

Performance Analysis with Vampir



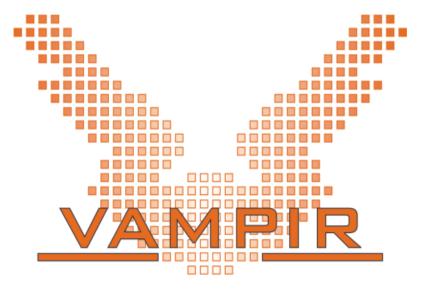
Outline

Part I: Welcome to the Vampir Tool Suite

- Mission
- Event Trace Visualization
- Vampir & VampirServer
- The Vampir Displays

Part II: Vampir Demo

Analyzing load imbalance over time in COSMO



Event Trace Visualization with Vampir

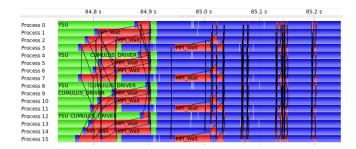
- Alternative and supplement to automatic analysis
- Show dynamic run-time behavior graphically at any level of detail
- Provide statistics and performance metrics

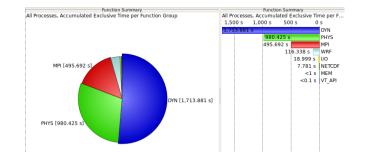
Timeline charts

Show application activities and communication along a time axis

Summary charts

Provide quantitative results for the currently selected time interval



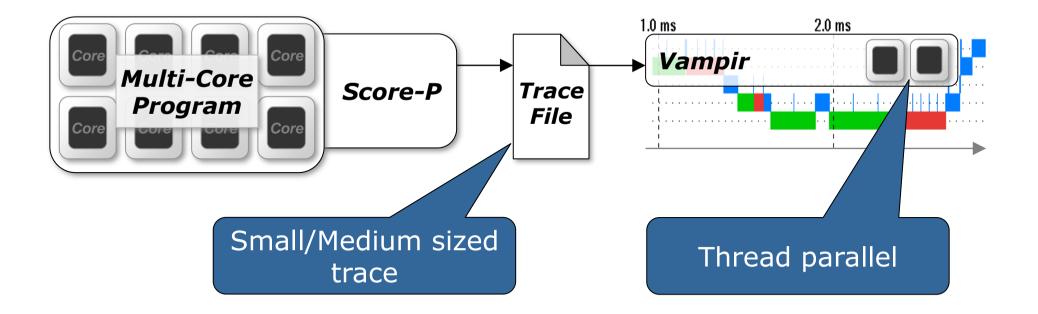


When is Vampir the right tool?

- Performance problems depend on information that changes over time or location
- Fine grained information is necessary:
 - What events occur simultaneously?
 - Which messages are being delayed?
 - Which threads are idle and why?

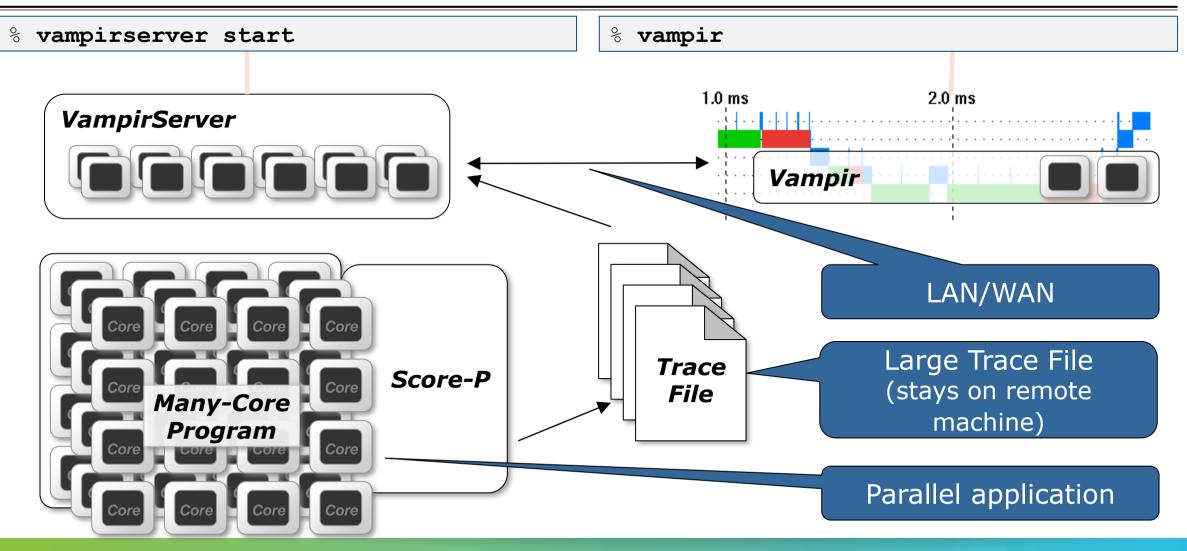
Visualization Modes (1) Directly on front end or local machine

