Challenges for Climate and Weather Prediction in the Era of Exascale Computer Architectures: Oscillatory Stiffness, Time-Parallelism, and the Role of Long-Time Dynamics

Beth Wingate

University of Exeter

Collaborators

Terry Haut

Lawrence Livermore National Laboratory

Adam Peddle

University of Exeter

Pedro Peixoto

University of Sao Paulo

Martin Schreiber

University of Exeter

Colin Cotter

Imperial College London

Jemma Shipton

Imperial College London

David Ham

Imperial College London

Chris Maynard & Others

UK Met Office

Mike Ashworth & Others

STFC Daresbury Laboratory

London

University of Exeter

London

University of Exeter

UK Met Office

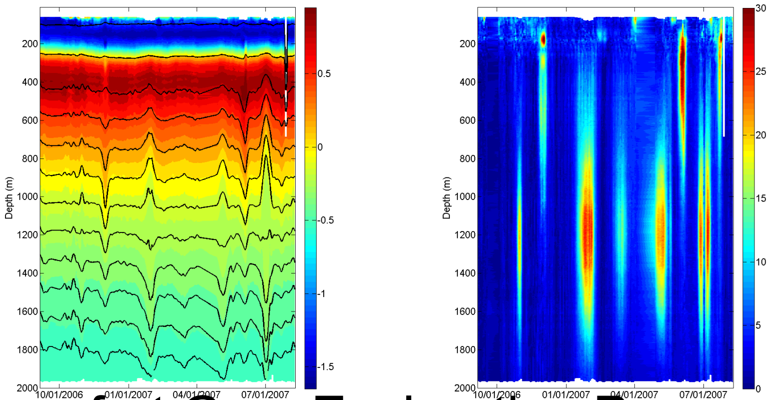
Stonehenge!

London

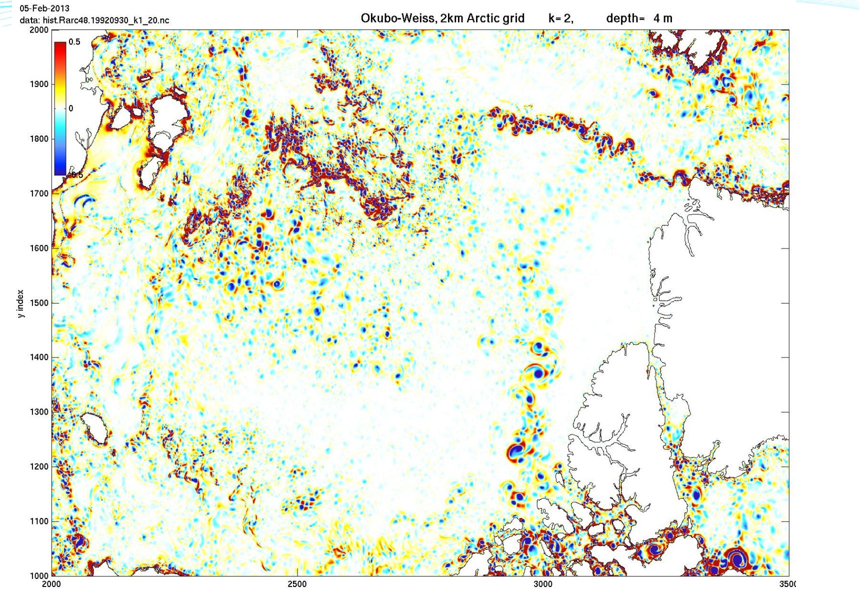
UK Met Office

University of Exeter

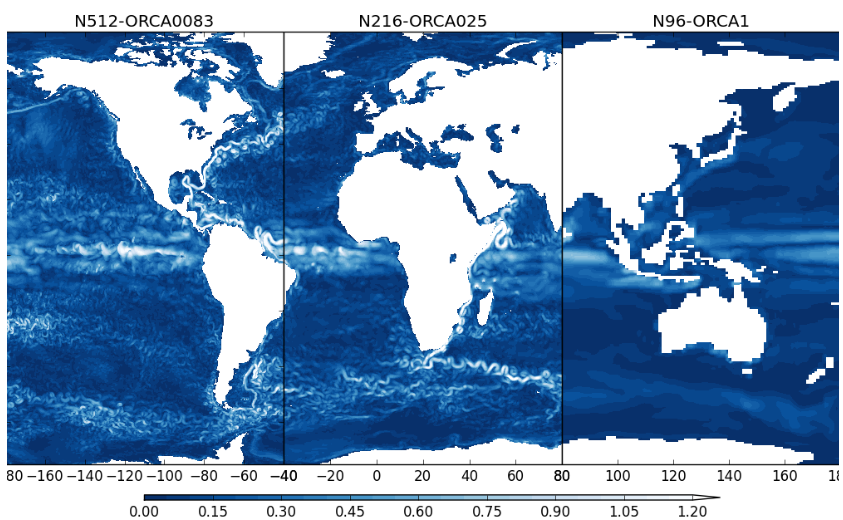
Dynamics in the ocean



Beaufort Gyre Exploration Program

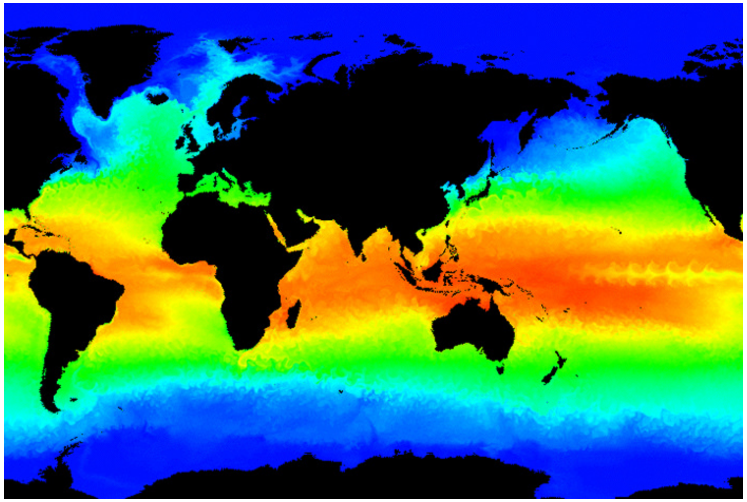


Arctic Ocean Eddies



NEMO: Met Office Hadley Centre

Mike Bell



Global temperature 1/10 deg - POP

Outline

- Ocean models (climate, weather) & HPC in the next decade
  ACME, ESCAPE, NextGenIO, ESiWACE, UK Met Office - Gung Ho, ECMWF, etc.
- Implications for physics with oscillatory and dissipative stiffness
  (including numerical models of weather and climate)
- Introduction to disruptive algorithms: time-parallelism
  Some examples: RDIC, results for the time-parallel matrix exponential (REXI), Parareal
- **Mathematics underpinning the algorithms: oscillatory and dissipative stiffness**
- Final thoughts

Meeting and projections about this topic are happening world-wide, what can we do with the new architectures?

