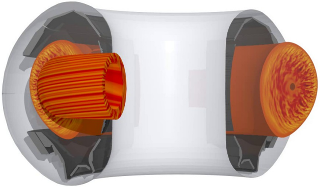
EXAHD – An Exa-Scalable Two-Level Approach for Higher-Dimensional Problems in Plasma Physics and Beyond

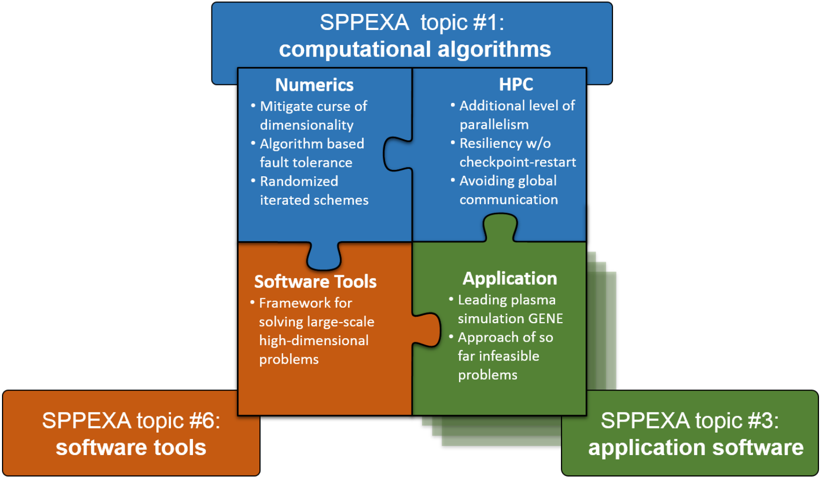
SPPEXA Annual Plenary Meeting

Dirk Pflüger, Universität Stuttgart
Hans-Joachim Bungartz, TU München
Michael Griebel, Universität Bonn, Fraunhofer SCAI
Tilman Dannert, Max Planck Computing & Data Facility (RZG)
Markus Hegland, Australian National University
Frank Jenko, University of California, Los Angeles

Garching, March 21, 2017

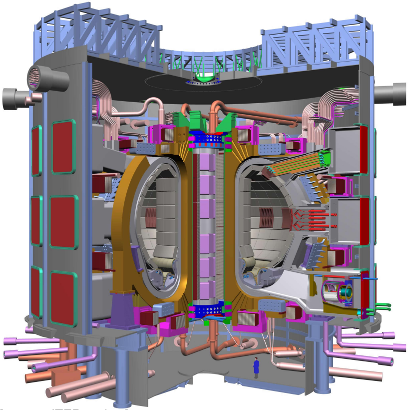


EXAHD – Project Goals

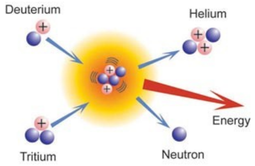


PDE: Turbulence simulations of hot fusion plasmas

- Idea: new, CO_2-free source of energy for the generations to come



[source: ITER project]



Practically Unlimited Ressources

Contents:

- **Deuterium** in bath tub full of water and **Lithium** in used laptop battery suffice for family over 50 years

Behind the Scenes

- Dilute/hot plasmas are (almost) collisionless
- Not magneto-hydrodynamic, but kinect description (Vlasov):

$$\left[\frac{\partial}{\partial t} + \vec{v}\frac{\partial}{\partial \vec{x}} + \frac{q}{m}\left(E + \frac{\vec{v}}{c} \times B\right)\frac{\partial}{\partial \vec{v}}\right] f(\vec{x}, \vec{v}, t) = 0$$

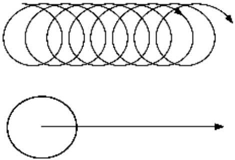
  - Distribution function $f(\vec{x}, \vec{v}, t)$
  - 6D in state space
  - Coupled to Maxwell equations

Behind the Scenes

- Dilute/hot plasmas are (almost) collisionless
- Not magneto-hydrodynamic, but kinect description (Vlasov):

$$\left[\frac{\partial}{\partial t} + \vec{v}\frac{\partial}{\partial \vec{x}} + \frac{q}{m}\left(E + \frac{\vec{v}}{c} \times B\right)\frac{\partial}{\partial \vec{v}}\right] f(\vec{x}, \vec{v}, t) = 0$$

  - Distribution function $f(\vec{x}, \vec{v}, t)$
  - 6D in state space
  - Coupled to Maxwell equations



