TECHNISCHE UNIVERSITÄT DRESDEN

Center for Information Services and High Performance Computing (ZIH)

HPC Methods for Coupling Spectral Cloud Microphysics with the COSMO Model

Max Planck Institute for Meteorology
17 March 2011

Matthias Lieber [a], Verena Grützun [b], Ralf Wolke [c], Matthias S. Müller [a], Wolfgang E. Nagel [a]

matthias.lieber@tu-dresden.de

[a] Center for Information Services and High Performance Computing (ZIH), TU Dresden, Germany

[b] Max Planck Institute for Meteorology (MPI-M), Hamburg, Germany

[c] Leibniz Institute for Tropospheric Research (IfT), Leipzig, Germany

LEIBNIZ INSTITUTE FOR TROPOSPHERIC RESEARCH

Max-Planck-Institut für Meteorologie

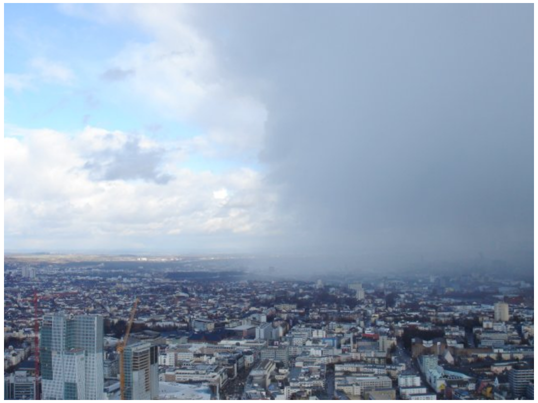
Center for Information Services & High Performance Computing

Outline

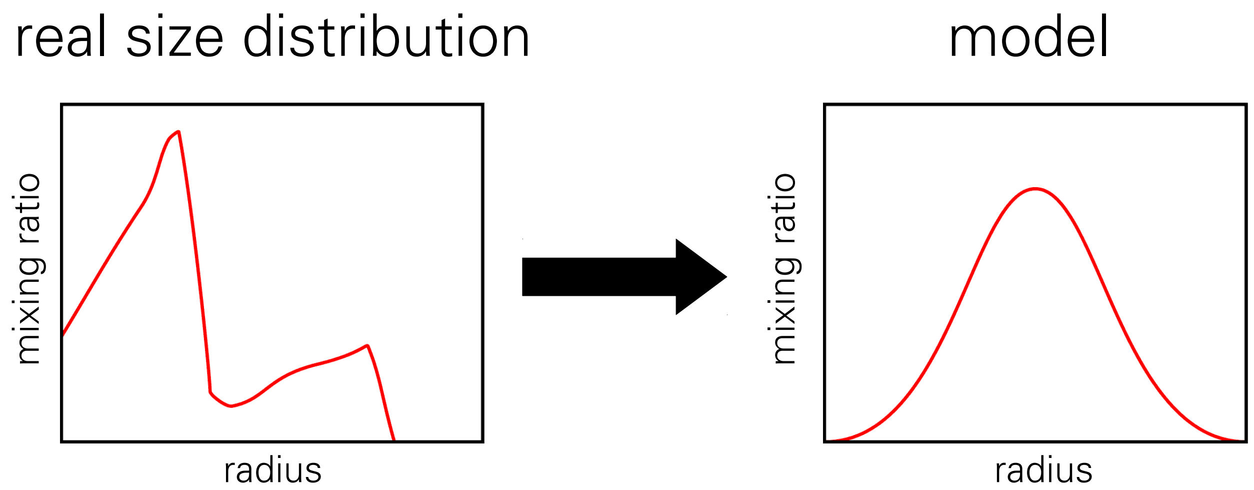
- Introduction
- Performance Problems of COSMO-SPECS
- New Coupling Scheme
  - Basic Idea
  - Framework FD4
  - COSMO-SPECS+FD4
- Performance Results
- Conclusion & Outlook

Introduction: Cloud Modeling

- Clouds play a major role for climate and weather
  - Influence on the radiation budget of the planet
  - Part of the hydrological cycle
  - Interaction with aerosol particles and pollution
- Clouds represent one of the major uncertainties in climate and weather models [IPCC07, CCSP09]
- Unsatisfying improvements in precipitation forecast during the last decades

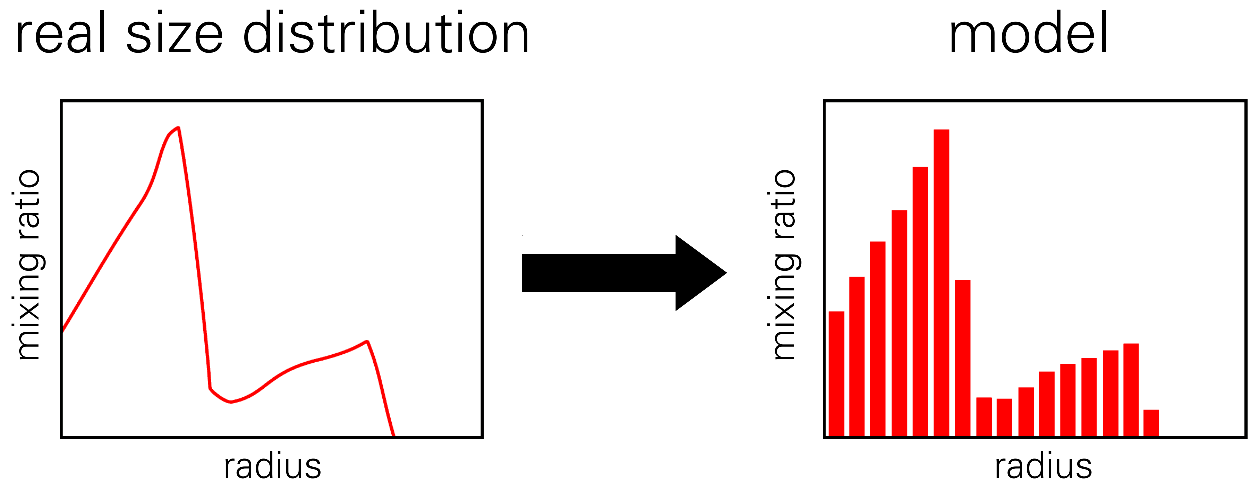


Introduction: Bulk Parameterization Schemes



- Cloud droplets are described by their bulk mass only
- Actual size distribution is neglected
- Computationally fast
- Used in most weather models

Introduction: Spectral Bin Microphysics Schemes



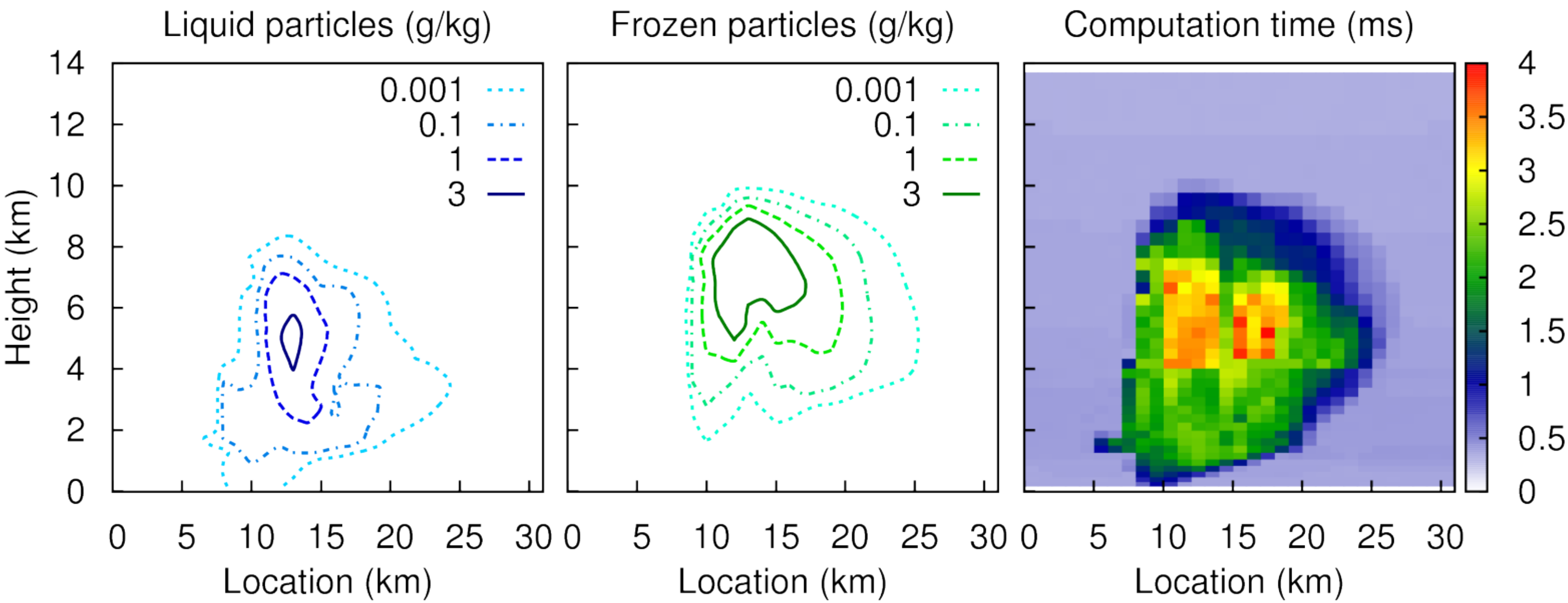
- Bin discretization of size distribution
- Allows more detailed modeling of interaction between aerosol particles, clouds, and precipitation
- Computationally very expensive (runtime & memory)
- Only used for process studies up to now (MM5, WRF, COSMO)

COSMO-SPECS: COSMO with SPECtral bin microphysicS

- COSMO Model: non-hydrostatic limited-area atmospheric model, formerly known as “Lokal-Modell” of DWD
  - http://www.cosmo-model.org

- COSMO-SPECS: Cloud parameterization scheme of COSMO replaced by the spectral bin microphysics model SPECS [Simmel06, Grützun08]
  - 11 new variables to describe 3 types of hydrometeors (water droplets, frozen particles, insoluble particles)
  - Discretized into a predefined number of size bins, e.g. 66 bins x 11 variables = 726 values!
  - Very high computational costs

COSMO-SPECS: SPECS Workload Measurement



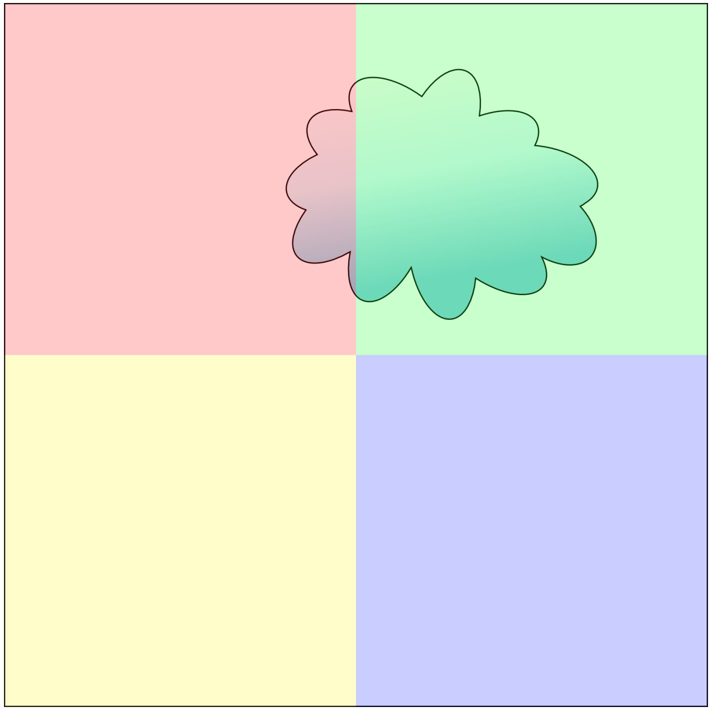
- SPECS computing time varies strongly depending on the range of the particle size distribution and presence of frozen particles
- Time splitting scheme makes SPECS even more costly, in this case: Δt_{COSMO}=10s, Δt_{SPECS}=0.5s

