Performance Analysis Exercises with Vampir

Matthias Weber, Holger Brunst, Hartmut Mix
Technische Universität Dresden

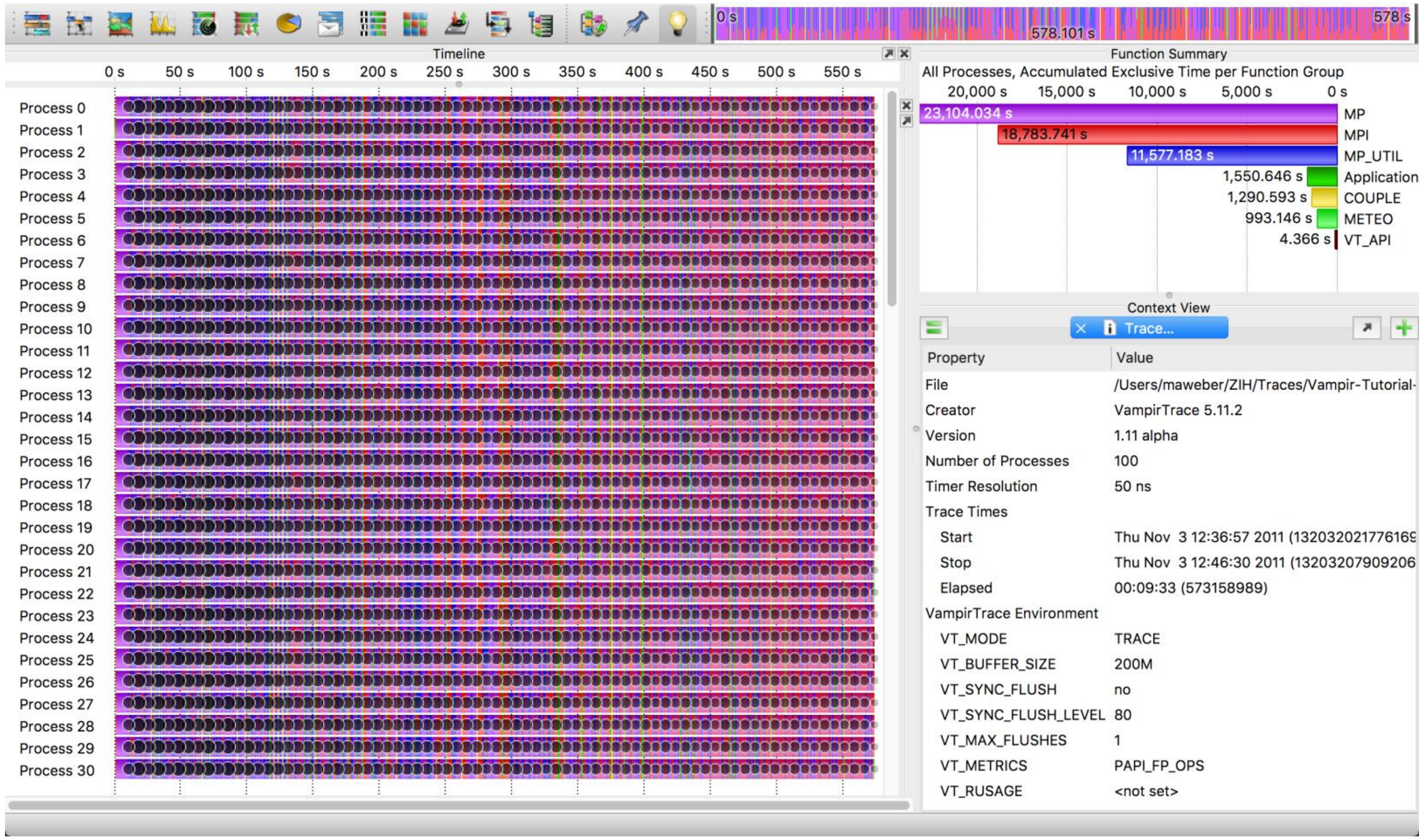
Exercise Trace Files

```
% ls $TW35/trace-examples/\
          Vampir-Tutorial-Analysis-Examples/

01-p100-cosmo-specs-orig
02-p100-cosmo-specs-fd4
03_wrf_deimos
04_sbmfd4_jugene
```

- Four trace files for exercising performance analysis with Vampir
- Traces show real application runs
- Do the traces contain performance problems?
- If yes, try to find their causes

01-p100-cosmo-specs-orig

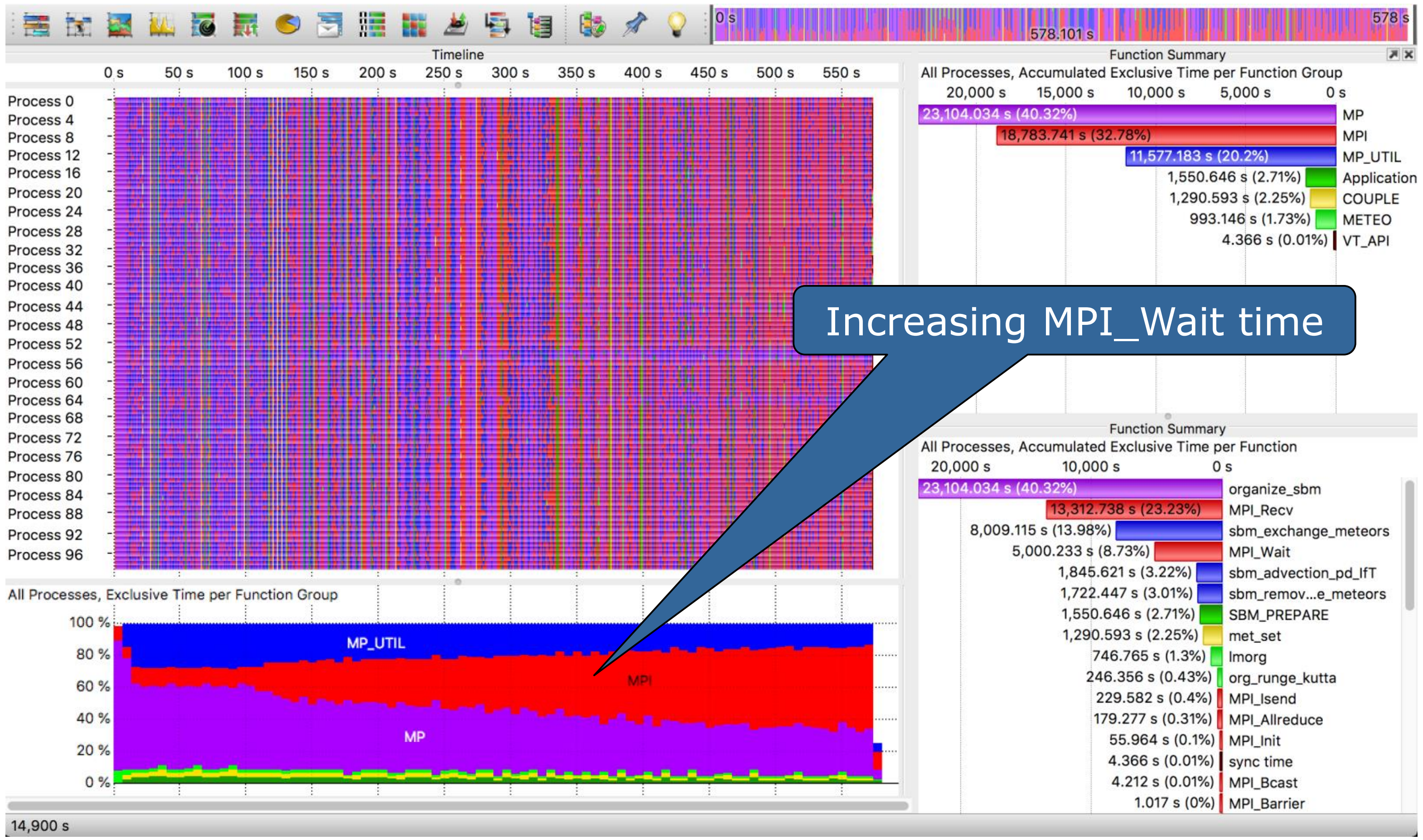


Function Group	Time
MP	23,104.034 s
MPI	18,783.741 s
MP_UTIL	11,577.183 s
Application	1,550.646 s
COUPLE	1,290.593 s
METEO	993.146 s
VT_API	4.366 s

Property	Value
File	/Users/maweber/ZIH/Traces/Vampir-Tutorial-
Creator	VampirTrace 5.11.2
Version	1.11 alpha
Number of Processes	100
Timer Resolution	50 ns
Trace Times	
Start	Thu Nov 3 12:36:57 2011 (132032021776169
Stop	Thu Nov 3 12:46:30 2011 (13203207909206
Elapsed	00:09:33 (573158989)
VampirTrace Environment	
VT_MODE	TRACE
VT_BUFFER_SIZE	200M
VT_SYNC_FLUSH	no
VT_SYNC_FLUSH_LEVEL	80
VT_MAX_FLUSHES	1
VT_METRICS	PAPI_FP_OPS
VT_RUSAGE	<not set>

- Weather forecast code COSMO-SPECS
- Run with 100 processes
- COSMO: weather model (METEO group)
- SPECS: microphysics for accurate cloud calculation (MP and MP_UTIL group)
- Coupling of both models done in COUPLE group

01-p100-cosmo-specs-orig



Function Summary

All Processes, Accumulated Exclusive Time per Function Group

Time	Function Group
23,104.034 s (40.32%)	MP
18,783.741 s (32.78%)	MPI
11,577.183 s (20.2%)	MP_UTIL
1,550.646 s (2.71%)	Application
1,290.593 s (2.25%)	COUPLE
993.146 s (1.73%)	METEO
4.366 s (0.01%)	VT_API