- SIAM CSE 2017
- ReCoVER - UK EPSRC
- DOE-ACME
- ESCAPE
- Horizon2020 - NextGenIO,
- Horizon2020 - ESiWACE
- UK Met Office - Gung Ho, ECMWF, etc.

Some attributes of Climate & weather models

- have "physics" models for clouds, land surface (trees!), sometimes even "economic forcing"
- assimilate data into the simulations
- Tightly coupled physics and numerics: the Gent McWilliams model example

Paper: **The Gent-McWilliams Parameterization: 20/20 hindsight, P. Gent,** Ocean Modelling, Vol 39, 2011

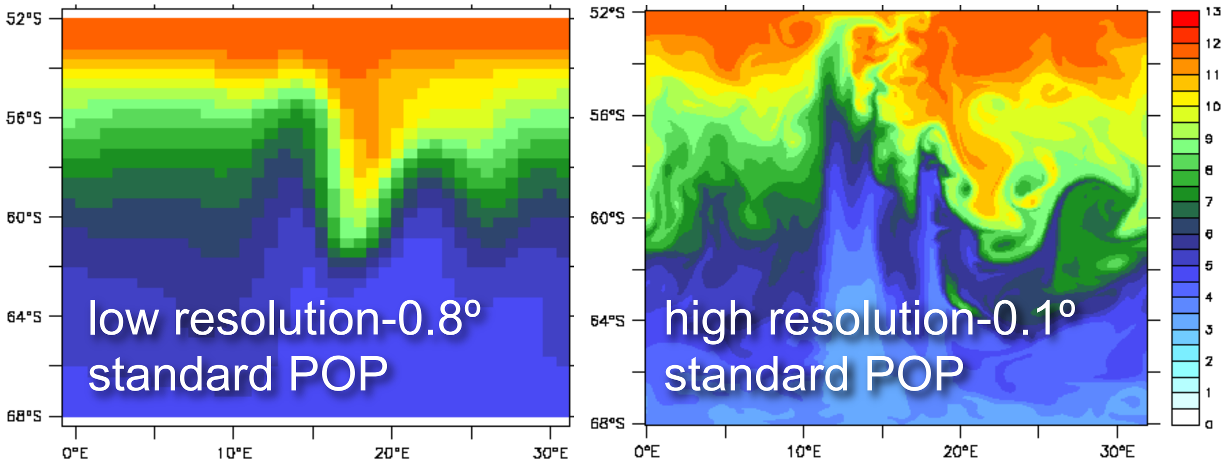
- Different resolutions require different physics models. A weather or climate model **that 'converges' as the grid spacing decreases** is generally much more complex than simple grid convergence studies

Climate and weather models are complex fusions of numerical and physical models.

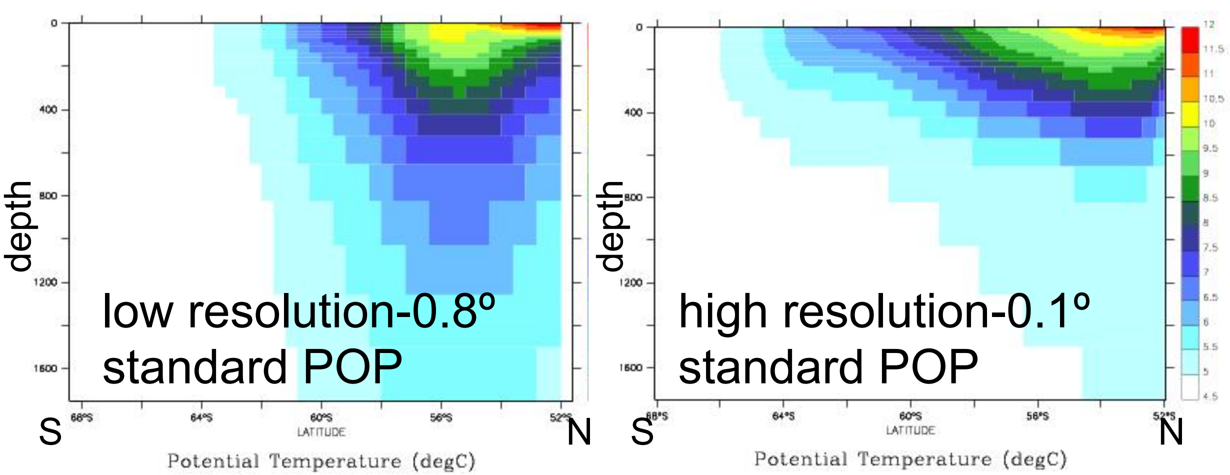
Baroclinic Instability – improvement in models
Can be more important than improvements In numerics!

UNIVERSITY OF EXETER

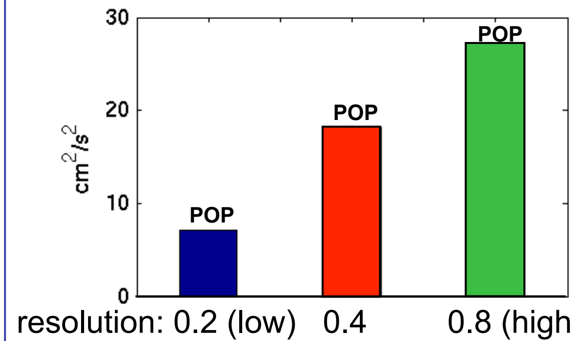
Surface temperature



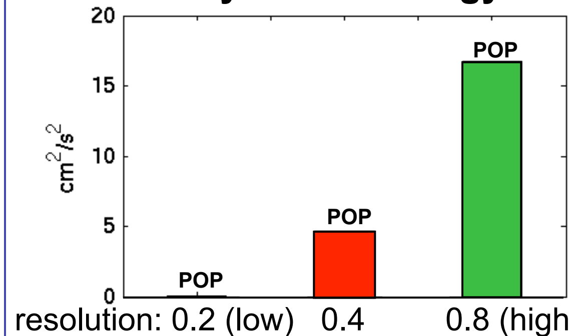
Potential temperature - vertical cross section



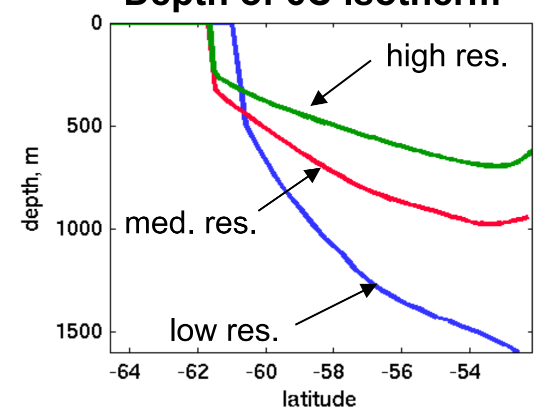
Kinetic energy



Eddy kinetic energy



Depth of 6C isotherm



Algorithms for climate and weather and new computing architectures

- **Fixed grid on N processors**

  For a fixed grid you may already have an optimal distribution of the grid on N processors. If you add more processors, more communication would be required.