% vampir



Visualization Modes (2)

On local machine with remote VampirServer



Local setup for Vampirserver Explicit SSH tunneling

% module use /home/vihps/software/modulefiles/ % module load vampir % module load mpi/openmpi/3.1.2-gcc-8.2.0 % vampirserver start Launching VampirServer... VampirServer 9.9.0 (14232279) Licensed to Goethe-Universitaet Frankfurt am Main Running 4 analysis processes... (abort with vampirserver stop 18299) VampirServer <18299> listens on: login02.cm.cluster:30018

% vampir &
% ssh -L 30000:login02.cm.cluster:30018 <user>@goethe.hhlr-gu.de

VIRTUAL INSTITUTE - HIGH PRODUCTIVITY SUPERCOMPUTING

The main displays of Vampir

- Timeline Charts:
 - 🚟 Master Timeline
 - Process Timeline
 - Counter Data Timeline
 - Performance Radar
- Summary Charts:
 - Summary
 - Message Summary
 - Process Summary
 - Communication Matrix View



Vampir Case Study: Optimizing COSMO-SPECS



COSMO-SPECS Original

	Martin Particular Part	5	6 🔁						10	•			ZX		8.101 s	E		
0)s 50 s	100 s	150 s	200 s	Time 250 s		00 s	350 s	400 s	450 s	500 s	550 s	~ ×			Function Summary Exclusive Time per Fu	nction Group	
Ū	1 1	100 3	100 3	200 3	100	, 0.	:	000 0	+00 5	400 3	:	:			,000 s	10,000 s 5,000	and the second states of the second states and	
ocess 0	ODDDDDD	DDDDDD	BBBBBB BBB			DDDD	00000	80888	00000				X	00 104 004 0			MP	,
ocess 1	ODDDDDD	DDDDDD	BBBBBB		100000			000000					•	18,792	741 6			<u></u>
ocess 2	ODDDDD	mm	BODDDDD		10000	2020	0000	833331	00000									
ocess 3	CODDDDD		DDDDDDD		199990			000000	2222							Maste	er Tim	le
ocess 4	CODDDDD	DDDDD	DDDDDD)))))									- /		
ocess 5	ODDDDD		DDDDDD	0000000	100001		m	000001	20000			000000				9		TEO
ocess 6		DDDDDD	0000001	0000000	nn		DDD:	88888	00000	0000000		0000000					4.366 s VT_	_API
ocess 7	0000000	DDDDDD	DDDDDDD	DDDDDDD	00000			00000	00000	0000000			•		/	-		
ocess 8	ODDDDDD	DDDD	BBBBBBB	DEBERGE	10000	0000	DDDD	888888	20000									
ocess 9	•BDBDDD	DDDDDD	0000000	0000000	22222	2022	10000	888888	00000						_	Context View	_	
ocess 10	ODDDDD	DDDDDD	DDDDDD			0000		888888	22220			0000	\sim		×	Trace		*
cess 11	ODDDDD										THE REAL			Property		Value		
cess 12	ODDDDD						and the second se				102-10		•	File		/Users/maweber/ZIH/	Traces/Vampir-1	Tuto
cess 13	ODDDDDD			Contractor Contractor Con									0	Creator		VampirTrace 5.11.2	rideed, rampi	, ato,
cess 14	OBDBDDD1											1000000	P	-				
ocess 15	ODDDDD												9	Version		1.11 alpha		
cess 16	OBUBDBBB		he fast set set set set set se	der bester besterbe			and shall be		(hérékéréké	00.00	0.000.000		•	Number of Process	es	100		
cess 17	03030301		helpelee printer					1-1-1-1-1-1-1-	0-0-0-0-0-0	and the set			0	Timer Resolution		50 ns		
cess 18	ODDDDD1	in the second		detailed details					0-0-0-0-0-0					Trace Times				
ocess 19	OBDBDBD	had deducted					A-A-A-A							Start		Thu Nov 3 12:36:57	2011 (132032021	1776
ocess 20	OBDBDDD									0000000				Stop		Thu Nov 3 12:46:30	2011 (132032079	9092
cess 21	ODDDDDD								33335					Elapsed		00:09:33 (57315898	1.0.0	
ocess 22	ODDDDDD									0000000		000000	2			00.09.33 (07510898	5)	
cess 23			he factor for the factor of		1-1-1-0-0		0.1111		μ					VampirTrace Enviro	nment			
ocess 24		te da de de de de de				and spin the state								VT_MODE		TRACE		
ocess 25														VT_BUFFER_SIZE		200M		
cess 26	ODDDDD						1000000000							VT_SYNC_FLUSH	ł	no		
cess 27	CODDDDD											Contraction of the second		VT_SYNC_FLUSH	LEVEL	80		
cess 28									a red role of red of					VT MAX FLUSHE	S	1		
ocess 29		ter	and in some division of the later.					inductor includes				and a subscription of		VT_METRICS		PAPI_FP_OPS		
ocess 30	ODDDDDD	2000				STO D BI							e	VT_RUSAGE		<not set=""></not>		