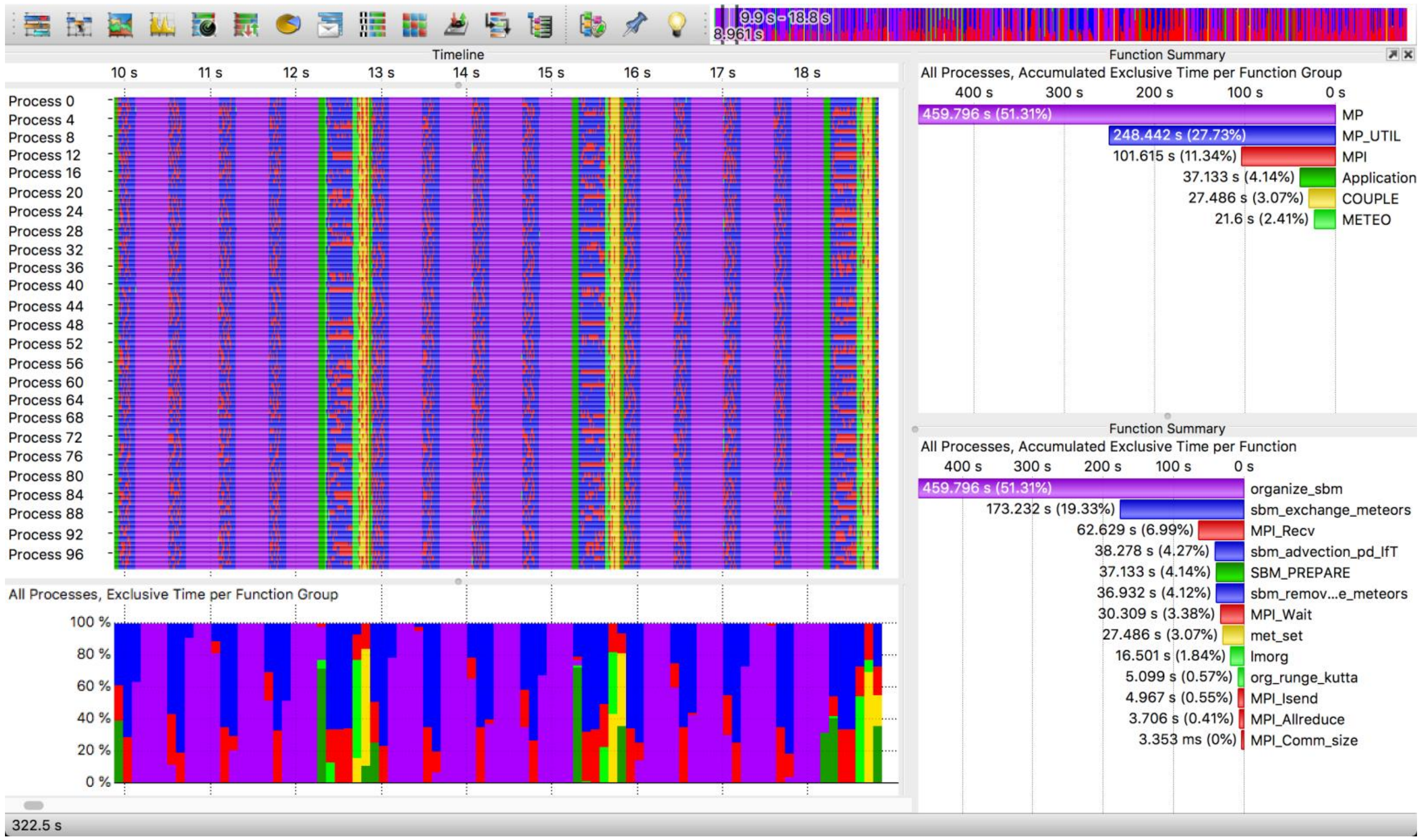
Function Summary

All Processes, Accumulated Exclusive Time per Function

Time	Function
23,104.034 s (40.32%)	organize_sbm
13,312.738 s (23.23%)	MPI_Recv
8,009.115 s (13.98%)	sbm_exchange_meteors
5,000.233 s (8.73%)	MPI_Wait
1,845.621 s (3.22%)	sbm_advection_pd_lfT
1,722.447 s (3.01%)	sbm_remov...e_meteors
1,550.646 s (2.71%)	SBM_PREPARE
1,290.593 s (2.25%)	met_set
746.765 s (1.3%)	lmorg
246.356 s (0.43%)	org_runge_kutta
229.582 s (0.4%)	MPI_Isend
179.277 s (0.31%)	MPI_Allreduce
55.964 s (0.1%)	MPI_Init
4.366 s (0.01%)	sync time
4.212 s (0.01%)	MPI_Bcast
1.017 s (0%)	MPI_Barrier

- Compared to METEO, MP and MP_UTIL are very compute intensive, however this is due to more complex calculations and no performance issue
- Problem: >32% of time spent in MPI
- MPI runtime share increases throughout the application run

01-p100-cosmo-specs-orig



Function Summary

All Processes, Accumulated Exclusive Time per Function Group

Time	Function Group
459.796 s (51.31%)	MP
248.442 s (27.73%)	MP_UTIL
101.615 s (11.34%)	MPI
37.133 s (4.14%)	Application
27.486 s (3.07%)	COUPLE
21.6 s (2.41%)	METEO

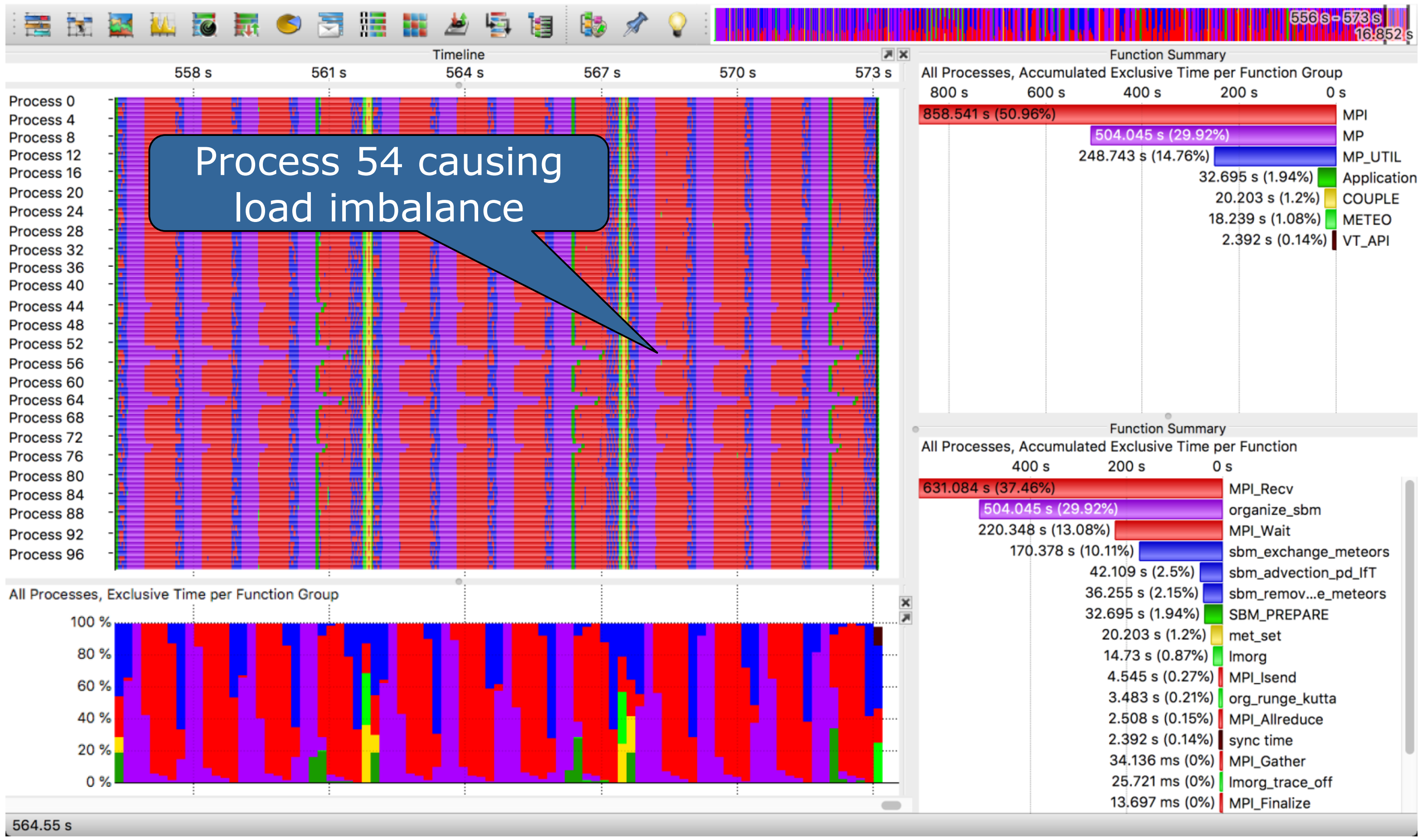
Function Summary

All Processes, Accumulated Exclusive Time per Function

Time	Function
459.796 s (51.31%)	organize_sbm
173.232 s (19.33%)	sbm_exchange_meteors
62.629 s (6.99%)	MPI_Recv
38.278 s (4.27%)	sbm_advection_pd_lfT
37.133 s (4.14%)	SBM_PREPARE
36.932 s (4.12%)	sbm_remov...e_meteors
30.309 s (3.38%)	MPI_Wait
27.486 s (3.07%)	met_set
16.501 s (1.84%)	lmorg
5.099 s (0.57%)	org_runge_kutta
4.967 s (0.55%)	MPI_Isend
3.706 s (0.41%)	MPI_Allreduce
3.353 ms (0%)	MPI_Comm_size

- Zoom into the first three iterations
- MP/MP_UTIL perform four sub-steps in one iteration
- Low MPI time share
- Everything is balanced and looks okay

01-p100-cosmo-specs-orig



Function Summary

All Processes, Accumulated Exclusive Time per Function Group

Time	Function Group
858.541 s (50.96%)	MPI
504.045 s (29.92%)	MP
248.743 s (14.76%)	MP_UTIL
32.695 s (1.94%)	Application
20.203 s (1.2%)	COUPLE
18.239 s (1.08%)	METEO
2.392 s (0.14%)	VT_API