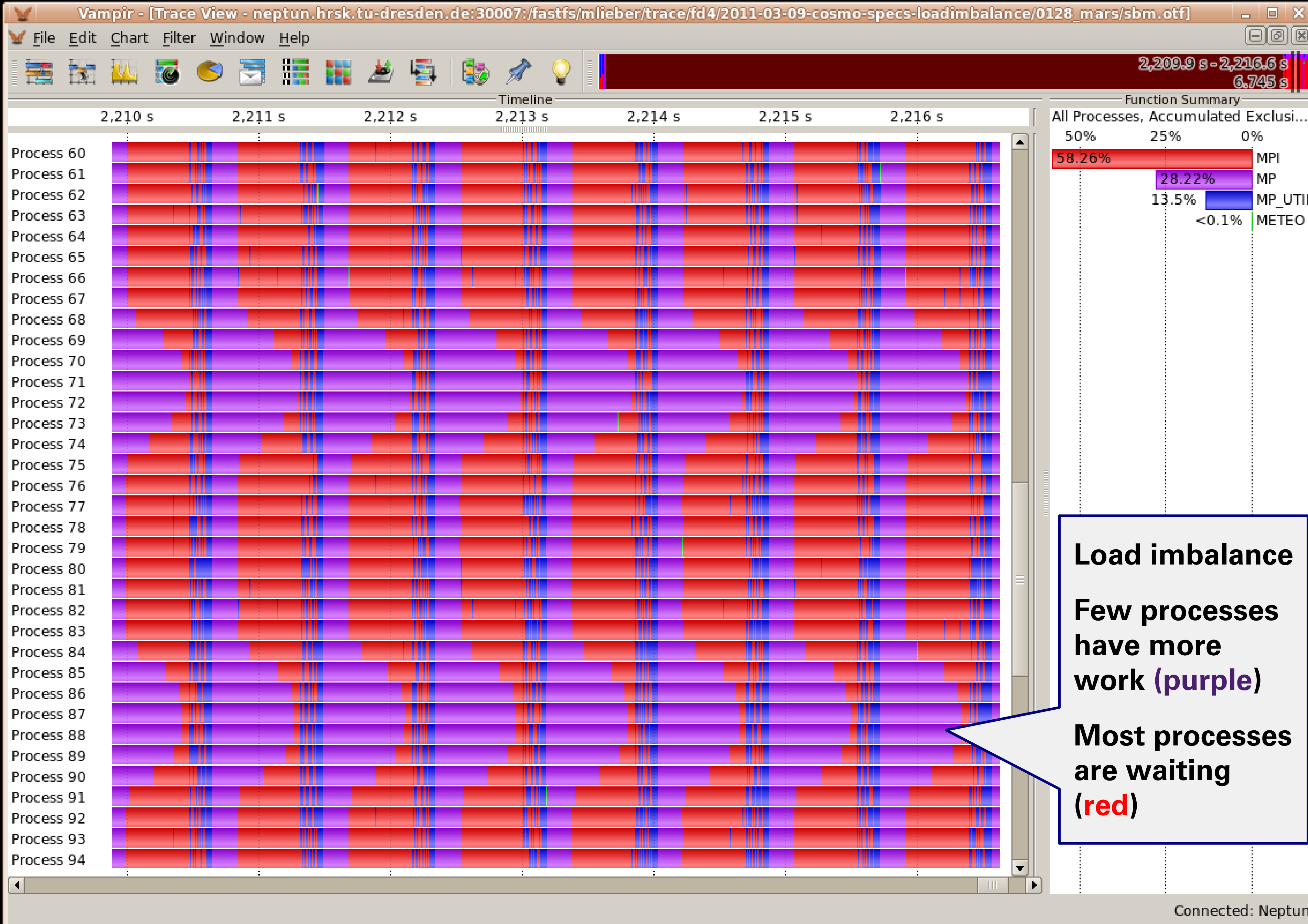
TECHNISCHE UNIVERSITÄT DRESDEN

ZIH
Center for Information Services & High Performance Computing

COSMO-SPECS: Parallel Implementation

- COSMO parallelization
  - 2D (horizontal) decomposition into MxN static partitions
  - Message Passing Interface
- Load imbalance due to SPECS computing time variations

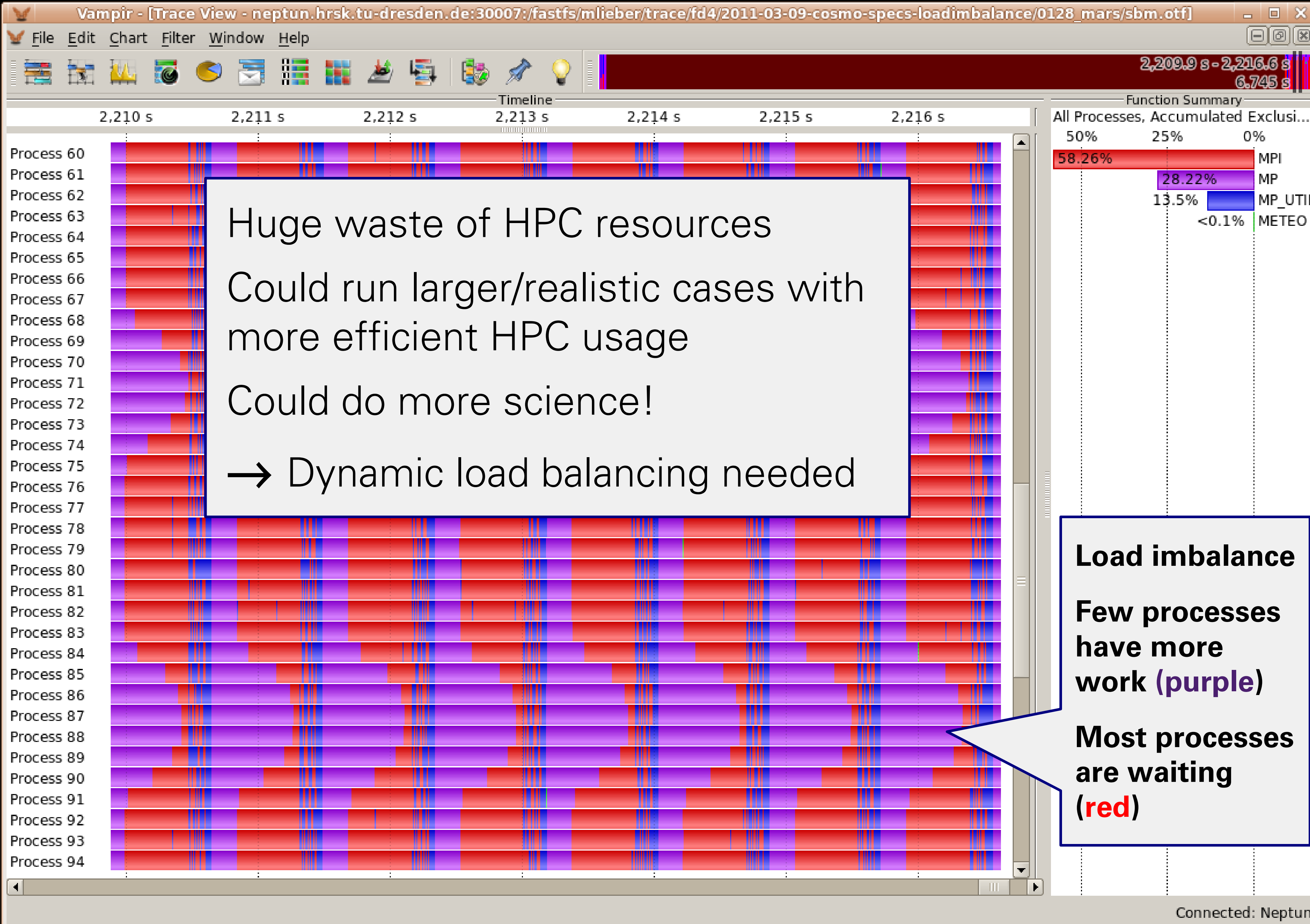




Load imbalance

Few processes have more work (purple)

Most processes are waiting (red)



Huge waste of HPC resources

Could run larger/realistic cases with more efficient HPC usage

Could do more science!

→ Dynamic load balancing needed

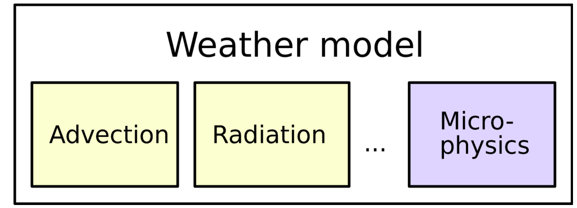
Load imbalance

Few processes have more work (purple)

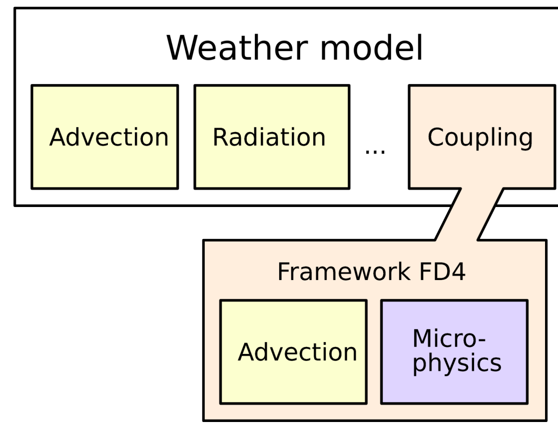
Most processes are waiting (red)

Basic Idea of Load-Balanced Coupling

- Present approaches:
  - Cloud model is implemented as a submodule within the weather model
  - No dynamic load balancing