VIHPS TUNING WEEK 37 (FRANKFURT/M., GERMANY, 10 DEC 2020)

Weather forecast code

COSMO: weather model

SPECS: microphysics for

accurate cloud calculation

(MP and MP_UTIL group)

Coupling of both models

done in COUPLE group

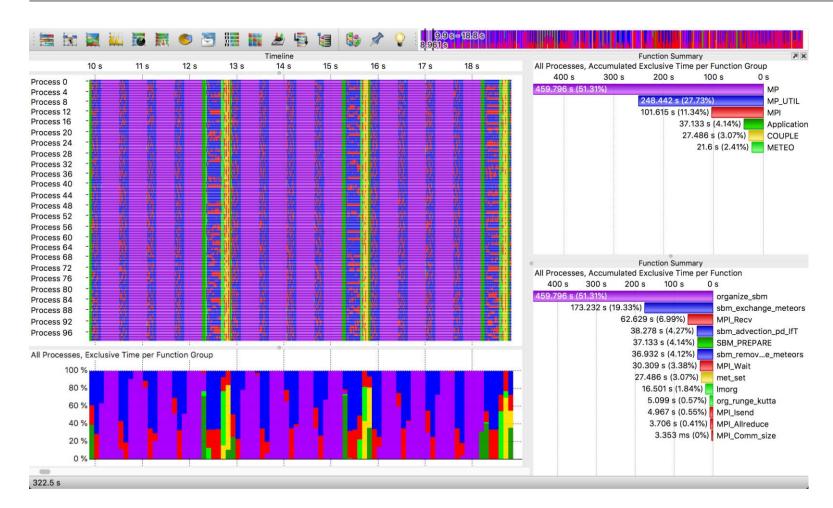
h with 100 processes

COSMO-SPECS

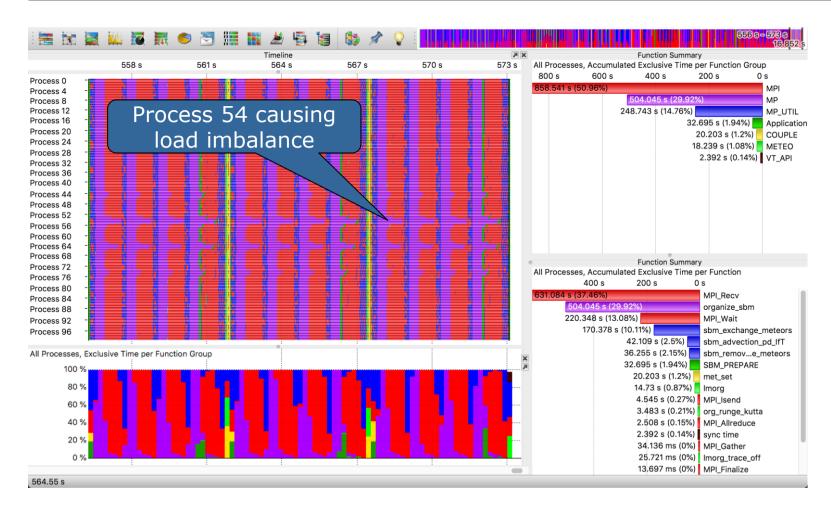
(METEO group)

	-					Timest	- 4		24				578.101 s
	0 s	50 s	100 s	150 s	200 s	Timeline 250 s	300 s	350 s	400 s	450 s	500 s	550 s	Function Summary All Processes, Accumulated Exclusive Time per Function Group
	0.3	50 5	:	150 3	200 3	230 3	500 5	550 3	400 3	450 3	500 3	550 3	20,000 s 15,000 s 10,000 s 5,000 s 0 s
ocess 0					i Liir-								23.104.034 s (40.32%)
ocess 4					14.315-								18,783.741 s (32.78%) MPI
cess o	-												11,577.183 s (20.2%) MP_U
cess 16	- 64				14388								1,550.646 s (2.71%) Applie
cess 20	- 60												1,290.593 s (2.25%) COUF
cess 24	- 10												993.146 s (1.73%)
cess 28	- 20												4.366 s (0.01%) VT_A
cess 32								- 1			· · · · · · · · · · · · · · · · · · ·		
ocess 36 ocess 40													Summary Timeline:
ocess 40	-												
ocess 48	- 20							468				Ind	creasing MPI Wait time
