ESM Algorithmic Acceleration for petascale systems

by

Amik St-Cyr

Computational and Information Systems Laboratory

NCAR

Outline

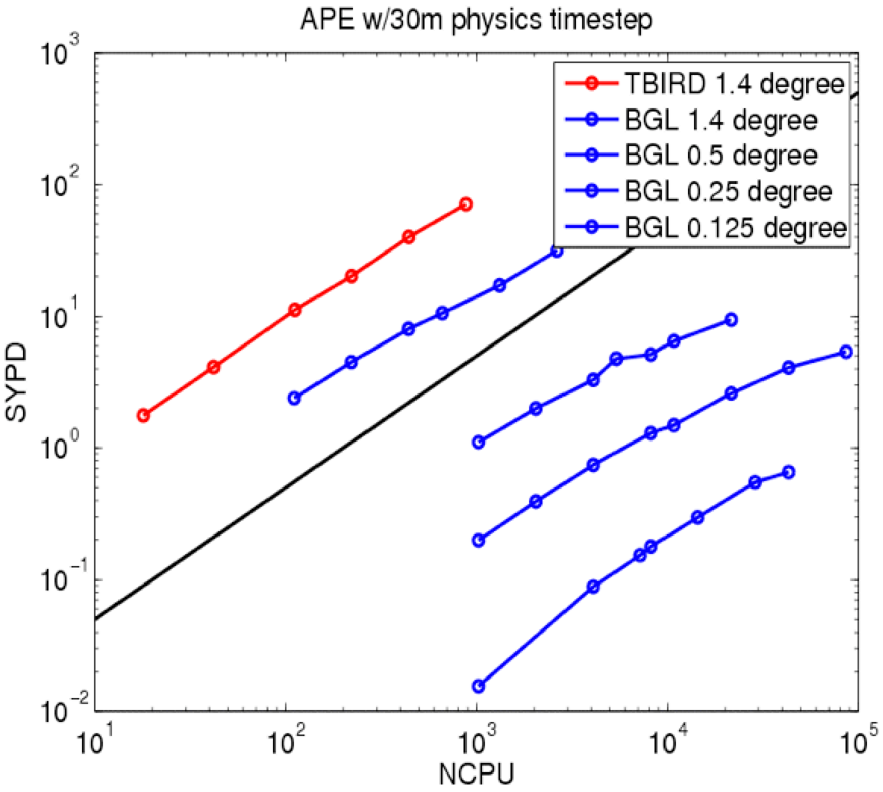
- The problem
- Algorithms in the geosciences
- Examples of "ESM" acceleration
  - Time-stepping
  - Solvers
  - AMR
- ESM algorithmic "vision" and current efforts
- Possible future directions

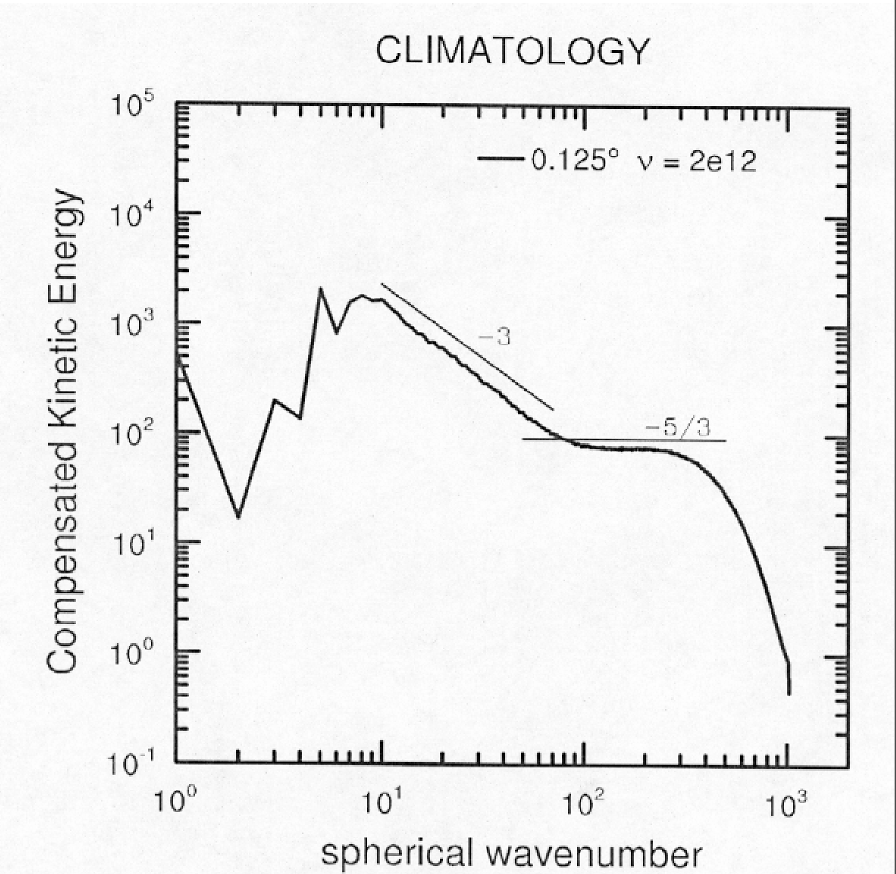
Goal

Goal

- *Establish a roadmap for most-productive use of petascale computing systems for improving our knowledge of important geophysical dynamical processes.*

Where are we now?



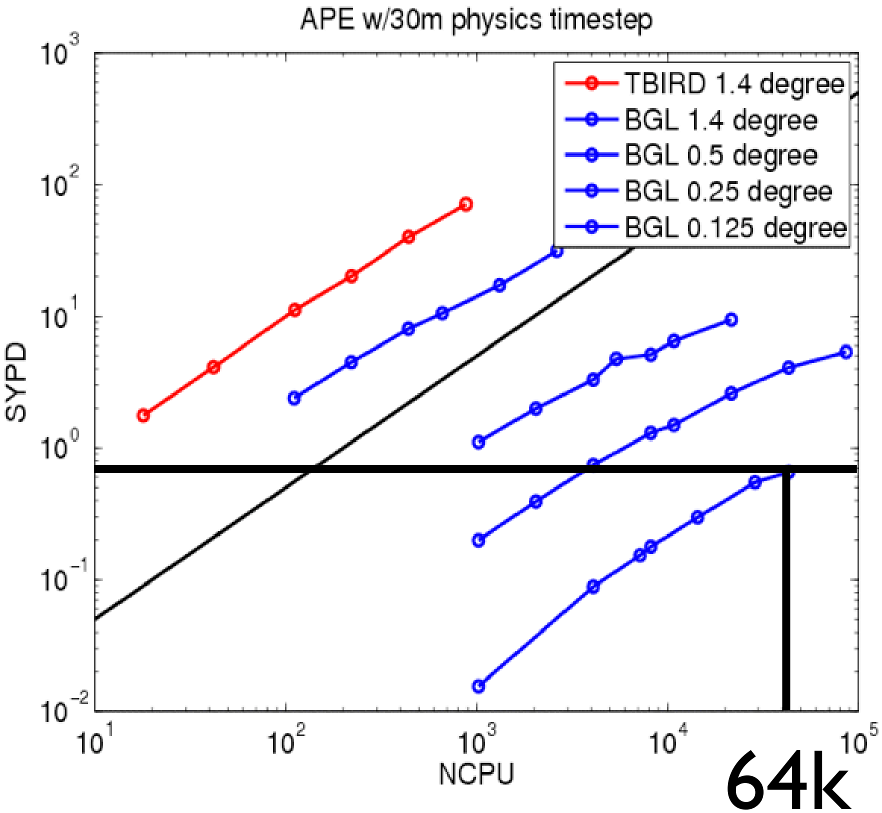


- HOMME-APE CAM 3.0 physics
- Hydrostatic formulation
- 13km resolution (global)
- (Also global WRF: nonhydrostatic)

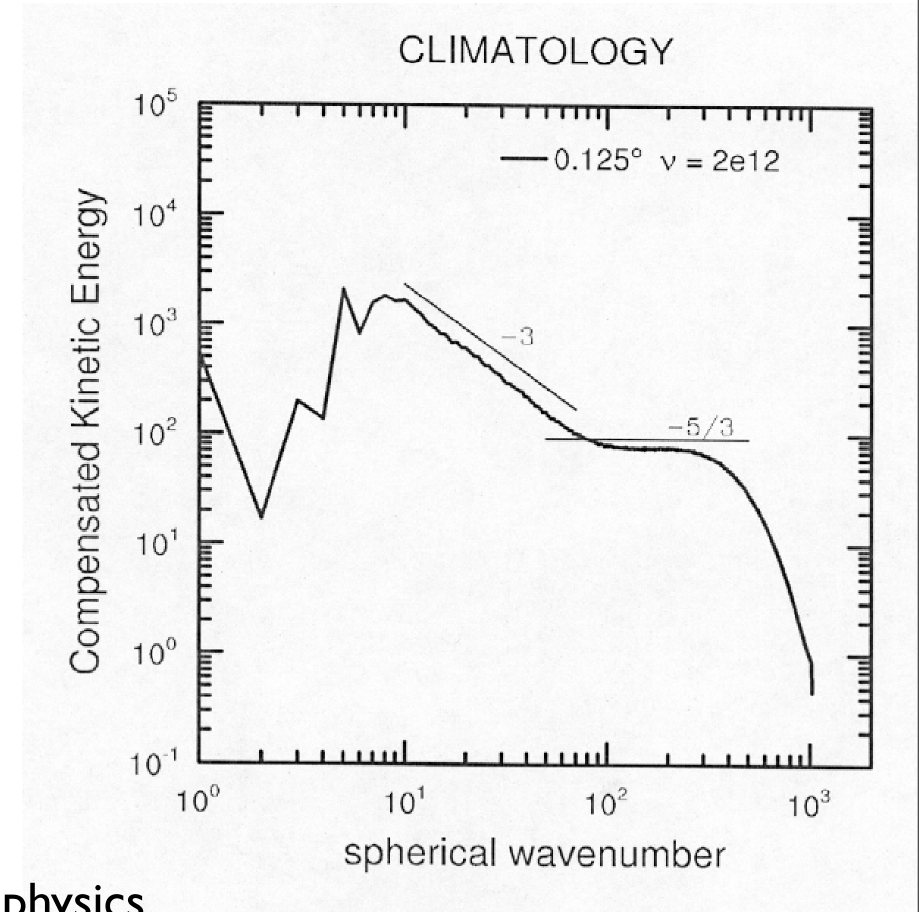
LLNL BG/L

Courtesy M. Taylor and J.P. Edwards

Where are we now?



.66 ypd



- HOMME-APE CAM 3.0 physics
- Hydrostatic formulation
- 13km resolution (global)
- (Also global WRF: nonhydrostatic)

LLNL BG/L

Courtesy M. Taylor and J.P. Edwards

Where are we now?

- Explicit / split-explicit time-stepping
- “Uniform” or structured meshes
- Science this enables:
    - “Seasonal” climate modeling
    - Next step “Weekly” climate modeling?
    - Next step “Daily” climate modeling??

Where are we now?

- Explicit / split-explicit time-stepping
- "Uniform" or structured meshes
- Science this enables:
  - "Seasonal" climate modeling
  - Next step "Weekly" climate modeling?
  - Next step "Daily" climate modeling??

Stuck in a **weak scaling** paradigm

Where are we now?

- Idealized model (no I/O, explicit ts, structured mesh...)
- Producing scientifically significant integration rates for climate (5 sypd @ 4km) will require increase in compute horsepower of the order of 10^4 to 10^6 !!!
- Moore's Law: will happen in 13 to 20 years...
- <u>What can we do?</u>

Improve algorithms used in geosciences...

Algorithms in geosciences

- Time-integration: explicit, split-explicit, semi-implicit, implicit, LMM, RK, Multi-rate, IMEX ...
- Space integration: SEM, DGM, FDM, FVM, RBF,...
- Elliptic problems: direct methods, iterative methods: KSP, multi-grid, preconditioning ...
- Optimization techniques

Time-stepping

- Assumptions for HOM:
  - Elements: E
  - Points/Elements: N^d
  - Tensor products cost: N^{d+1}
  - Matrix vector cost: EN^{d+1}
  - Fast diagonalization where possible

Time-stepping

- Potential for acceleration:

Method:	Cost:
Explicit	$O(k_e E N^{d+1})$
Semi-implicit	$O(k_s[EN^{d+1} + k_k E(N/s)^{d+1} + E(N/s)^{d+1}])$
Fully implicit	$O(k_i[REN^{d+1} + EN^{3d}\frac{\pi^4}{90}])$

Time-stepping

- Potential for acceleration:

Method:	Acceleration:
Explicit	$a \leq \frac{1}{EN^2}$
Semi-implicit	$a \leq \frac{s^{d+1}}{\tilde{c}[(1+s^{d+1})/m+1]}$
Fully implicit	$a \leq \frac{\Delta t_i}{\frac{\tilde{c}}{\nu}[R/N^2+1.2N^{2d-3}]}$

(For gravity waves) N=300 in 2D and N=8 in 3D

Examples of acceleration

SISL

- Semi-Implicit + Semi-Lagrangian
- Gravity waves and advective time-scale
- Proposed by A. Robert (81)
- Parallel issues in its classical version...
- Use idea of Maday et al. (90)
- N-L version for sw: (A and Thomas 05)
- Acceleration is 4 wrt explicit version
- Problem solved! ...