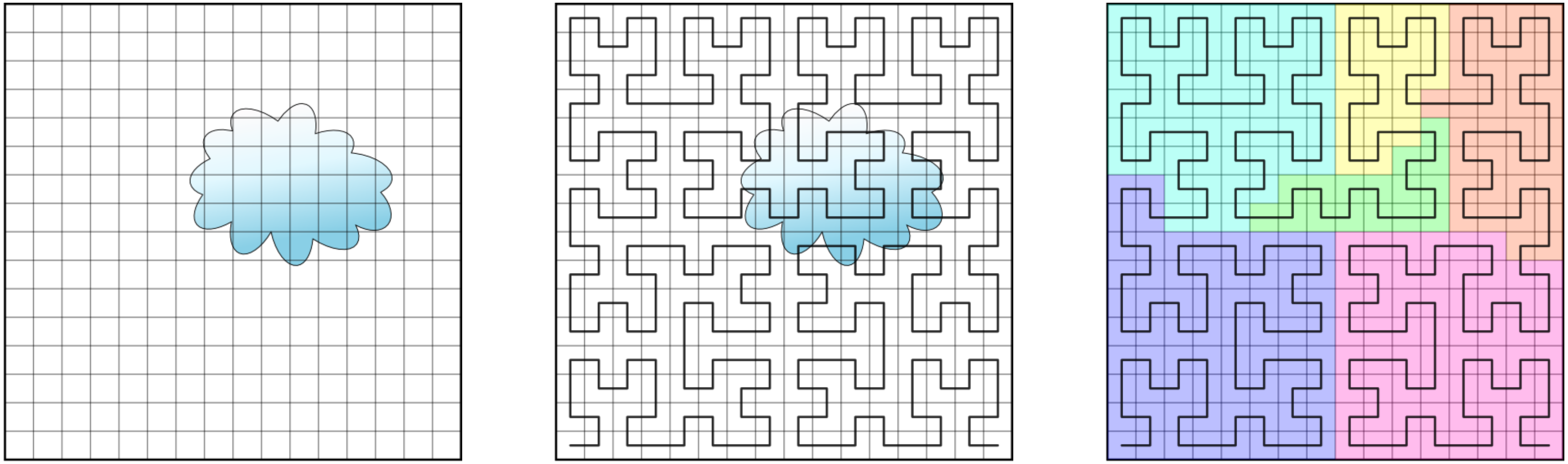


- Our idea:
  - Separate cloud model data from static weather model data structures
  - Independent domain decompositions
  - Dynamic load balancing for cloud model
  - (Re)couple weather and cloud model



- Implement these features in a new framework called FD4
- FD^4 = Four-Dimensional Distributed Dynamic Data structures

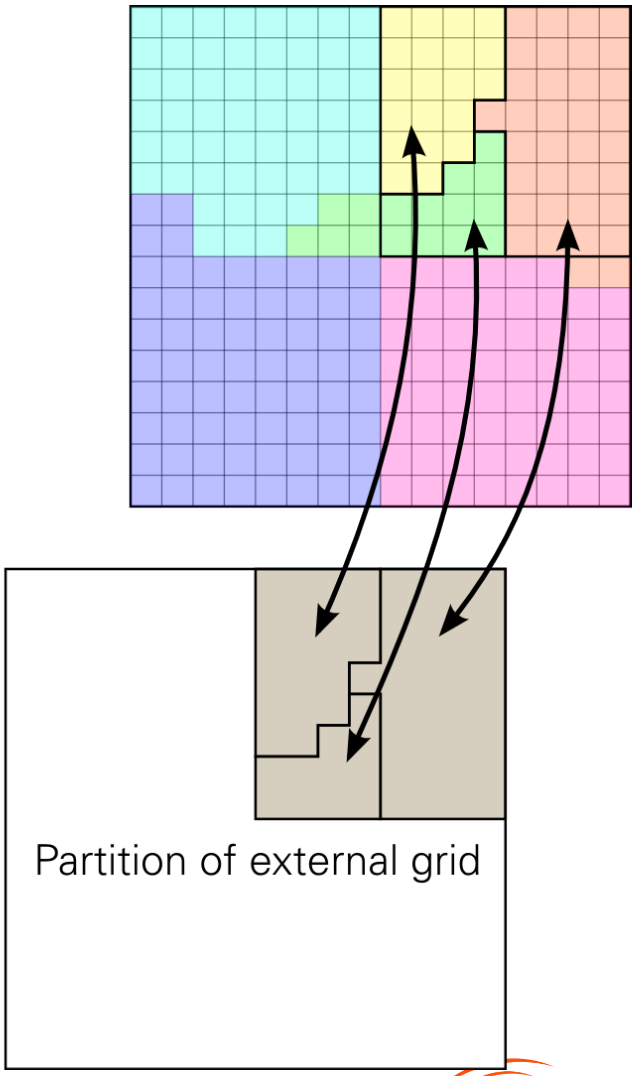
Framework FD4: Dynamic Load Balancing



- Based on 3D block decomposition of grid (#blocks > #processes)
- Task: Assign blocks to processes in a balanced way which also minimizes ghost communication costs. And do this quickly.
- Hilbert space-filling curve (SFC) partitioning [Sagan94, Teresco06]
  - SFC reduces 3D partitioning problem to 1D [Pinar04]
  - Locality properties of SFC lead to moderate ghost comm. costs
  - Very fast algorithm

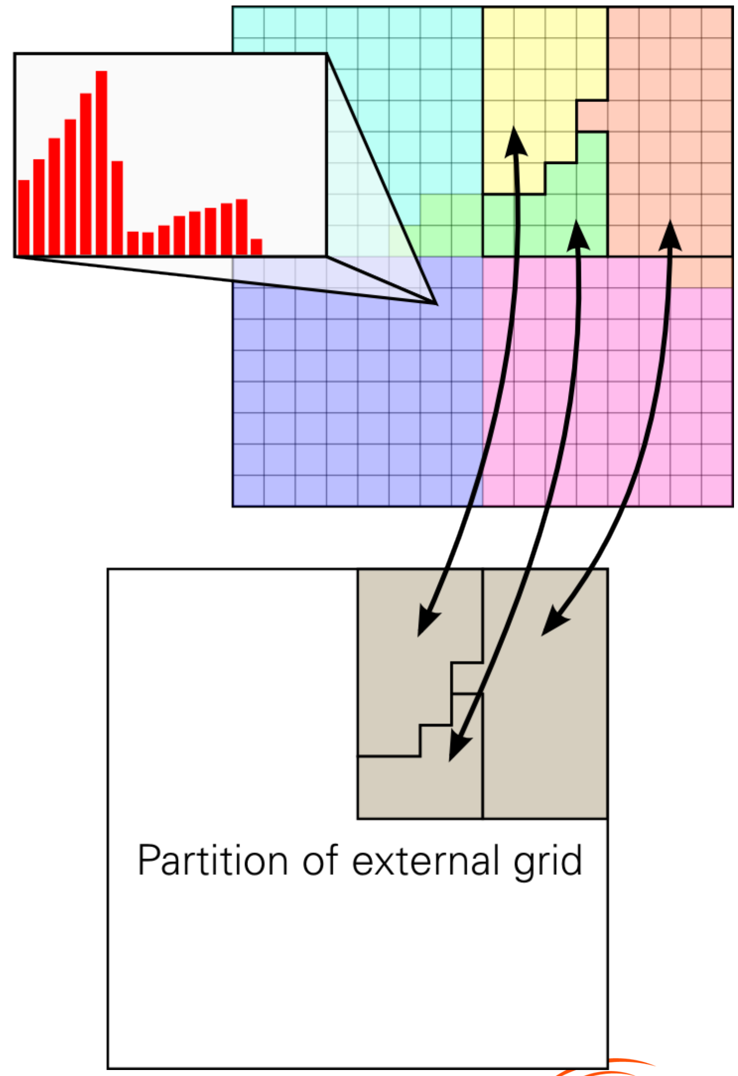
Framework FD4: Model Coupling

- Data exchange between FD4 based model and an external model
  - E.g. weather or CFD model
- Direct data transfer between overlap-ping parts of of the partitions
- Transfer in both directions
- No grid transformation / interpolation
  - External model must provide data matching the grid used in FD4



Framework FD4: 4th Dimension

- Extra dimension of grid variables
- E.g. size resolving models, array of gas phase tracers
- FD4 is optimized for a large 4th dimension
- COSMO-SPECS requires
  2 x 11 x 66 ~ 1500 values



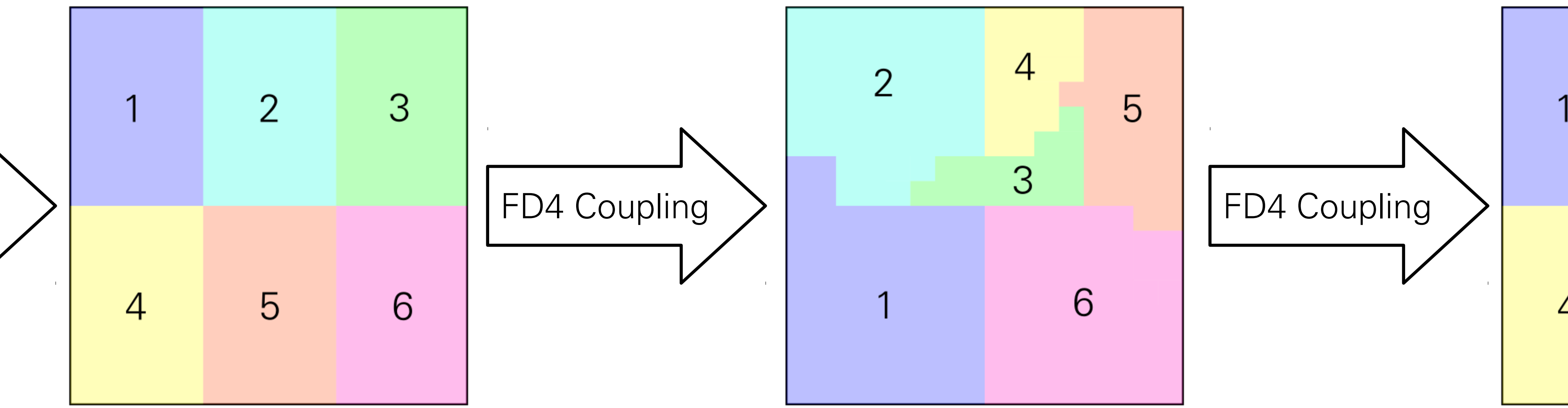
Framework FD4: Implementation

- Implemented in Fortran 95
- MPI-based parallelization
- Open Source Software
- www.tu-dresden.de/zih/clouds

```fortran
! MPI initialization
call MPI_Init(err)
call MPI_Comm_rank(MPI_COMM_WORLD, rank, err)
call MPI_Comm_size(MPI_COMM_WORLD, nproc, err)
! create the domain and allocate memory
call fd4_domain_create(domain, nb, size,     &
     vartab, ng, peri, MPI_COMM_WORLD, err)
call fd4_util_allocate_all_blocks(domain, err)
! initialize ghost communication
call fd4_ghostcomm_create(ghostcomm, domain, &
     4, vars, steps, err)
! loop over time steps
do timestep=1,nsteps
  ! exchange ghosts
  call fd4_ghostcomm_exch(ghostcomm, err)
  ! loop over local blocks
  call fd4_iter_init(domain, iter)
  do while(associated(iter%cur))
    ! do some computations
    call compute_block(iter)
    call fd4_iter_next(iter)
  end do
  ! dynamic load balancing
  call fd4_balance_readjust(domain, err)
end do
```