ocess 52	-											TIIC	creasing in i_wait time
ocess 56	-				id di P			is is it	n i i i i u	ار <u>و</u> در با			
ocess 60	-												
ocess 64												1917	
ocess 68 ocess 72													Function Summary
ocess 72					1111								All Processes, Accumulated Exclusive Time per Function
cess 80	- 23											/	20,000 s 10,000 s 0 s
cess 84	-												23,104.034 s (40.32%) organize_sbm
cess 88	- 22				1918						/		13,312.738 s (23.23%) MPI_Recv
cess 92	- 20										/ /		8,009.115 s (13.98%) sbm_exchange_meteo
cess 96	-										1 /		5,000.233 s (8.73%) MPI_Wait
		:											1,845.621 s (3.22%) sbm_advection_pd_IfT
Processes	s, Exclus	ive Time	per Functio	on Group									1,722.447 s (3.01%) sbm_remove_meteor
100) %									/			1,550.646 s (2.71%) SBM_PREPARE
00)%				MP_UTIL								1,290.593 s (2.25%) met_set
80	/0												746.765 s (1.3%) Imorg
60)%								MPI				246.356 s (0.43%) org_runge_kutta 229.582 s (0.4%) MPL lsend
40)%												179.277 s (0.31%) MPI_Isend
40	/0					/P						the second second	55.964 s (0.1%) MPI_Allreduce
20	%												4.366 s (0.01%) sync time
20						-	-						
)%						And Address of the Ad					and the second s	4.212 s (0.01%) MPI_Bcast