SISL

ODE resulting from SEM discretization (MOL)

$$\frac{du(t)}{dt} = S(u(t)) + F(u(t)), \quad t \in [0, T]$$

with initial condition $u(0) = u_0$

<u>Problem</u>: find integrating factor, $Q_S^{t^*}(t)$ such that $Q_S^{t^*}(t^*) = I$,

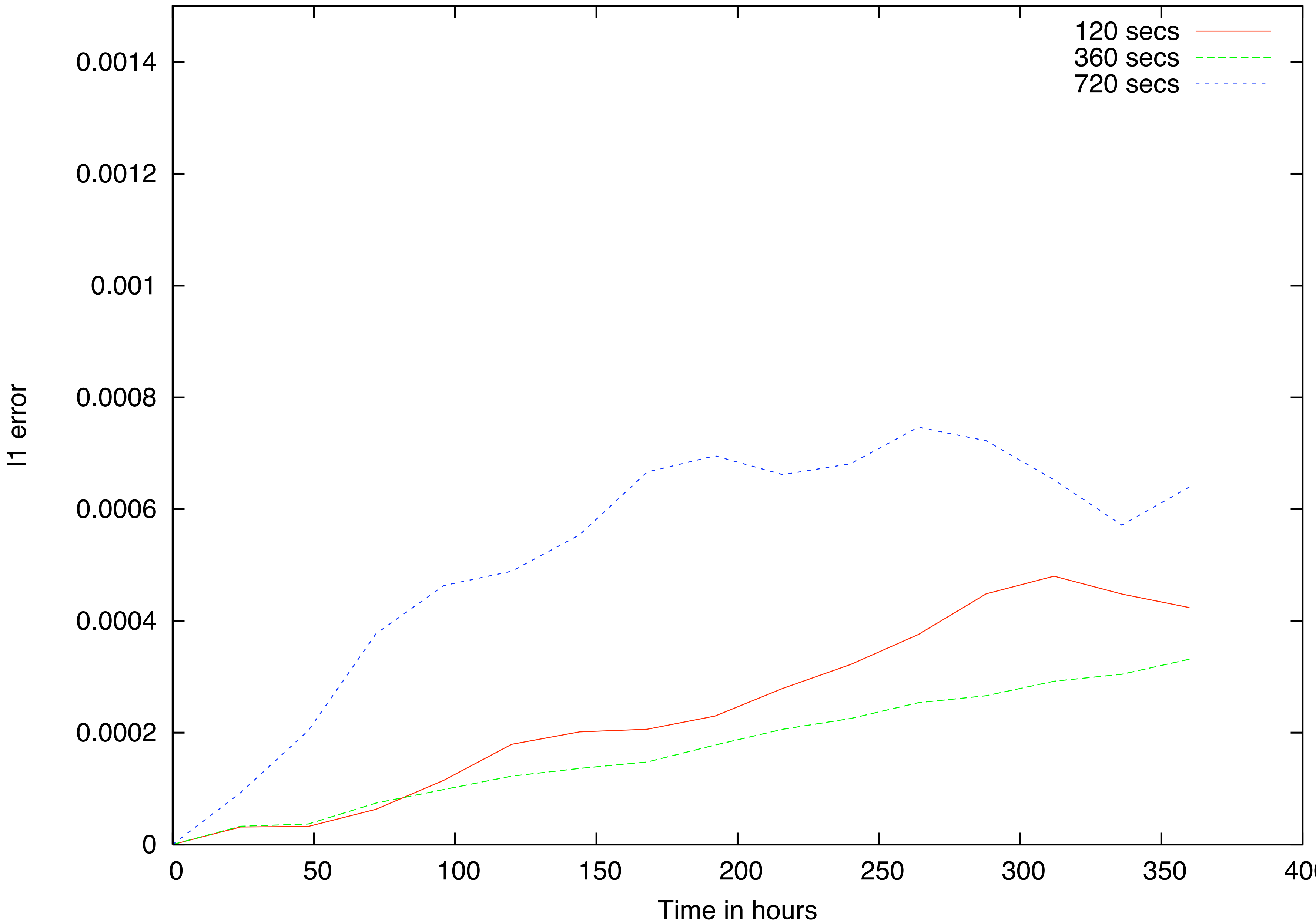
$$\frac{d}{dt} Q_S^{t^*}(t) \cdot u = Q_S^{t^*}(t) \cdot F(u).$$

To find the action of $Q_S^{t^*}(t)$ solve:

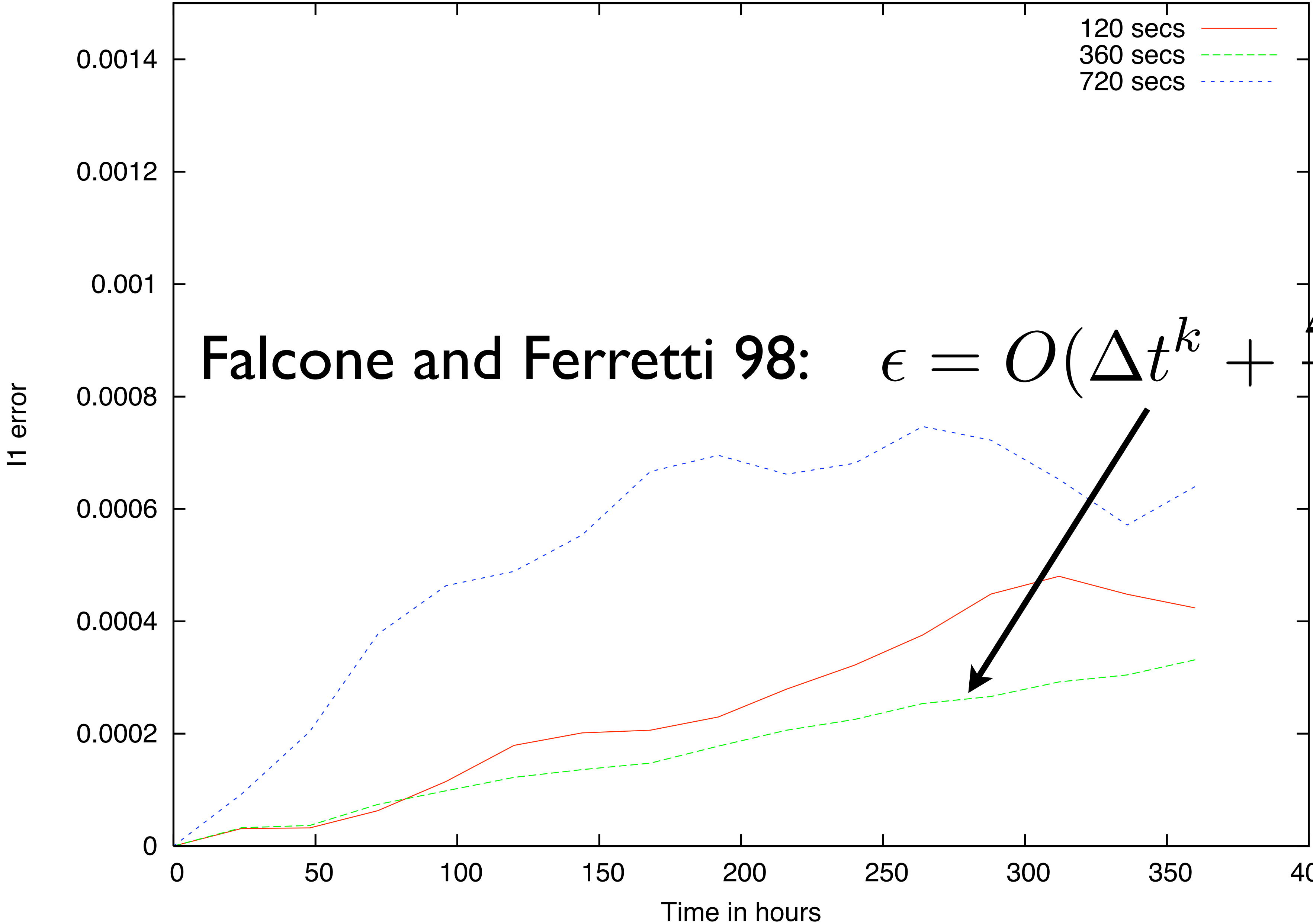
$$\frac{dv^{(t^*,t)}(s)}{ds} = S(v^{(t^*,t)}), \quad 0 \leq s \leq t - t^*$$

with initial condition $v^{(t^*,t)}(0) = u(t)$

Comparison with reference solution from NCAR pseudo spectral core



Comparison with reference solution from NCAR pseudo spectral core



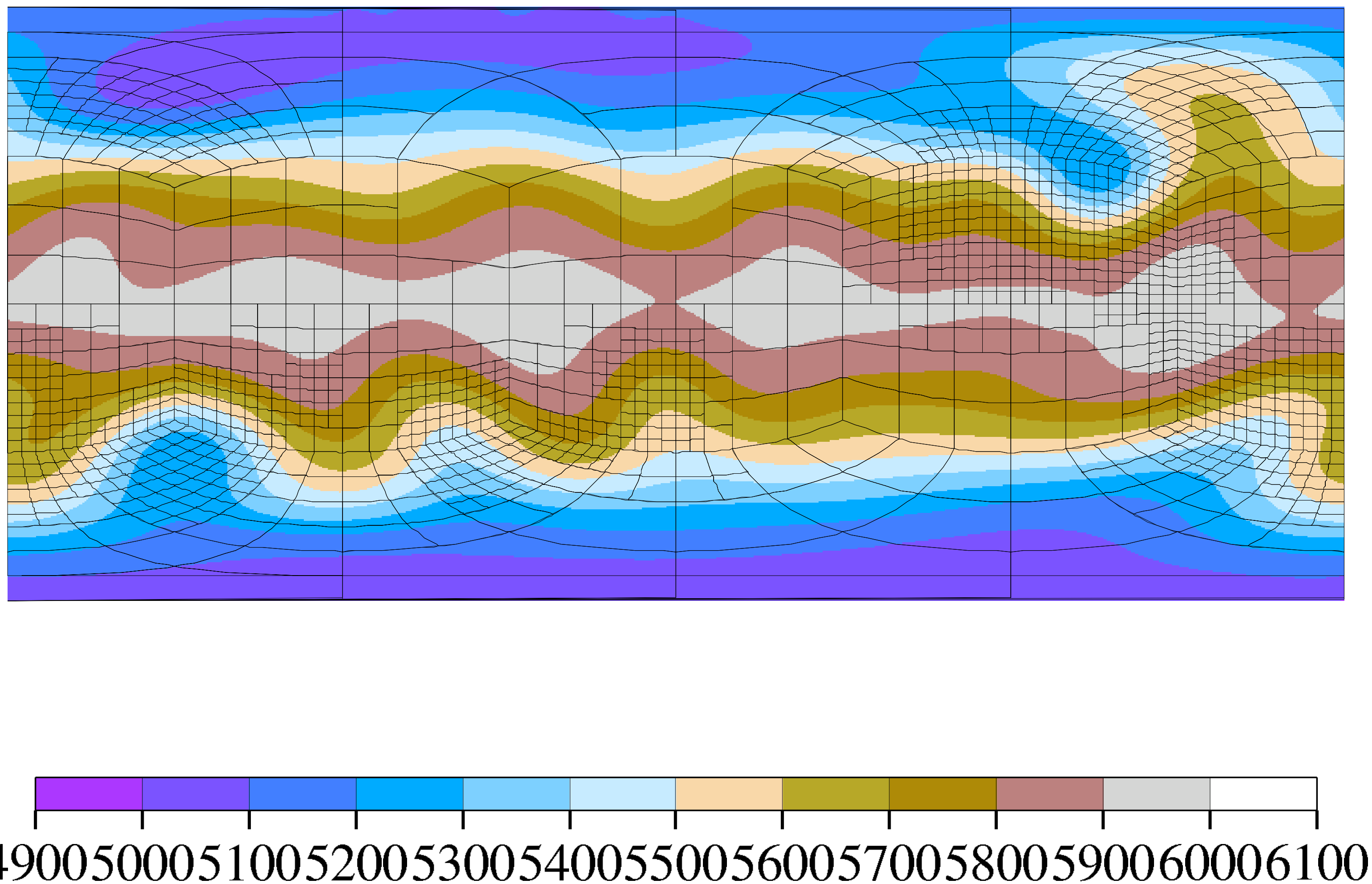
Falcone and Ferretti 98: $\epsilon = O(\Delta t^k + \frac{\Delta x^p}{\Delta t})$

Why is this important?