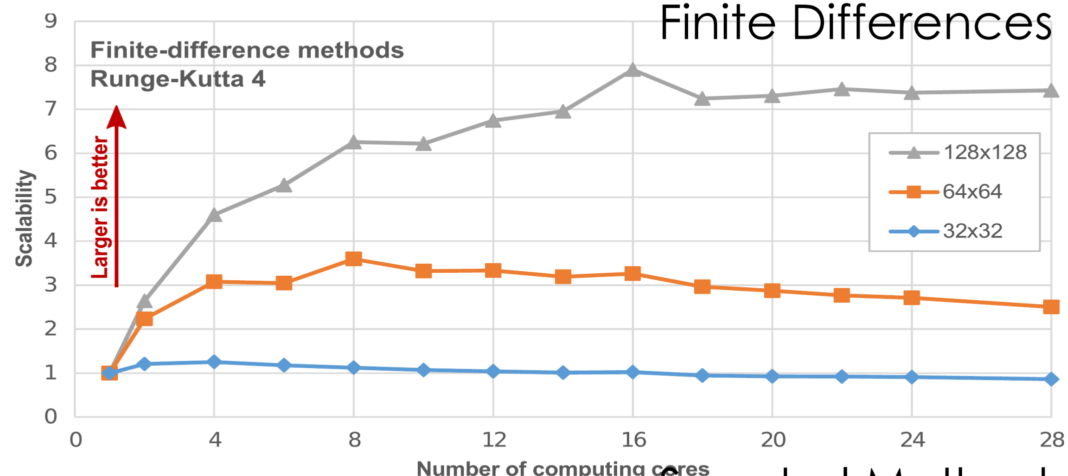
- **Grid refinement** (we'll still have to wait for each time step)

  Because current algorithms need to reduce their maximum time step as the number of grid points increases, these new machines may not significantly reduce wall clock time. You may be able to have a higher resolution grid but you will still wait a longer time for each time step to complete.

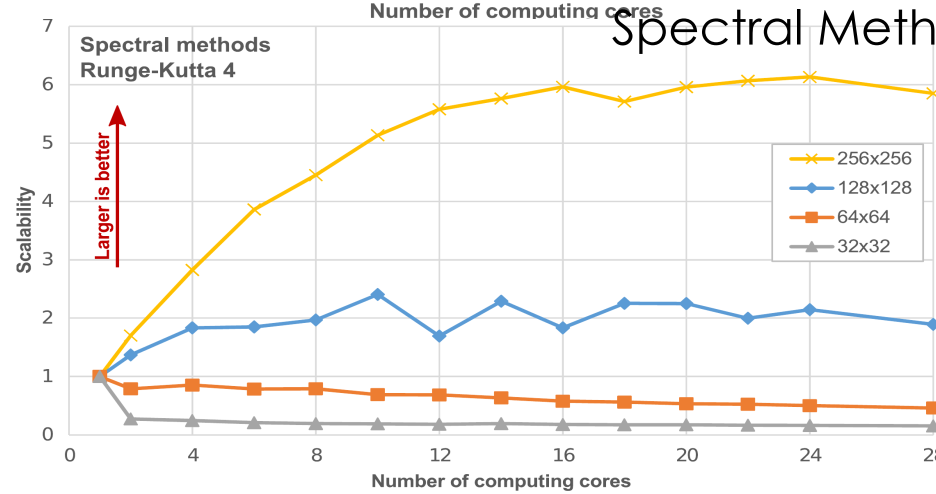
Example: Linear Rotating Shallow-water equations

- 2x 14 cores, Intel Xeon(R) CPU E5-2697, no hyperthreading, compact affinities
- Shared-memory parallelization only, no distributed-memory communication overheads
- Scalability limited

Finite Differences

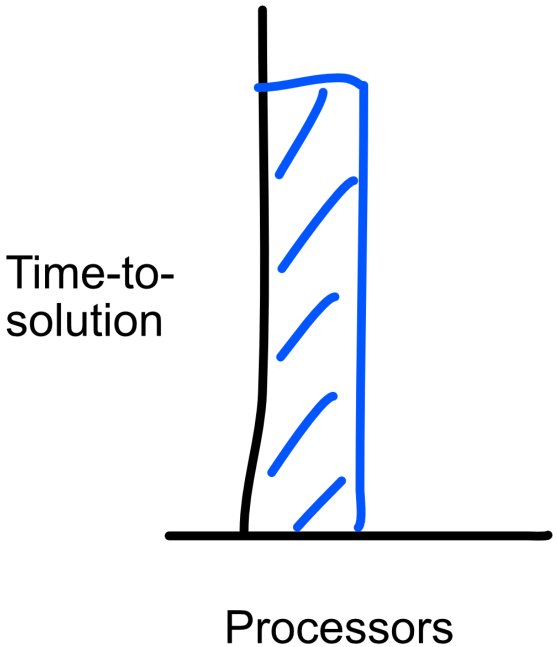


Spectral Methods

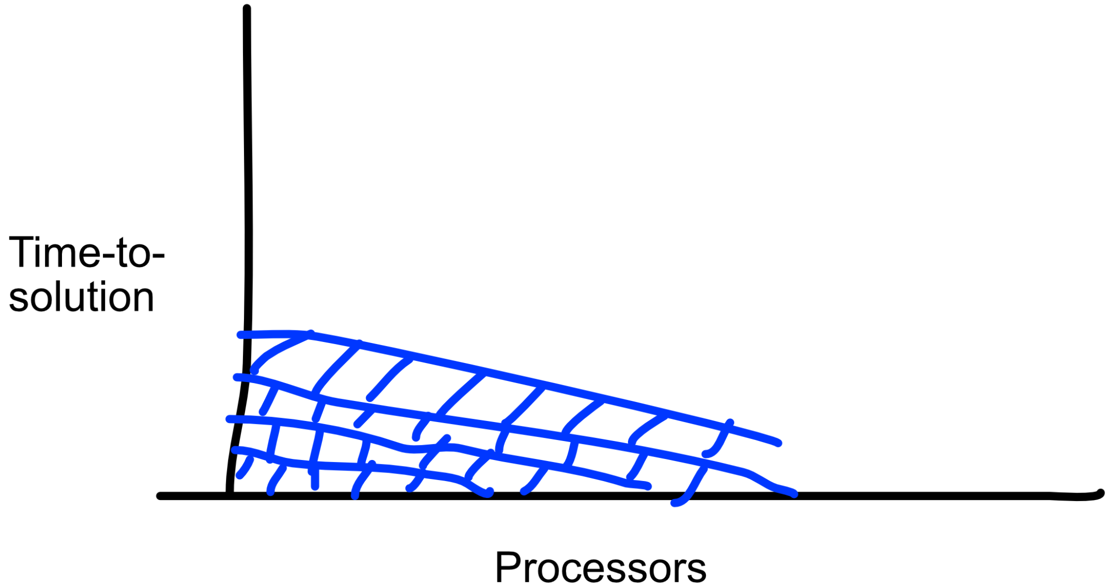


Schreiber, Peixoto, Haut & Wingate, **Beyond spatial scalability limitations with a massively parallel method for linear oscillatory problems** *accepted* International Journal of High Performance Computing Applications 2017

Time-parallel performance models? We need these kinds of models.