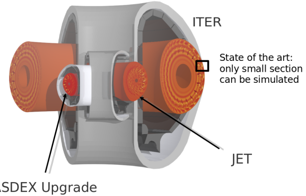
- Gyrokinetics: remove fast gyromotion (smallest scale)

$$\left[\frac{\partial}{\partial t} + \tilde{\vec{v}} \cdot \frac{\partial}{\partial \vec{x}} + \tilde{F}\frac{\partial}{\partial v_{||}}\right] f(\vec{x}, v_{||}, \mu, t) = \Delta(f)$$

  - 5D
  - $\tilde{\vec{v}}$ and $\tilde{F}$ are complex expressions, contain evaluation of E and B

Numerical Simulations for Actual Tokamaks with GENE

Goal: global simulation with physical realism



Gyrokinetic Electromagnetic Numerical Experiment

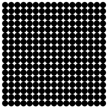
`http://www.genecode.org`

- Szenario for simulation of “numerical ITER”
  - Global, non-linear runs
  - At least 10^{11} grid points, 10^6 time steps
  - $>$1 TB just to store single result in memory (complex)
- Possible at all?

Sparse Grids – Hierarchical Approach

- High-dimensional problems suffer “curse of dimensionality”
  - Effort $\mathcal{O}((2^n)^d)$

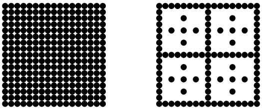
	full grid
5d, level 10	$> 10^{15}$



Sparse Grids – Hierarchical Approach

- High-dimensional problems suffer "curse of dimensionality"
  - Effort $\mathcal{O}((2^n)^d)$
- Therefore: hierarchical discretization
  - Sparse grids: $\mathcal{O}(2^n \cdot n^{d-1})$ [Zenger 91]
  - Makes high-dimensional discretizations possible

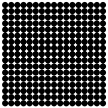
	full grid	sparse grid
5d, level 10	$> 10^{15}$	25,416,705

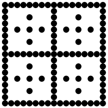


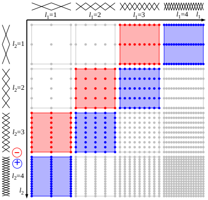
Sparse Grids – Hierarchical Approach

- High-dimensional problems suffer “curse of dimensionality”
  - Effort $\mathcal{O}((2^n)^d)$
- Therefore: hierarchical discretization
  - Sparse grids: $\mathcal{O}(2^n \cdot n^{d-1})$ [Zenger 91]
  - Makes high-dimensional discretizations possible

	full grid	sparse grid	sg combination technique
5d, level 10	$> 10^{15}$	25,416,705	1,876 $\times$ 82,000



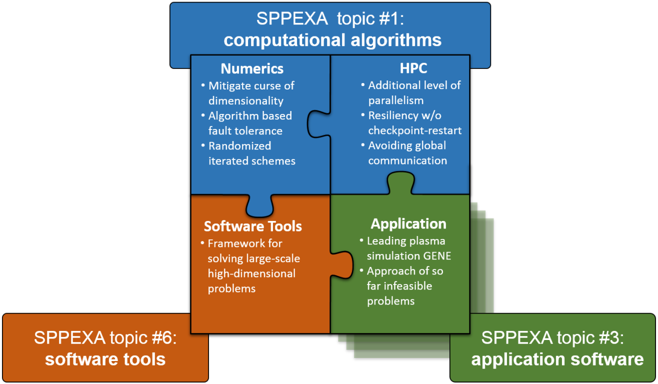




- Combination technique (multivariate extrapolation-style scheme)
  - Multiple, but smaller grids: $\mathcal{O}(d \cdot n^{d-1})$ problems of size $\mathcal{O}(2^n)$

EXAHD – Project Goals

Scalability



Scalability

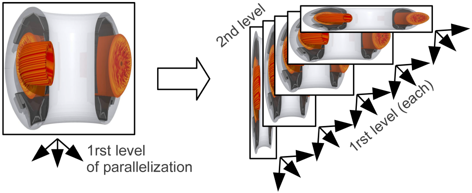
Problem of standard solver: global communication within each time-step

Scalability

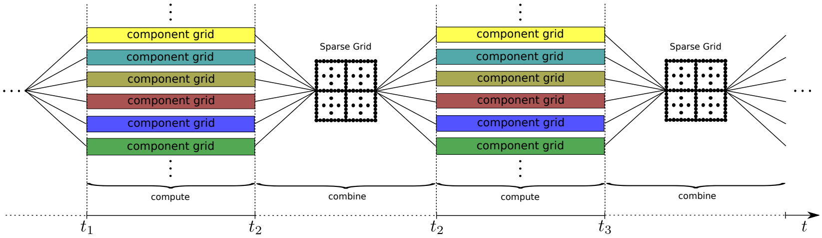
Problem of standard solver: global communication within each time-step