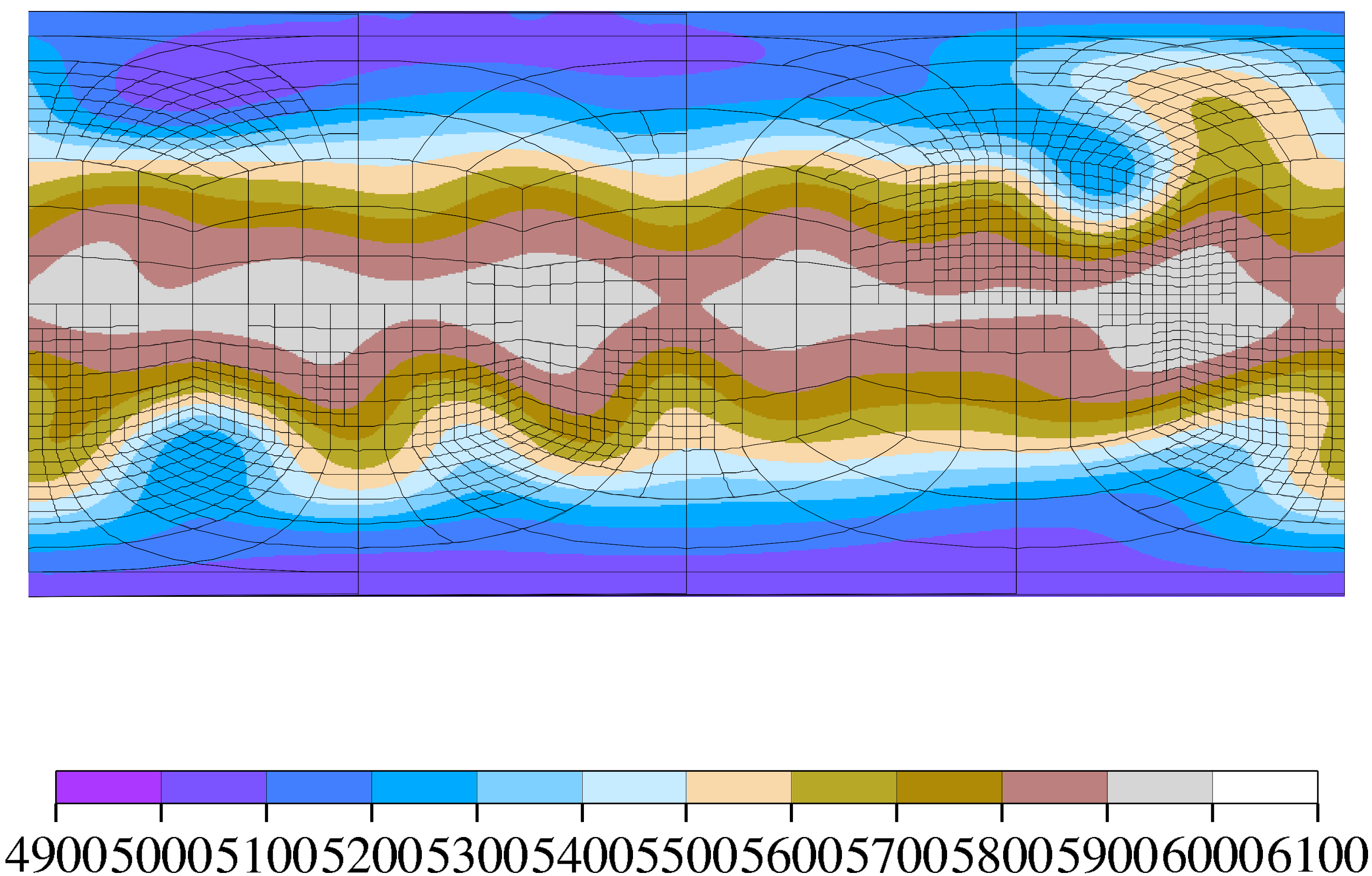
We want to model all scales in time and space...

Combining SISL with AMR leads to a contradiction!

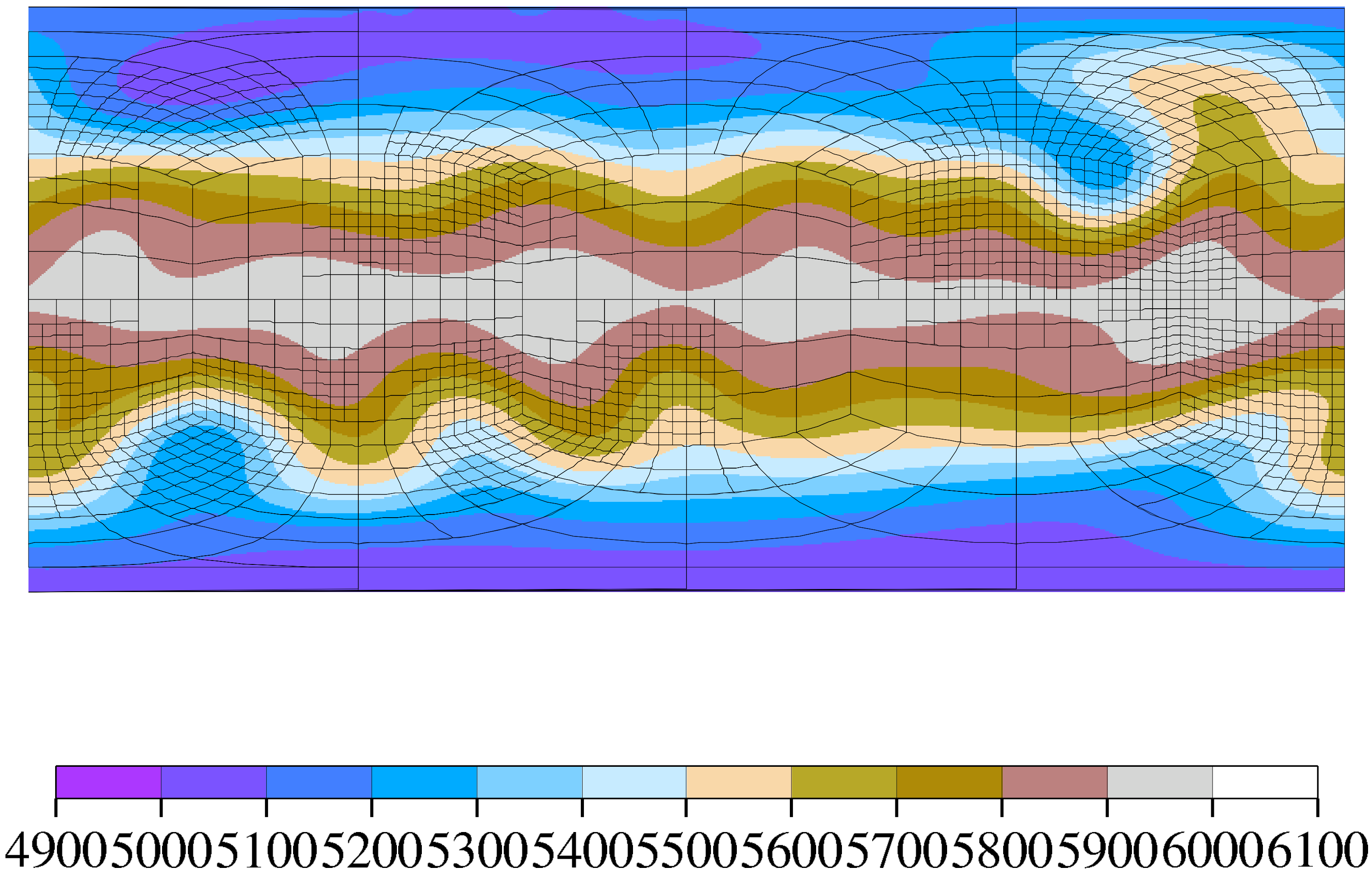
dt=120s


4900 5000 5100 5200 5300 5400 5500 5600 5700 5800 5900 6000 6100

dt=360s

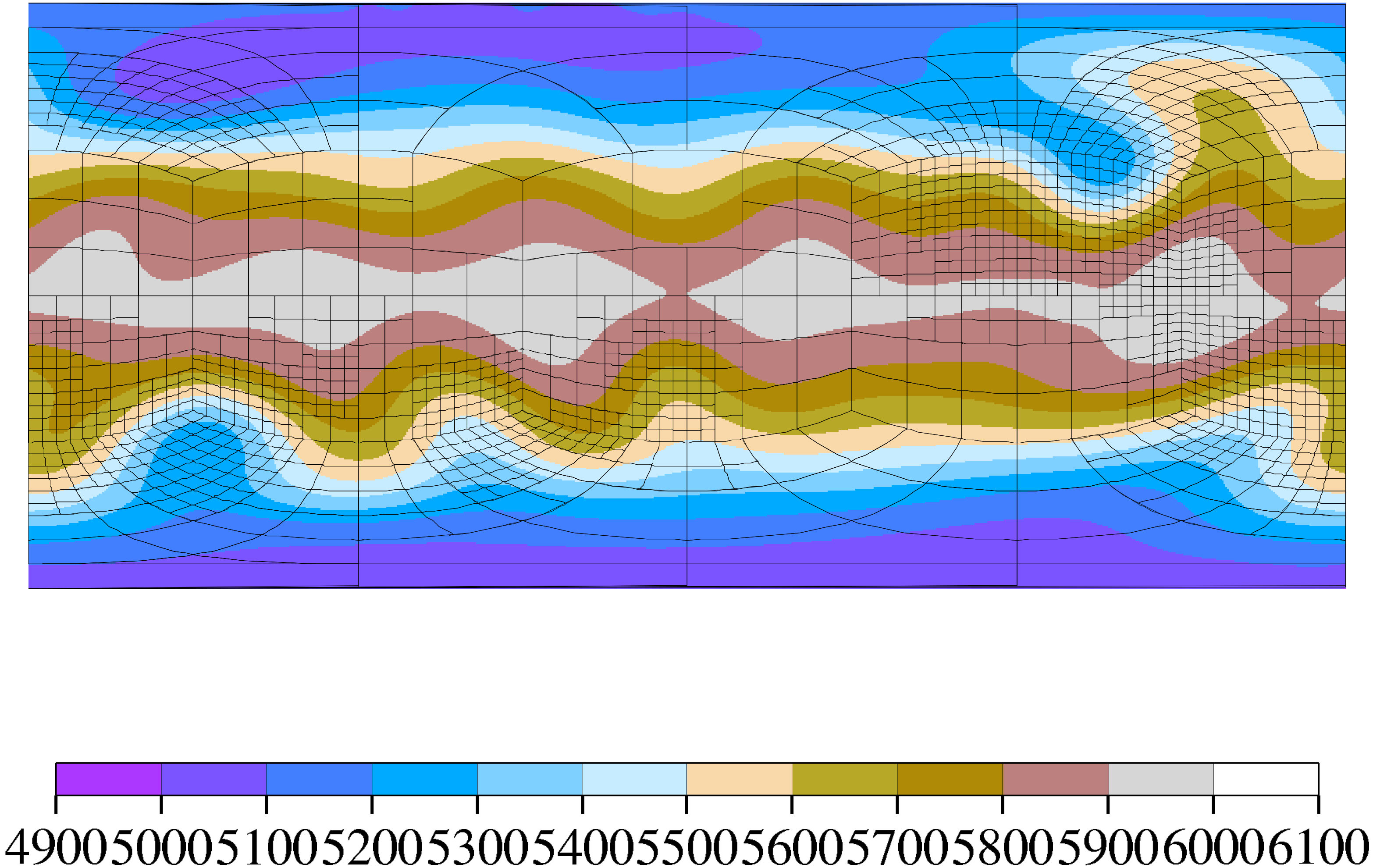

4900
5000
5100
5200
5300
5400
5500
5600
5700
5800
5900
6000
6100

dt=720s



4900
5000
5100
5200
5300
5400
5500
5600
5700
5800
5900
6000
6100

dt=120s



Conclusion: not a true multiscale algorithm ...

Fully implicit

- DG in space for Euler equations: WRF form
- New Rosenbrock W-method
- No non-linear cycles (Newton)
- No Jacobians: Jacobian free
- Low Mach preconditioning
- Element block Jacobi
- Results on benchmark tests
- Acceleration: 3 to 45 wrt to explicit version

Effects of low Mach

solver tolerance ~ 1E-6, (Nx,Nz)=(16,8), p=7, 180 meters resolution (approx.)