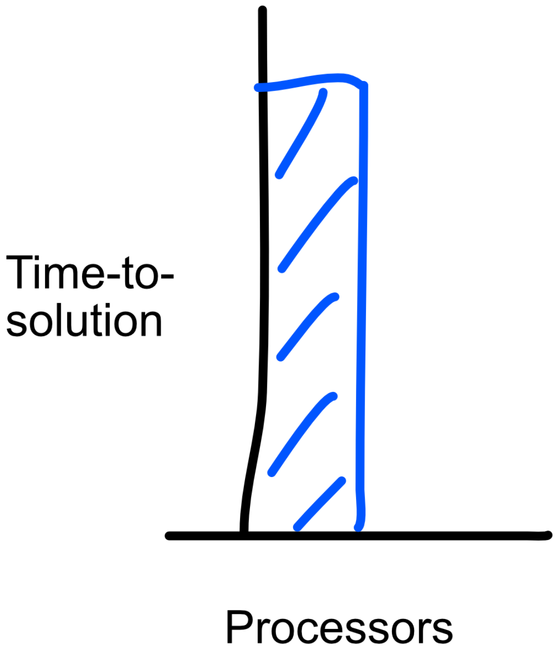


“Monolithic serial”

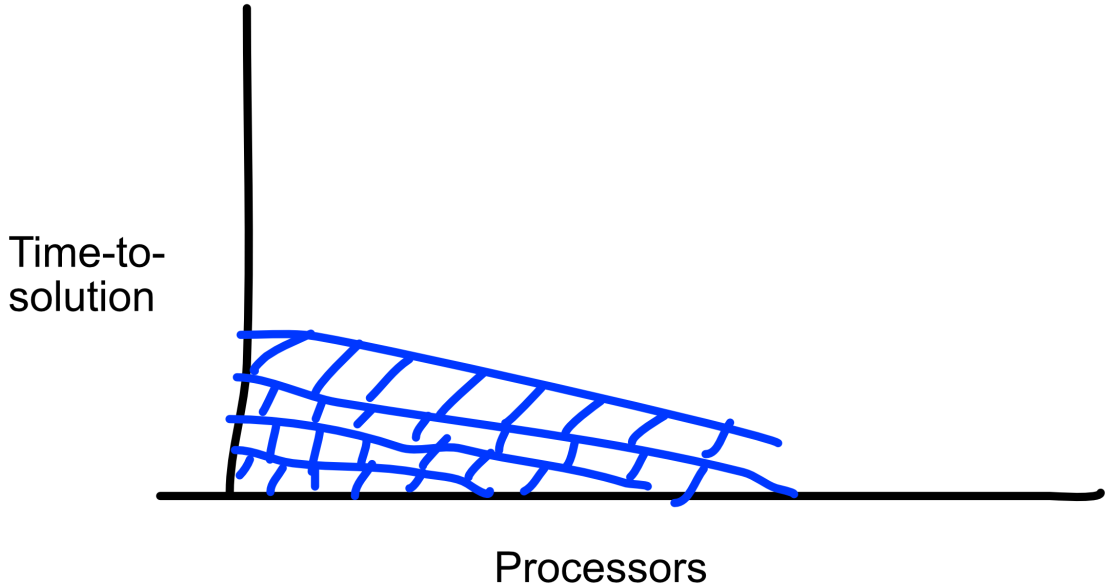


“Sliding Window time-parallelism”

Time-parallel performance models? We need these kinds of models.



"Monolithic serial"

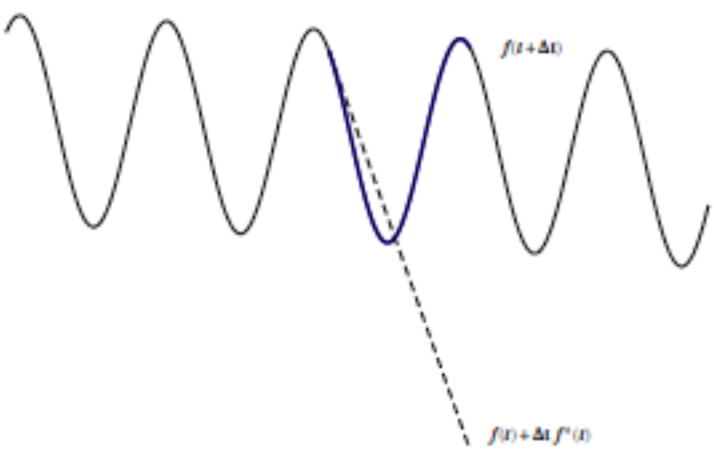


"Sliding Window time-parallelism"

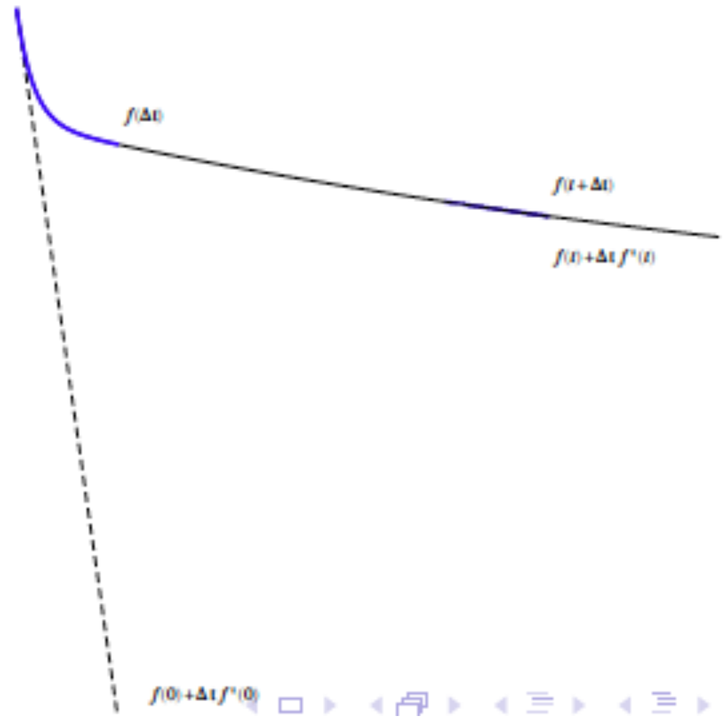
Are there models like this for different types of time-parallelism, simple equation sets and new architectures?