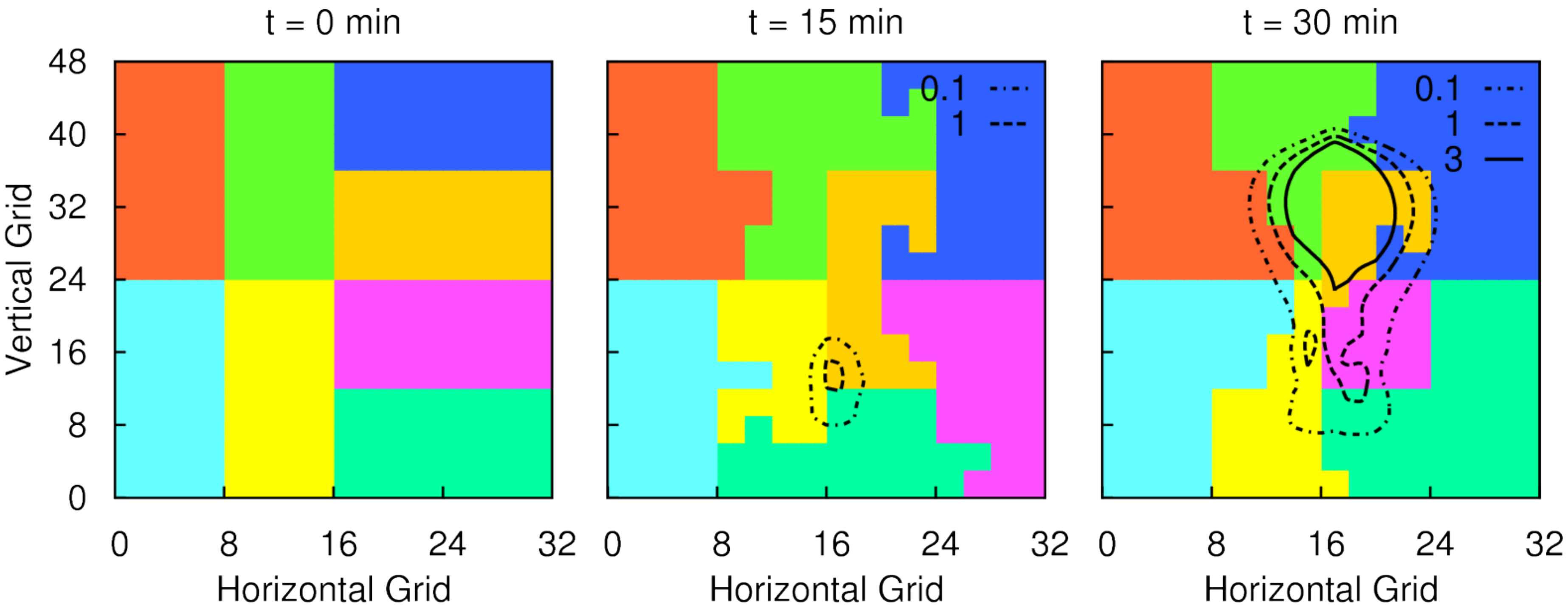
COSMO-SPECS+FD4: Coupling Scheme

COSMO-SPECS+FD4 = COSMO-SPECS using FD4 for coupling



COSMO	FD4	SPECS	FD4
Computes dynamics	Send data to FD4 data structures:	Computes Microphysics	Receive data from FD4 data structures:
Static MxN partitioning	u, v, w, T, p, ρ, q_v	Data dynamically balanced by FD4	ΔT, q_v, q_c, q_i

COSMO-SPECS+FD4: Visualization of Load Balancing

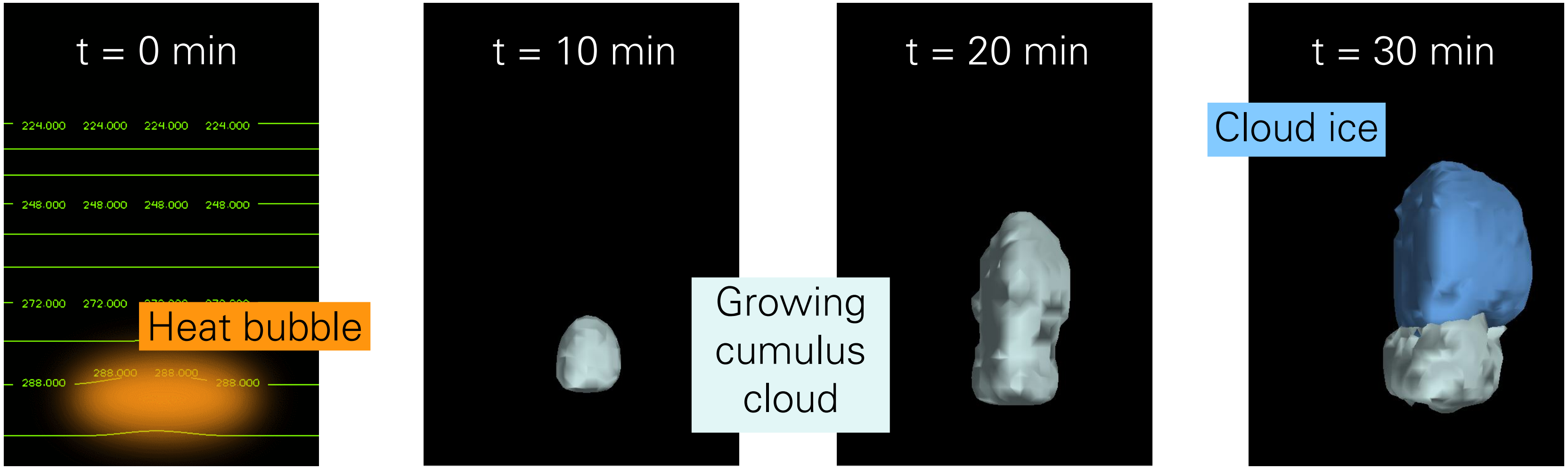


- Small 2D case, 32 x 48 grid, 16 x 16 FD4 blocks
- Colors: 8 partitions for SPECS balanced by FD4
- Contours: liquid and frozen particle concentration (g/kg)

TECHNISCHE UNIVERSITÄT DRESDEN

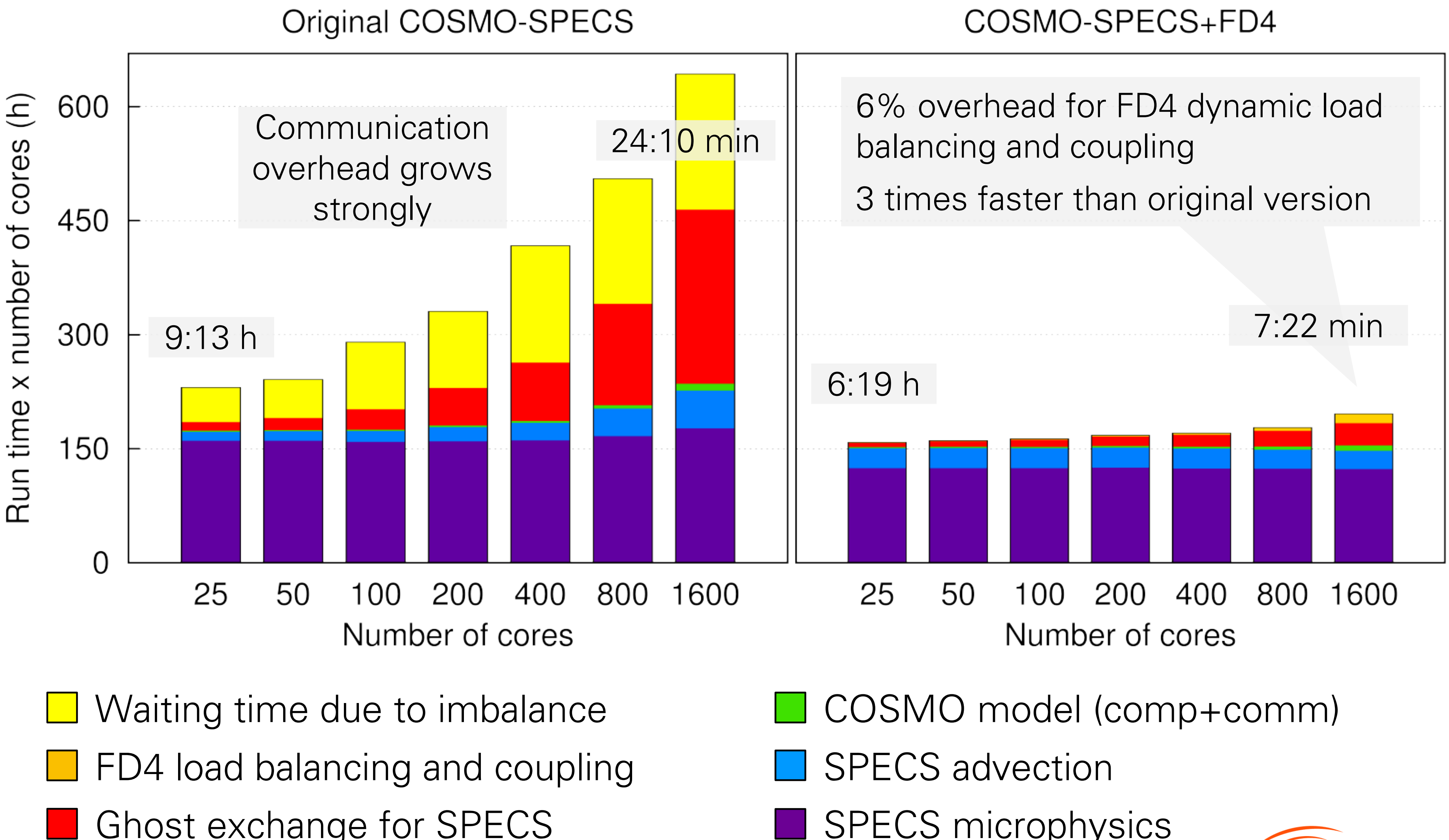
ZIH
Center for Information Services & High Performance Computing

COSMO-SPECS vs. COSMO-SPECS+FD4 Benchmark



- Artificial Scenario:
  - Heat bubble results in growth of idealized cumulus cloud
- 80 x 80 horizontal grid with 1 km resolution
- 48 nonuniform height levels (up to 18 km)
- 19 200 FD4 blocks (2 x 2 x 4 = 16 grid cells)
- 30 min forecast time

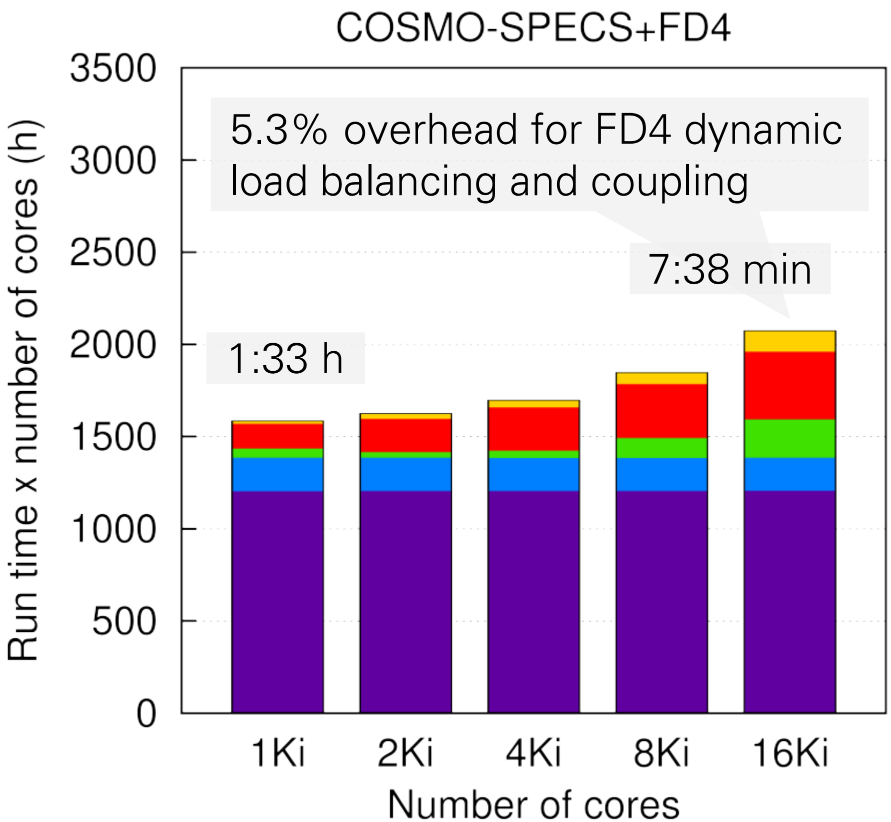
Performance Comparison on HLRB-II (SGI Altix 4700)