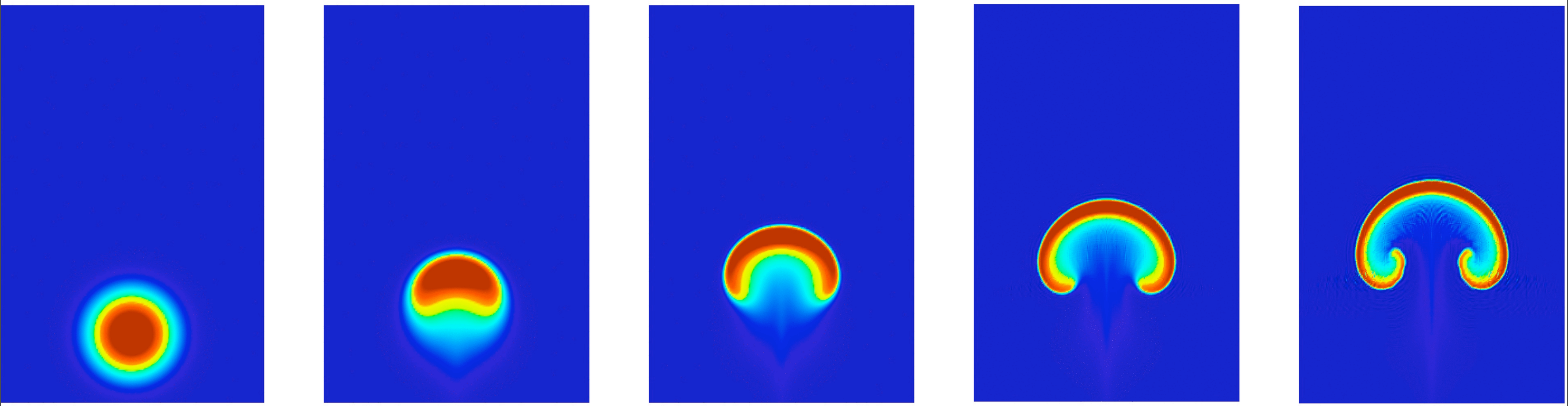
time	W LM	accel	WO LM	accel
1.0s	30	3.2	33	2.8
2.0s	36	5.1	45	4.1
10.0s	69	13.5	103	9.1
50.0s	207	22.7	493	10.2

"Wicker" Bubble: Wicker and Skamarock MWR02

Rising bubble

5 meters resolution, $p = 7$, $T_f = 600$ secs

Rising bubble



5 meters resolution, p =7, Tf = 600 secs

Inertia Gravity wave

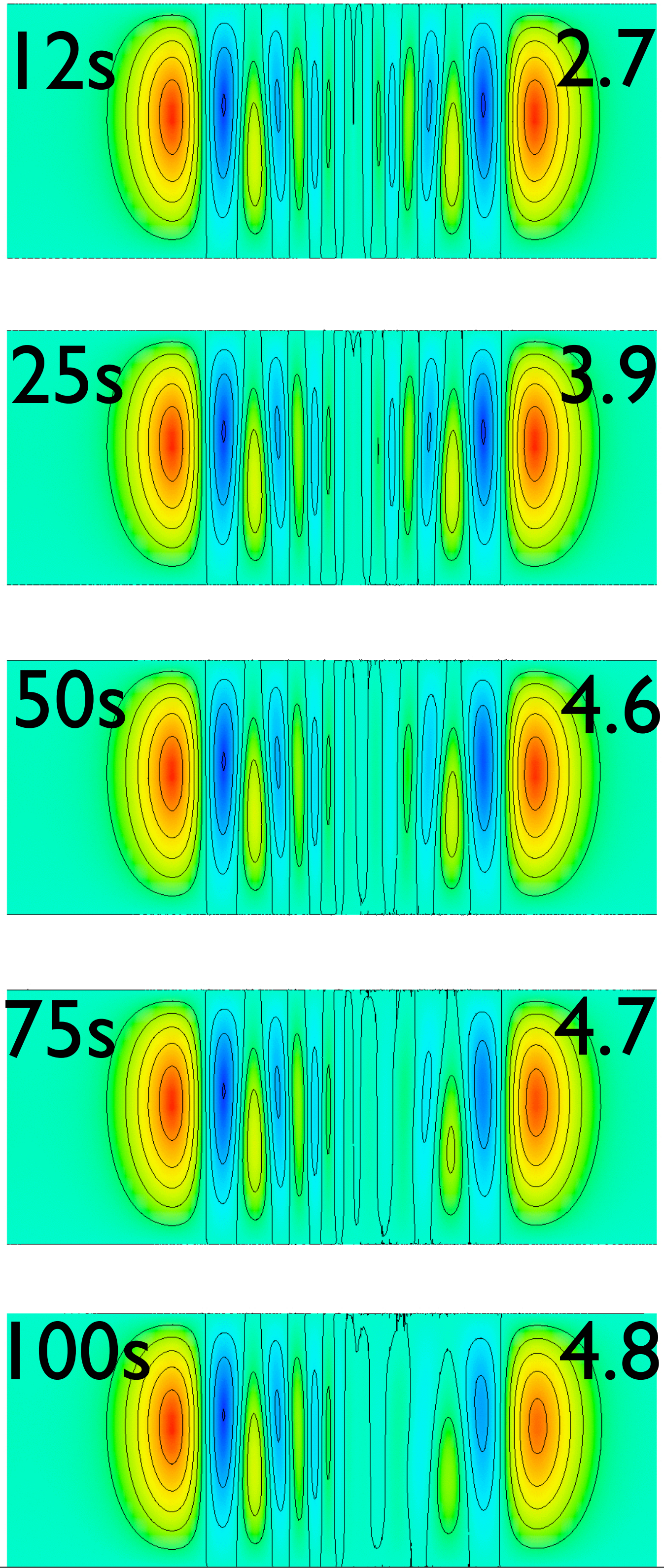
- Inertia gravity wave in channel + bg flow
- dx=dz=500m, poly order 8, nez=3, nex=90
- dt=12, 25, 50, 75, 100 seconds
- Accelerations: 2.7, 3.9, 4.6, 4.7, 4.8 wrt explicit
- 20 m/s to the right

Skamarock and Klemp 02

Inertia Gravit

- Inertia gravity wave in channel + bg flow
- dx=dz=500m, poly order 8, nez=3, nex=90
- dt=12, 25, 50, 75, 100 seconds
- Accelerations: 2.7, 3.9, 4.6, 4.7, 4.8 wrt explicit
- 20 m/s to the right

Skamarock and Klemp 02



Inertia Gravity wave

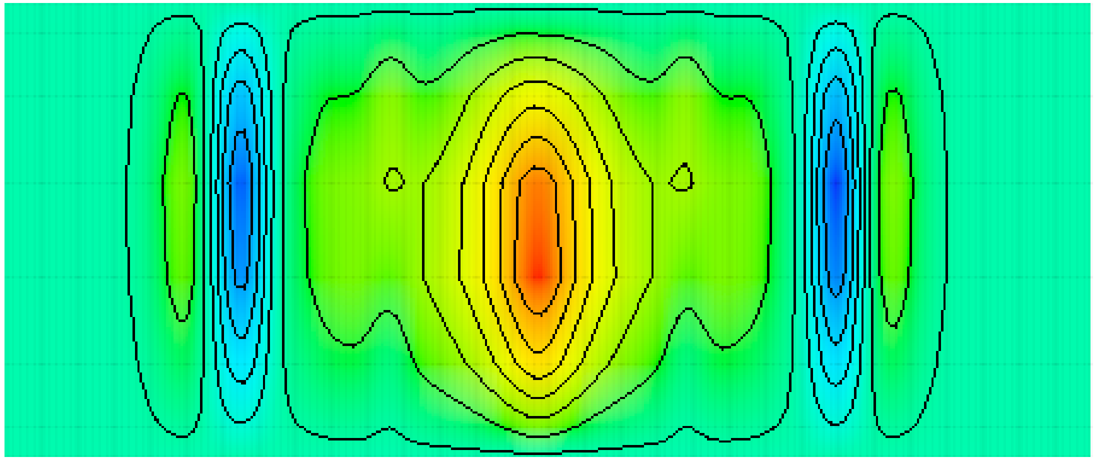
- Eady model (one more equation + Coriolis)
- Very thin channel (hydrostatic: shallow atm)
- 1 element in the vertical
- 600 in the horizontal (1km x 1km resolution)
- p=7
- accel > 45

Skamarock and Klemp 02

Inertia Gravity wave

- Eady model (one more equation + Coriolis)
- Very thin channel (hydrostatic: shallow atm)
- 1 element in the vertical
- 600 in the horizontal (1km x 1km resolution)
- p=7
- accel > 45

Skamarock and Klemp 02



Semi (or full) implicitness leads to a matrix to invert...

Solvers

- Storing full LU impossible:
- Iterative method unavoidable
- Multigrid: is $O(N)$
- Two-levels Schwarz is optimal
- If parabolic PDE no coarse solver needed (semi-implicit behaves this way!)
- For Laplacian high-order has a N^4 growth
- Non-overlapping optimized Schwarz

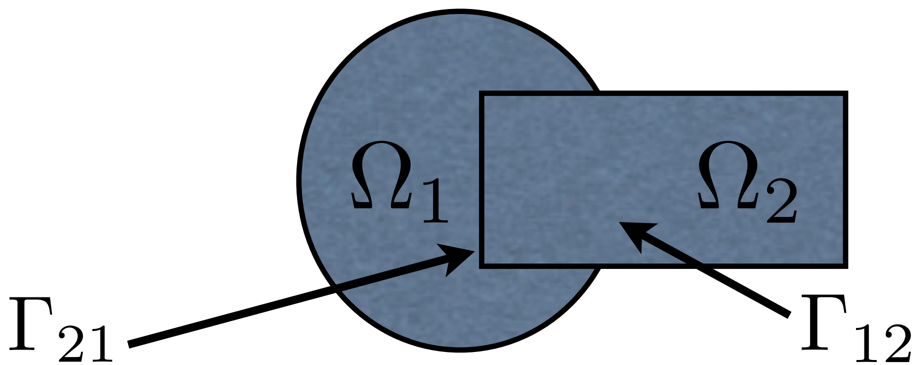
Classical Schwarz

Suppose we need to solve:

$$\mathcal{L}u = f \quad \text{in } \Omega, \qquad \mathcal{B}u = g \quad \text{on } \partial\Omega$$

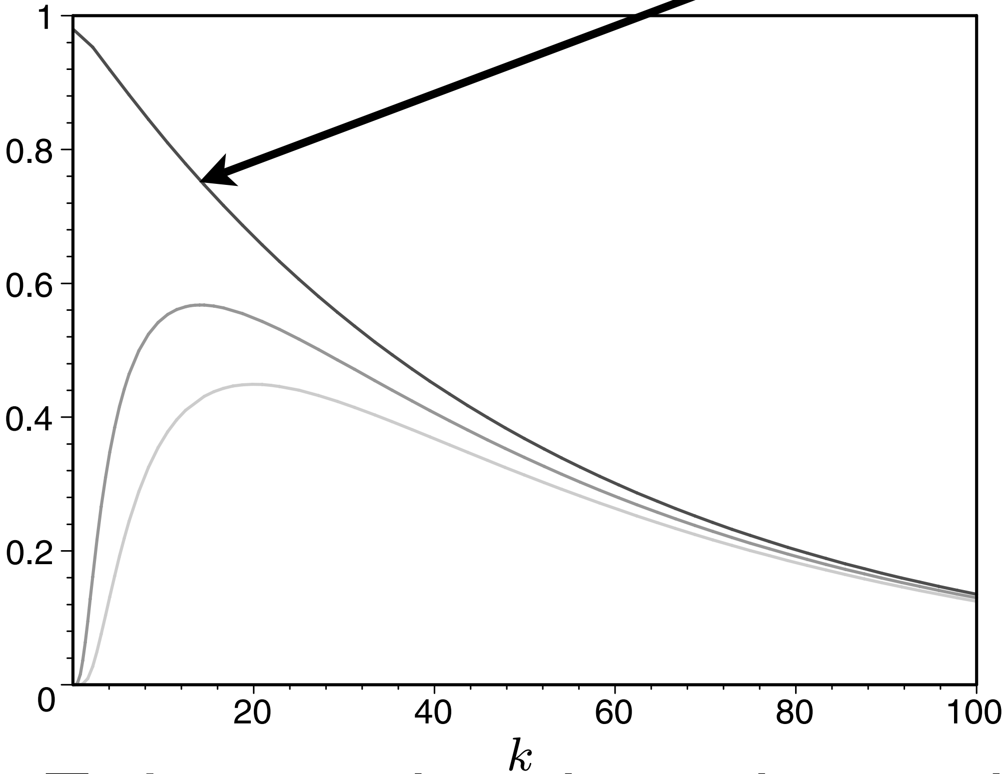
Partition the original domain into 2 domains:

$$\begin{aligned} \mathcal{L}u_1^{n+1} &= f & \text{in } \Omega_1, \qquad & \mathcal{L}u_2^{n+1} &= f & \text{in } \Omega_2, \\ \mathcal{B}(u_1^{n+1}) &= g & \text{on } \partial\Omega_1, \qquad & \mathcal{B}(u_2^{n+1}) &= g & \text{on } \partial\Omega_2, \\ u_1^{n+1} &= u_2^n & \text{on } \Gamma_{12}, \qquad & u_2^{n+1} &= u_1^n & \text{on } \Gamma_{21}. \end{aligned}$$