Challenges for Parareal and Climate/Weather models: Stiffness in Parareal – dissipative & oscillatory stiffness

(a) Stiffness from oscillations



(b) Stiffness from dissipation



FAST singular limit
Taylor Series only useful on small time scales

SLOW singular limit
Taylor Series useful

Disruptive Algorithms: Parellelisation in time is 50 years old

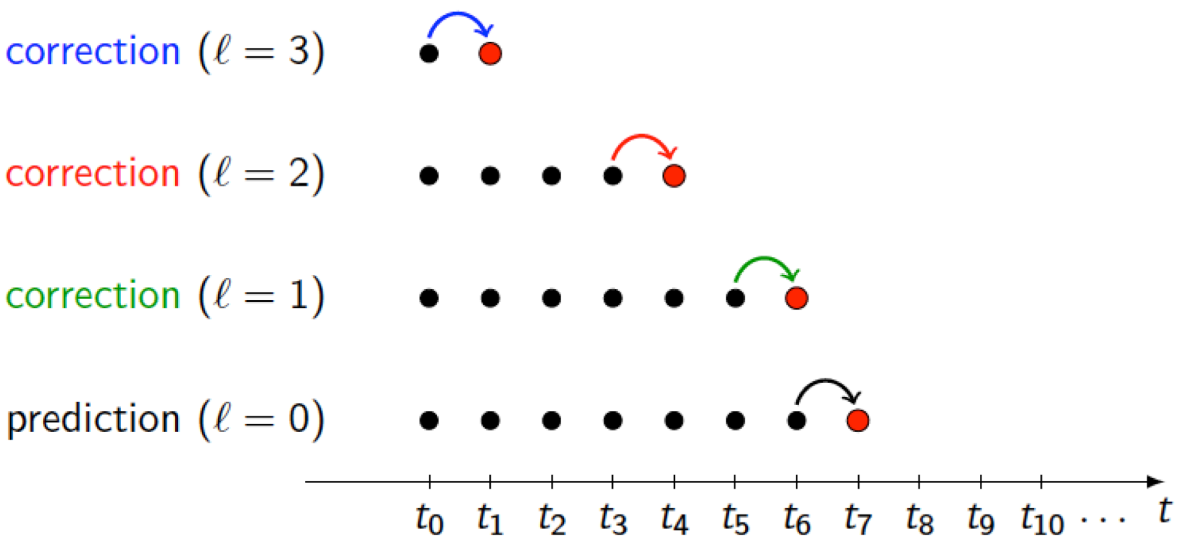
- RIDC – Revisionist Integral Deferred Correction

  Pereyra, 1966 & Ong et al 2010
- Shooting type methods – Parareal – *Lions, Turinici, Maday, 2001*
- Multi-grid type methods – *Emmett and Minion 2012*
- exp(t L) – exponential integrators
- Iterative and direct

50 years of time parallel time integration, Gander, M.J. In Carraro, T., Geiger, M., Korkel, S., Rannacher, R (Eds). *Multiple Shooting and Time Domain Decomposition.* Springer-Verlag, 2015

Friendly Example: RIDC Revisionist Integral Deferred Correction

- Based on the idea of promoting a lower order scheme to a higher one, then using ideas from predictor corrector
- Small-scale parallelism – you compute the iterates in parallel
- ***Pereyra, 1966 & Ong et al 2010***



Serial Time Stepping versus Parallel Exp

Exponential Integrators for Weather
Clancy C and Pudykiewicz J, 2013
Garcia F, Bonaventura L, Net M et al. 2014

The problem:

$$\frac{d\ \mathbf{u}(t)}{dt} = \mathcal{L}\ \mathbf{u}(t), \qquad \mathbf{u}(0) = \mathbf{u}_0$$

Has matrix exponential solution:

$$\mathbf{u}(t) = \mathbf{e}^{\mathbf{t}\mathcal{L}}\mathbf{u}_0$$

Has serial time stepping solution:

$$\mathbf{u}^n = (\mathbf{I} + \boldsymbol{\Delta}\mathbf{T}\mathcal{L})^n\,\mathbf{u}_0$$

Serial standard time stepping versus **Parallel** Exp

UNIVERSITY OF EXETER

Exponential Integrators:
C. Moler, "19 Dubious ways to compute the exponential of a matrix", 1978, 2003
M. Hochbruck and A. Ostermann, "Exponential Integrators", 2010

- **Parallel** Matrix Exp

$$\mathbf{u}(t) = \mathbf{e}^{\mathbf{t}\mathcal{L}}\mathbf{u}_0$$

- **Serial** time stepping

$$\mathbf{u}^n = \left(\mathbf{I} + \mathbf{\Delta T}\mathcal{L}\right)^n \mathbf{u}_0$$