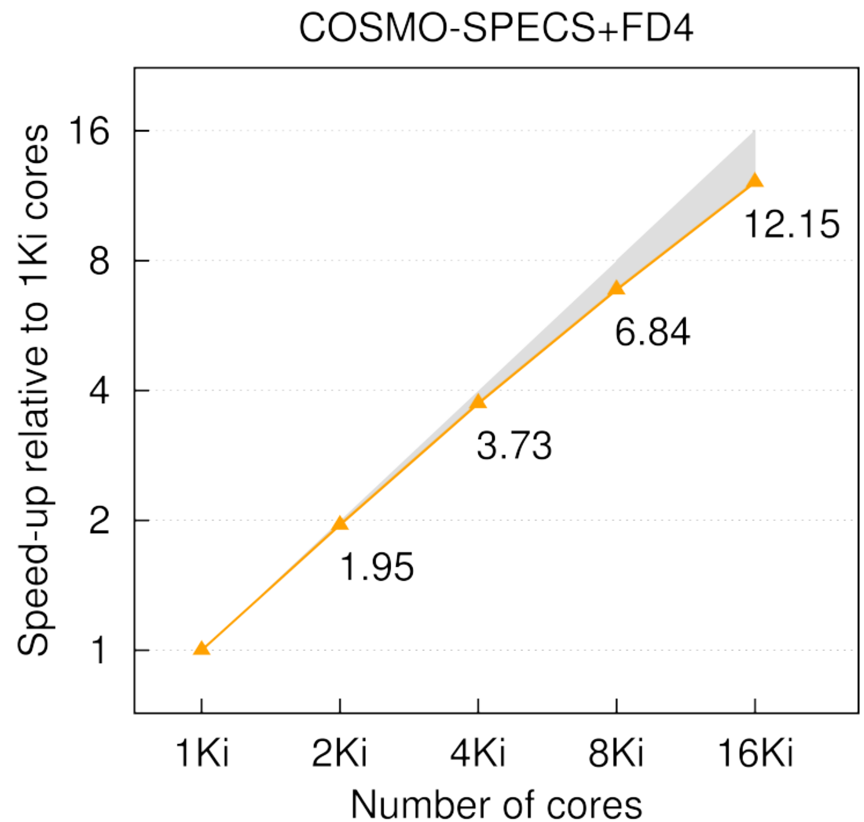


TECHNISCHE UNIVERSITÄT DRESDEN

ZIH Center for Information Services & High Performance Computing

High Scalability Benchmark on Jaguar (Cray XT5)





COSMO model (comp+comm)
FD4 load balancing and coupling
SPECS advection
Ghost exchange for SPECS
SPECS microphysics

256 x 256 grid, 196 608 FD4 blocks, 64 clouds

TECHNISCHE UNIVERSITÄT DRESDEN

ZIH Center for Information Services & High Performance Computing

Conclusion & Outlook

- FD4 provides an efficient and scalable way of coupling spectral cloud microphysics with atmospheric models
- COSMO-SPECS performance increased significantly by FD4
- Large-scale studies with spectral cloud microphysics feasible
- FD4 not limited to spectral cloud microphysics
- FD4 is freely available at http://www.tu-dresden.de/zih/clouds

- Outlook:
  - Investigate adaptive time stepping in COSMO-SPECS+FD4
  - Parallel I/O in FD4
  - Simulate a real-case scenario with COSMO-SPECS+FD4

Acknowledgments

- COSMO Model: German Weather Service (DWD)
- Access to Jugene: Jülich Supercomputing Centre (JSC)
- Access to HLRB-II: Leibniz Supercomputing Centre Munich (LRZ)
- Access to Jaguar: Oak Ridge National Lab (ORNL)
- Funding: German Research Foundation (DFG)

Thank you for your attention!

Refences

[Grützun08] V. Grützun, O. Knoth, and M. Simmel. *Simulation of the influence of aerosol particle characteristics on clouds and precipitation with LM–SPECS: Model description and first results.* Atmos. Res., 90:233–242, 2008.

[CCSP09] *Atmospheric Aerosol Properties and Climate Impacts*, U.S. Climate Change Science Program and the Sub-committee on Global Change Research, 2009.

[IPCC07] S. Solomon et al. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Forth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)*, 2007.

[Lieber10] M. Lieber, V. Grützun, R. Wolke, M.S. Müller, and W.E. Nagel. *FD4: A Framework for Highly Scalable Load Balancing and Coupling of Multiphase Models*, AIP Conference Proceedings 1281(1), 1639-1642, 2010

[Pinar04] A. Pinar and C. Aykanat. *Fast optimal load balancing algorithms for 1D partitioning.* J. Parallel Distrib. Comput., 64(8):974-996, 2004.

[Sagan94] H. Sagan. *Space-filling curves*, Springer, 1994

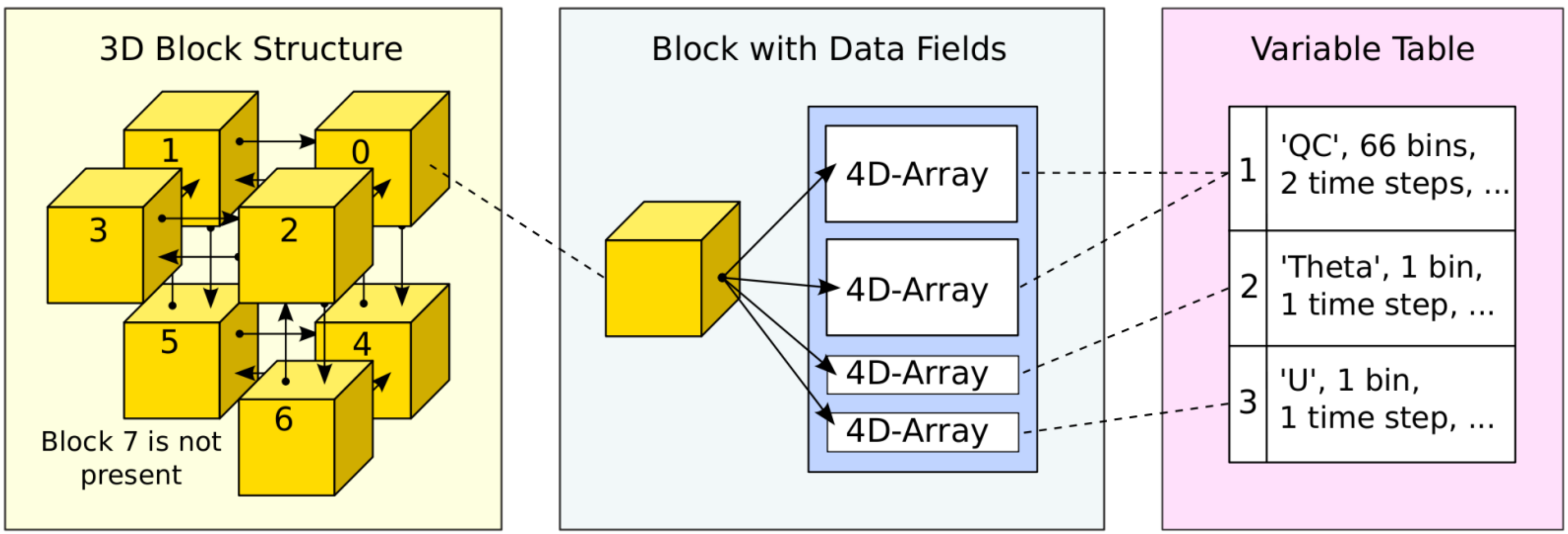
[Simmel06] M. Simmel and S. Wurzler. *Condensation and activation in sectional cloud microphysical models*, Atmos. Res., 80:218-236, 2006.

[Teresco06] J.D. Teresco, K.D. Devine, and J.E. Flaherty. *Partitioning and Dynamic Load Balancing for the Numerical Solution of Partial Differential Equations,* in Numerical Solution of Partial Differential Equations on Parallel Computers, pages 55-88, Springer, 2006.

Backup slides

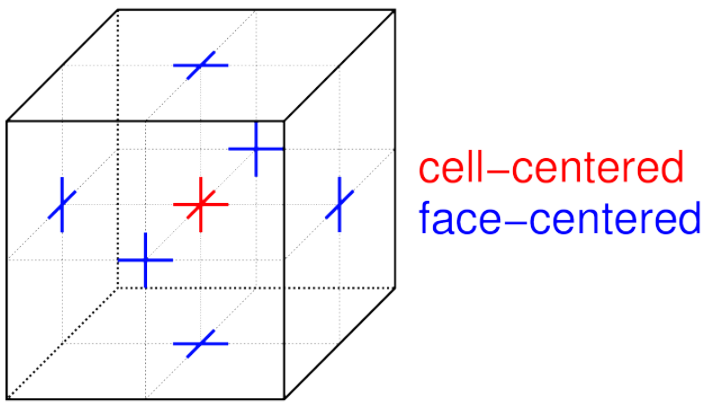
Framework FD4: Basic Data Structure

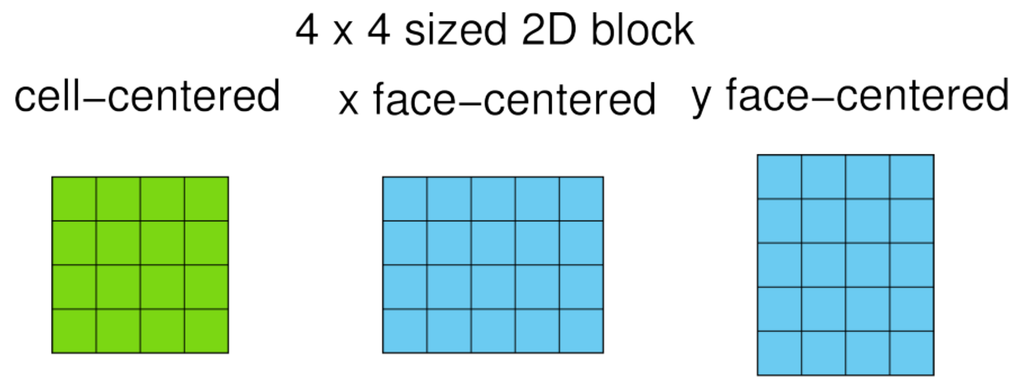
- 3D Regular grid is decomposed into blocks
- FD4 allocates data fields in the blocks based on variable table
  - Variable name
  - Length of 4th dimension
  - Number of time steps
  - Discretization, i.e. cell-centered or face-centered