Function Summary

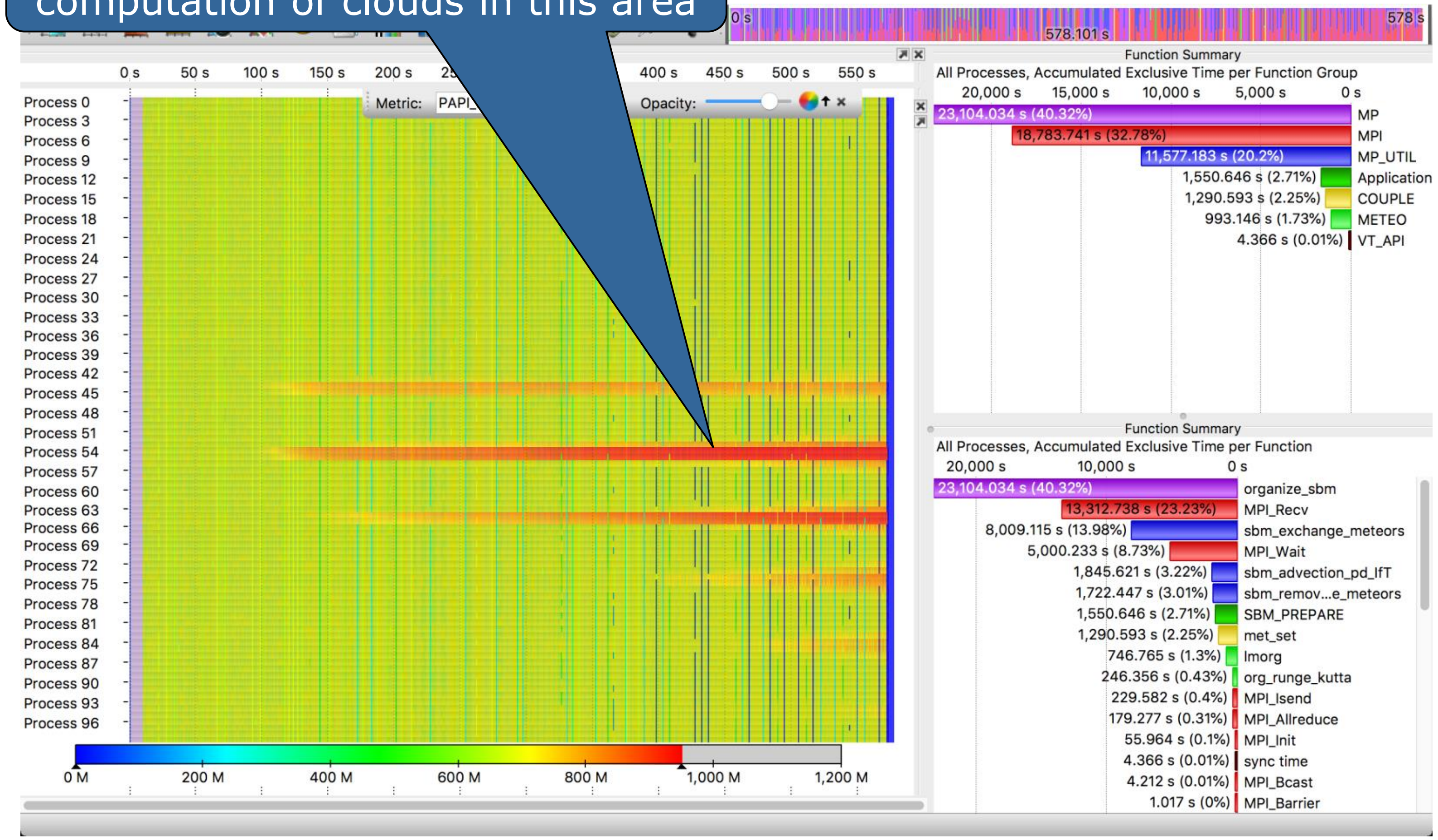
All Processes, Accumulated Exclusive Time per Function

Time	Function
631.084 s (37.46%)	MPI_Recv
504.045 s (29.92%)	organize_sbm
220.348 s (13.08%)	MPI_Wait
170.378 s (10.11%)	sbm_exchange_meteors
42.109 s (2.5%)	sbm_advection_pd_lfT
36.255 s (2.15%)	sbm_remov...e_meteors
32.695 s (1.94%)	SBM_PREPARE
20.203 s (1.2%)	met_set
14.73 s (0.87%)	lmorg
4.545 s (0.27%)	MPI_Isend
3.483 s (0.21%)	org_runge_kutta
2.508 s (0.15%)	MPI_Allreduce
2.392 s (0.14%)	sync time
34.136 ms (0%)	MPI_Gather
25.721 ms (0%)	lmorg_trace_off
13.697 ms (0%)	MPI_Finalize

- Zoom into the last three iterations
- Very high MPI time share (>50%)
- Large load imbalance caused by MP functions around **`Process 54`** and **`Process 64`**

01-p100-cosmo-specs-orig

High FLOPs rates due to computation of clouds in this area



Function Summary
All Processes, Accumulated Exclusive Time per Function Group

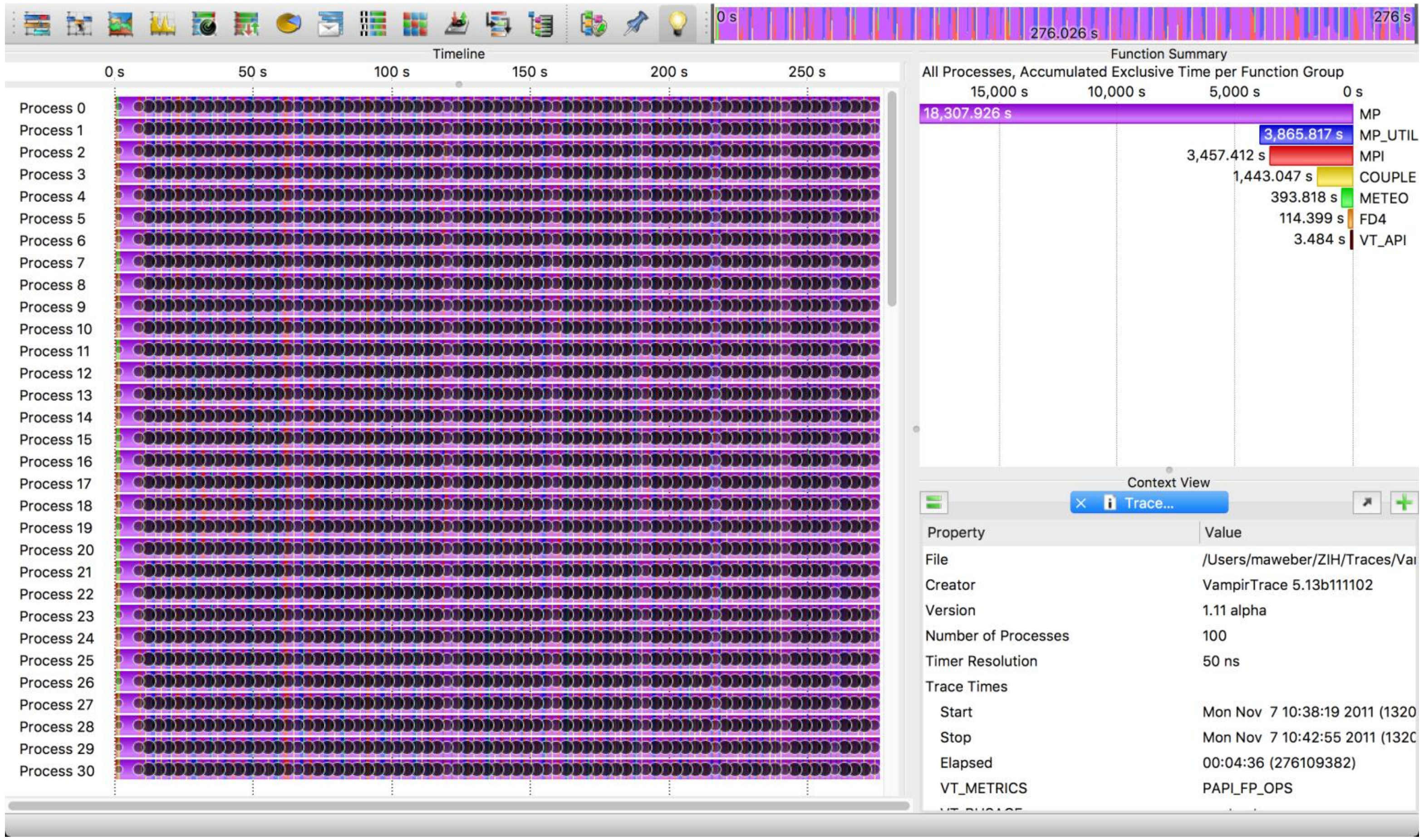
23,104.034 s (40.32%)	MP
18,783.741 s (32.78%)	MPI
11,577.183 s (20.2%)	MP_UTIL
1,550.646 s (2.71%)	Application
1,290.593 s (2.25%)	COUPLE
993.146 s (1.73%)	METEO
4.366 s (0.01%)	VT_API

Function Summary
All Processes, Accumulated Exclusive Time per Function

23,104.034 s (40.32%)	organize_sbm
13,312.738 s (23.23%)	MPI_Recv
8,009.115 s (13.98%)	sbm_exchange_meteors
5,000.233 s (8.73%)	MPI_Wait
1,845.621 s (3.22%)	sbm_advection_pd_lfT
1,722.447 s (3.01%)	sbm_remov...e_meteors
1,550.646 s (2.71%)	SBM_PREPARE
1,290.593 s (2.25%)	met_set
746.765 s (1.3%)	lmorg
246.356 s (0.43%)	org_runge_kutta
229.582 s (0.4%)	MPI_Isend
179.277 s (0.31%)	MPI_Allreduce
55.964 s (0.1%)	MPI_Init
4.366 s (0.01%)	sync time
4.212 s (0.01%)	MPI_Bcast
1.017 s (0%)	MPI_Barrier

- **`PAPI_FP_OPS`** counter showing higher FLOPs rates on processes causing the imbalance
- Reason for imbalance: Static grid used for distribution of processes. Depending on the weather, expensive cloud computations (MP group) may be only necessary on some processes