Use hierarchical ansatz

- Two-level approach
- Numerics: decoupling into locally coupled problems
- Algorithms: second level of parallelism
- First level: no need to scale to exascale



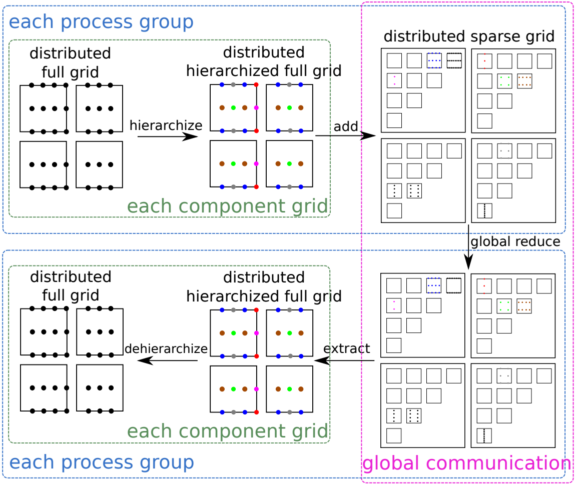
Time-Dependent PDEs



- Gather-scatter steps every time-interval
- Remaining reduced global communication

Global Communication

Optimal communication schemes



Global Communication

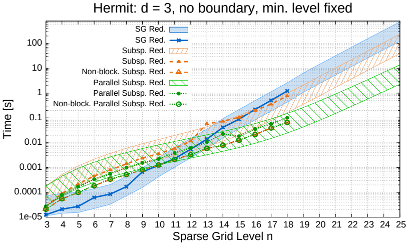
- Minimize number of communications (Range Query Trees):

$$\mathcal{O}(\log(dn^{d-1})) \times \mathcal{O}(2^n n^{d-1})$$

- Minimize package size

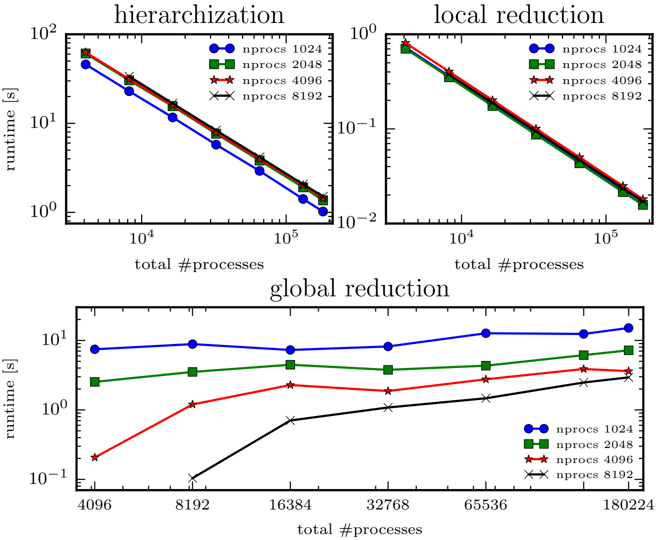
$$\mathcal{O}(2n \cdot n^{d-1}) \times \mathcal{O}(2^{n-1})$$

- Derivation in BSP/PEM model



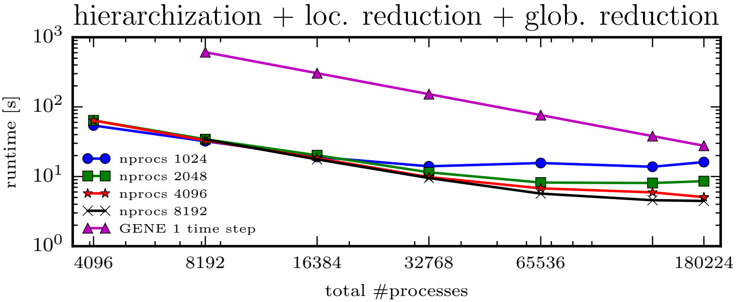
[joint work with R. Jacob (ITU, Algorithm Engineering)]

