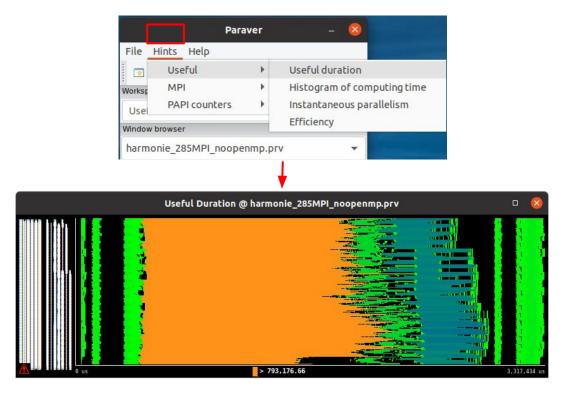
• "File > Unload Traces..." and select the trace that you want to unload.

|                    |        |   |                | Unload Traces       |
|--------------------|--------|---|----------------|---------------------|
|                    |        |   | Select the tra | ces to unload:      |
| Para               | aver – | 8 | Marmonie       | 285MPI_noopenmp.prv |
| File Hints Help    |        |   |                |                     |
| Load Trace         | Ctrl+O |   |                |                     |
| Previous Traces    | •      |   |                |                     |
| Unload Traces      |        |   |                |                     |
| Load Configuration |        |   |                |                     |
|                    |        |   |                |                     |
|                    |        |   |                | Cancel·la D'acor    |

Centro Nacional de Supercomputación

### **Use of Paraver hints**

- Paraver provides some **hints**, which are views and histograms of the performance metrics **included** in the trace.
- In "Hints", feel free to load whatever option you want.





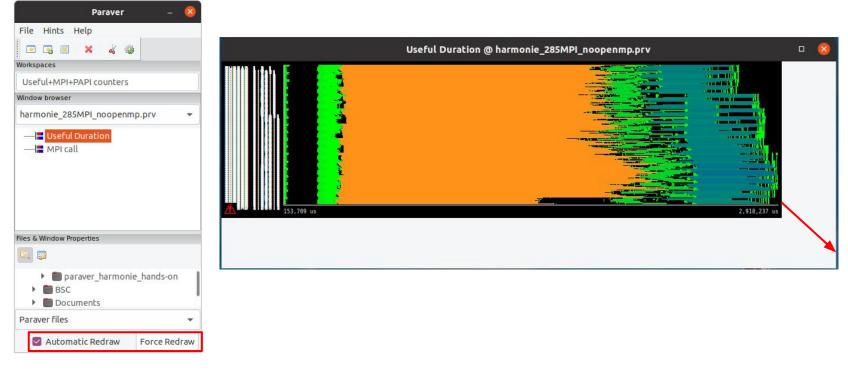






## **Basics of Paraver: resize**

- When you resize the window of a view, Paraver automatically redraws it.
- **Depending** on the **situation**, this might be useful (such in our case), or it might be a problem due to the time of rendering the view again.





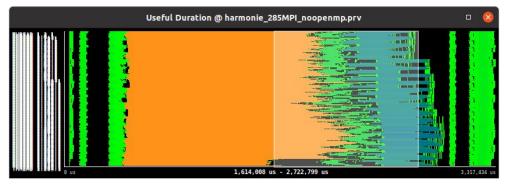
Barcelona Supercomputing Center Centro Nacional de Supercomputación





# **Basics of Paraver: (undo) zoom**

 To make zoom into a period of time, just click on the desired area. If you also interested on zooming in MPI processes, press Ctrl while clicking the desired area.