02-p100-cosmo-specs-fd4

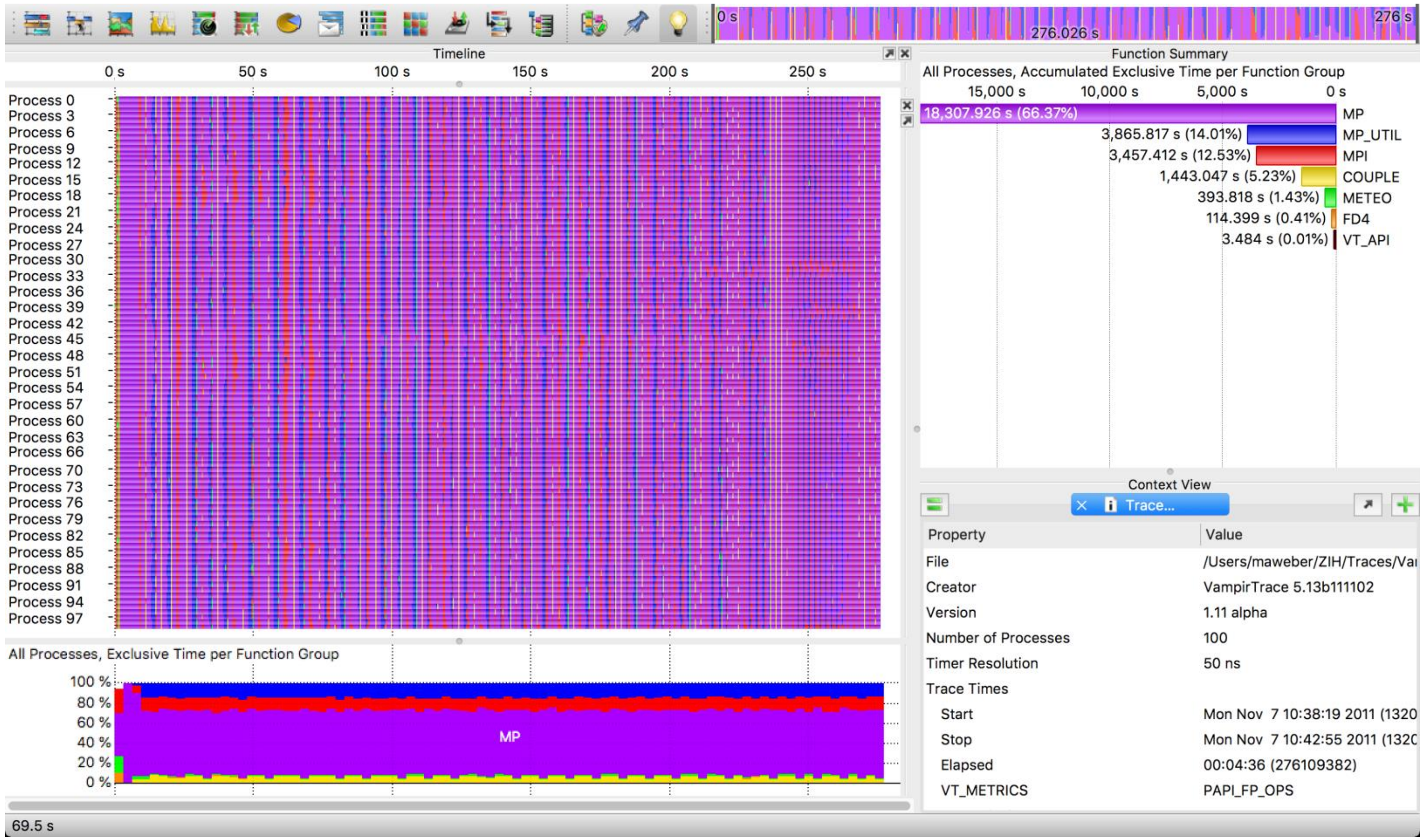


18,307.926 s	MP
3,865.817 s	MP_UTIL
3,457.412 s	MPI
1,443.047 s	COUPLE
393.818 s	METEO
114.399 s	FD4
3.484 s	VT_API

Property	Value
File	/Users/maweber/ZIH/Traces/Va
Creator	VampirTrace 5.13b111102
Version	1.11 alpha
Number of Processes	100
Timer Resolution	50 ns
Trace Times	
Start	Mon Nov 7 10:38:19 2011 (1320
Stop	Mon Nov 7 10:42:55 2011 (132
Elapsed	00:04:36 (276109382)
VT_METRICS	PAPI_FP_OPS

- Weather forecast code COSMO-SPECS
- Run with 100 processes
- COSMO: weather model (METEO group)
- SPECS: microphysics for accurate cloud calculation (MP and MP_UTIL group)
- Coupling of both models done in COUPLE group
- Dynamic load balancing (FD4 group)

02-p100-cosmo-specs-fd4

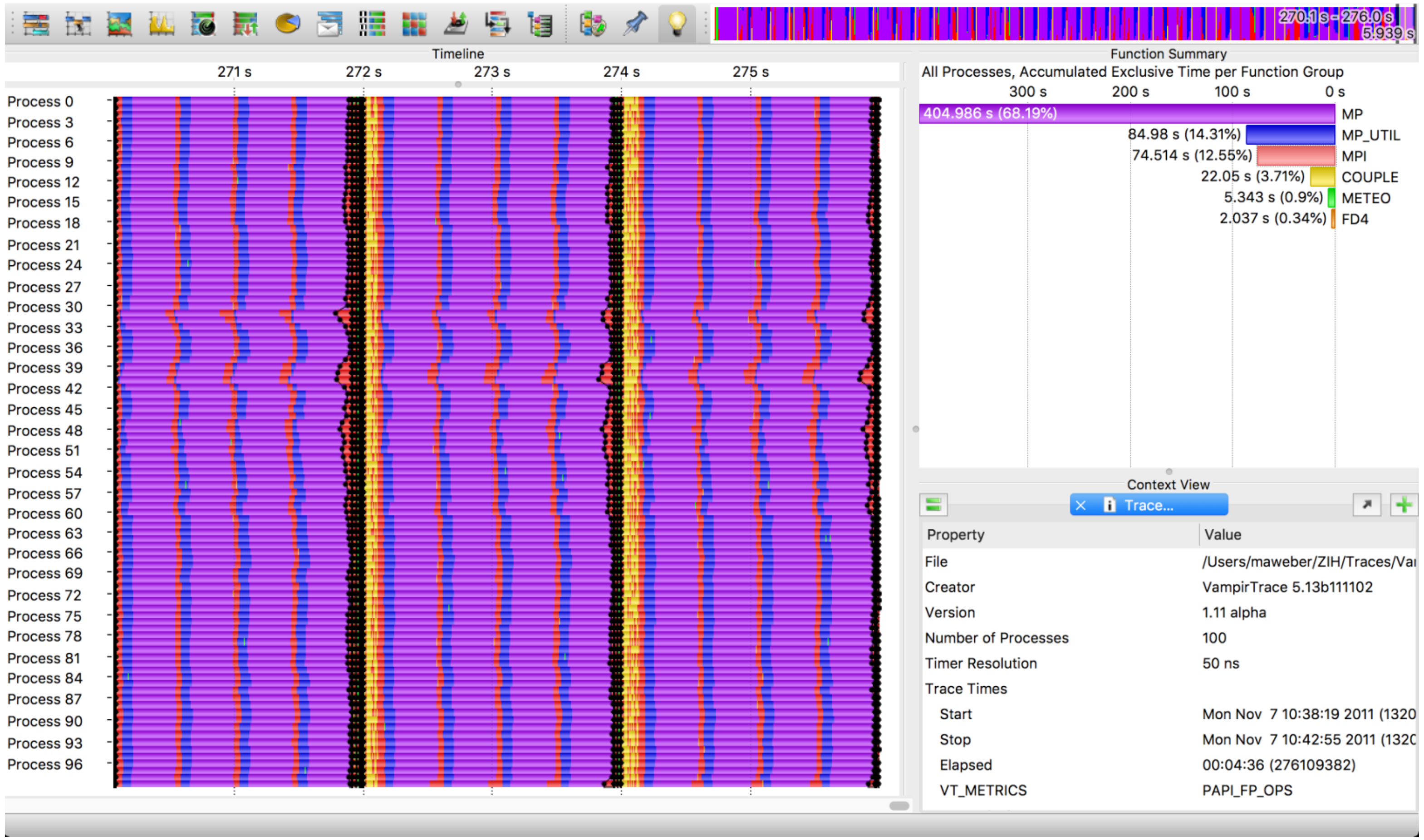


18,307.926 s (66.37%)	MP
3,865.817 s (14.01%)	MP_UTIL
3,457.412 s (12.53%)	MPI
1,443.047 s (5.23%)	COUPLE
393.818 s (1.43%)	METEO
114.399 s (0.41%)	FD4
3.484 s (0.01%)	VT_API

Property	Value
File	/Users/maweber/ZIH/Traces/Va
Creator	VampirTrace 5.13b111102
Version	1.11 alpha
Number of Processes	100
Timer Resolution	50 ns
Trace Times	
Start	Mon Nov 7 10:38:19 2011 (1320
Stop	Mon Nov 7 10:42:55 2011 (1320
Elapsed	00:04:36 (276109382)
VT_METRICS	PAPI_FP_OPS

- Dynamic load balancing mitigates the balance problems of the original COSMO-SPECS version
- MPI time share is reduced to <13%
- MPI time share stays constant throughout the application runtime
- Runtime reduced by factor of 2.1, from initially 578s to 276s