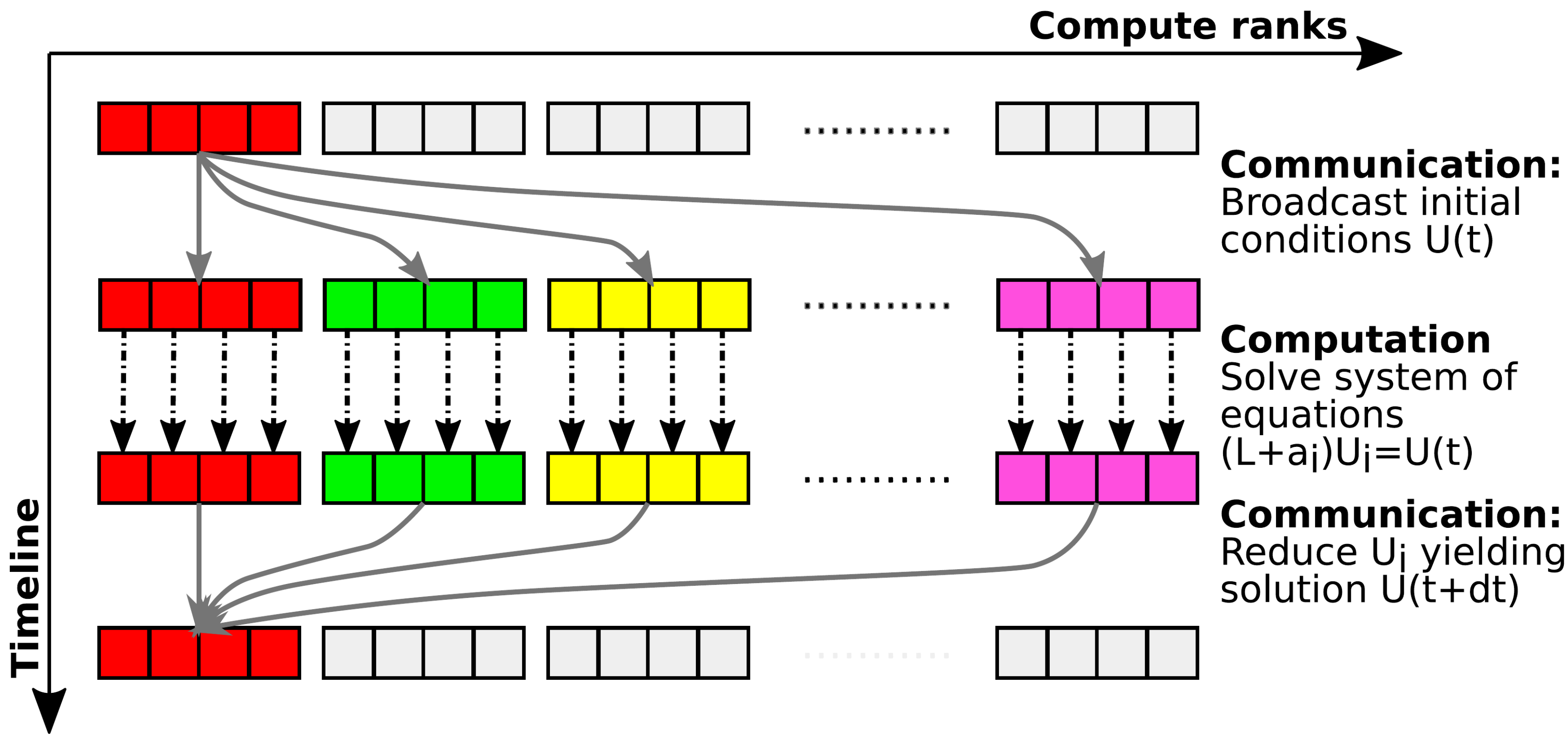
- REXI

REXI

$$e^{t\mathcal{L}} \approx \sum_{m=-M}^{M} b_m \left(t\mathcal{L} - \alpha_m\right)^{-1}$$

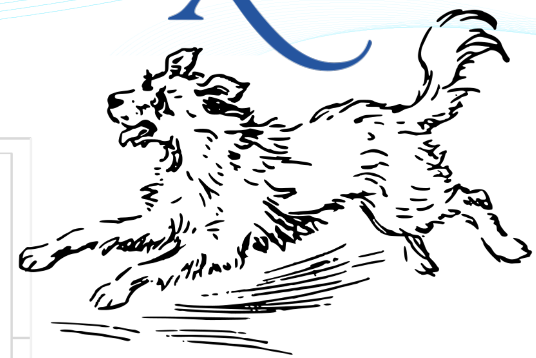
Paper: **A high-order time-parallel scheme for solving wave propagation problems via the direct construction of an approximate time-evolution operator, Haut, Babb, Martinssen, Wingate**, IMA J. Numer. Anal, 2015

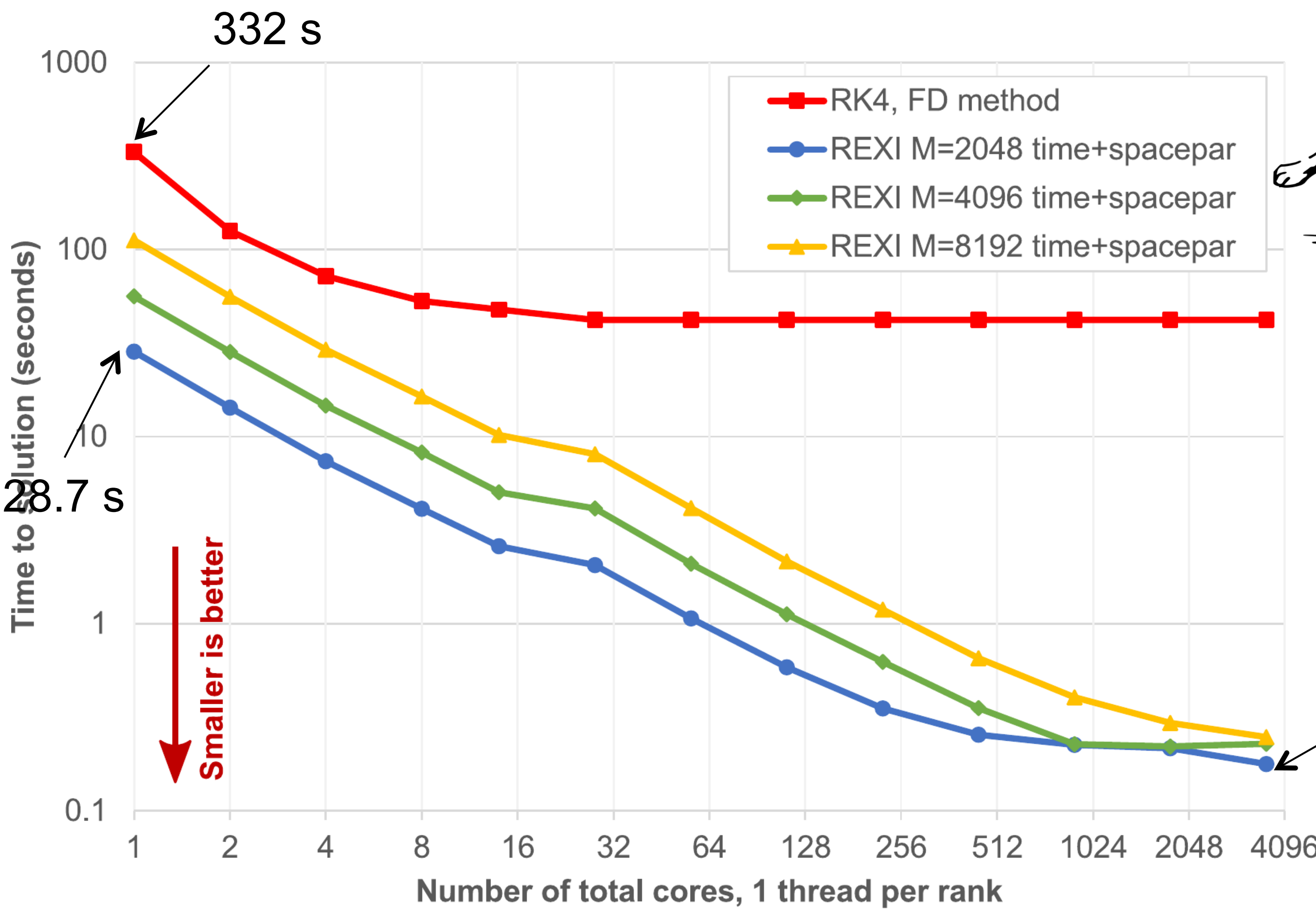
Parallelization pattern



Paper: **Beyond spatial scalability limitations with a massively parallel method for linear oscillatory problems, Schreiber, Peixoto, Haut and Wingate**, submitted to Intl J. of High Perf. Comp. IMA J. Numer. Anal, 2016