Runtimes on Hazel Hen



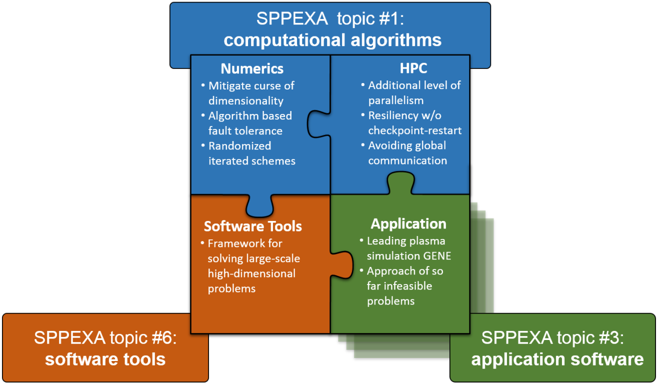
Runtimes on Hazel Hen

Total time



EXAHD – Project Goals

Algorithm-Based Fault Tolerance



Resilience for the Exa-Age

Hard Faults

- Errors that trigger signals to the user
- Node, OS, network or process failure
- Software crashes

⇒ Default MPI response: abort application

Resilience for the Exa-Age

Hard Faults

- Errors that trigger signals to the user
- Node, OS, network or process failure
- Software crashes

$\Rightarrow$ Default MPI response: abort application

Soft Faults

- No signal to user
- Faults unnoticed unless searched for
- Most common type: Silent Data Corruption (SDC)
  Errors in arithmetic operations, memory corruption, bit flips

1	0	1	0	1	1	1	0	1

1	0	1	0	0	1	1	0	1

Resilience for the Exa-Age

Hard Faults

- Errors that trigger signals to the user
- Node, OS, network or process failure
- Software crashes

⇒ Default MPI response: abort application

Soft Faults

- No signal to user
- Faults unnoticed unless searched for
- Most common type: Silent Data Corruption (SDC)
  Errors in arithmetic operations, memory corruption, bit flips

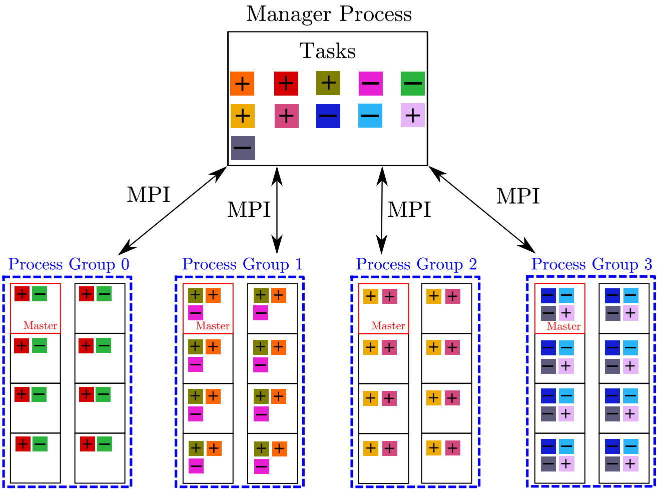
1	0	1	0	1	1	1	0	1

1	0	1	0	0	1	1	0	1

Assumption: checkpoint/restart infeasible ⇒ ABFT

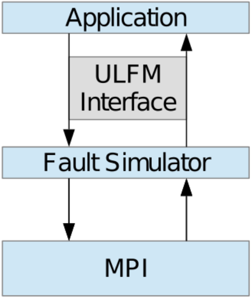
Communication Scheme

Master-worker model

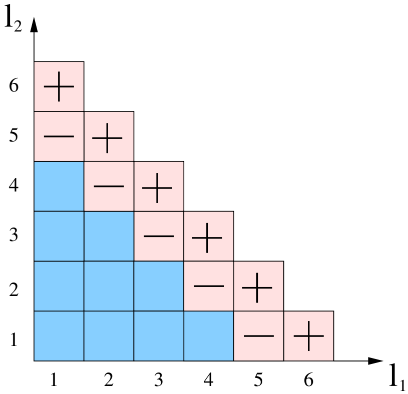


Software Stack

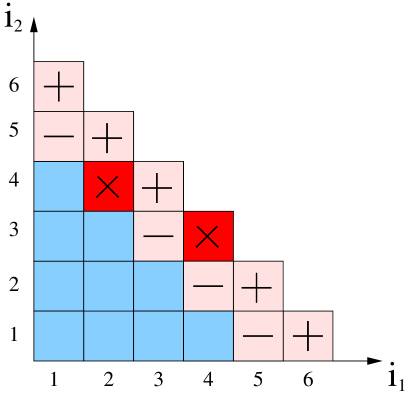
- Fault simulation layer
- Implements interface of ULFM plus `kill_me()` functionality