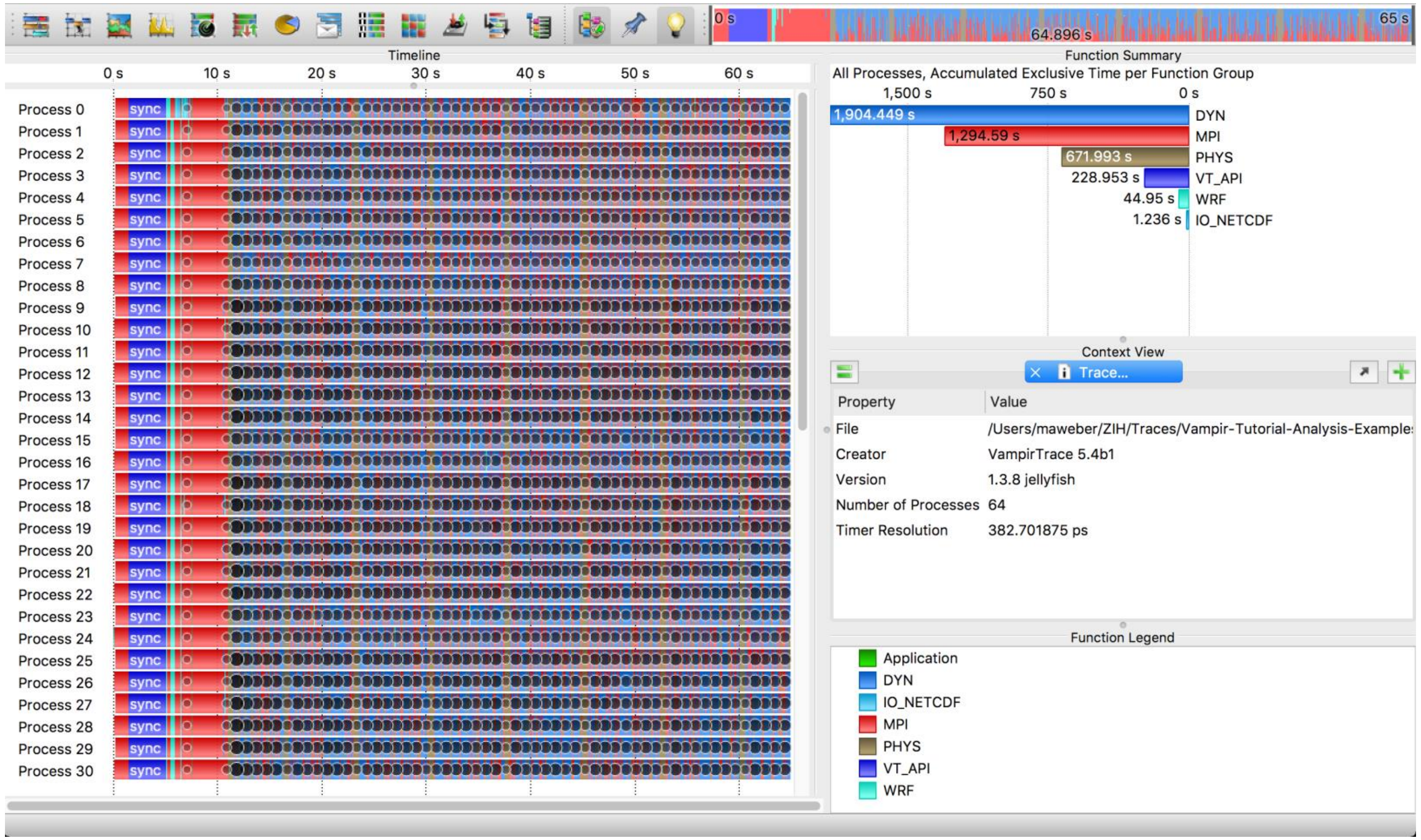
02-p100-cosmo-specs-fd4



Property	Value
File	/Users/maweber/ZIH/Traces/Va
Creator	VampirTrace 5.13b111102
Version	1.11 alpha
Number of Processes	100
Timer Resolution	50 ns
Trace Times	
Start	Mon Nov 7 10:38:19 2011 (1320
Stop	Mon Nov 7 10:42:55 2011 (132
Elapsed	00:04:36 (276109382)
VT_METRICS	PAPI_FP_OPS

- Zoom into last three iterations
- FD4 balances MP load (precipitation processes in clouds) across all available processes

03_wrf_deimos

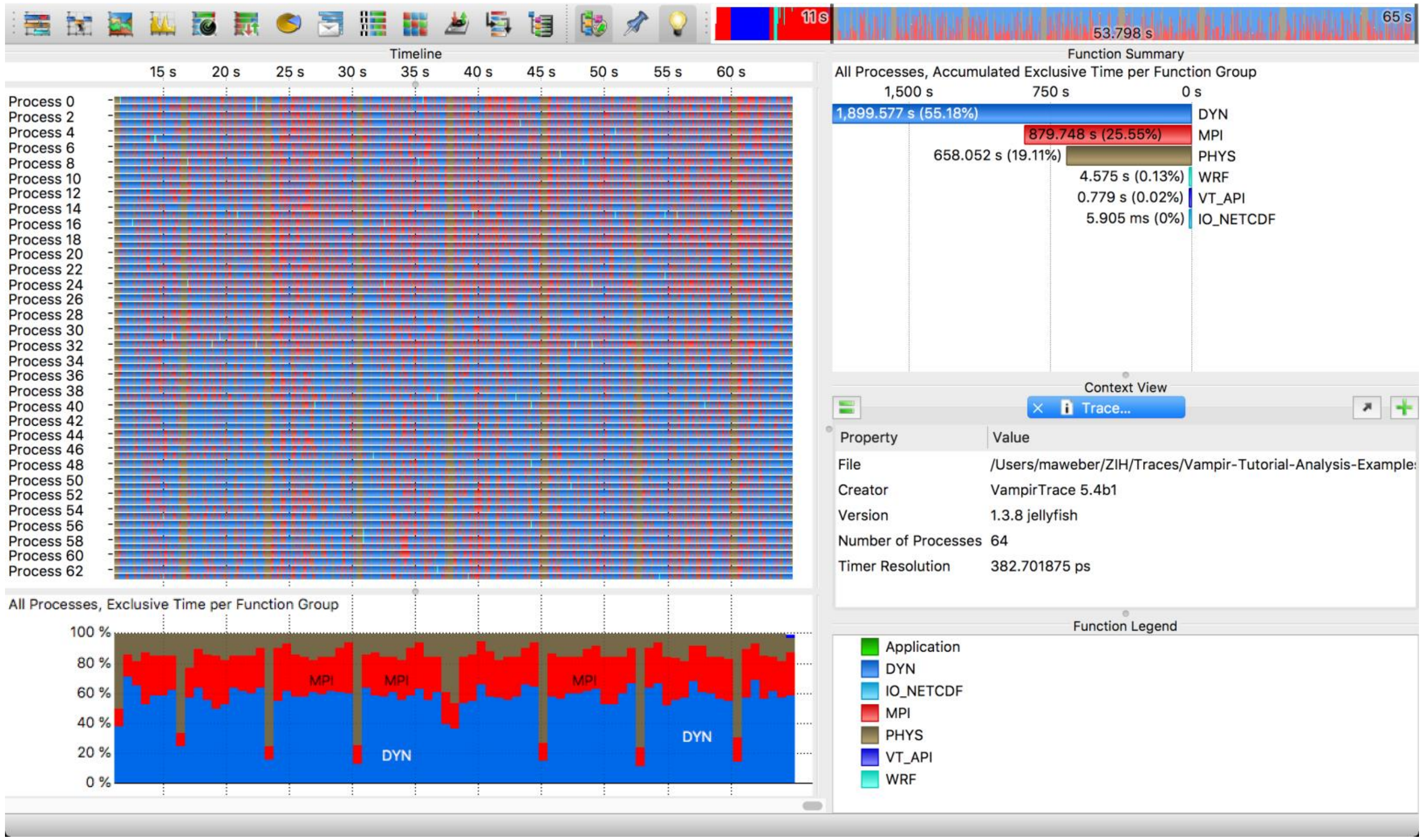


Property	Value
File	/Users/maweber/ZIH/Traces/Vampir-Tutorial-Analysis-Example
Creator	VampirTrace 5.4b1
Version	1.3.8 jellyfish
Number of Processes	64
Timer Resolution	382.701875 ps

Function Legend:
- 🟩 Application
- 🟦 DYN
- 🟦 IO_NETCDF
- 🟥 MPI
- 🟫 PHYS
- 🟦 VT_API
- 🟦 WRF

- Weather forecast code WRF
- Run with 64 processes
- *Dynamical core*: e.g., density, temperature, pressure, and winds in the atmosphere (DYN group)
- *Physical parameterization: e.g.,* clouds, rain, and radiation (PHYS group)

03_wrf_deimos



All Processes, Accumulated Exclusive Time per Function Group

1,500 s | 750 s | 0 s

1,899.577 s (55.18%)	DYN
879.748 s (25.55%)	MPI
658.052 s (19.11%)	PHYS
4.575 s (0.13%)	WRF
0.779 s (0.02%)	VT_API
5.905 ms (0%)	IO_NETCDF

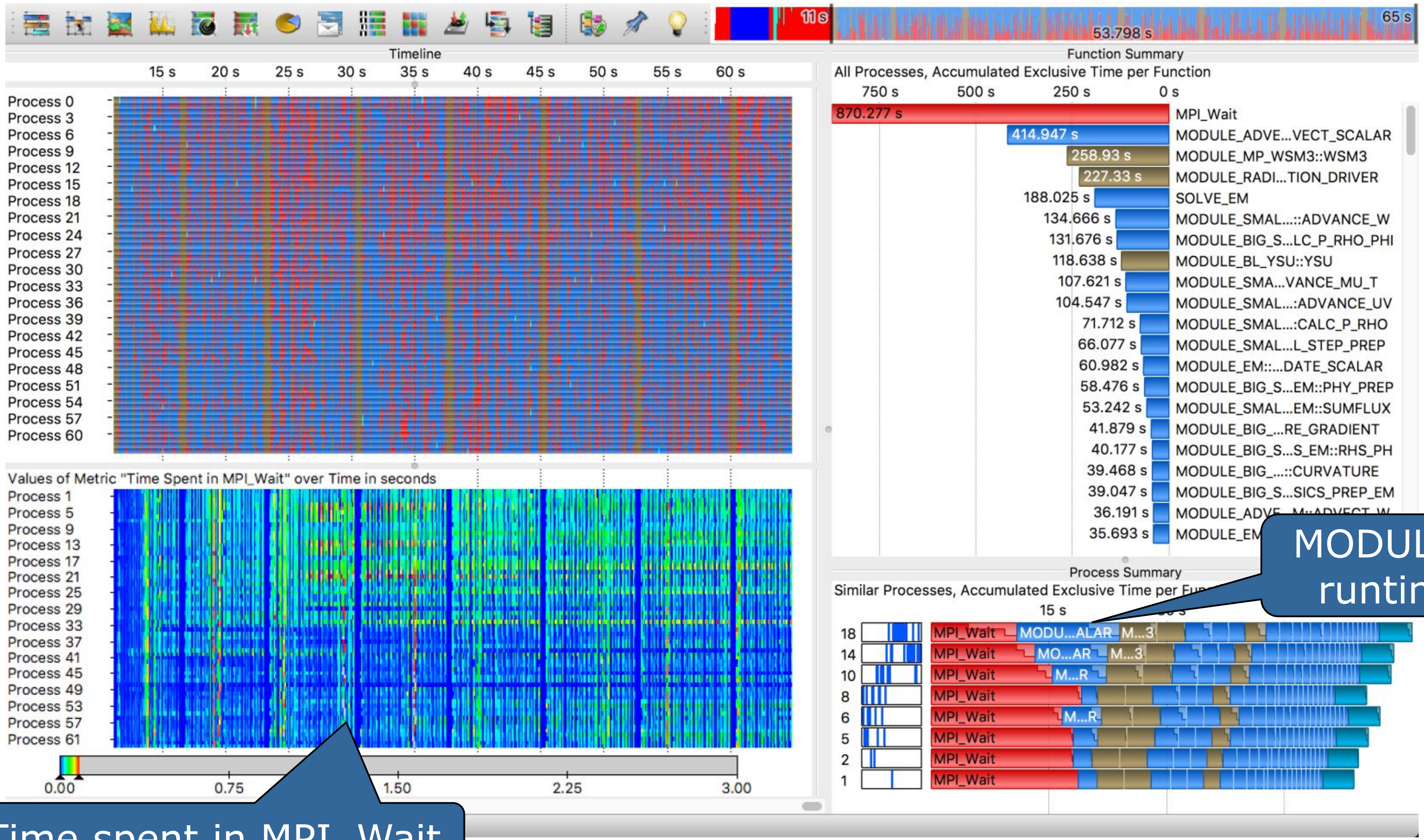
Property	Value
File	/Users/maweber/ZIH/Traces/Vampir-Tutorial-Analysis-Example:
Creator	VampirTrace 5.4b1
Version	1.3.8 jellyfish
Number of Processes	64
Timer Resolution	382.701875 ps