Performance: Finite Difference vs. (T)REXI
Time parallelism only





0.22 s

N=128x128

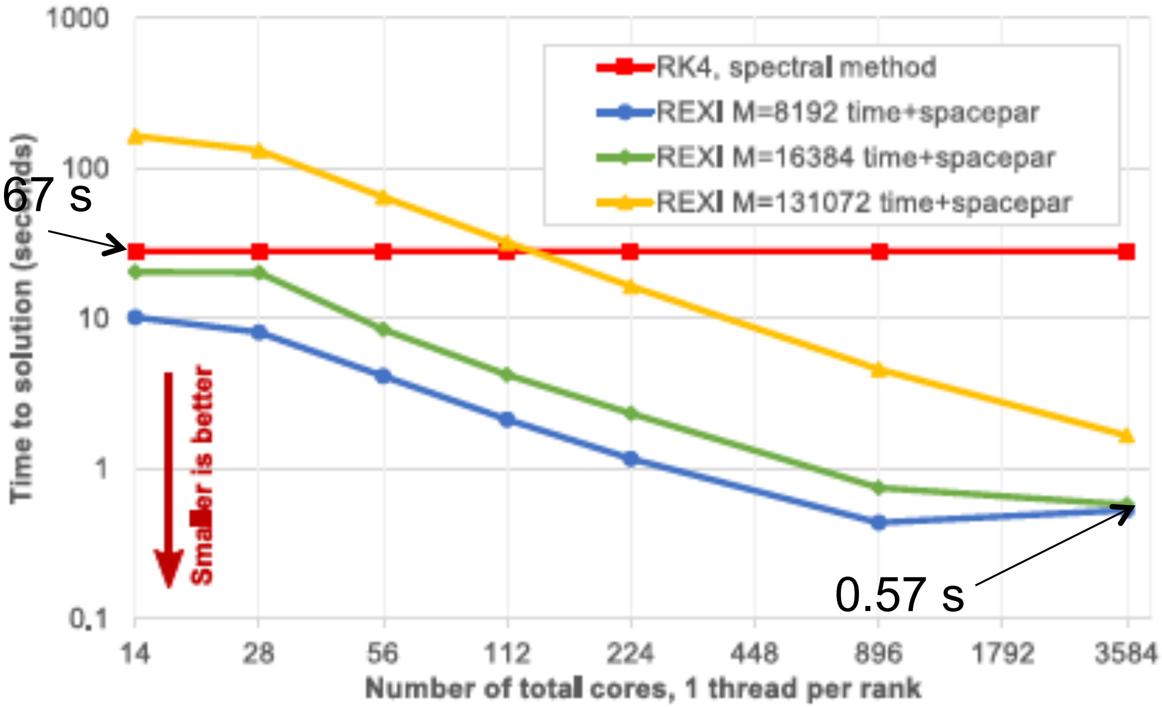
Reduction in time:
322.19/0.22 =
1503.0 X faster

Helmholtz equation is directly solved in spectral space

Computed on Linux Cluster, LRZ / Technical University of Munich

Performance: Spectral Methods vs. (T)REXI

Time parallelism only



N=128x128

Faster by a factor of 67/.57=118

Helmholtz equation is directly solved in spectral space

Computed on Linux Cluster, LRZ / Technical University of Munich

Nonlocal form in a Hilbert Space

Embid and Majda, 1996, 1997

Schochet, 1994

Klainerman and Majda 1981

$$\mathbf{u} = \begin{pmatrix} \mathbf{v} \\ \rho \end{pmatrix}$$

Hilbert Space X of vector fields **u** in L^2 that are divergence free and equipped with the L^2 norm.

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{1}{Ro} L_{Ro}(\mathbf{u}) + \frac{1}{Fr} L_{Fr}(\mathbf{u}) + \mathcal{N}(\mathbf{u}, \mathbf{u}) = \frac{1}{Re} D(\mathbf{u})$$
$$\mathbf{u}|_{t=0} = \mathbf{u_0}(\mathbf{x})$$

$$L_{Ro}(\mathbf{u}) = \begin{pmatrix} \widehat{\mathbf{z}} \times \mathbf{v} + \nabla \Delta^{-1} \omega_3 \\ 0 \end{pmatrix} \quad L_{Fr}(\mathbf{u}) = \begin{pmatrix} \widehat{\mathbf{z}}\ \rho + \nabla \Delta^{-1} (\frac{\partial \rho}{\partial z}) \\ -w \end{pmatrix}$$

$$\mathcal{N}(\mathbf{u}, \mathbf{u}) = \begin{pmatrix} \mathbf{v} \cdot \nabla \mathbf{v} - \nabla \Delta^{-1} (\nabla \cdot \mathbf{v} \cdot \nabla \mathbf{v}) \\ \mathbf{v} \cdot \nabla \rho \end{pmatrix} \quad D(\mathbf{u}) = \begin{pmatrix} \Delta \mathbf{v} \\ 1/Pr \Delta \rho \end{pmatrix}$$

Separation of time scales and the eigenfrequencies of the fast linear operator

Quasi Geostrophy $Ro \to 0 \quad Fr \to 0 \quad Fr/Ro = f/N = \text{finite}$

$$\omega(\mathbf{k}) = \pm \frac{(Fr^2 m^2 + Ro^2 |\mathbf{k_H}|)^{1/2}}{RoFr|\mathbf{k}|}$$

$$\omega(\mathbf{k}) = 0 \quad \text{(double)}$$

Two kinds of frequencies:

1) "slow" zero frequency for all **k** which contribute to the potential vorticity
2) "fast" dispersive waves with zero potential vorticity

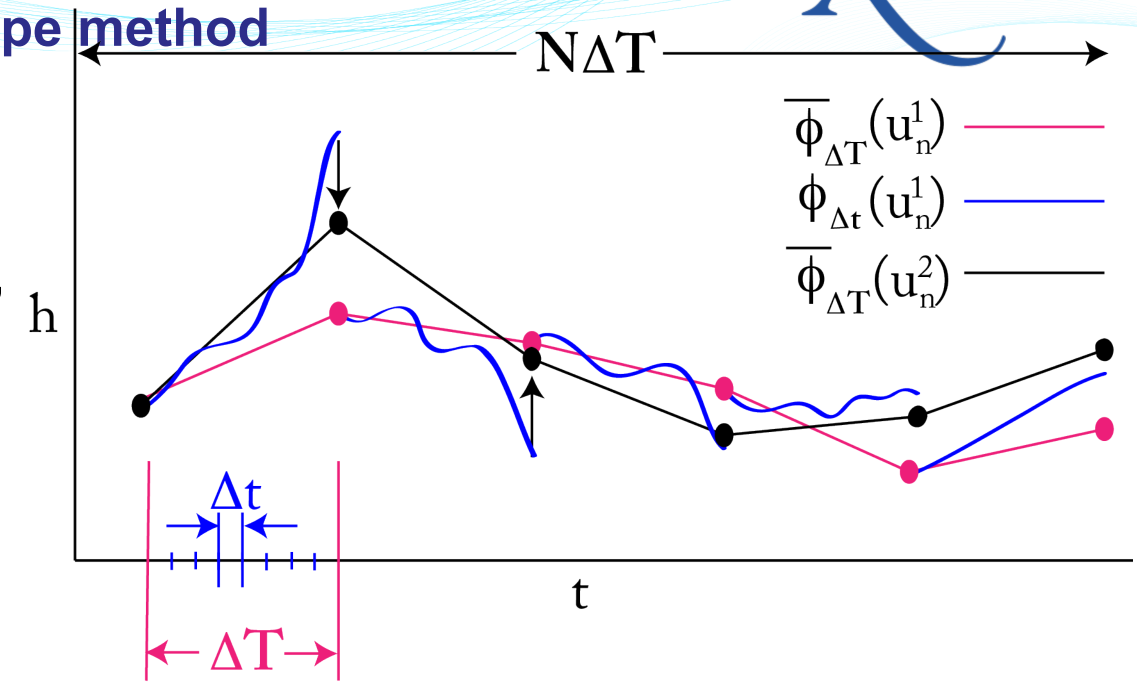
Oscillatory Stiffness in the PDE

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{1}{\epsilon}\mathcal{L}\mathbf{u} + \mathcal{N}\left(\mathbf{u},\mathbf{u}\right) = \mathcal{D}\mathbf{u}, \quad \mathbf{u}\left(0\right) = \mathbf{u}_0,$$