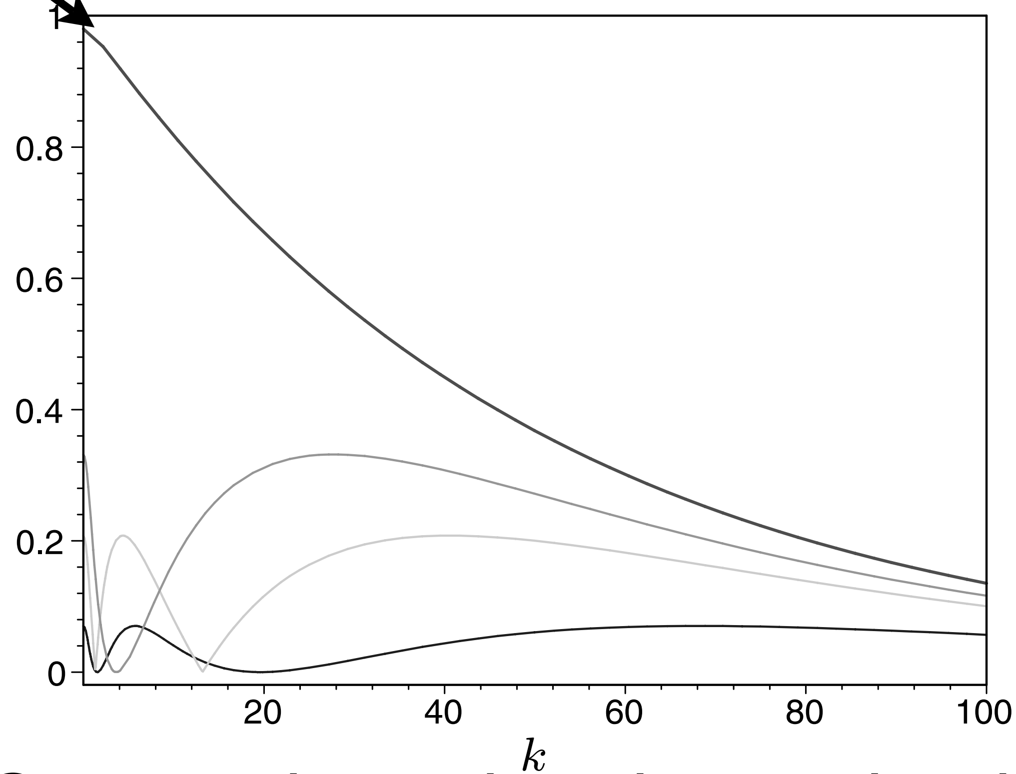


Convergence rates

Classical Schwarz



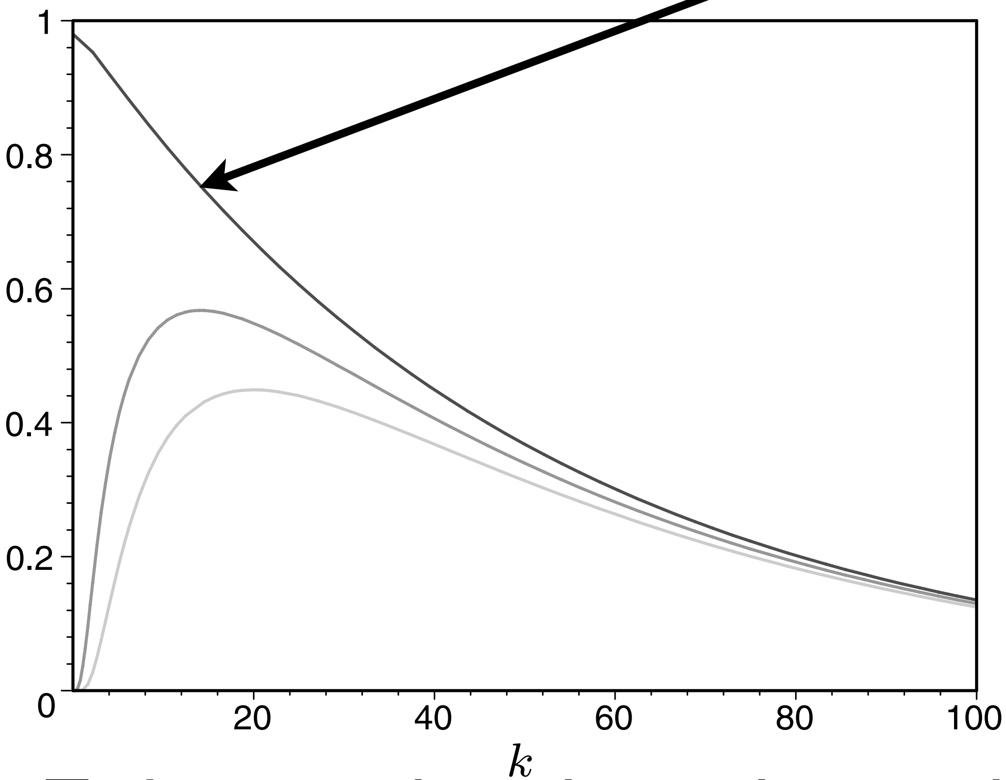
Taylor zeroth order and second order



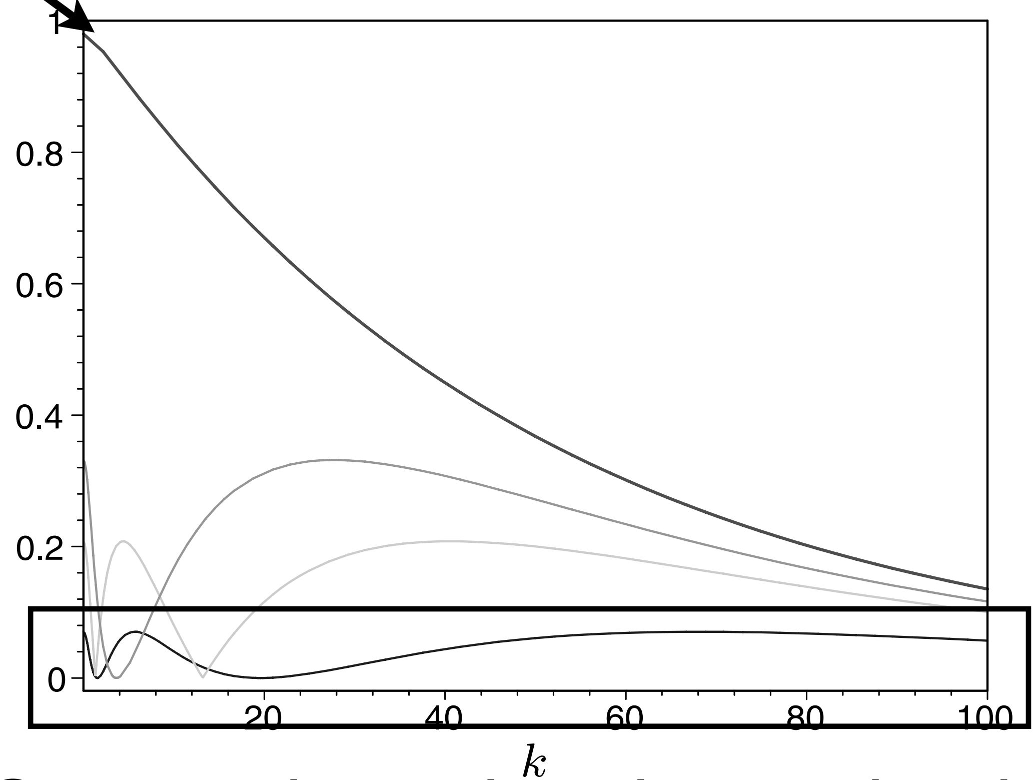
Optimized zeroth and second order with two-sided zeroth order