- 🟩 Application
- 🟦 DYN
- 🟦 IO_NETCDF
- 🟥 MPI
- 🟫 PHYS
- 🟦 VT_API
- 🟦 WRF

- Problem: 25% MPI run time fraction during the iterations of the model
- Behaviour is constant throughout all iterations

- Question: Which user function causes the problem? And why?

03_wrf_deimos



Function Summary

All Processes, Accumulated Exclusive Time per Function

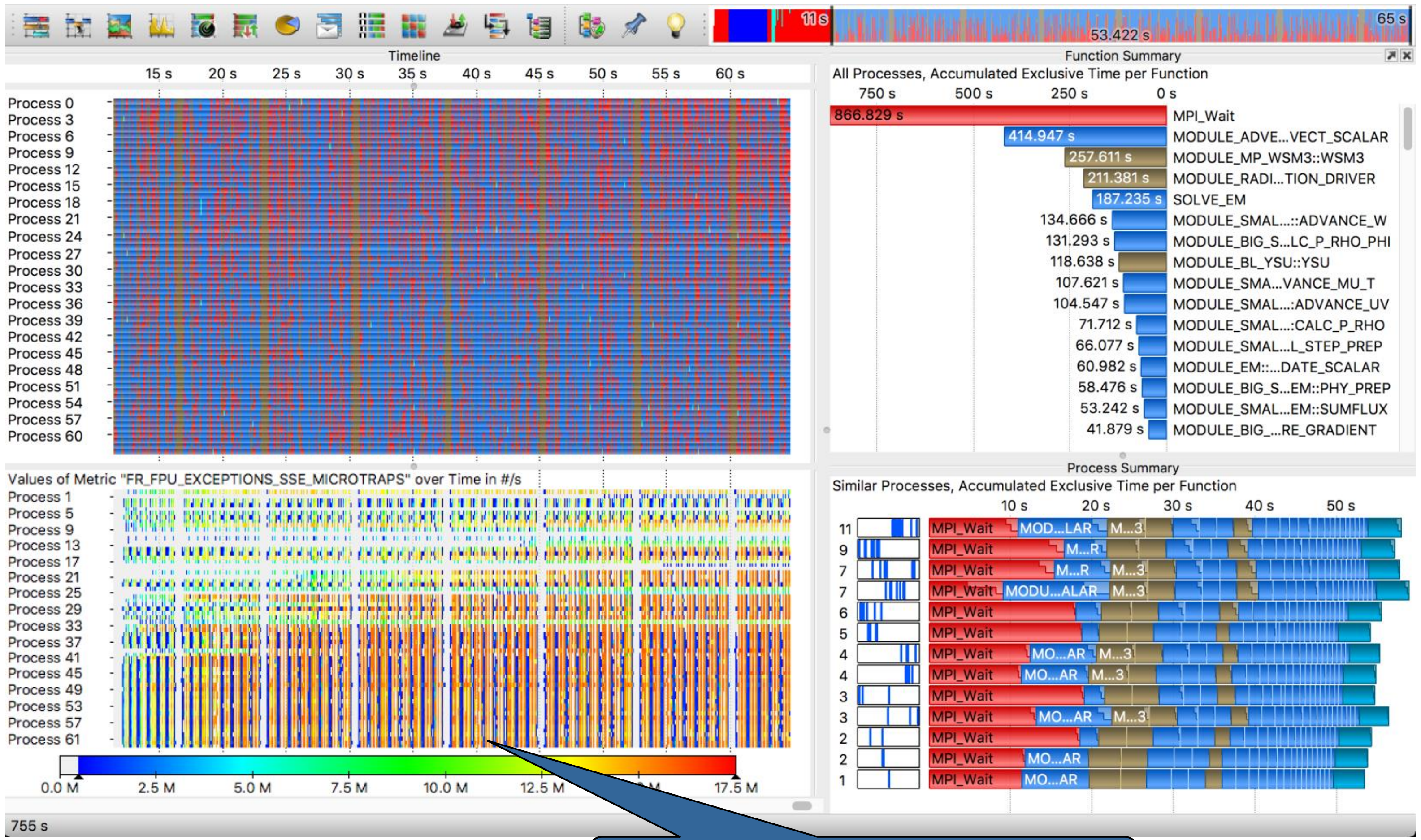
```
750 s      500 s      250 s      0 s
870.277 s   MPI_Wait
414.947 s   MODULE_ADVE...VECT_SCALAR
258.93 s    MODULE_MP_WSM3::WSM3
227.33 s    MODULE_RADI...TION_DRIVER
188.025 s   SOLVE_EM
134.666 s   MODULE_SMAL...::ADVANCE_W
131.676 s   MODULE_BIG_S...LC_P_RHO_PHI
118.638 s   MODULE_BL_YSU::YSU
107.621 s   MODULE_SMA...VANCE_MU_T
104.547 s   MODULE_SMAL...:ADVANCE_UV
71.712 s    MODULE_SMAL...:CALC_P_RHO
66.077 s    MODULE_SMAL...L_STEP_PREP
60.982 s    MODULE_EM::...DATE_SCALAR
58.476 s    MODULE_BIG_S...EM::PHY_PREP
53.242 s    MODULE_SMAL...EM::SUMFLUX
41.879 s    MODULE_BIG_...RE_GRADIENT
40.177 s    MODULE_BIG_S...S_EM::RHS_PH
39.468 s    MODULE_BIG_...:CURVATURE
39.047 s    MODULE_BIG_S...SICS_PREP_EM
36.191 s    MODULE_ADVE...M::ADVECT_W
35.693 s    MODULE_EM
```

MODULE_ADVECT_EM::ADVECT_SCALAR runtime increases in bottom processes

Time spent in MPI_Wait

- Most time is spent in **`MPI_Wait`**
- Top processes spent more time in **`MPI_Wait`** than bottom processes
- Load imbalance in DYN group

03_wrf_deimos



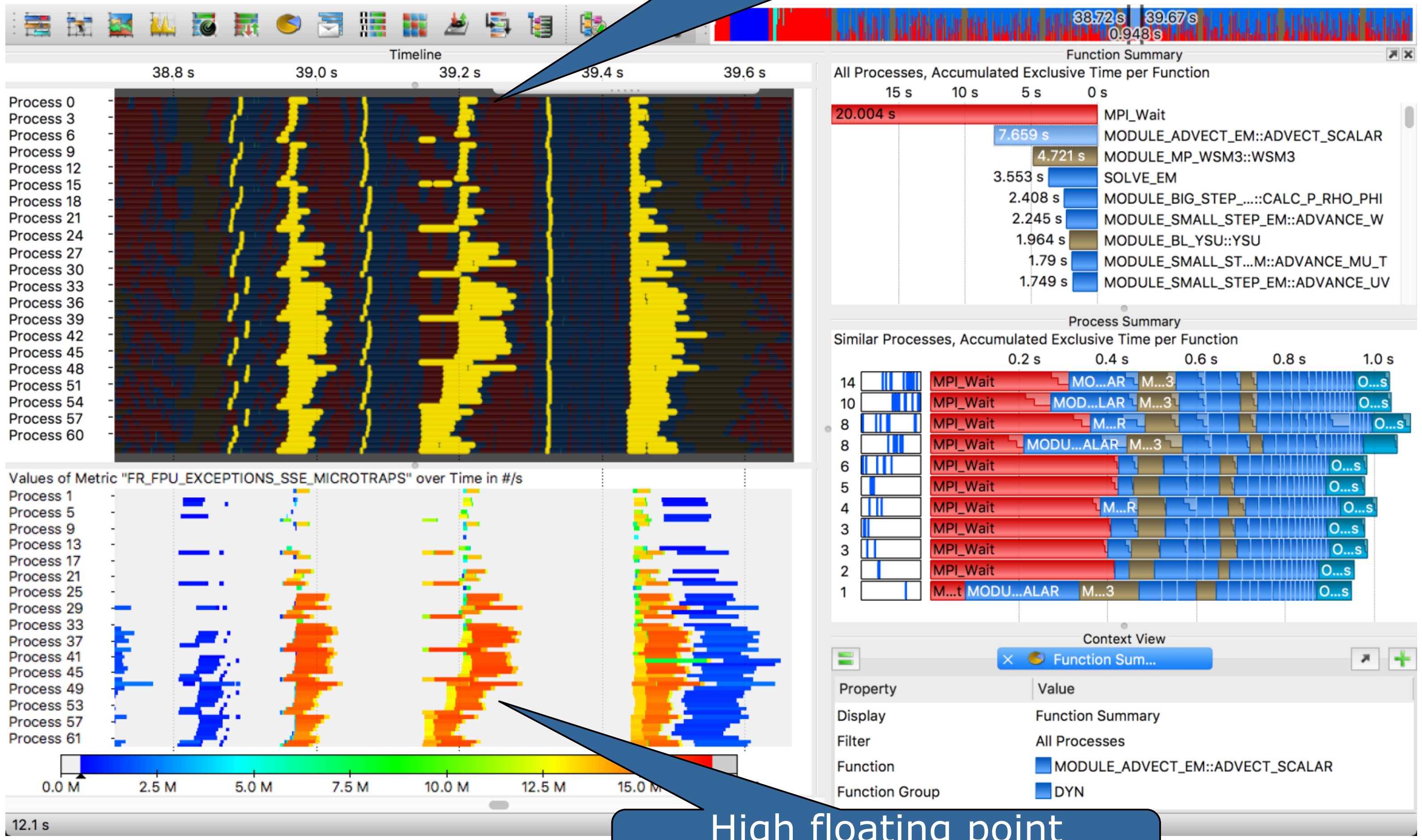
866.829 s	MPI_Wait
414.947 s	MODULE_ADVE...VECT_SCALAR
257.611 s	MODULE_MP_WSM3::WSM3
211.381 s	MODULE_RADI...TION_DRIVER
187.235 s	SOLVE_EM
134.666 s	MODULE_SMAL...::ADVANCE_W
131.293 s	MODULE_BIG_S...LC_P_RHO_PHI
118.638 s	MODULE_BL_YSU::YSU
107.621 s	MODULE_SMA...VANCE_MU_T
104.547 s	MODULE_SMAL...:ADVANCE_UV
71.712 s	MODULE_SMAL...:CALC_P_RHO
66.077 s	MODULE_SMAL...L_STEP_PREP
60.982 s	MODULE_EM::...DATE_SCALAR
58.476 s	MODULE_BIG_S...EM::PHY_PREP
53.242 s	MODULE_SMAL...EM::SUMFLUX
41.879 s	MODULE_BIG_...RE_GRADIENT