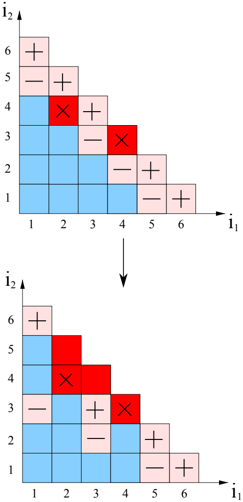
2D Example



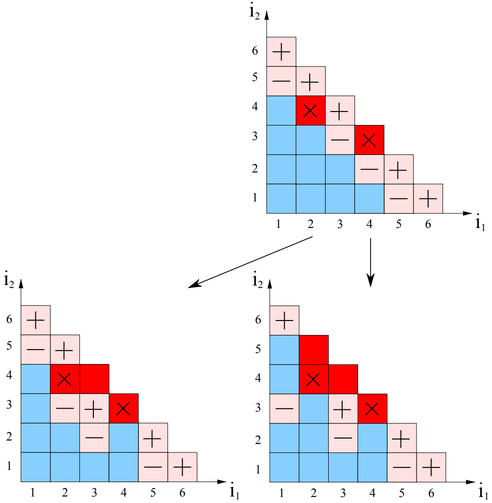
2D Example



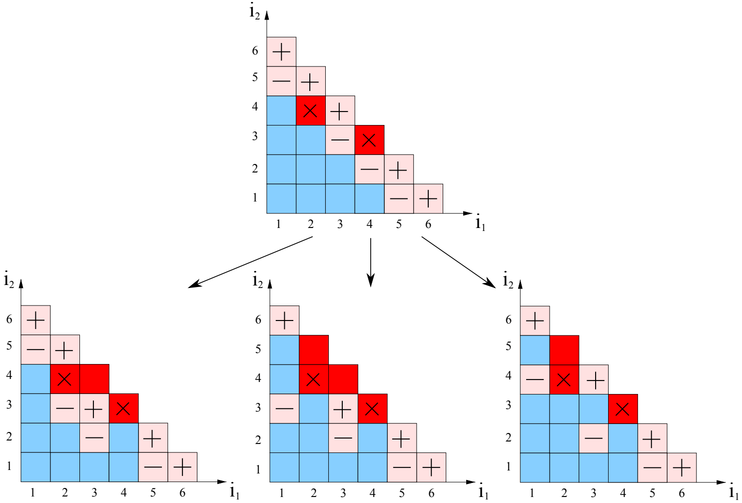
GCP: 2D Example



GCP: 2D Example

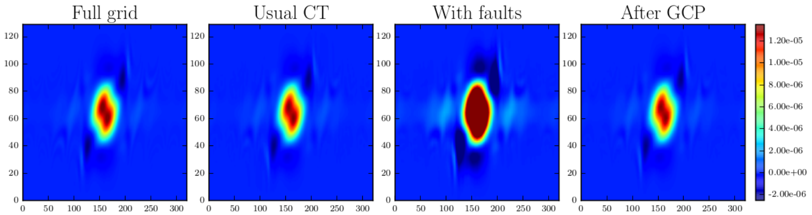


GCP: 2D Example



Results Using GENE

Example:

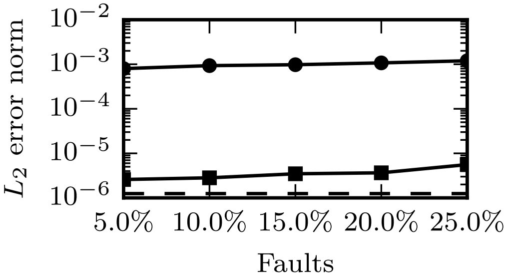


Good reconstruction (visual inspection)

Results Using GENE (2)

Small (reduced) problem

- 4D: $x, z, \mu, v_{||}$
- $\vec{l}_{\min} = [2, 3, 2, 4]$, $\vec{l} = [6, 7, 6, 8] \Rightarrow$ 69 combination grids

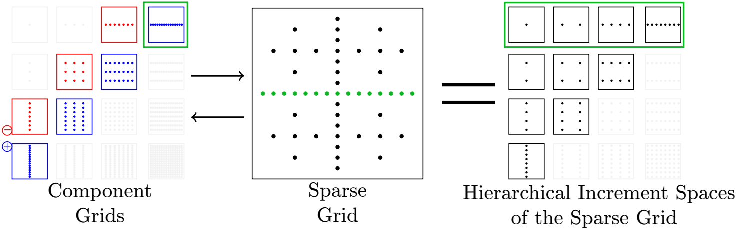


Excellent recovery properties!