Framework FD4: Variable Discretization Types

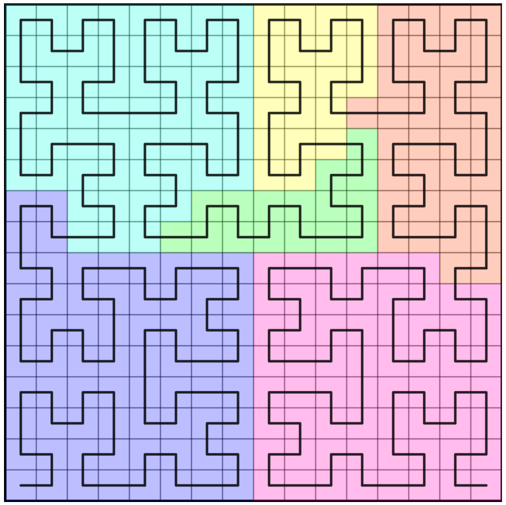
- Cell-centered variables
- Face-centered variables

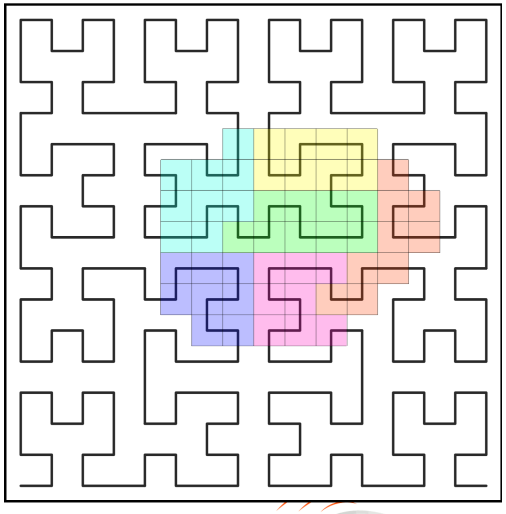




Framework FD4: Dynamic Load Balancing

- When blocks are added or removed (adaptive block mode)
- When load balance decreases below a certain limit
- User can assign each block a weight, e.g. computation time of the block
- Two partitioning methods:
  - Hilbert space-filling curve (SFC) partitioning [Sagan94]
  - Graph partitioning using ParMETIS
- SFC preferred because graph partitioning has much higher overhead



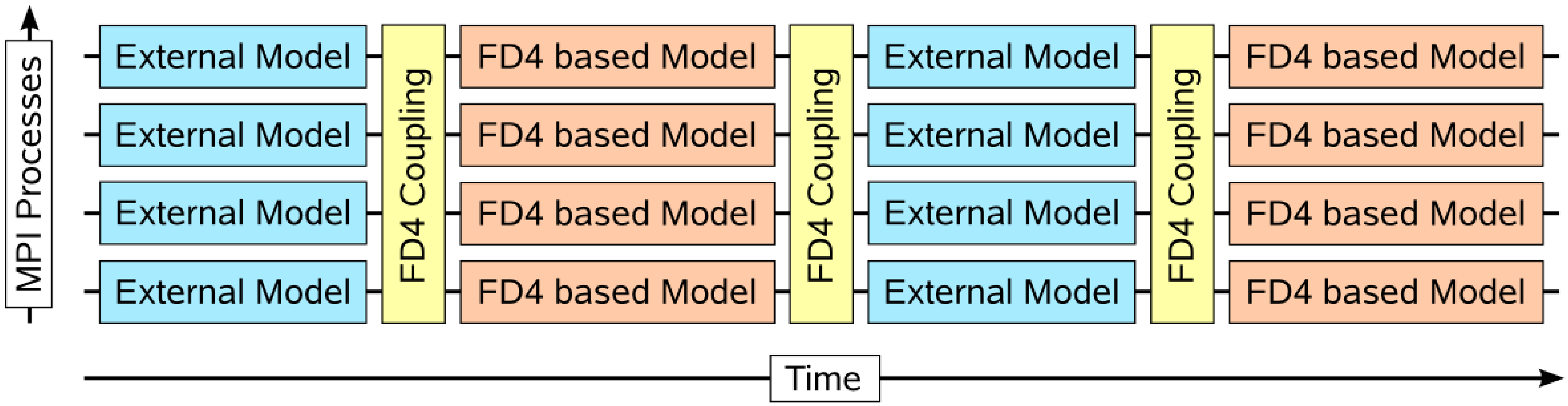


Framework FD4: SFC Partitioning Algorithm

- Space-filling curve reduces 3D partitioning problem to 1D
- Partitioning of an 1D array of block weights
- Goal: Reduce the maximum load (*bottleneck value*)
- Determination of minimal *bottleneck value* is not trivial
  - Only lower and upper bound are known a priori [Pinar04]
- Fast parallel heuristic implemented in FD4:
  - All processes exchange their block weights
  - Each process checks for a different *bottleneck value* within the known interval if a partitioning exists for it
  - Minimal value determined with MPI_Allreduce
  - Each process determines its own partition based on the minimal *bottleneck value* found

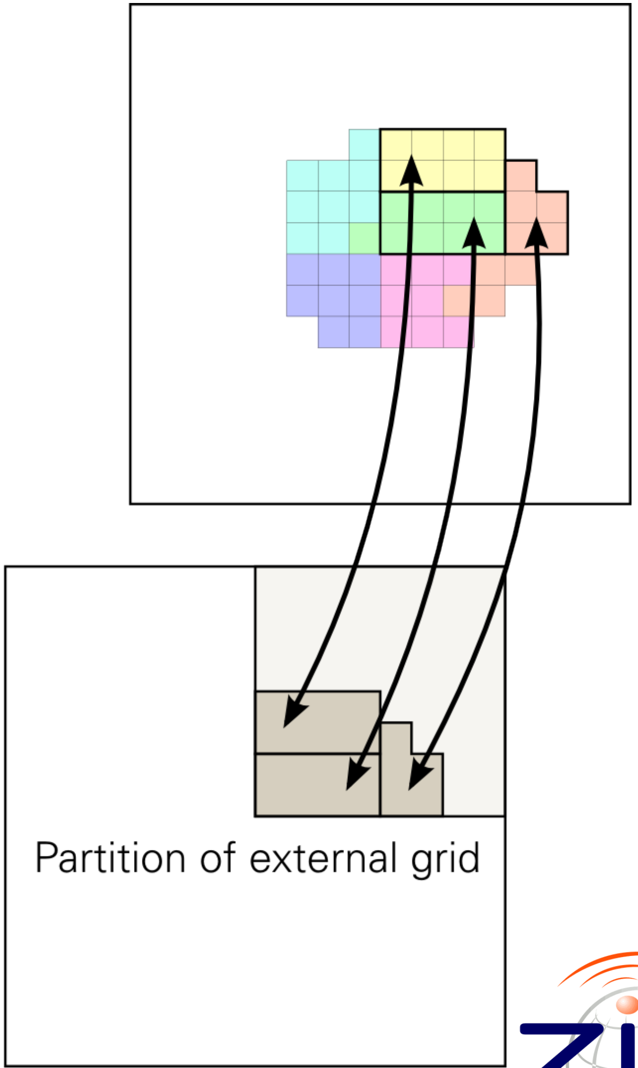
Framework FD4: Model Coupling

- *Sequential coupling*: all processes perform computations for both models alternately [Lieber08]



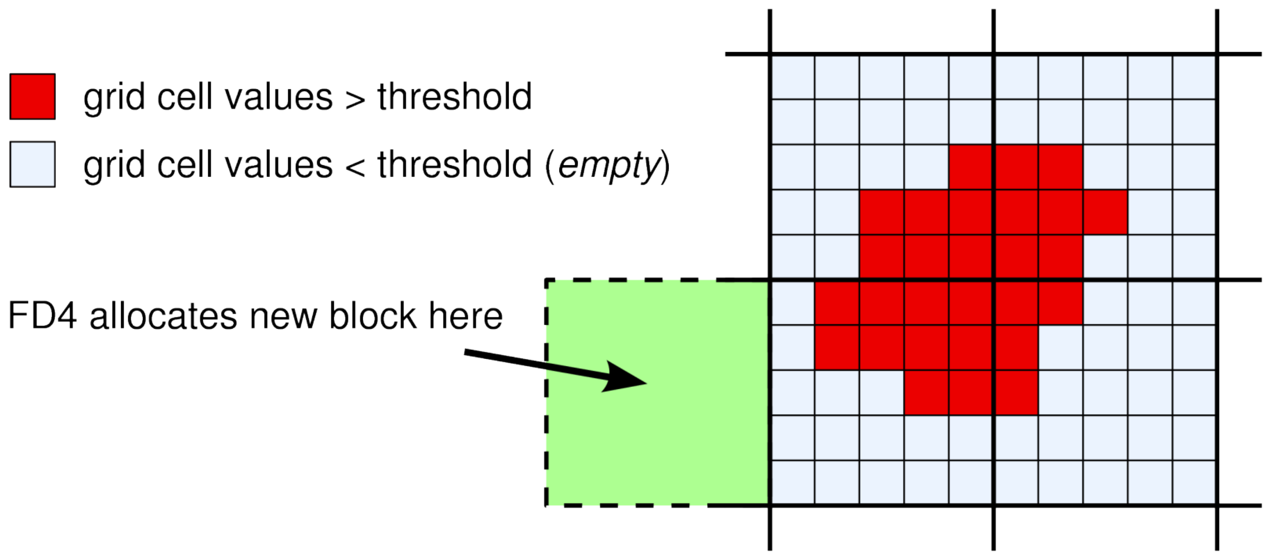
Framework FD4: Adaptive Block Mode

- Save memory in case data and computations are required for a spatial subset only
- Suitable for multiphase problems like drops, clouds, flame fronts