• To undo the zoom, right click on the window and "Undo Zoom"

|   | Сору           | Ctrl+C |  |
|---|----------------|--------|--|
|   | Paste          | +      |  |
|   | Clone          |        |  |
|   | Undo Zoom      | Ctrl+U |  |
| _ | Redo Zoom      | Ctrl+R |  |
|   | Fit Time Scale |        |  |







# **Basics of Paraver: copy & paste**

- It is possible to **copy** and **paste** duration, time, size, etc between windows.
- To do so, right click on the window and "Copy" or "Paste > X".

| and the second second                 | Useful Duration @ har | monie_285MPI_noop | enmp.prv       |                               |
|---------------------------------------|-----------------------|-------------------|----------------|-------------------------------|
| i likh 🛛 🖉 🗾                          | Сору                  | Ctrl+C            |                |                               |
|                                       | Paste                 | Þ                 |                |                               |
|                                       | Clone                 |                   |                |                               |
|                                       | Undo Zoom             | Ctrl+U            |                |                               |
|                                       | Redo Zoom             |                   |                | areasting and a second second |
|                                       | Fit Time Sca          |                   |                |                               |
|                                       | Fit Semanti           | c Scale           |                |                               |
|                                       | Fit Objects           |                   |                | -ondi -                       |
|                                       | Select Object         | cts               |                |                               |
| THREAD 1.1.1                          |                       |                   |                |                               |
| THREAD 1.72.1                         | Сору                  | Ctrl+C            |                |                               |
| THREAD 1.143.1                        | Paste                 | ÷.                | Default        | Ctrl+V                        |
| THREAD 1.214.1<br>THREAD 1.285.1 0 us | Clone                 |                   | Time           | r                             |
| 0 us                                  | Undo Zoom             | Ctrl+U            | Objects        |                               |
|                                       | Redo Zoom             | Ctrl+R            | Size           |                               |
|                                       | Fit Time Scale        |                   | Duration       |                               |
| a FRA                                 | Fit Semantic Scale    | •                 | Semantic Scale |                               |
| 3 1 1 1                               | Fit Objects           |                   | Filter         | •                             |
| A. C. State                           | Select Objects        |                   | Paste Special  | 32                            |



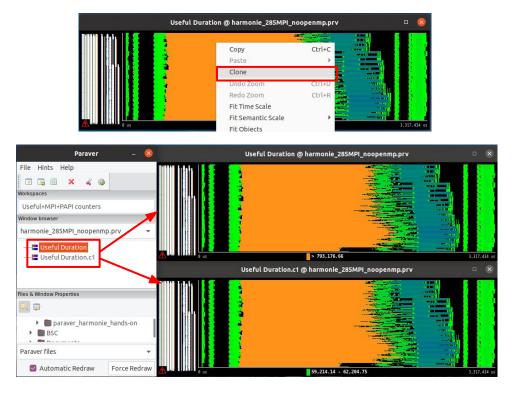
Barcelona Supercomputing Center Centro Nacional de Supercomputación





## **Basics of Paraver: clone**

- In some situations, it might be useful to clone the same window to add a complementary visualization, compare different metrics, etc.
- To do so, right click on the window and "Clone"



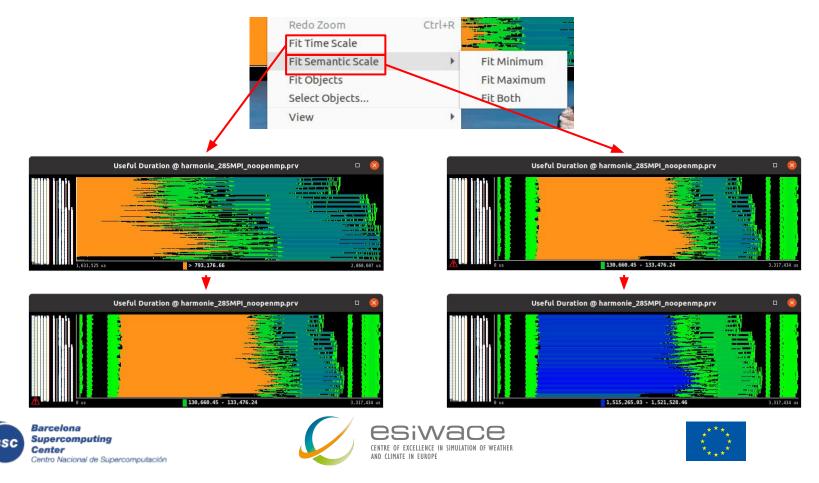






# **Basics of Paraver: time and semantic scales**

It is useful to **adjust** both the **time scale** (x axis) and the **semantic scale** (colors). The latter is important to paint events taking into account the minimum and maximum values.