Convergence rates

Classical Schwarz

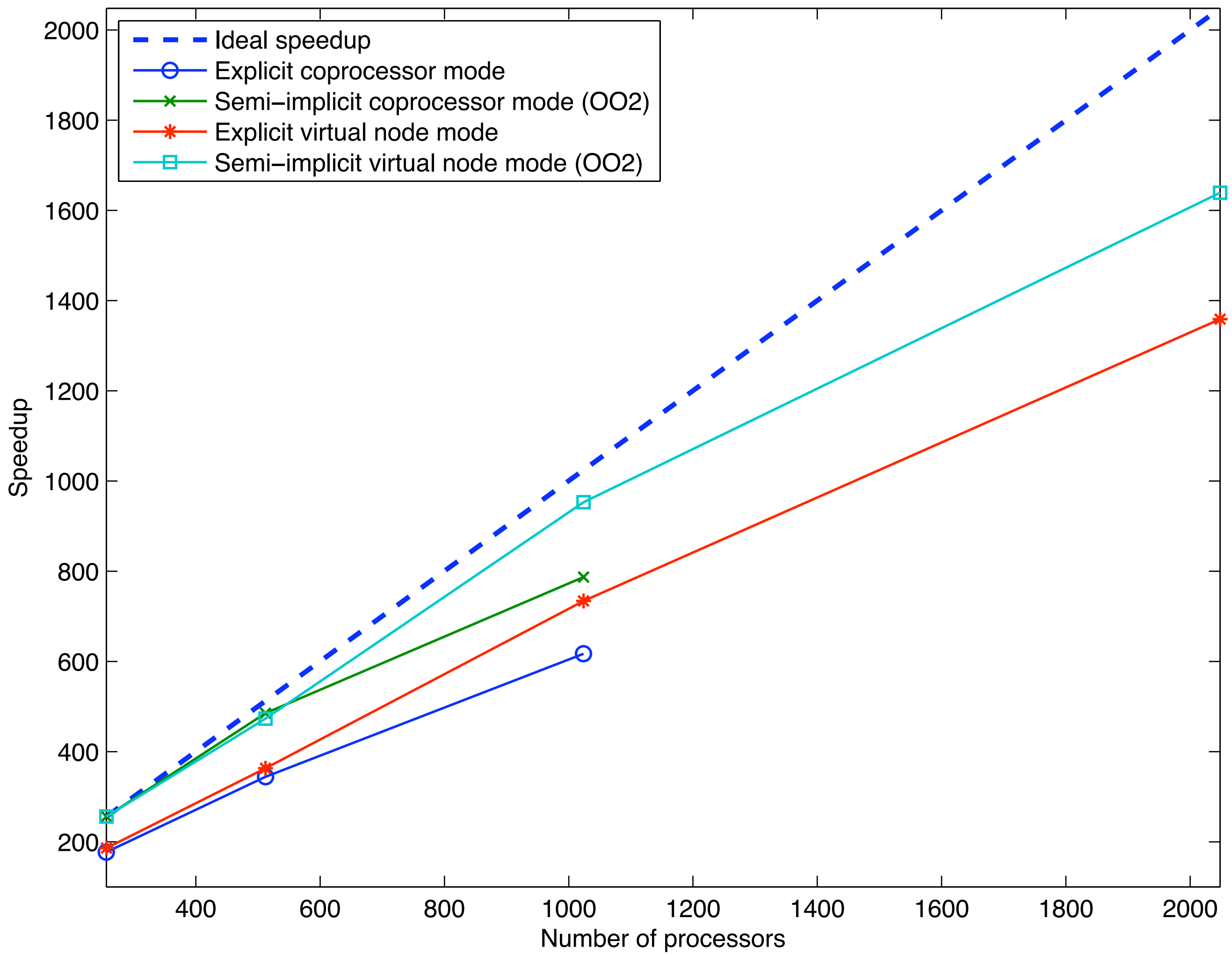


Taylor zeroth order and second order



Optimized zeroth and second order with two-sided zeroth order

Si vs Exp: Blue gene

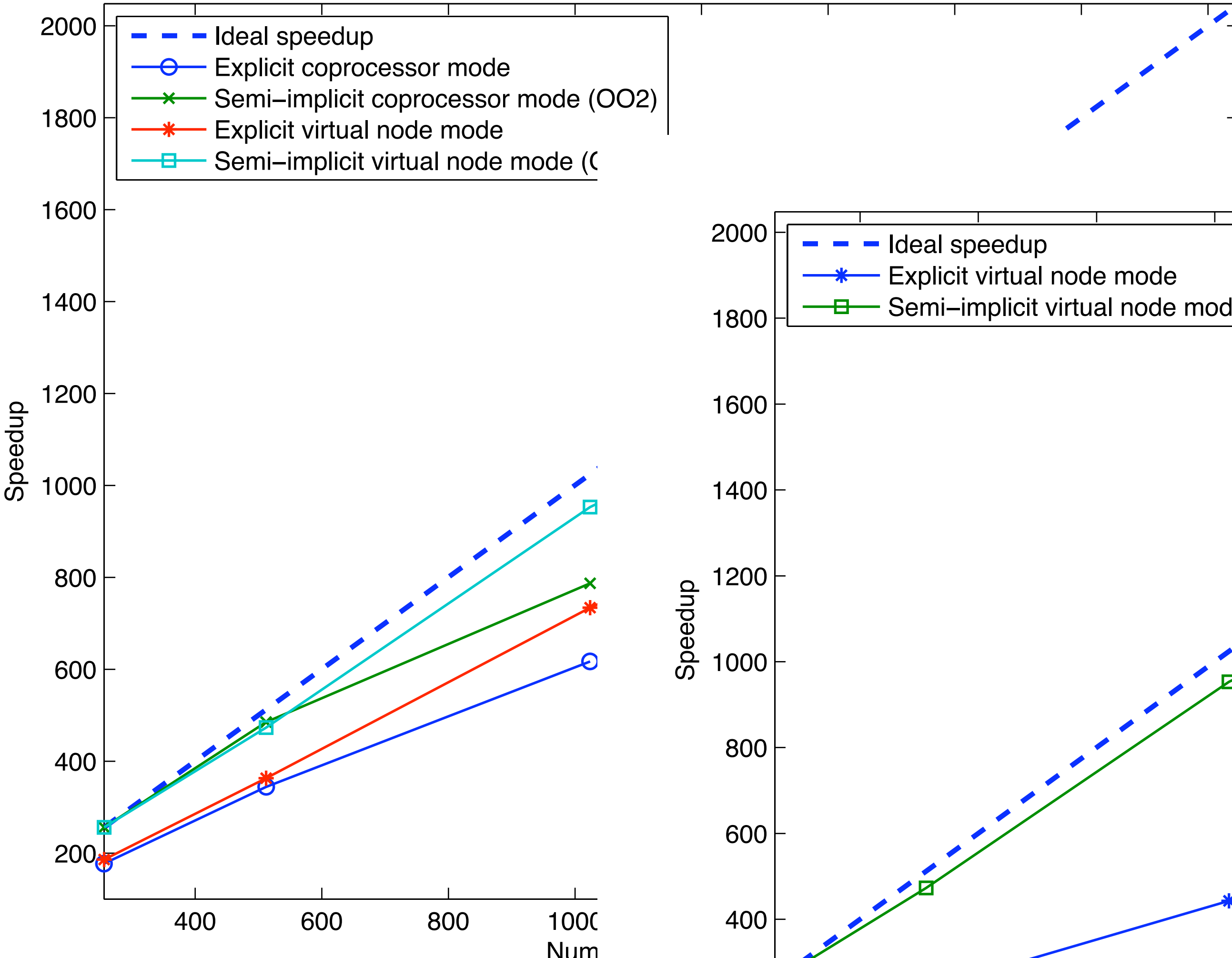


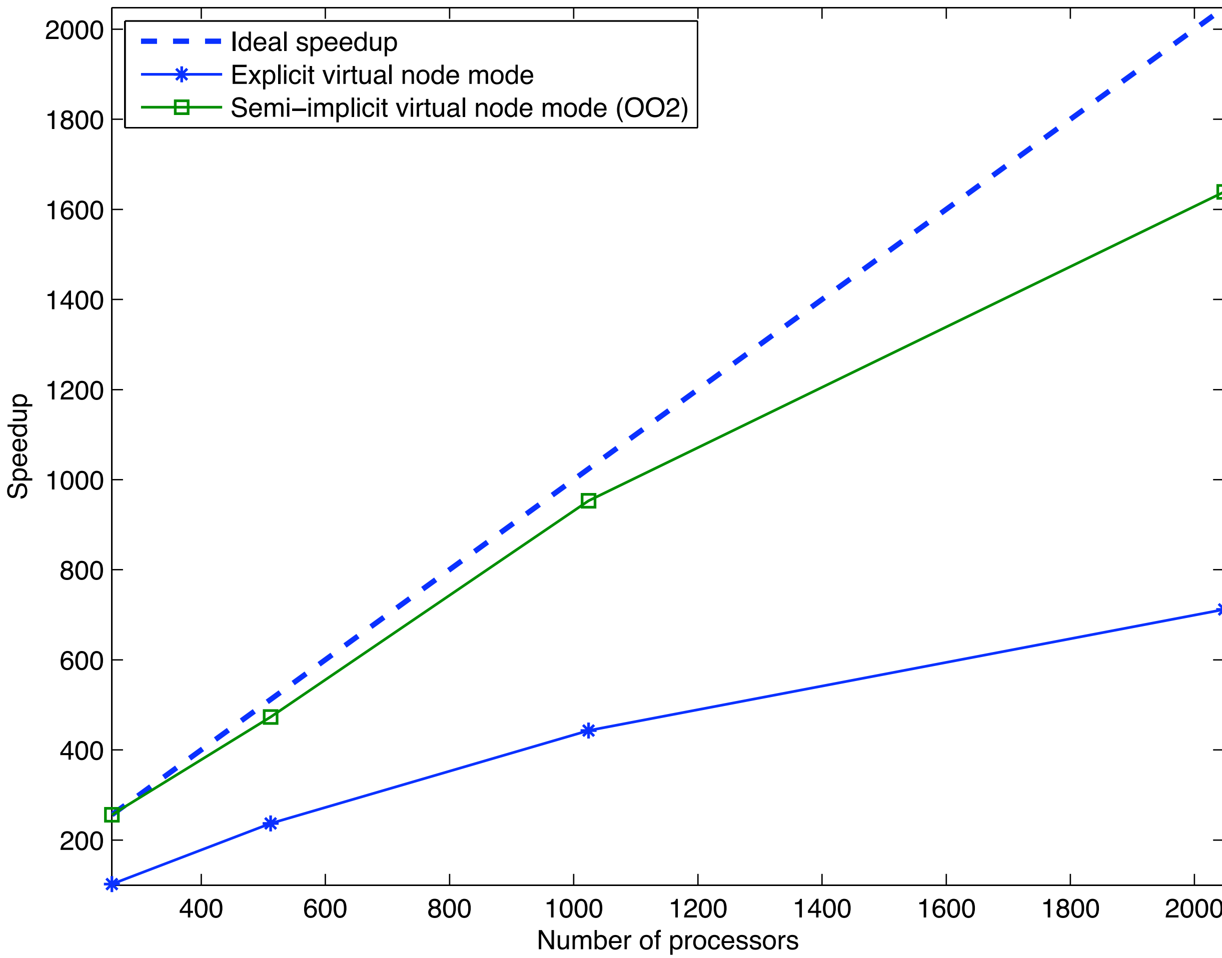
SGT 07 SISC

ne=32, 40km

Si vs Exp: Blue gene

SGT 07 SISC

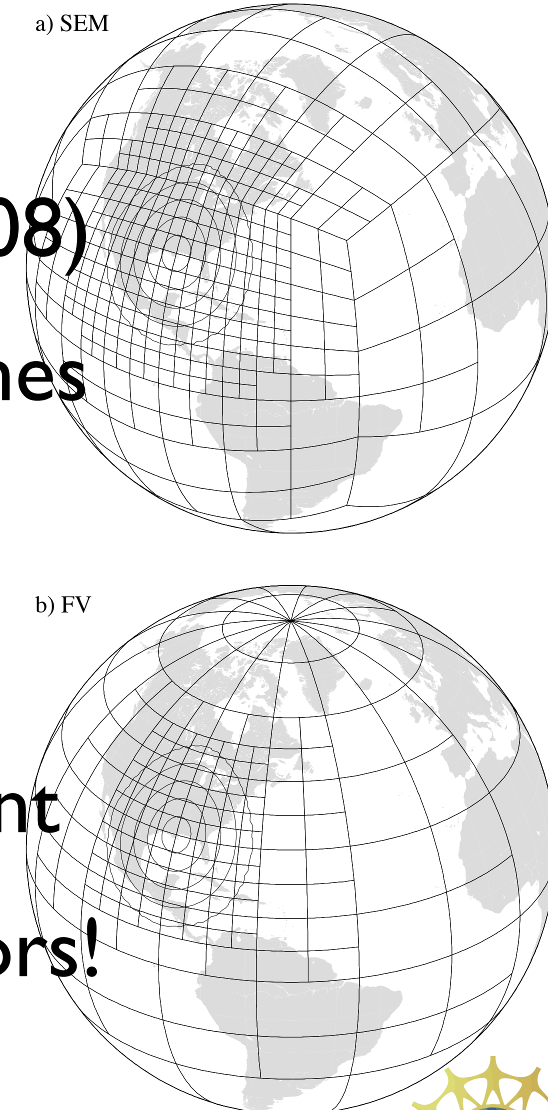




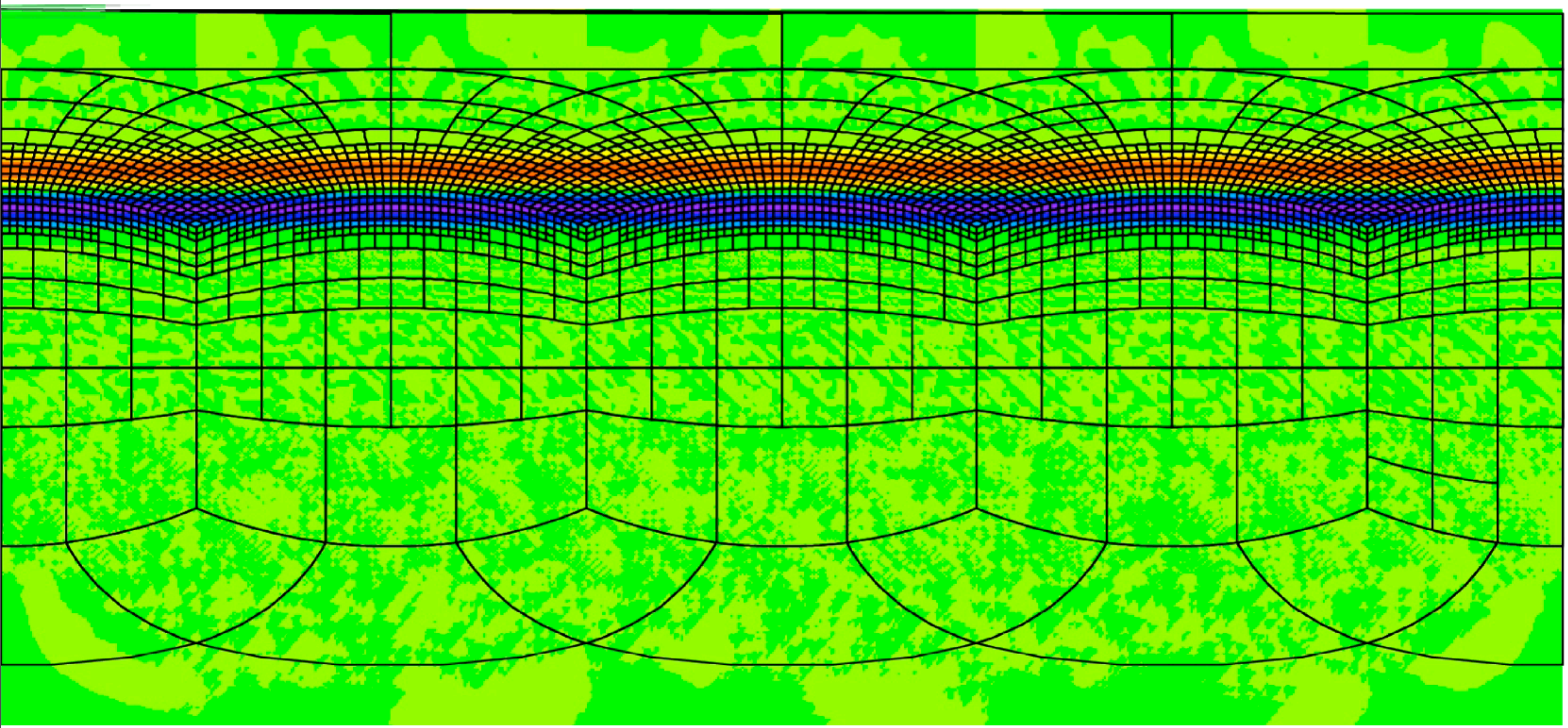
ne=32, 40km

AMR

- Comparison of SEM with FVM (SJDTT 08)
- Both non-conforming dynamic approaches
- Uses tests from literature
- Cubed sphere (SEM) lat-lon (FVM)
- At comparable errors SEM more efficient
- Runs below 1/3 degrees on 16 processors!
- <u>SEM 6.5 times faster on realistic test</u>



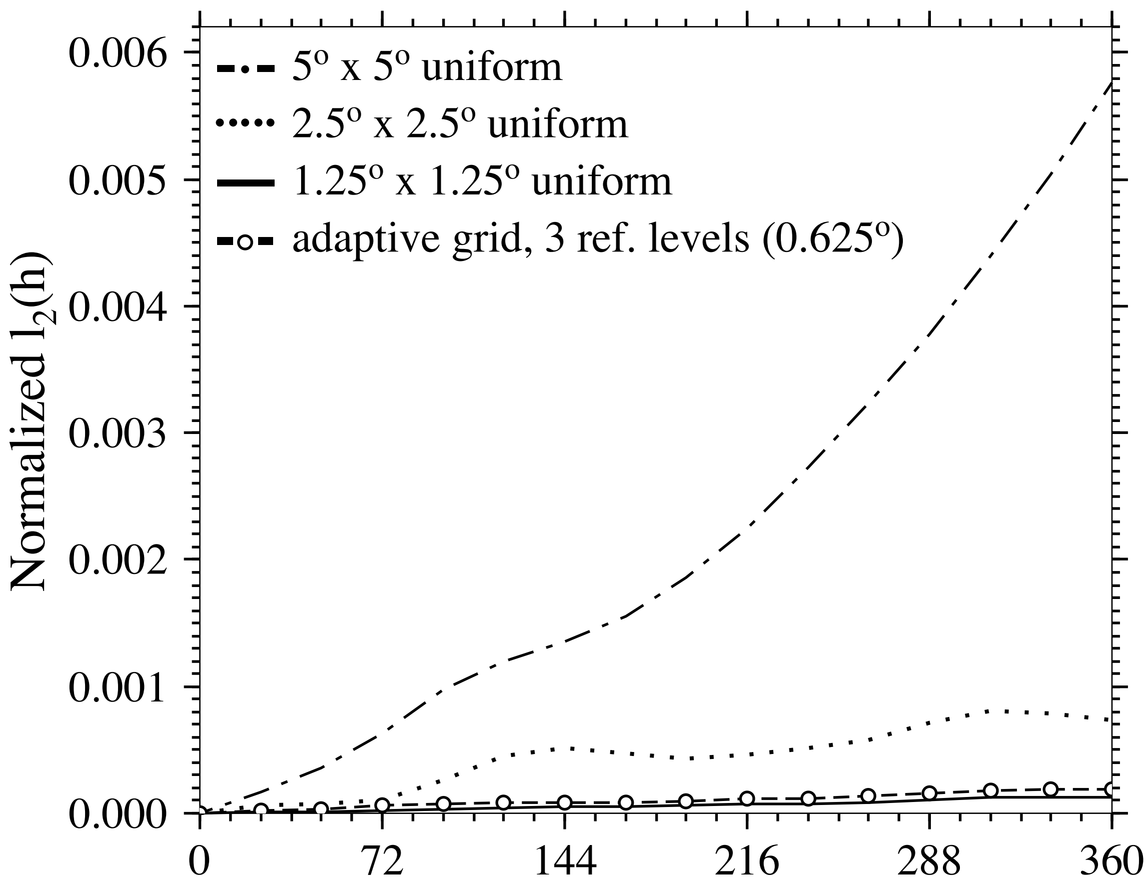
Galewski et al. 2004 test



0.3125 degrees... $|\zeta| \geq 3 \times 10^{-5}\ \mathrm{s}^{-1}$

Flow impinging a mountain

High-resolution solution DWD (German weather service)