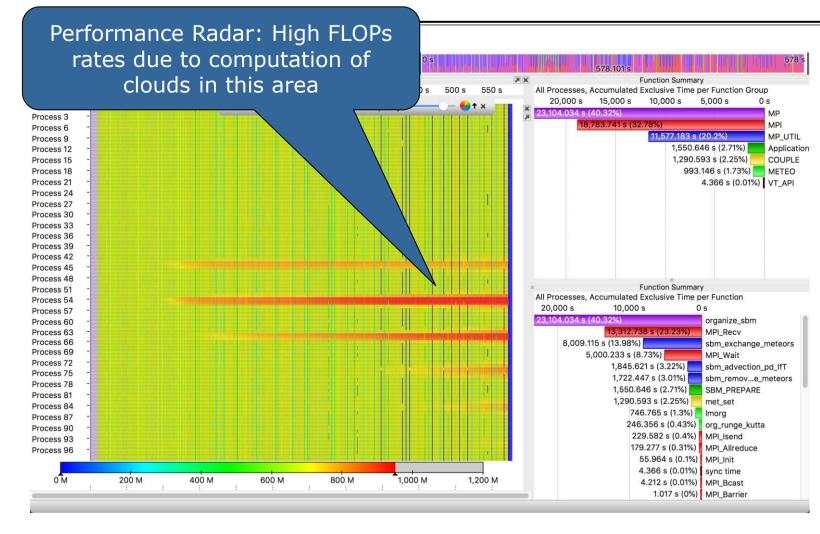
- Compared to METEO, MP and MP_UTIL are very compute intensive, however this is due to more complex calculations and no performance issue
 Problem: >32% of time
- spent in MPI
- MPI runtime share increases throughout the application run



- 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



- 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



- 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

VIRTUAL INSTITUTE - HIGH PRODUCTIVITY SUPERCOMPUTING

COSMO-SPECS FD4

	- 100-100 - 100-100 - 100-100 - 100-100 - 100-100-		Timeline				Function Summary	
C)s 50 s	100 s	150 s	200 s	250 s	All Processes, Accumulated		tion Group 0 s
rocess 0						15,000 s 10 18,307.926 s	,000 s 5,000 s	U S MP
rocess 1				SCOLOGICCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO		18,307.920 \$	3	865.817 s MP U
rocess 2	CODDDDDDDDDDDDDD	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		SOLOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	000000000000000000000000000000000000000		3,457,412 s	MP_0
rocess 3	• • • • • • • • • • • • • • • • • • •	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	000000000000000000000000000000000000000	ICCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		1,443.	
rocess 4		CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	100000000000000000000000000000000000000	COCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		3	93.818 s METE
rocess 5	• • • • • • • • • • • • • • • • • • •	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	I DODDDDDDDDDDDDDDDDD	COOCOCOCOCOCOCOCOCO				114.399 s FD4
rocess 6		CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ICCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	COCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC			3.484 s VT_A
rocess 7	000000000000000000000000000000000000000	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	I COCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	ICCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC			
rocess 8	• • • • • • • • • • • • • • • • • • •	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	000000000000000000000000000000000000000	BOLLEGELEGELEGELEGELEGELEGELEGE LEGELEGELE	000000000000000000000000000000000000000			
rocess 9	• • • • • • • • • • • • • • • • • • •	00000000000000000		COCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC			
rocess 10	• • • • • • • • • • • • • • • • • • • 	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	COCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	COCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC			
ocess 11	• • • • • • • • • • • • • • • • • • • 	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ID COLOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOC	ICCCCCCCCCCCCCCC	DDDDDDDDDDDDDDD			
rocess 12	• • • • • • • • • • • • • • • • • • •		000000000000000000000000000000000000000	COCCECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC			
rocess 13	• • • • • • • • • • • • • • • • • • •	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	I DODDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD	I DE	CE			
rocess 14	OBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB		I COCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	COCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC			
rocess 15		CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		ICCCCCCCCCCCCCCCCCC				
rocess 16			ICOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	DECOCOCOCOCOCOCOCO				
rocess 17			100000000000000000000000000000000000000				Context View	
rocess 18		111					Trace	×
rocess 19						Property	Value	
rocess 20			DDDDDDDDDDDDDDDDDDD			File	/Users/may	veber/ZIH/Traces,
rocess 21			000000000000000000000000000000000000000	1	and the second states and	Creator		ce 5.13b111102
rocess 22						Version	1.11 alpha	0.10011102
rocess 23			000000000000000000000000000000000000000	Option of the design of the factor of the second states			10	
rocess 24						Number of Processes	100	
rocess 25			000000000000000000000000000000000000000			Timer Resolution	50 ns	
rocess 26						Trace Times		
rocess 27						Start	Mon Nov 7	10:38:19 2011 (1:
rocess 28						Stop	Mon Nov 7	10:42:55 2011 (1:
rocess 29						Elapsed	00:04:36 (N 102 107 10
rocess 30								