Floating point exceptions

- Load imbalance is caused by floating point (FP) exceptions in WRF
- Counter **FR_FPU_EXCEPTIONS_ SSE_MICROTRAPS** shows FP exceptions

03_wrf_deimos

MODULE_ADVECT_EM::ADVECT_SCALAR occurrences shown in yellow



Function Summary

All Processes, Accumulated Exclusive Time per Function

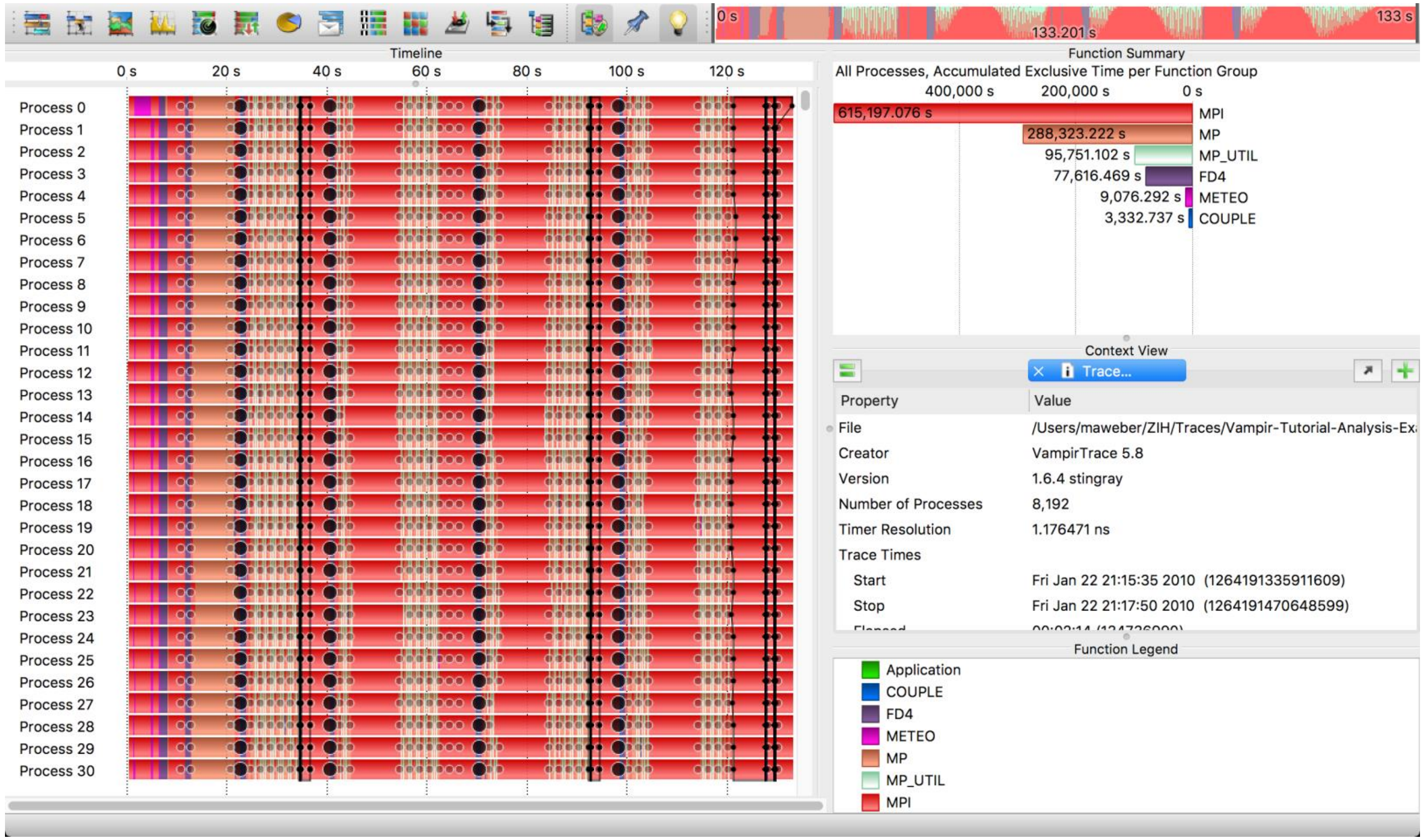
Time	Function
20.004 s	MPI_Wait
7.659 s	MODULE_ADVECT_EM::ADVECT_SCALAR
4.721 s	MODULE_MP_WSM3::WSM3
3.553 s	SOLVE_EM
2.408 s	MODULE_BIG_STEP_...::CALC_P_RHO_PHI
2.245 s	MODULE_SMALL_STEP_EM::ADVANCE_W
1.964 s	MODULE_BL_YSU::YSU
1.79 s	MODULE_SMALL_ST...M::ADVANCE_MU_T
1.749 s	MODULE_SMALL_STEP_EM::ADVANCE_UV

Property	Value
Display	Function Summary
Filter	All Processes
Function	MODULE_ADVECT_EM::ADVECT_SCALAR
Function Group	DYN

High floating point exceptions

- Zoom into one iteration
- Function invocations of **`MODULE_ADVECT_EM::ADVECT_SCALAR`** (shown in yellow) match high floating point exception occurrences indicated by the counter at the bottom

04_sbmfd4_jugene



Function Summary

All Processes, Accumulated Exclusive Time per Function Group

Time	Function Group
615,197.076 s	MPI
288,323.222 s	MP
95,751.102 s	MP_UTIL
77,616.469 s	FD4
9,076.292 s	METEO
3,332.737 s	COUPLE

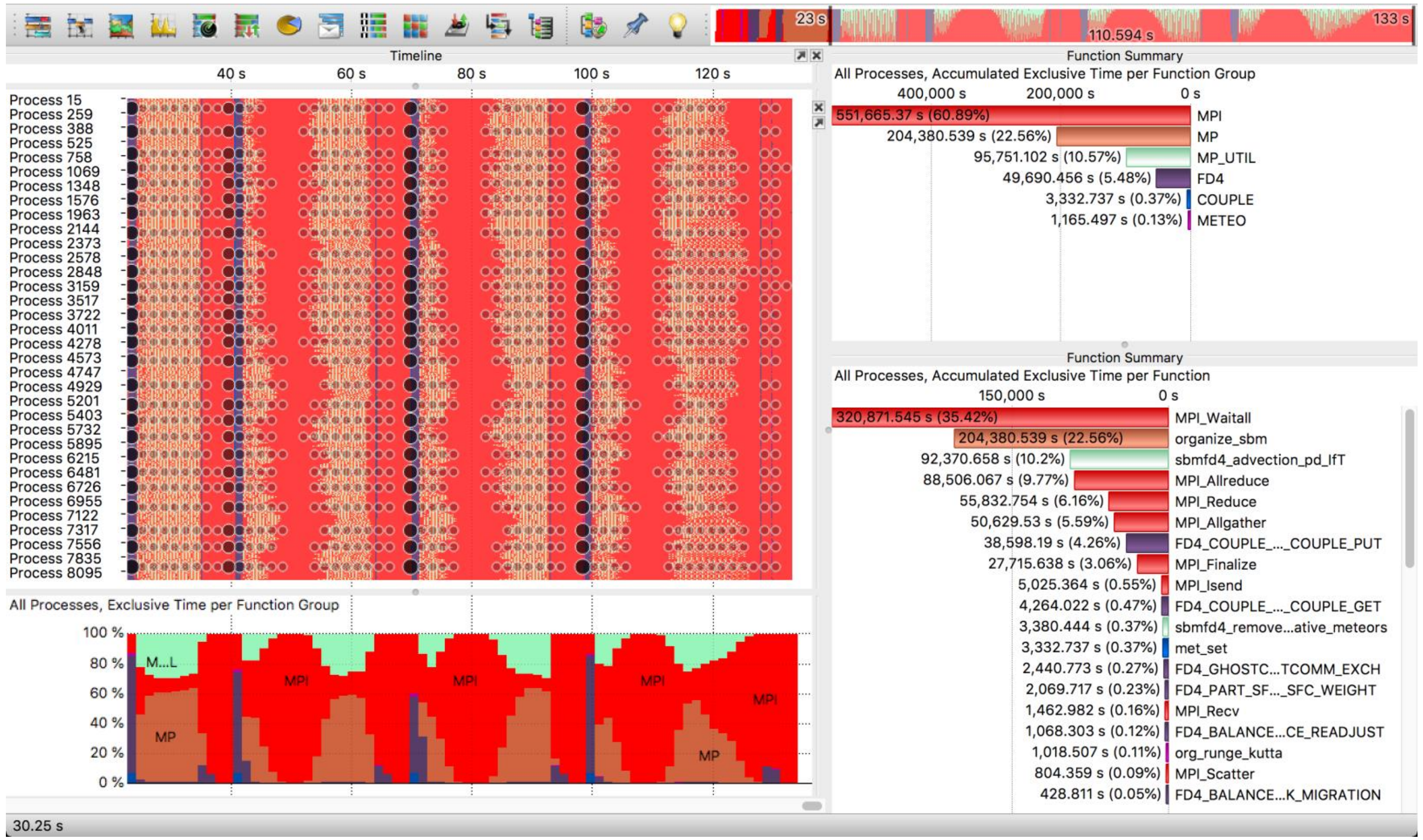
Property	Value
File	/Users/maweber/ZIH/Traces/Vampir-Tutorial-Analysis-Ex
Creator	VampirTrace 5.8
Version	1.6.4 stingray
Number of Processes	8,192
Timer Resolution	1.176471 ns
Trace Times	
Start	Fri Jan 22 21:15:35 2010 (1264191335911609)
Stop	Fri Jan 22 21:17:50 2010 (1264191470648599)