SEM error 10 times lower than FVM

To generate same error one more ref in FVM: 6.5 slower than SEM

Take home message:

- Efficient time-stepping
- Unstructured adaptive meshes
- Optimal solvers

Take home message:

- Efficient time-stepping
- Unstructured adaptive meshes
- Optimal solvers

⇒ Can enable us to beat Moore's barrier

ESMs algorithm vision

- AMR h-p + unstructured meshes:
    - Resolutions of 1km possible on petascale machine: 10 km easily
    - Below 1km if exascale considered:
    - Adaptation criteria based on a goal: global temperature
- Multi-method time-stepping:
    - Can mix any time-stepping scheme with provable error
    - Yields load balancing in space and time
    - All processes performed at their ideal time-scale
    - One big step with macro steps in-bedded
    - Multigrid p-based solvers with demonstrated efficiency
- Algorithmic implementation flexibility:
    - Unstructured meshes: mesh planet, mesh ocean, mesh Sun, mesh mantle's core, mesh cities, mesh boxes, mesh mesh mesh... HOM makes this viable
    - Jacobian free: change the equations but same algorithms

NSF cdi

Current efforts

DGNH+AMR

- With D. Neckels ESMF
- Curvilinear elements
- Overhead of parallel AMR at each time-step: less than 1%

Idea based on Fischer,
Kruse, Loth (02)

$$\sum_k \frac{d}{dt} \int_{\Omega_k} u = -\sum_m \int_{\Gamma_m} (\mathbf{F}_{num}(\mathcal{I}_M^N u_l, u_r) - \mathbf{F}_{num}(u_r, \mathcal{I}_M^N u_l)) \cdot \hat{\mathbf{n}} = 0$$

DGNH+AMR

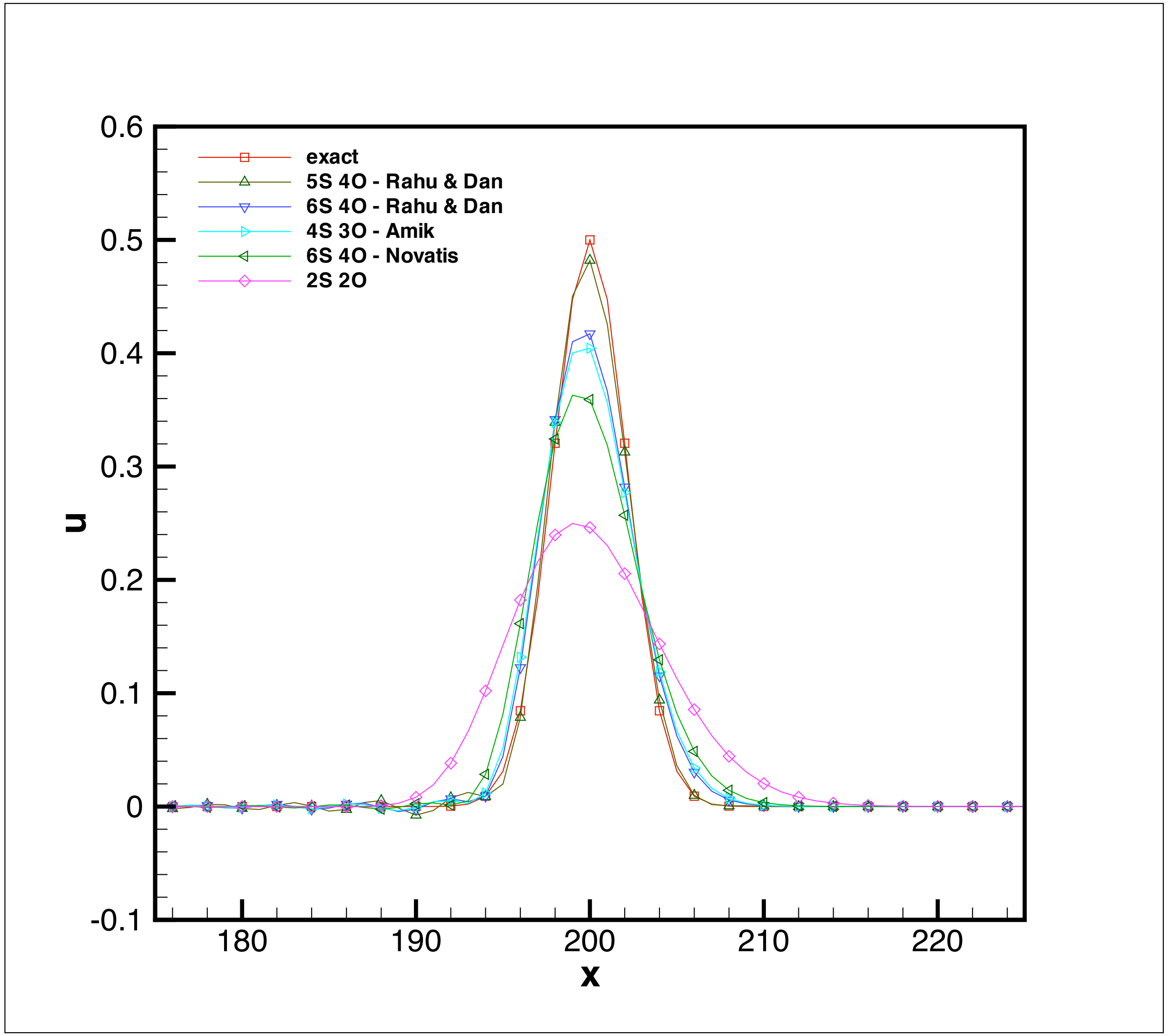
- With D. Neckels ESMF
- Curvilinear elements
- Overhead of parallel AMR at each time-step: less than 1%

Idea based on Fischer, Kruse, Loth (02)

$$\sum_k \frac{d}{dt}\int_{\Omega_k} u = -\sum_m \int_{\Gamma_m} \left(\mathbf{F}_{num}(\mathcal{I}_M^N u_l, u_r) - \mathbf{F}_{num}(u_r, \mathcal{I}_M^N u_l)\right) \cdot \hat{\mathbf{n}} = 0$$

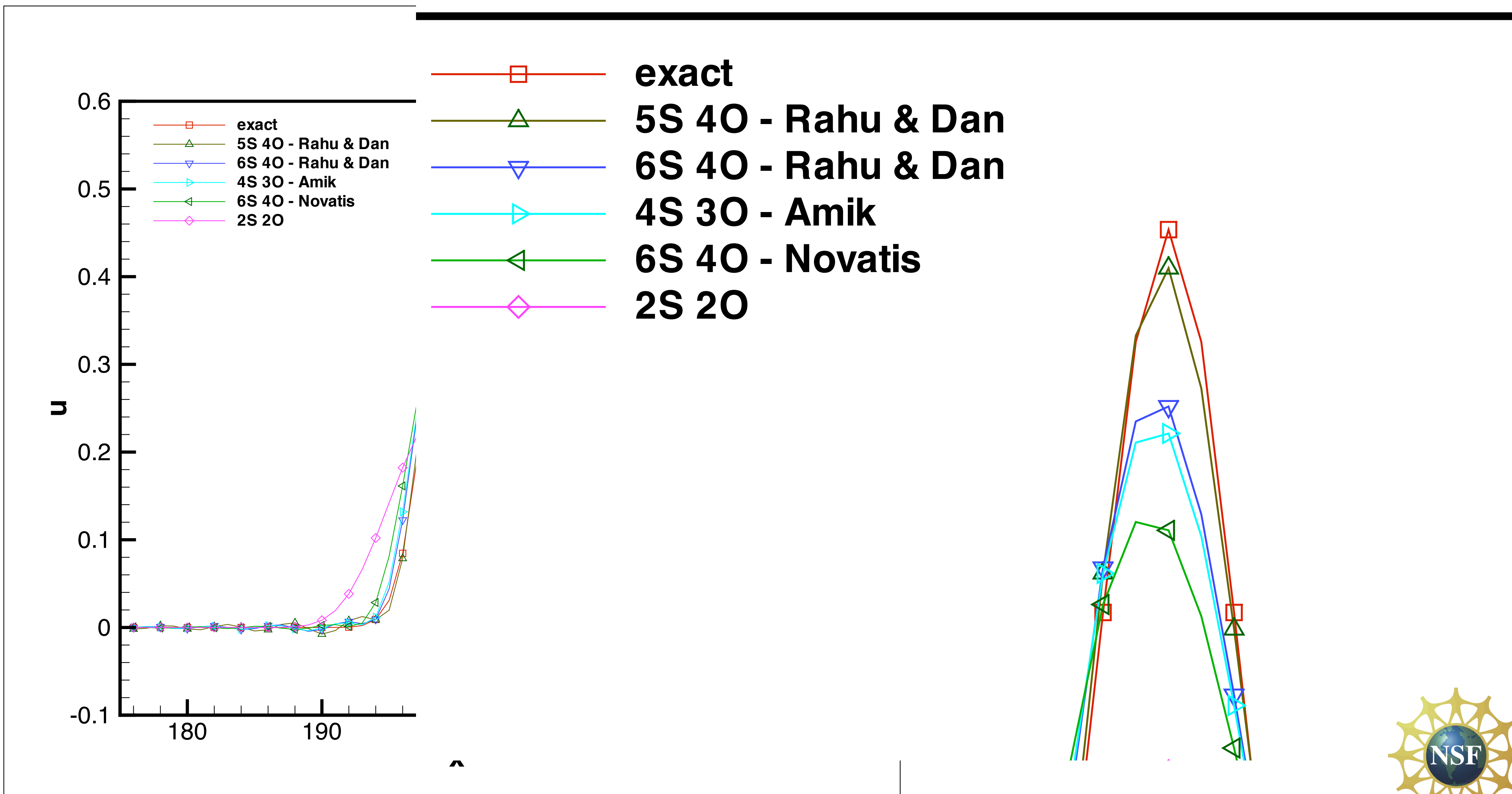
New Rosenbrock W

- With Dan Stanescu: University of Wyoming



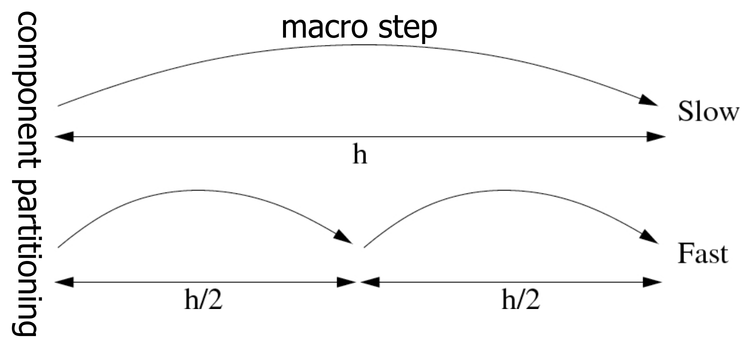
New Rosenbrock W

- With Dan Stanescu: University of Wyoming



Consider a two dimensional partitioning of the problem in numerical integration of large-scale models

- Multiphysics: additive partitioning
  - different physics have different dynamics and integrators with appropriate properties are required
- Multiscale: component partitioning
  - mesh refinement and variable wave speed restrict the global timestep



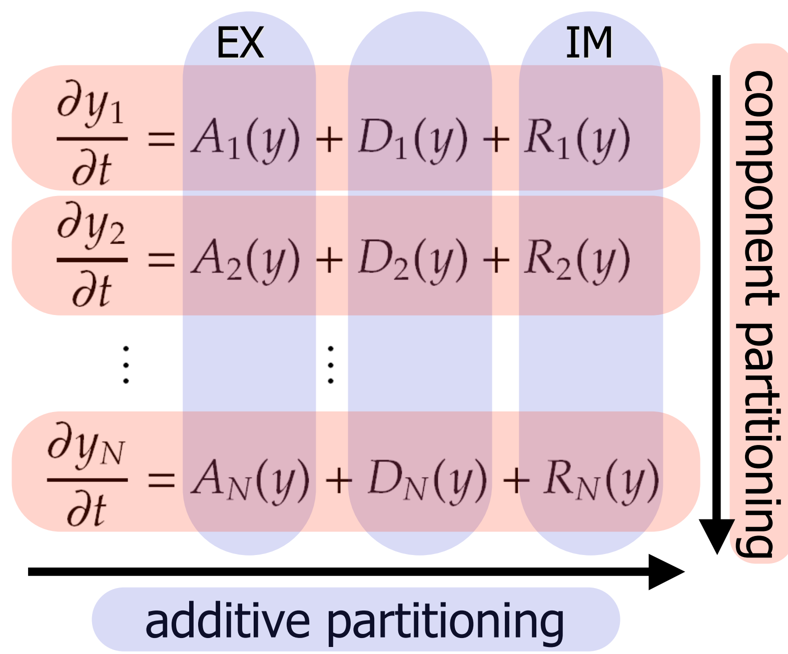
- Advection-Diffusion-Reaction:

$$\frac{\partial y}{\partial t} = -u\nabla\, y + \frac{1}{\rho}\nabla(\rho K \nabla y) + \frac{1}{\rho} C(\rho y)$$