# View (communication lines and event flags)

There are two interesting options to draw the **communication lines** and/or the **event flags**.









# Window properties section

- In the Paraver main window, the Window
   Properties section provides detailed information about what it is represented in the current view.
- It allows to change different values, such as the time scale, the semantic scale, which communications to draw, what types of events, etc.

|                  |                     | Paraver – 🥸              |  |  |
|------------------|---------------------|--------------------------|--|--|
| Fil              | e Hints Help        |                          |  |  |
|                  | • 🕞 🗐 🗙             | 4 4                      |  |  |
| ÷                | kspaces             |                          |  |  |
|                  | seful+MPI+PAPI cou  | oters                    |  |  |
|                  | dow browser         | licers                   |  |  |
|                  |                     |                          |  |  |
| ha               | rmonie_285MPI_no    | openmp.prv 🔻             |  |  |
| -                | MPI call            |                          |  |  |
|                  |                     |                          |  |  |
|                  |                     |                          |  |  |
|                  | & Window Properties |                          |  |  |
| Eq               | <b>1</b>            |                          |  |  |
| 1                | Name                | MPI call                 |  |  |
| 1                | Begintime           | 0.00 us                  |  |  |
| 1                | End time            | 3,317,433.79 us          |  |  |
|                  | Semantic Minimum    | 0                        |  |  |
| Semantic Maximum |                     | 70                       |  |  |
|                  | Level               | Thread                   |  |  |
| ł                | Time unit           | Microseconds             |  |  |
|                  | Filter              |                          |  |  |
|                  | Communication       | S                        |  |  |
|                  | Logical             | $\checkmark$             |  |  |
|                  | Physical            |                          |  |  |
|                  | Comm from           | All;                     |  |  |
|                  | From/To Op          | And                      |  |  |
|                  | E Comm to           | All;                     |  |  |
|                  | 🗄 Comm tag          | All;                     |  |  |
|                  | Tag/Size Op         | And                      |  |  |
|                  | ⊞ Comm size         | All;                     |  |  |
|                  | 🕀 Comm bandwidt     | All;                     |  |  |
| 1                | Events              |                          |  |  |
|                  | Event type          | [x,y]; 50000001;50000005 |  |  |
|                  | Function            | [x,y]                    |  |  |
|                  | Types               | 5000001;50000005         |  |  |
|                  | Type/Value Op       | And                      |  |  |
|                  | Event value         | All;                     |  |  |
| ± :              | Semantic            |                          |  |  |
|                  |                     |                          |  |  |







# Info panel

- An **Info Panel** might be enabled by right-clicking on the window and choosing "Info Panel".
- It contains the legend of the window, timing, duration, etc.

| MPI call @ ha  | rmonie_285MPI_noopenm   | ıp.prv           | • 😣          |
|--|---|------------------|--------------|
| THREAD 1.1.1<br>THREAD 1.36.1<br>THREAD 1.71.1   | Copy<br>Paste<br>Clone  | Ctrl+C           |              |
| THREAD 1.106.1         THREAD 1.141.1         THREAD 1.176.1         THREAD 1.211.1         THREAD 1.246.1         THREAD 1.283.1         g us | Undo Zoom<br>Redo Zoom<br>Fit Time Scale<br>Fit Semantic Scale<br>Fit Objects<br>Select Objects<br>View | Ctrl+R           | 3,317,434 us |
| What / Where     Timing     Colors       Outside MPI       MPI_Recv       MPI_Isend       MPI_Irecv  | Paint As<br>Drawmode<br>Pixel Size<br>Object Labels<br>Object Axis                                      | ><br>><br>><br>> |              |
| MPI_Wait<br>MPI_Alltoallv<br>MPI_Comm_size   | Run<br>Synchronize<br>Save  | •<br>•           |              |
| MPI_Waitany  | Timing  | Ctrl+T           |              |