- 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)

WIRTUAL INSTITUTE - HIGH PRODUCTIVITY SUPERCOMPUTING

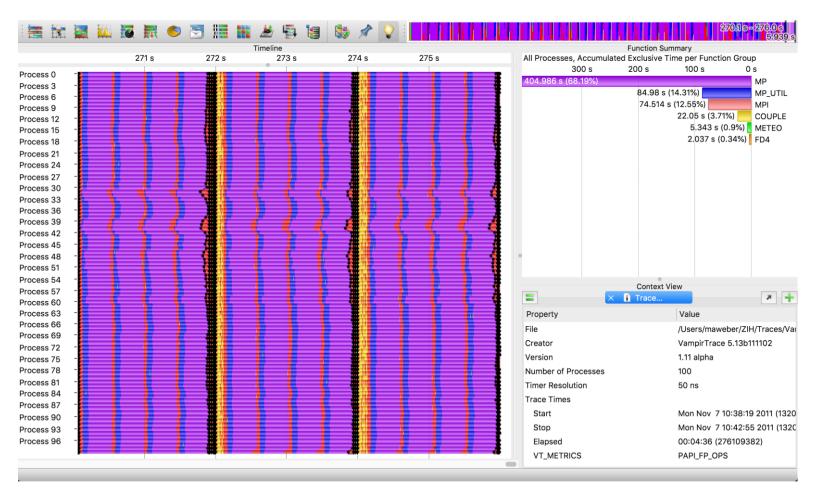
COSMO-SPECS FD4

	50	Timeline	150			X	Function Su		
0 s	50 s	100 s	150 s	200 s	250 s		Accumulated Exclusive T 10,000 s		up 0 s
ocess 0 -						15,000 s	Contraction of the second s	5,000 s	
ocess 3 -						× 18,307.926 s (44.0400	MP
ocess 6 -							3,865.817 s		MP_UT
ocess 12							3,457.412 s		MPI
cess 15 -							1,44	3.047 s (5.23%)	COUPL
ocess 18 📑								393.818 s (1.43%)	METEO
ocess 21 -								114.399 s (0.41%)	FD4
ocess 24 -								3.484 s (0.01%)	VT API
ocess 27									
cess 33									
cess 36									
ocess 39									
ocess 42									
ocess 45 -									
cess 51									
cess 54									
ocess 57 -									
ocess 60 -									
ocess 63 -									
ocess 66 -									
ocess 70							Context \	/iew	
ocess 76						=	× Trace		×
ocess 79									
ocess 82 -						Property		Value	
ocess 85 -						File		/Users/maweber/Z	IH/Traces
ocess 88 -						Creator		VampirTrace 5.13b	
ocess 94 -						Version		1.11 alpha	111102
ocess 97						Number of Pro	cesses	100	
Processes, Exclusive Tim	e per Function Group)				Timer Resoluti	on	50 ns	
100 %						Trace Times			
80 %		Non-Section of the local division of the loc	State of the local division of the local div			Start		Mon Nov 7 10:38:	19 2011 (1
60 %			MP			Stop		Mon Nov 7 10:42:	
40 % 20 %						··· Elapsed		00:04:36 (276109)	
0 %			-	and the second s	And the second second second	VT_METRICS		PAPI_FP_OPS	502)

- 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

V VIRTUAL INSTITUTE - HIGH PRODUCTIVITY SUPERCOMPUTING

COSMO-SPECS FD4



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



Summary and Conclusion



Summary

- Vampir & VampirServer
 - Interactive trace visualization and analysis
 - Intuitive browsing and zooming
 - Scalable to large trace data sizes (20 TiByte)
 - Scalable to high parallelism (200,000 processes)
- Vampir for Linux, Windows, and Mac OS X

VI-HPS



http://www.vampir.eu

vampirsupport@zih.tu-dresden.de