Silent/Soft Faults

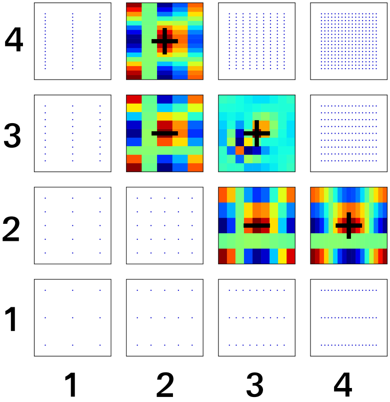
Exploit hierarchical approach

- Similar discretizations lead to similar results
- Exploit redundancy and hierarchical representation to check for faults
- Detection of outliers possible
- Direct integration into communication schemes possible (Subspace Reduce)



SDC Check: Compare Pairs of Solutions

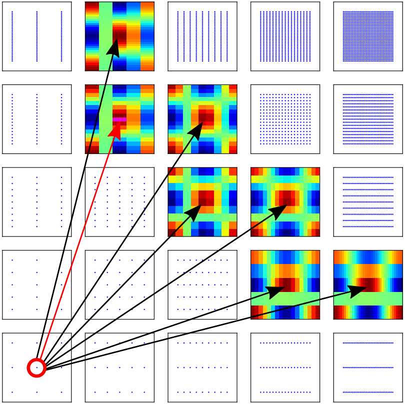
- Similar discretizations should lead to similar results



Pair	$\hat{\beta}^{(\underline{s},\underline{t})}$
(2, 4) (2, 3)	3.98e-01
(3, 2) (2, 4)	1.11e+00
(4, 2) (2, 3)	1.11e+00
(3, 2) (2, 3)	6.32e-01
(3, 3) (4, 2)	**9.85e+05**
(3, 3) (2, 3)	**1.07e+06**
(3, 2) (4, 2)	3.98e-01
(2, 4) (4, 2)	1.27e+00
(3, 2) (3, 3)	**1.07e+06**
(2, 4) (3, 3)	**9.85e+05**

$$\hat{\beta}^{(\underline{s},\underline{t})} := \max_{\underline{l}\leq\underline{s}\wedge\underline{t}} \max_{\underline{j}\in\mathcal{I}_{\underline{l}}} \frac{\left|\alpha^{(\underline{t})}_{\underline{l},\underline{j}} - \alpha^{(\underline{s})}_{\underline{l},\underline{j}}\right|}{\min\left\{\left|\alpha^{(\underline{t})}_{\underline{l},\underline{j}}\right|, \left|\alpha^{(\underline{s})}_{\underline{l},\underline{j}}\right|\right\}}$$

SDC Check: Outlier detection



$$\bar{u}(0,0) = [1.002, \mathbf{5.356}, 0.998, 1.002, 1.001, 1.001, .999]$$

2D Example

Advection equation

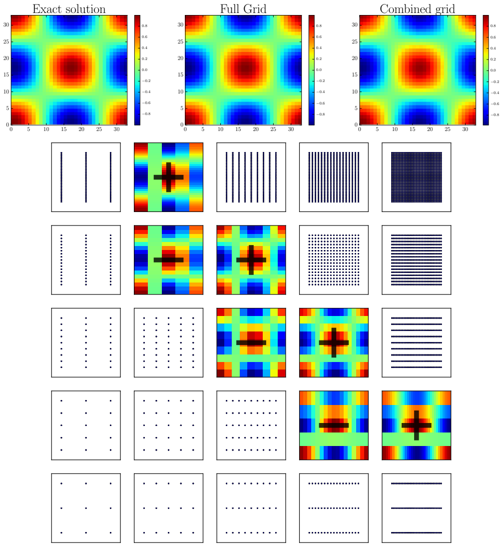
$$\frac{\partial u}{\partial t} + c_x \frac{\partial u}{\partial x} + c_y \frac{\partial u}{\partial x} = 0 \quad \Omega = [0,1]^2$$

- Periodic boundary conditions
- Constant advection velocities c_x, c_y
- Initial condition $u(x, y, t = 0) = \sin(2\pi x)\sin(2\pi y)$
- Lax-Wendroff scheme (2nd order space + time)
- Error/solution at $t = 0.5$ compared to analytical solution