Function Legend

- 🟩 Application
- 🟦 COUPLE
- 🟪 FD4
- 🟪 METEO
- 🟧 MP
- 🟩 MP_UTIL
- 🟥 MPI

- Weather forecast code COSMO-SPECS
- Run with 8192 processes
- COSMO: weather model (METEO group)
- SPECS: microphysics for accurate cloud calculation (MP and MP_UTIL group)
- Coupling of both models done in COUPLE group
- Dynamic load balancing (FD4 group)

04_sbmfd4_jugene



Function Summary

All Processes, Accumulated Exclusive Time per Function Group

```
400,000 s        200,000 s        0 s
551,665.37 s (60.89%)    MPI
204,380.539 s (22.56%)   MP
95,751.102 s (10.57%)    MP_UTIL
49,690.456 s (5.48%)     FD4
3,332.737 s (0.37%)      COUPLE
1,165.497 s (0.13%)      METEO
```

Function Summary

All Processes, Accumulated Exclusive Time per Function

```
150,000 s        0 s
320,871.545 s (35.42%)   MPI_Waitall
204,380.539 s (22.56%)   organize_sbm
92,370.658 s (10.2%)     sbmfd4_advection_pd_lfT
88,506.067 s (9.77%)     MPI_Allreduce
55,832.754 s (6.16%)     MPI_Reduce
50,629.53 s (5.59%)      MPI_Allgather
38,598.19 s (4.26%)      FD4_COUPLE_..._COUPLE_PUT
27,715.638 s (3.06%)     MPI_Finalize
5,025.364 s (0.55%)      MPI_Isend
4,264.022 s (0.47%)      FD4_COUPLE_..._COUPLE_GET
3,380.444 s (0.37%)      sbmfd4_remove...ative_meteors
3,332.737 s (0.37%)      met_set
2,440.773 s (0.27%)      FD4_GHOSTC...TCOMM_EXCH
2,069.717 s (0.23%)      FD4_PART_SF..._SFC_WEIGHT
1,462.982 s (0.16%)      MPI_Recv
1,068.303 s (0.12%)      FD4_BALANCE...CE_READJUST
1,018.507 s (0.11%)      org_runge_kutta
804.359 s (0.09%)        MPI_Scatter
428.811 s (0.05%)        FD4_BALANCE...K_MIGRATION
```

- Problem: Large MPI runtime fraction (>60%) during iterations
- Especially in **`MPI_Waitall`** and **`MPI_Allreduce`**
- Behaviour is constant throughout all iterations

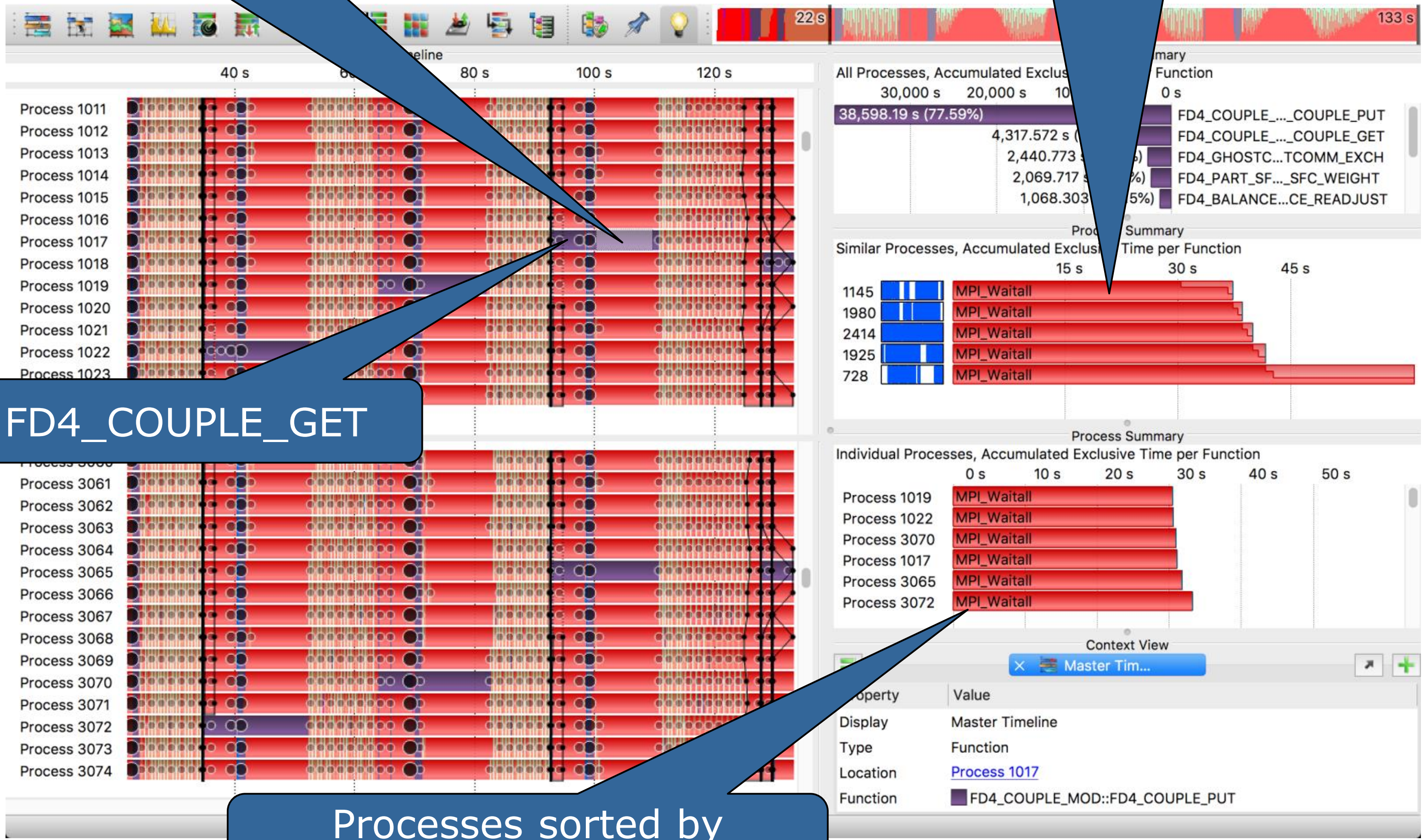
04_sbmfd4_jugene

High MPI_Allreduce variance between processes

FD4_COUPLE_PUT

FD4_COUPLE_GET

Processes sorted by MPI_Allreduce timeshare



All Processes, Accumulated Exclus... Function

```
30,000 s   20,000 s   10...   0 s
38,598.19 s (77.59%)   FD4_COUPLE_..._COUPLE_PUT
4,317.572 s (          FD4_COUPLE_..._COUPLE_GET
2,440.773 ...%)        FD4_GHOSTC...TCOMM_EXCH
2,069.717 ...%)        FD4_PART_SF..._SFC_WEIGHT
1,068.303 ...5%)       FD4_BALANCE...CE_READJUST
```

Similar Processes, Accumulated Exclus... Time per Function

```
      15 s     30 s     45 s
1145  MPI_Waitall
1980  MPI_Waitall
2414  MPI_Waitall
1925  MPI_Waitall
728   MPI_Waitall
```

Process Summary

Individual Processes, Accumulated Exclusive Time per Function

```
              0 s   10 s   20 s   30 s   40 s   50 s
Process 1019  MPI_Waitall
Process 1022  MPI_Waitall
Process 3070  MPI_Waitall
Process 1017  MPI_Waitall
Process 3065  MPI_Waitall
Process 3072  MPI_Waitall
```

Context View — Master Tim...

Property	Value
Display	Master Timeline
Type	Function
Location	Process 1017
Function	FD4_COUPLE_MOD::FD4_COUPLE_PUT

- Large runtime variation in **`MPI_Allreduce`**
- Sorted profile reveals processes with small **`MPI_Waitall`** timeshare
- Reason: Load imbalance in **`FD4_COUPLE_PUT`** and **`FD4_COUPLE_GET`**
- Most processes need to wait at **`MPI_Allreduce`** and **`MPI_Waitall`** (asynchronously)

Summary

- Performance flaws can lead to significant runtime overheads
- Use resources efficiently
- Analyze your code
- Performance analysis tools are there to help you

SPONSORED BY THE Federal Ministry of Education and Research

INDIANA UNIVERSITY

SILC

OAK RIDGE National Laboratory

LMAC

READEX Runtime Exploitation of Application Dynamism for Energy-efficient eXascale computing

eeClust Energy-Efficient Cluster Computing

nextgenio

Score-E

VAMPIR

European Commission | Horizon 2020 European Union funding for Research & Innovation

Microsoft®

ITEA 2 INFORMATION TECHNOLOGY FOR EUROPEAN ADVANCEMENT

SEVENTH FRAMEWORK PROGRAMME

//H4H

//ParMA

http://www.vampir.eu

vampirsupport@zih.tu-dresden.de