- MOL: time dependent PDE => ODE

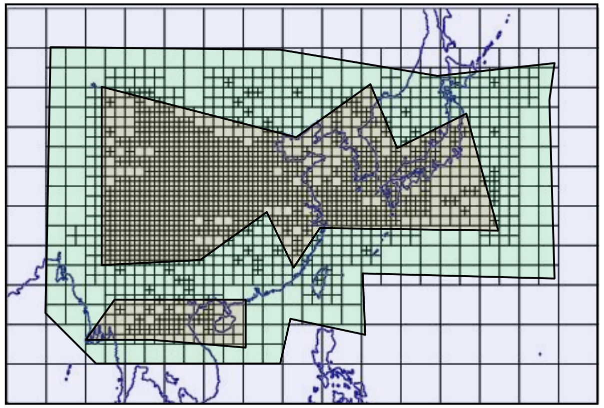
$$\frac{\partial y}{\partial t} = A(y) + D(y) + R(y)$$

$$y = [y_1, y_2, \cdots, y_N]^T$$



(Courtesy of Adrian Sandu VTU)

Propose multirate partitioned Runge-Kutta methods for component partitioning – MPRK2



- Consider the following (base) Runge-Kutta method:

$$
\begin{array}{c|c} c & A \\ \hline & b^T \end{array}
\quad \Longrightarrow \quad
\begin{array}{c|cc} \gamma_1 c & \gamma_1 A & 0 \\ \gamma_1 \mathbb{1} + \gamma_2 c & \gamma_1 \mathbb{1} b^T & \gamma_2 A \\ \hline & \gamma_1 b & \gamma_1 b \end{array}
\qquad
\begin{array}{c|cc} c & A & 0 \\ c & 0 & A \\ \hline & \gamma_1 b & \gamma_2 b \end{array}
$$

- Multirate partitioned Runge-Kutta method MPRK2:

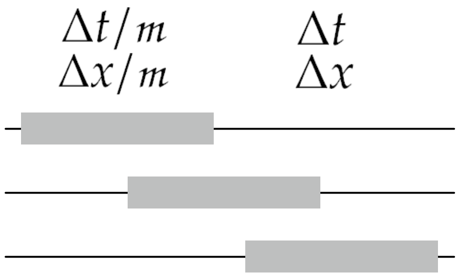
$$
\begin{array}{c|cccc}
\frac{1}{m} c & \frac{1}{m} A & & & \\
\frac{1}{m} \mathbb{1} + \frac{1}{m} c & \frac{1}{m} \mathbb{1} b^T & \frac{1}{m} A & & \\
\vdots & \vdots & \ddots & \ddots & \\
\frac{m-1}{m} \mathbb{1} + \frac{1}{m} c & \frac{1}{m} \mathbb{1} b^T & \cdots & \frac{1}{m} \mathbb{1} b^T & \frac{1}{m} A \\
\hline
 & \frac{1}{m} b^T & \frac{1}{m} b^T & \ldots & \frac{1}{m} b^T
\end{array}
\qquad
\begin{array}{c|cccc}
c & A & & & \\
c & & A & & \\
\vdots & & & \ddots & \\
c & & & & A \\
\hline
 & \frac{1}{m} b^T & \frac{1}{m} b^T & \ldots & \frac{1}{m} b^T
\end{array}
$$

Fast component: CFL = m x c Slow component: CFL = c

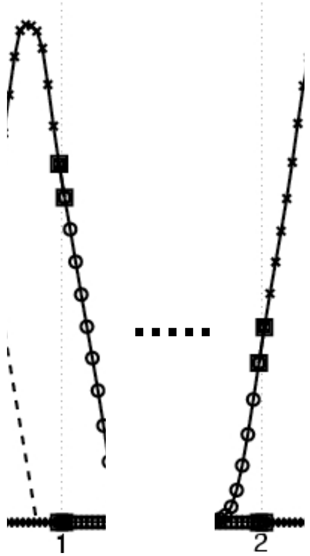
- **Emil M. Constantinescu** and Adrian Sandu, Multirate timestepping methods for hyperbolic conservation laws; Journal of Scientific Computing, Vol. 33(3), pp 239-278, 2007.

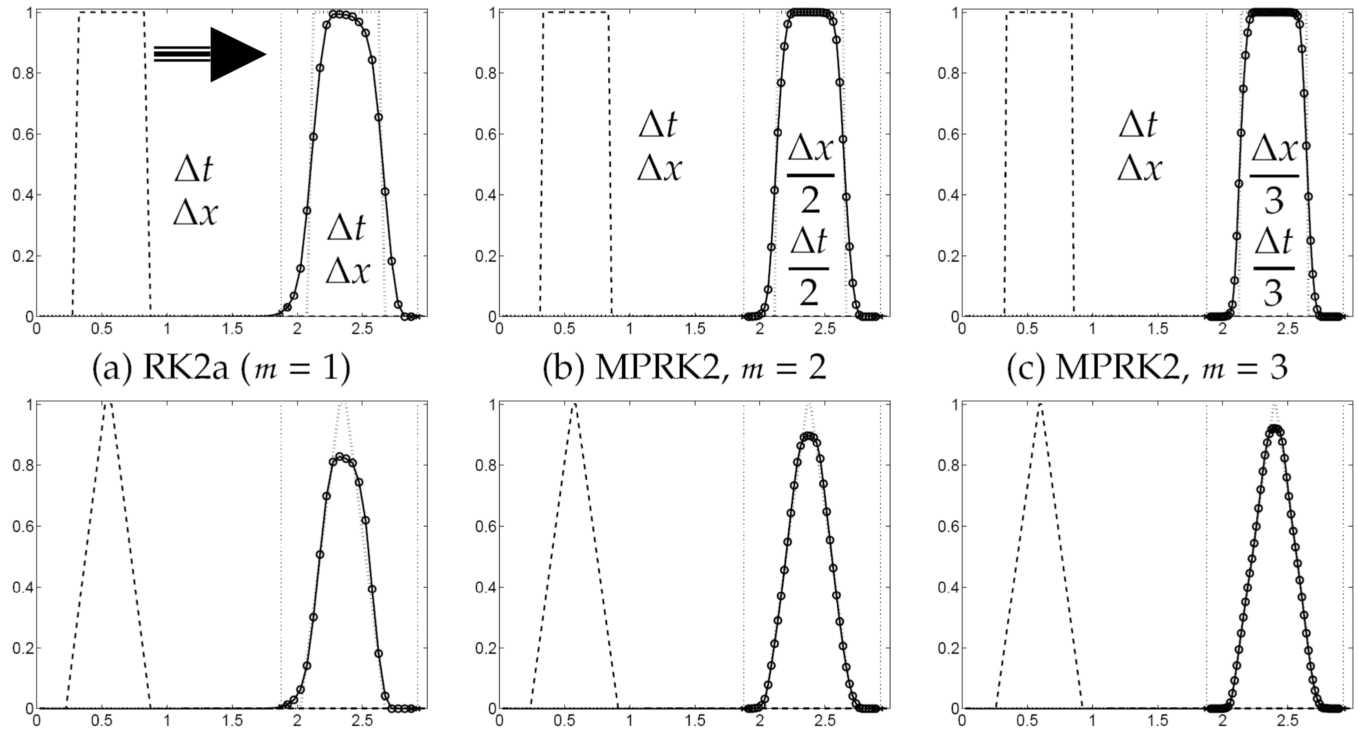
Numerical experiments confirm the theoretical results for a linear example

- The solution of advection equation – the fine grid is moving along with the wave profile: $\frac{\partial y}{\partial t} + \frac{\partial y}{\partial x} = 0$



(wave passing through a fixed interface)





(a) RK2a ($m = 1$) (b) MPRK2, $m = 2$ (c) MPRK2, $m = 3$

(d) RK2a ($m = 1$) (e) MPRK2, $m = 2$ (f) MPRK2, $m = 3$

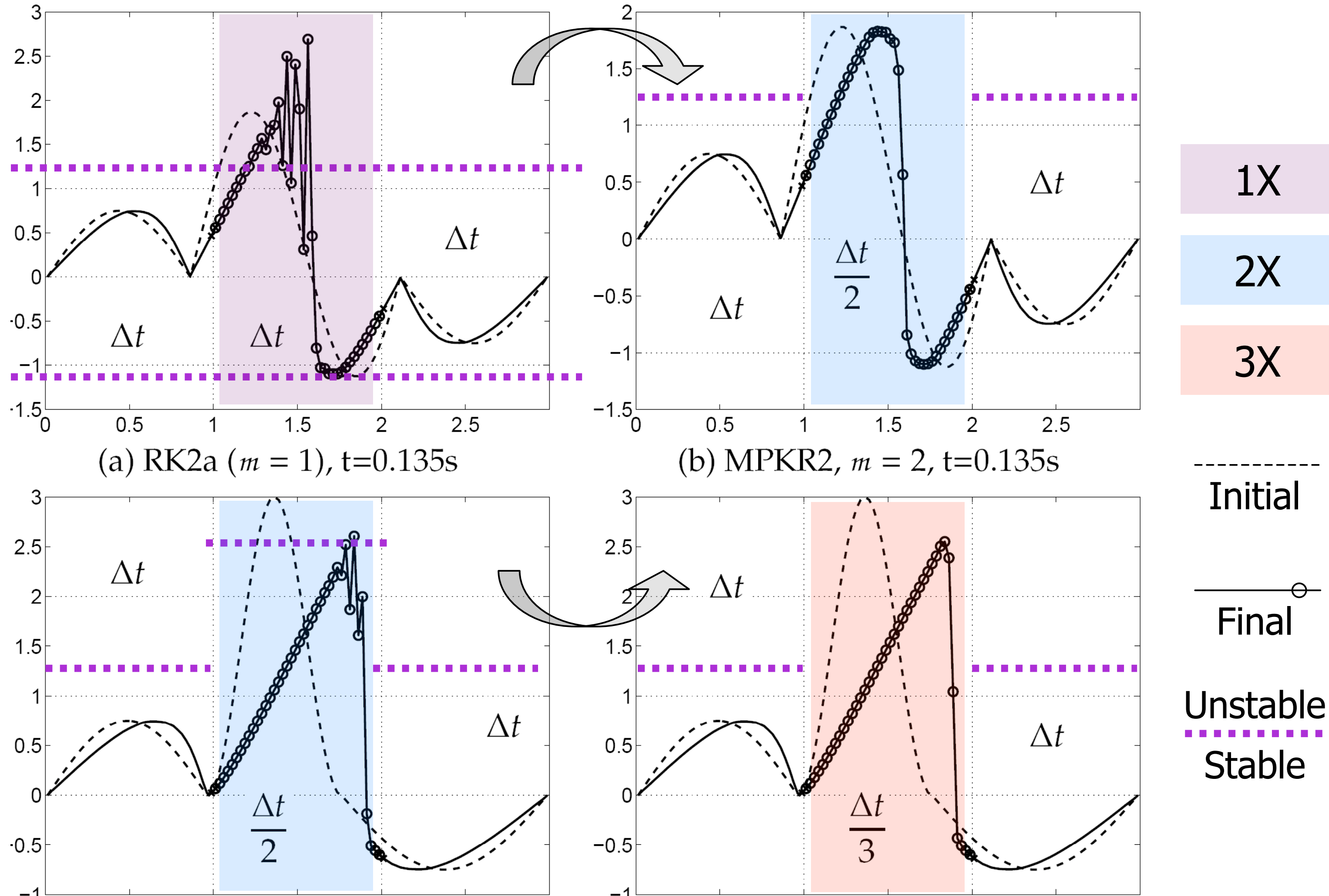
(Courtesy of Adrian Sandu VTU)

Numerical experiments confirm the theoretical results for a nonlinear example

Burgers' equation

$$\frac{\partial y}{\partial t} + \frac{1}{2}\frac{\partial y^2}{\partial x} = 0$$

$$\frac{\partial y}{\partial t} + y\,\frac{\partial y}{\partial x} = 0$$



(a) RK2a (m = 1), t=0.135s

(b) MPKR2, m = 2, t=0.135s

(c) MPKR2, m = 2, t=0.225s

(d) MPKR2, m = 3, t=0.225s

(Courtesy of Adrian Sandu VTU)

CDI-type 2 proposal

- Multi-institutional: VTU, UW, U Geneva, U Louvain la Neuve, U Nice Sophia-Antipolis
- Expertise in: time-stepping, optimal solvers, high/low-order methods, software engineering, HPC, Krylov subspace methods, adjoints, TLM.
- Goal: The discovery of efficient computational methods for multiscale adaptive, multidisciplinary physics on petascale system
- Build an all scales simulation framework

CDI-type 2 proposal

- NCAR: J. Tribbia, P. Smolarkiewicz and A. St-Cyr, D. Rosenberg, D. Neckels and A. Wyszogrodzki
- UW: D. Mavriplis and D. Stanescu
- VTU: A. Sandu
- U Geneva: M. J. Gander
- Sophia-Nice-Antipolis: V. Dolean

Thank you!