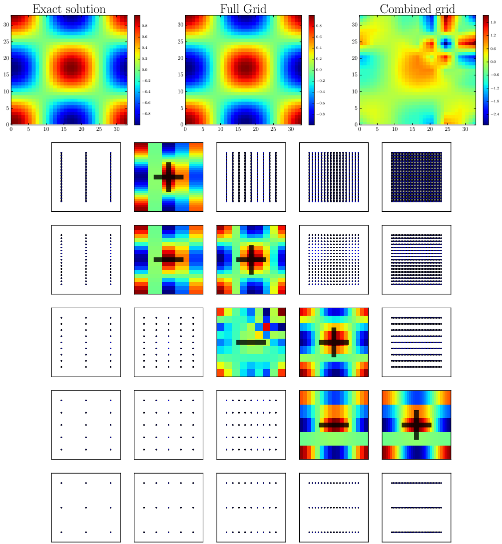
$$u(x, y, t) = \sin(2\pi(x - c_x t))\sin(2\pi(y - c_Y t))$$

- Corruption of one single data point in initial condition

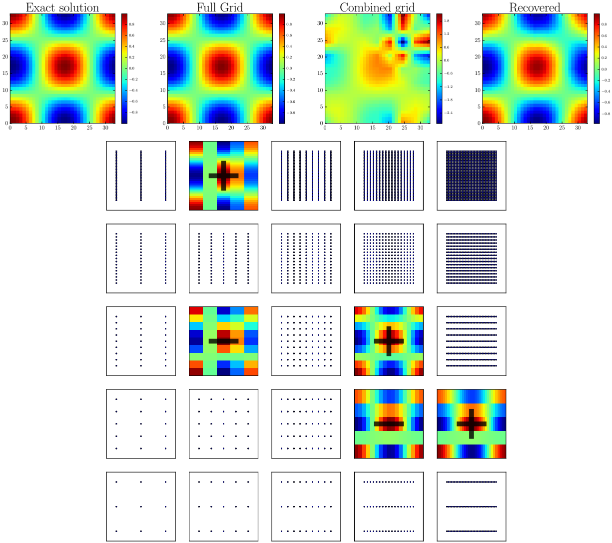
2D Example



2D Example

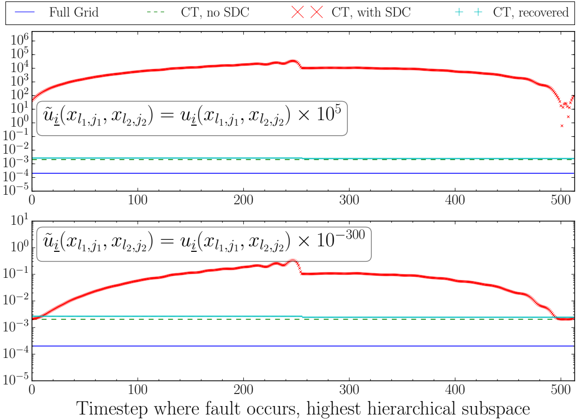


2D Example



2D Example: Simulated Soft Faults

- Inserting one soft fault
- Measuring L2-error at the end



Higher-D: Advection-Diffusion Equation

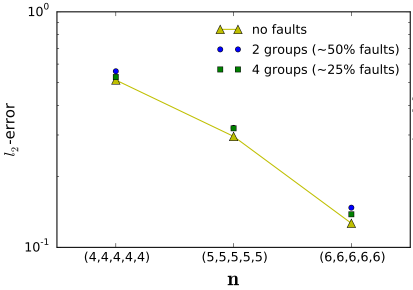
$$
\begin{aligned}
\partial_t u - \Delta u + \vec{a} \cdot \nabla u &= f && \text{in } \Omega \times [0, T) \\
u(\cdot, t) &= 0 && \text{in } \partial\Omega \\
u(\cdot, 0) &= u_0 && \text{in } \Omega
\end{aligned}
$$

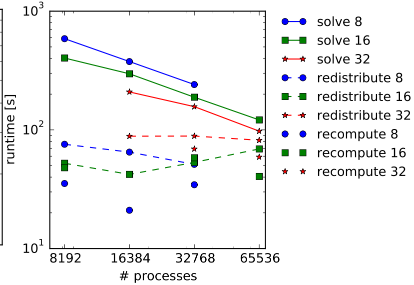
$$
\Omega = [0, 1]^d, \vec{a} = (1, \ldots, 1)^T, u_0 = e^{-100 \sum_{i=1}^{d} (x_i - 0.5)^2}
$$

- Implemented in DUNE-PDELab, joint thesis with Steffen Müthing
- FVM, explicit time integration

Results

- Fault in second time step
- Relative error w.r.t. full-grid solution ($\mathbf{n} = \mathbf{11}$ in 2D, $\mathbf{n} = \mathbf{7}$ in 5D)
- Computations on Hazel Hen (HLRS)
- 5D:

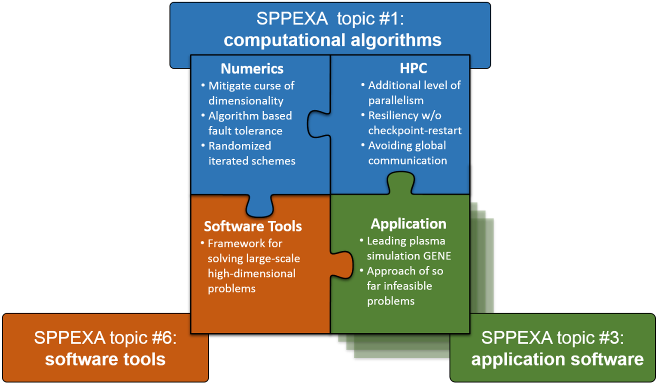




Again: excellent recovery properties!

EXAHD – Project Goals

What I did not have time to talk about. . .



More . . .

FT on first level

- Hard faults
- Development of `libSpina`
- Integrate standard MPI
- Spare processes to replace faulty ones
- Make use of domain knowledge to reduce checkpointing data

More . . .

FT on first level

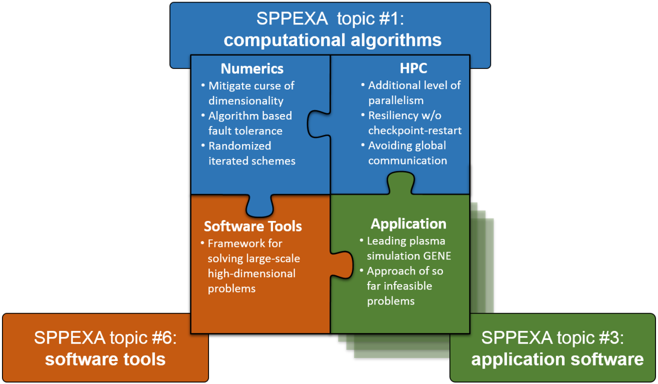
- Hard faults
- Development of `libSpina`
- Integrate standard MPI
- Spare processes to replace faulty ones
- Make use of domain knowledge to reduce checkpointing data

Higher-Order Interpolation Schemes

- Necessary due to peculiarities of application code
- 4th order FD *and* FFT
- Basis sets for each dimension independently

EXAHD – Project Goals

Activities, Collaborations, Outreach



Activities, Collaborations, Outreach

- Parallel multigrid for higher-dimensional PDEs on anisotropic grids in DUNE (with EXA-DUNE)
- IPAM workshop “Big Data meets Computation”
- Conference “Sparse Grids and Applications” (Miami)
- Resilience workshop @ Euro-Par
- Mutual exchange stays with international partners
- Several student thesis

Mario Heene, Alfredo Parra Hinojosa, Hans-Joachim Bungartz, and Dirk Pflüger.
A massively-parallel, fault-tolerant solver for time-dependent PDEs in high dimensions.
In *Euro-Par 2016*, Grenoble, June 2016.
Accepted.

Alfredo Parra Hinojosa, Brendan Harding, Hegland Markus, and Hans-Joachim Bungartz.
Handling silent data corruption with the sparse grid combination technique.
In *Proceedings of the SPPEXA Symposium*, Lecture Notes in Computational Science and Engineering. Springer-Verlag, February 2016.

Philipp Hupp, Mario Heene, Riko Jacob, and Dirk Pflüger.
Global communication schemes for the numerical solution of high-dimensional PDEs.
Parallel Computing, 52:78 – 105, 2016.

Mario Heene and Dirk Pflüger.
Scalable algorithms for the solution of higher-dimensional PDEs.
In *Software for Exascale Computing-SPPEXA 2013-2015*, pages 165–186. Springer International Publishing, 2016.

Alfredo Parra Hinojosa, Christoph Kowitz, Mario Heene, Dirk Pflüger, and Hans-Joachim Bungartz.
Towards a fault-tolerant, scalable implementation of GENE.
In *Recent Trends in Computational Engineering-CE2014*, pages 47–65. Springer International Publishing, 2015.

Dirk Pflüger, Hans-Joachim Bungartz, Michael Griebel, Frank Jenko, Tilman Dannert, Mario Heene, Alfredo Parra Hinojosa, Christoph Kowitz, and Peter Zaspel.
EXAHD: An exa-scalable two-level sparse grid approach for higher-dimensional problems in plasma physics and beyond.
In *Euro-Par 2014 Workshop, Part II*, volume 8806 of *Lecture Notes in Computer Science*, pages 566–577. Springer-Verlag, December 2014.