Barcelona Supercomputing Center Centro Nacional de Supercomputación





#### Save a snapshot

- In case you are interested in saving the view of a window, it is possible to save it.
- To do so, right-click on the window and "Save > Image...".

| Synchronize | •                                       |               |  |
|-------------|---|---------------|--|
| Save        | •                                       | Configuration |  |
| Timing      | Ctrl+T                                  | Image         |  |
| Info Panel  |   | Image Legend  |  |
|             | ALL | Text          |  |

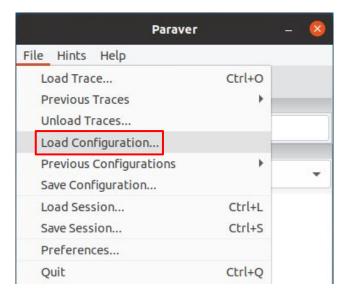






# Load a configuration

- Paraver allows studying a lot of different performance metrics by loading configuration files (.cfg) that tell the what information to read and how to represent it.
- To load a specific configuration file, go to "File > Load Configuration..." and search for the file that you want to load.









### Load/save a session

- Sometimes you might have many different views open, and you would like to **save your session** to continue the work at another time.
- To **load** a session, go to "File > Load Session..." and search for the session.
- To **save** an ongoing session, go to "File > Save Session..." and save it.

| Paraver                 |        | - 🤇 |
|-------------------------|--------|-----|
| le Hints Help           |        |     |
| Load Trace              | Ctrl+O |     |
| Previous Traces         | +      |     |
| Unload Traces           |        |     |
| Load Configuration      |        |     |
| Previous Configurations | •      | 22  |
| Save Configuration      |        |     |
| Load Session            | Ctrl+L |     |
| Save Session            | Ctrl+S |     |
| Preferences             |        |     |
| Quit                    | Ctrl+Q |     |

|  |        | Load                      | session       |              |                              |   |
|--|--------|---------------------------|---------------|--------------|------------------------------|---|
|  | Kavier | harmonie_traces           | esiwace_summe | r_school     | cfgs                         |   |
| Llocs  | Nom    | -<br>                     |               | Mida         | Modificat                    |   |
| Q Cerca<br>③ Utilitzats                          | ha ha  | rmonie_trace_all_cf<br>gs | gs_session    |              | 13/7/20<br>13/7/20           |   |
| cfgs xavieryepes Escriptori Sistema d HDD_Xavier | E ha   | rmonie_trace_all_cf       | gs.session    | 127 bytes    | i 13/7/20                    |   |
| + -  |        |                           | Pa            | raver sessio | on <mark>(*.session</mark> ) |   |
|  |        |                           |               | Cancel       | la Obre                      | 2 |







#### Notes

The following slides show **views** of the **different configuration files** used for the **hands-on**.

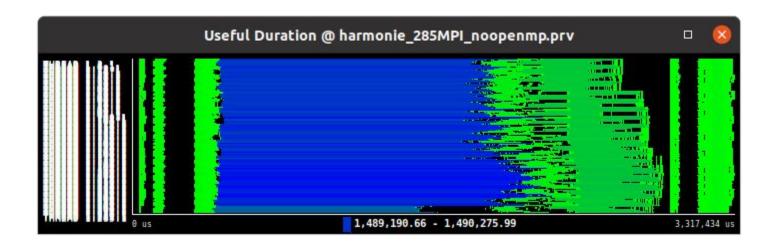






### **Useful duration**

useful\_duration.cfg: It shows the duration of the computational chunks.



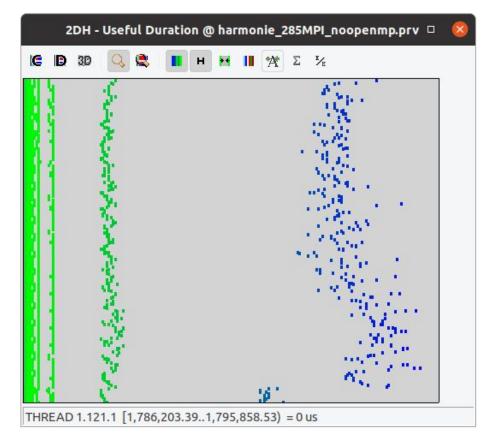






# **Histogram of useful duration**

**useful\_duration\_histogram.cfg:** It shows a histogram of the duration of the computational chunks to study the load balance.