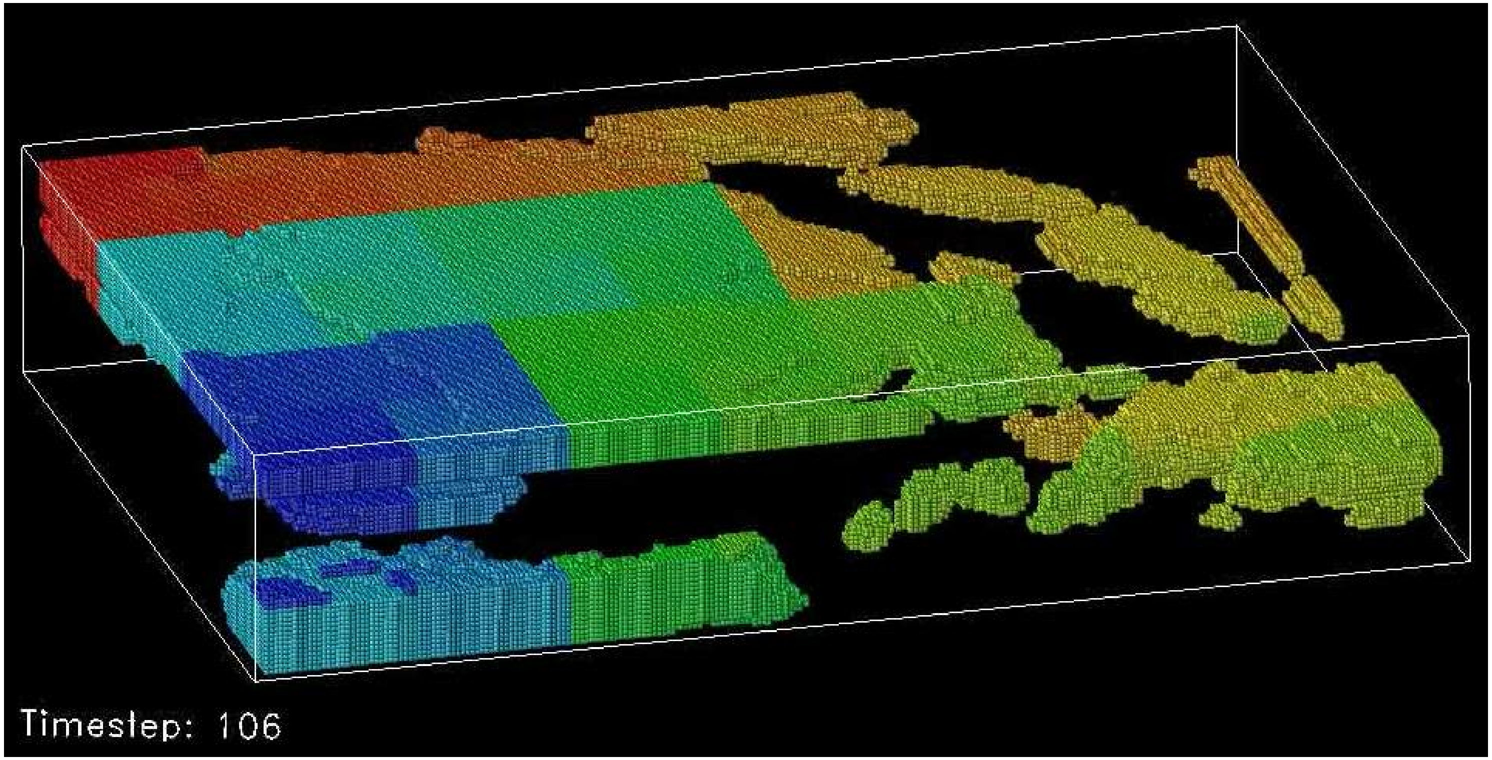


Framework FD4: Adaptive Block Mode

- Grid allocation adapts to spatial structure of simulated problem
- *Empty* blocks are not allocated
  - Defined by a threshold value for specified variables
  - I.e. block does not contain any quantities of a certain phase
- FD4 ensures existence of all blocks required for correct stencil operations

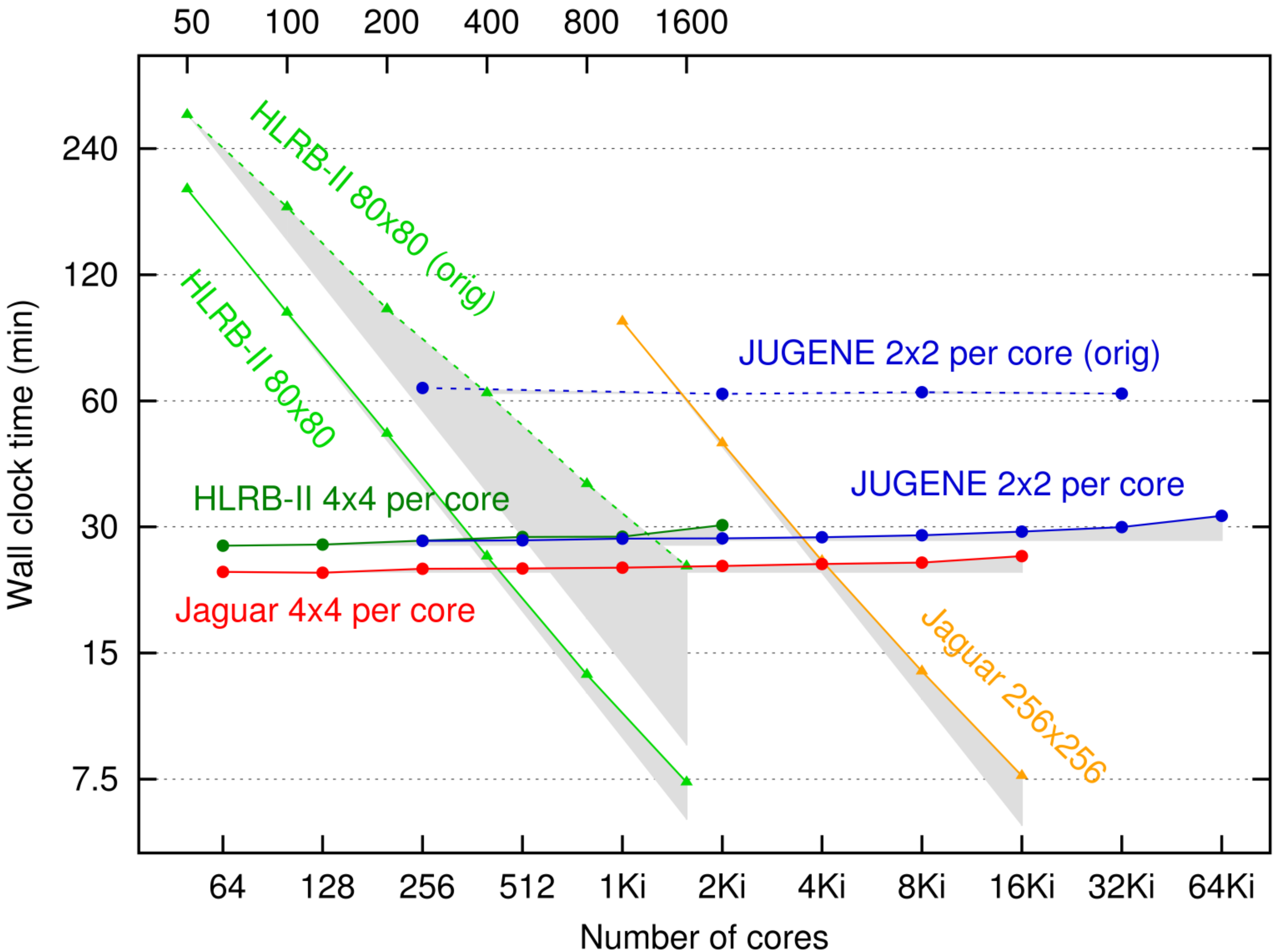


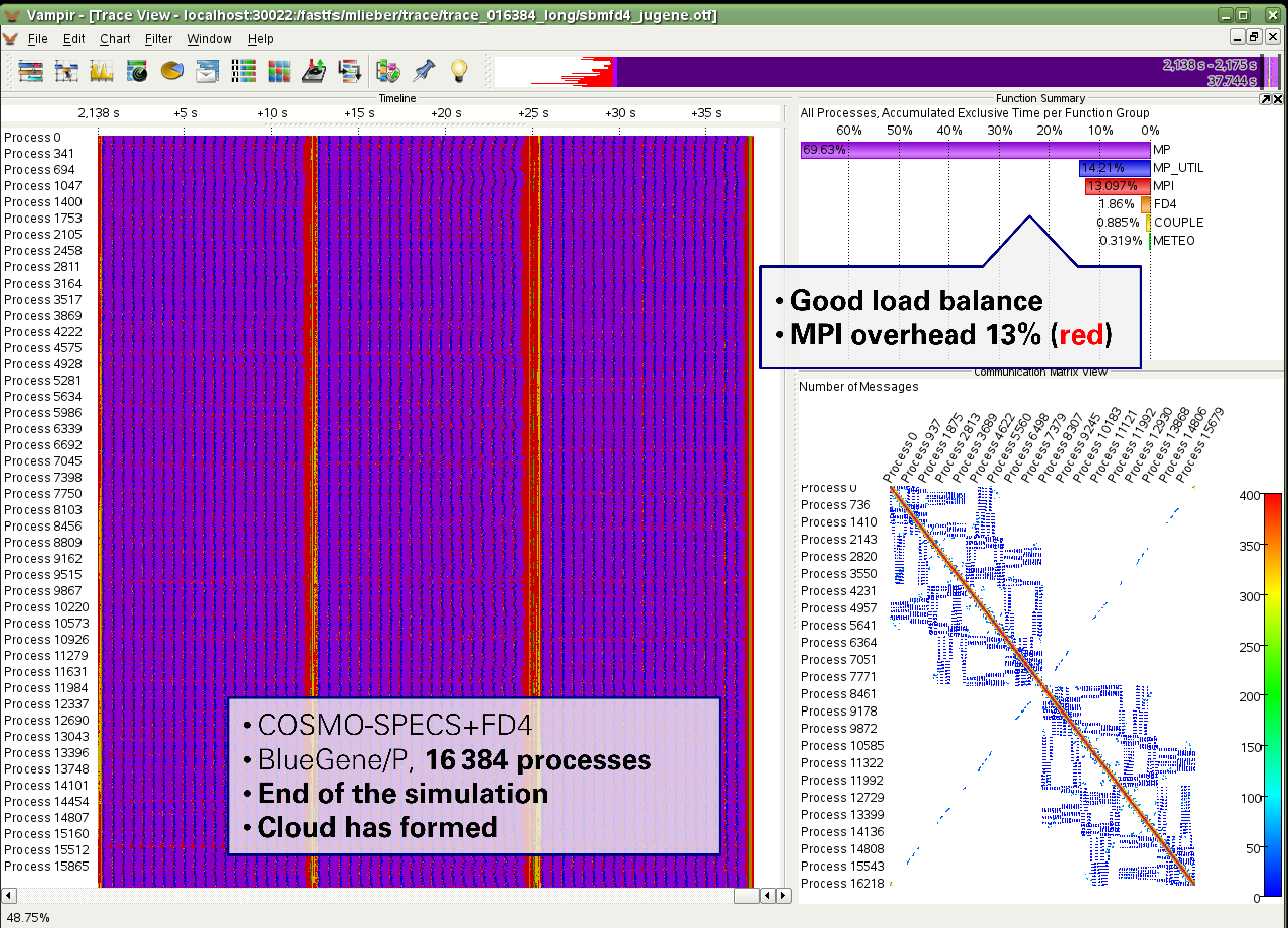
Framework FD4: Dynamic Load Balancing Movie



- Overhead test of adaptive block mode and load balancing [Lieber10]
- FD4 adapts to cloud formation in COSMO weather model
- Real-life scenario, 249 x 174 x 50 grid, 256 processes