#### **MPI calls**

**mpi\_call.cfg:** It shows the different calls to MPI functions along the execution.









#### **MPI profile**

mpi\_stats.cfg: It shows MPI statistics, to know the load balance, parallel efficiency, etc.

|         | Outside MPI | MPI_Recv | MPI_Isend | MPI_Irecv | MPI_Wait   | MPI_Alltoallv | MPI_Comm_size | MPI_Waitany |
|---------|-------------|----------|-----------|-----------|------------|---------------|---------------|-------------|
| Total   | 20,751.31 % | 2.16 %   | 44.77 %   | 13.23 %   | 1,058.01 % | 4,197.54 %    | 11.15 %       | 2,421.84 %  |
| Average | 72.81 %     | 0.01 %   | 0.16%     | 0.05 %    | 3.71 %     | 14.73 %       | 0.04 %        | 8.50 %      |
| Maximum | 83.79 %     | 0.08 %   | 0.91 %    | 0.18 %    | 11.58 %    | 19.93 %       | 0.05 %        | 23.27 %     |
| Minimum | 57.97 %     | 0.00 %   | 0.02 %    | 0.01 %    | 0.65 %     | 11.23 %       | 0.03 %        | 1.03 %      |
| StDev   | 5.00 %      | 0.01 %   | 0.13 %    | 0.03 %    | 2.04 %     | 2.48 %        | 0.00 %        | 3.86 %      |
| Avg/Max | 0.87        | 0.11     | 0.17      | 0.26      | 0.32       | 0.74          | 0.82          | 0.37        |

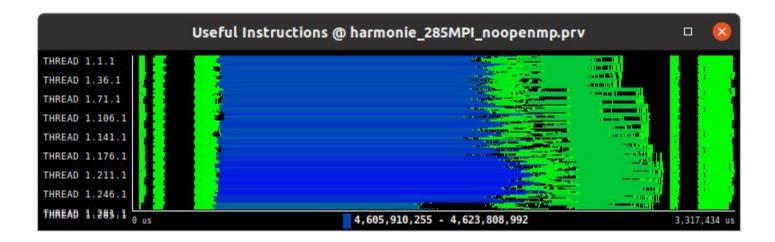






## **Useful instructions**

useful\_instructions.cfg: It shows the number of useful instructions executed.



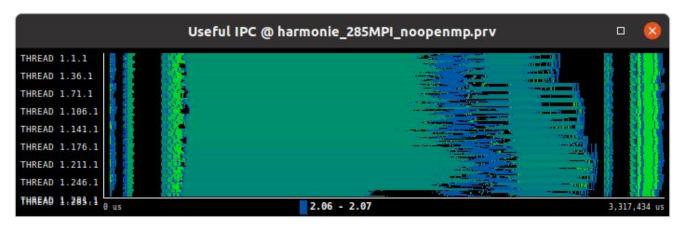






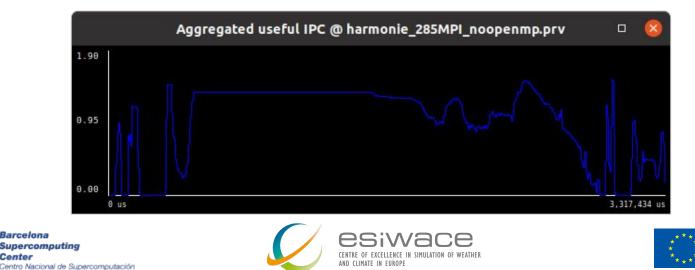
#### **IPC**

• useful\_ipc.cfg: It shows the useful Instructions per Cycle (IPC).



• aggregated\_useful\_ipc.cfg: It shows the aggregated useful IPC.

Center

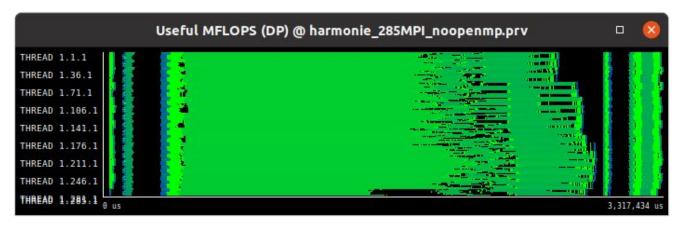


#### **MFLOPS**

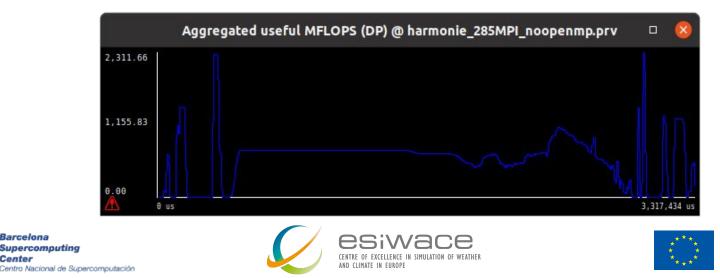
• useful\_mflops.cfg: It shows the useful MFLOPS

Barcelona

Center

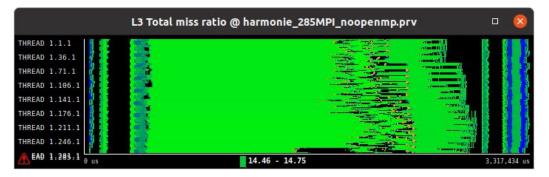


• aggregated\_useful\_mflops.cfg: It shows the aggregated useful MFLOPS.



## L3 cache misses

• useful\_L3\_misses\_per\_1000\_instructions.cfg: It shows the ratio of L3 cache misses per 1000 useful instructions.

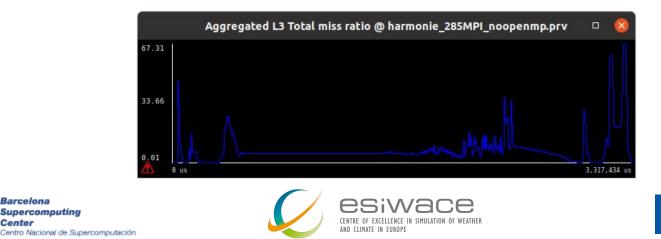


• aggregated\_useful\_L3\_misses\_per\_1000\_instructions.cfg: It shows the

aggregated ratio of L3 cache misses per 1000 useful instructions.

Barcelona

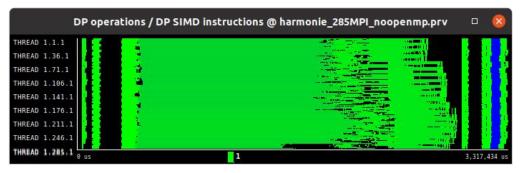
Center



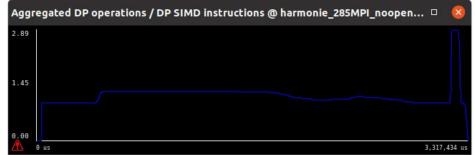


# **Vectorization efficiency**

• useful\_double-precision\_vectorization\_ratio.cfg: It shows the percentage of double precision floating-point operations per SIMD instruction.



 aggregated\_useful\_double-precision\_vectorization\_ratio.cfg: It shows the aggregated percentage of double precision floating-point operations per SIMD instruction.









#### Bonus

- If you feel encouraged to explore more options of Paraver, feel free to "play" with any functionality.
- We suggest you to **explore different configuration files** included in the **Paraver folder**: wxparaver-4.8.2-Linux\_x86\_64/cfgs
- Note that not all configuration files might be useful for the trace of this hands-on since it depends on the available events in the trace.









Barcelona Supercomputing Center Centro Nacional de Supercomputación



# Thank you



1-1-1





#### miguel.castrillo@bsc.es