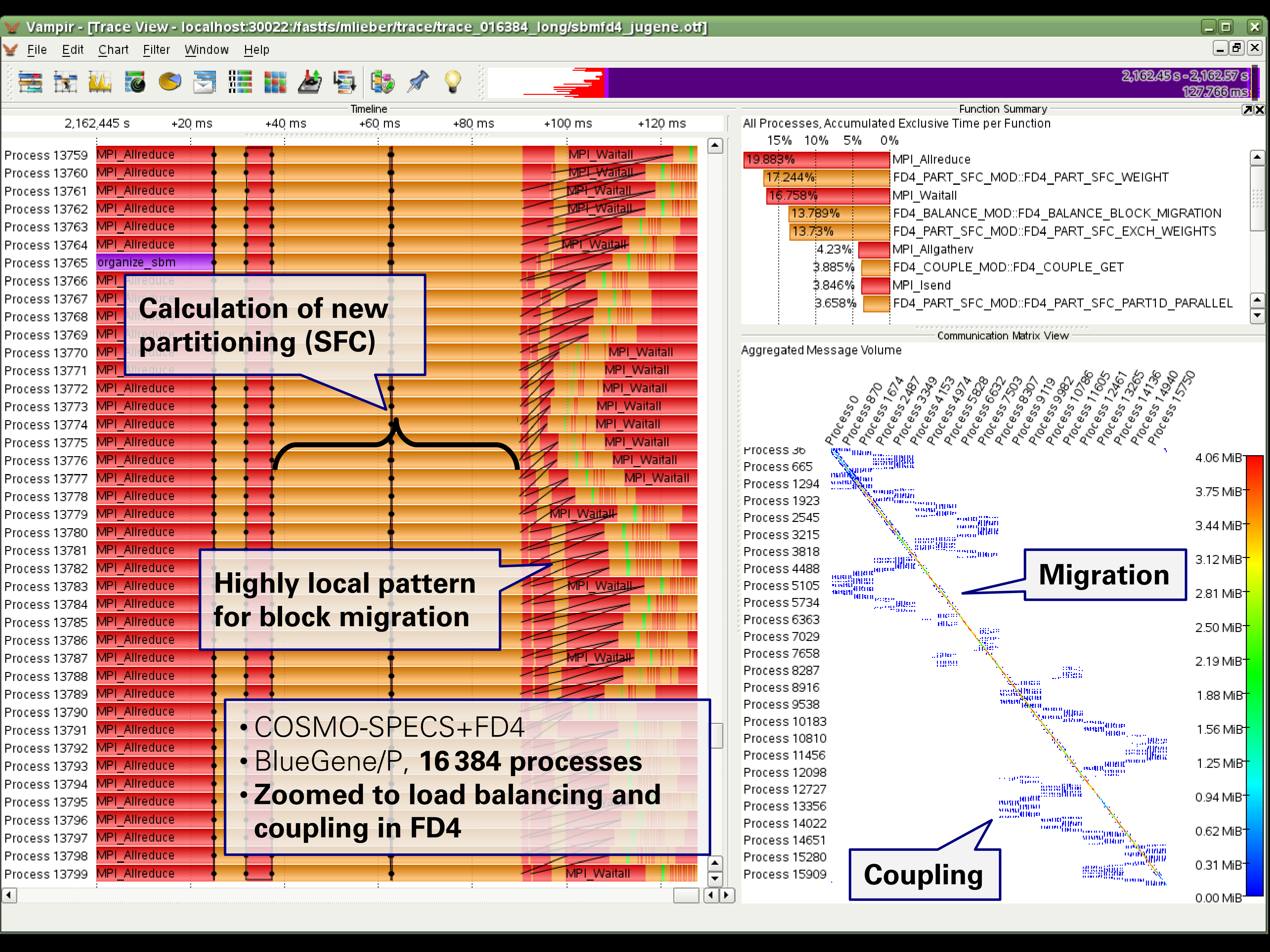
COSMO-SPECS(+FD4) Benchmarks Overview

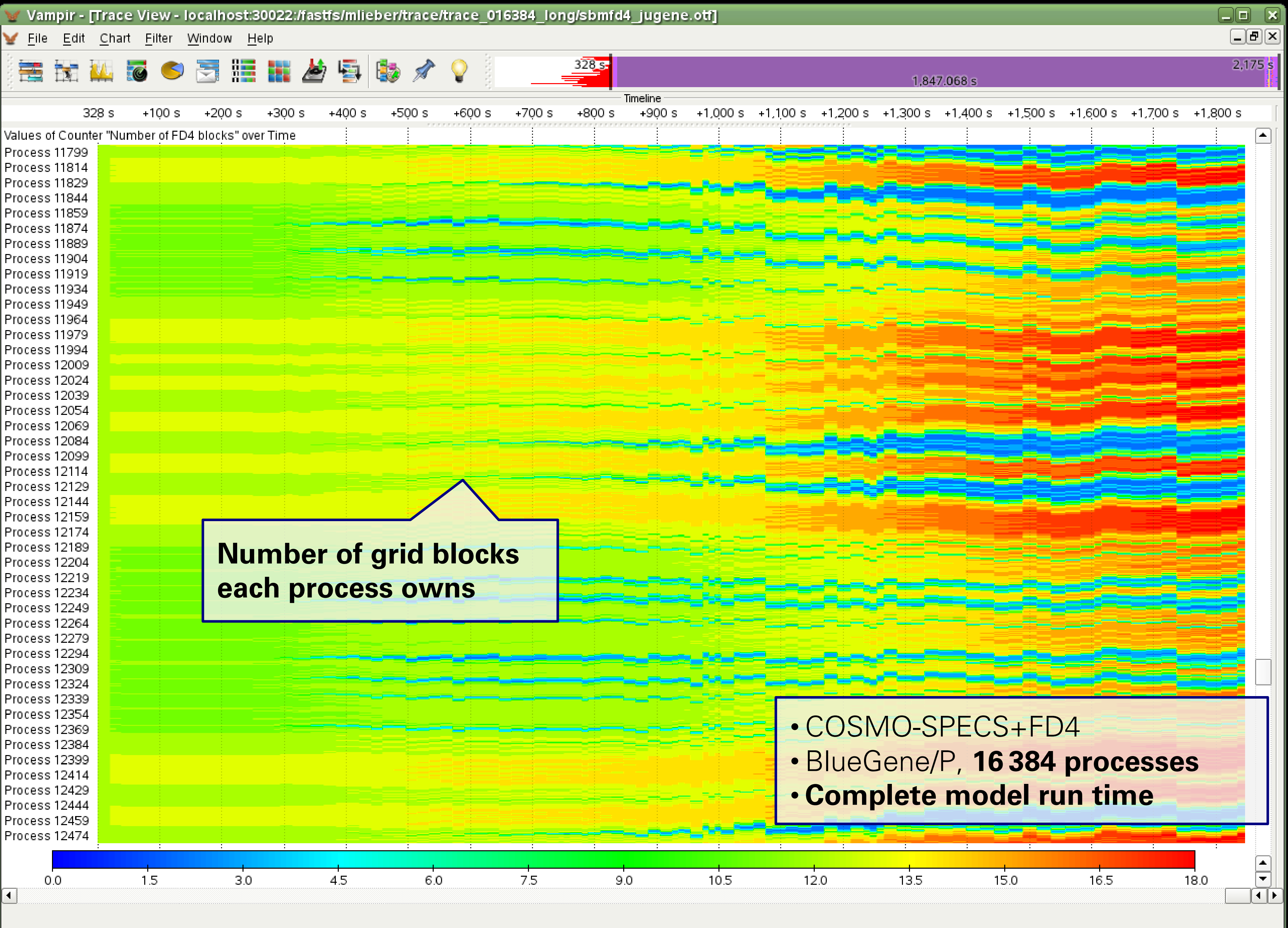




- Good load balance
- MPI overhead 13% (red)

- COSMO-SPECS+FD4
- BlueGene/P, **16 384 processes**
- **End of the simulation**
- **Cloud has formed**


Vampir - [Trace View - localhost:30022:/fastfs/mlieber/trace/trace_016384_long/sbmfd4_jugene.otf]
File Edit Chart Filter Window Help
2,162.45 s - 2,162.57 s
127.766 ms
Timeline
2,162.445 s +20 ms +40 ms +60 ms +80 ms +100 ms +120 ms
Process 13759
Process 13760
Process 13761
Process 13762
Process 13763
Process 13764
Process 13765
Process 13766
Process 13767
Process 13768
Process 13769
Process 13770
Process 13771
Process 13772
Process 13773
Process 13774
Process 13775
Process 13776
Process 13777
Process 13778
Process 13779
Process 13780
Process 13781
Process 13782
Process 13783
Process 13784
Process 13785
Process 13786
Process 13787
Process 13788
Process 13789
Process 13790
Process 13791
Process 13792
Process 13793
Process 13794
Process 13795
Process 13796
Process 13797
Process 13798
Process 13799
MPI_Allreduce
organize_sbm
MPI_Waitall
Calculation of new partitioning (SFC)
Highly local pattern for block migration
• COSMO-SPECS+FD4
• BlueGene/P, 16 384 processes
• Zoomed to load balancing and coupling in FD4
Function Summary
All Processes, Accumulated Exclusive Time per Function
15% 10% 5% 0%
19.883% MPI_Allreduce
17.244% FD4_PART_SFC_MOD::FD4_PART_SFC_WEIGHT
16.758% MPI_Waitall
13.789% FD4_BALANCE_MOD::FD4_BALANCE_BLOCK_MIGRATION
13.73% FD4_PART_SFC_MOD::FD4_PART_SFC_EXCH_WEIGHTS
4.23% MPI_Allgatherv
3.885% FD4_COUPLE_MOD::FD4_COUPLE_GET
3.846% MPI_Isend
3.658% FD4_PART_SFC_MOD::FD4_PART_SFC_PART1D_PARALLEL
Communication Matrix View
Aggregated Message Volume
Process 0
Process 870
Process 1674
Process 2487
Process 3349
Process 4153
Process 4974
Process 5828
Process 6632
Process 7503
Process 8307
Process 9119
Process 9982
Process 10786
Process 11605
Process 12461
Process 13265
Process 14136
Process 14940
Process 15750
Process 36
Process 665
Process 1294
Process 1923
Process 2545
Process 3215
Process 3818
Process 4488
Process 5105
Process 5734
Process 6363
Process 7029
Process 7658
Process 8287
Process 8916
Process 9538
Process 10183
Process 10810
Process 11456
Process 12098
Process 12727
Process 13356
Process 14022
Process 14651
Process 15280
Process 15909
4.06 MiB
3.75 MiB
3.44 MiB
3.12 MiB
2.81 MiB
2.50 MiB
2.19 MiB
1.88 MiB
1.56 MiB
1.25 MiB
0.94 MiB
0.62 MiB
0.31 MiB
0.00 MiB
Migration
Coupling


Vampir - [Trace View - localhost:30022:/fastfs/mlieber/trace/trace_016384_long/sbmfd4_jugene.otf]
File Edit Chart Filter Window Help
328 s
1,847.068 s
2,175 s
Timeline
328 s
+100 s
+200 s
+300 s
+400 s
+500 s
+600 s
+700 s
+800 s
+900 s
+1,000 s
+1,100 s
+1,200 s
+1,300 s
+1,400 s
+1,500 s
+1,600 s
+1,700 s
+1,800 s
Values of Counter "Number of FD4 blocks" over Time
Process 11799
Process 11814
Process 11829
Process 11844
Process 11859
Process 11874
Process 11889
Process 11904
Process 11919
Process 11934
Process 11949
Process 11964
Process 11979
Process 11994
Process 12009
Process 12024
Process 12039
Process 12054
Process 12069
Process 12084
Process 12099
Process 12114
Process 12129
Process 12144
Process 12159
Process 12174
Process 12189
Process 12204
Process 12219
Process 12234
Process 12249
Process 12264
Process 12279
Process 12294
Process 12309
Process 12324
Process 12339
Process 12354
Process 12369
Process 12384
Process 12399
Process 12414
Process 12429
Process 12444
Process 12459
Process 12474
Number of grid blocks each process owns
• COSMO-SPECS+FD4
• BlueGene/P, 16 384 processes
• Complete model run time
0.0
1.5
3.0
4.5
6.0
7.5
9.0
10.5
12.0
13.5
15.0
16.5
18.0

Application of FD4: Load Balance Measurement