- The $\epsilon^{-1}\mathcal{L}$ skew-Hermitian operator results in temporal oscillations on a time scale of $\mathcal{O}\left(\epsilon\right)$
- Standard numerical time-stepping methods must use time steps $\Delta t = \mathcal{O}\left(\epsilon\right)$ for accuracy.

Parareal – shooting-type method

- Nievergelt (1964)
- Lions, Maday, Turinici, (2001) 'Parareal'

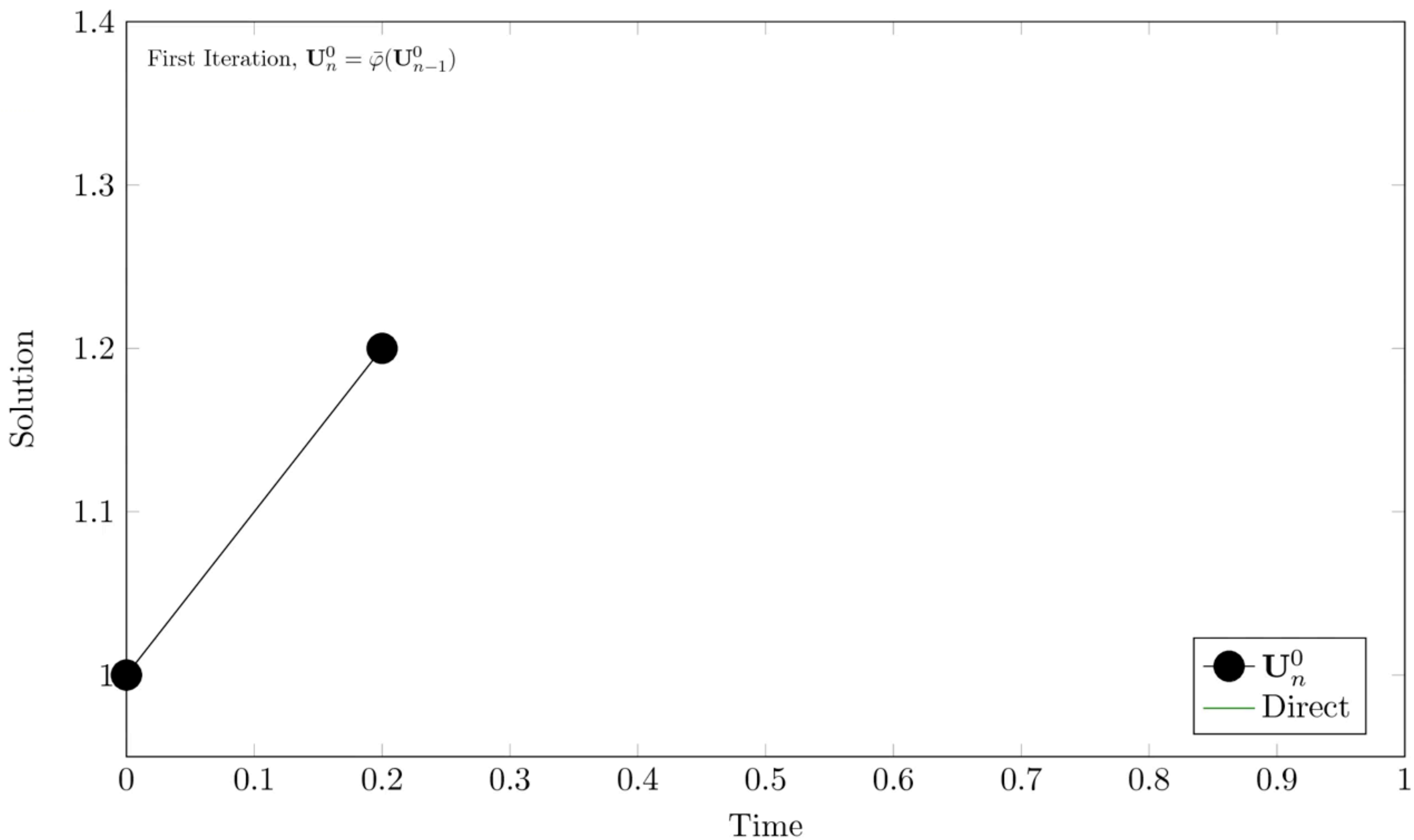


$$U_n^k = \overline{\phi}_{\Delta T}(U_{n-1}^k) + \left(\phi_{\Delta T}(U_{n-1}^{k-1}) - \overline{\phi}_{\Delta T}(U_{n-1}^{k-1})\right)$$

Newton Institute Lecture: **Time-parallel algorithms for weather prediction and climate simulation; Jean Côté** under the AMM program in September 2012

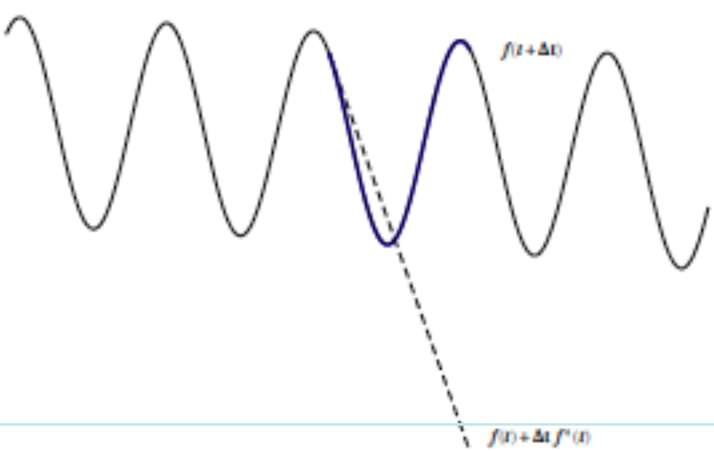
Paper: **Nonlinear Convergence Analysis of the Parareal Method, Gander and Hairer**, Domain Decomposition Methods in Science and Engineering XVII, Springer 2008

Demonstration of parareal algorithm



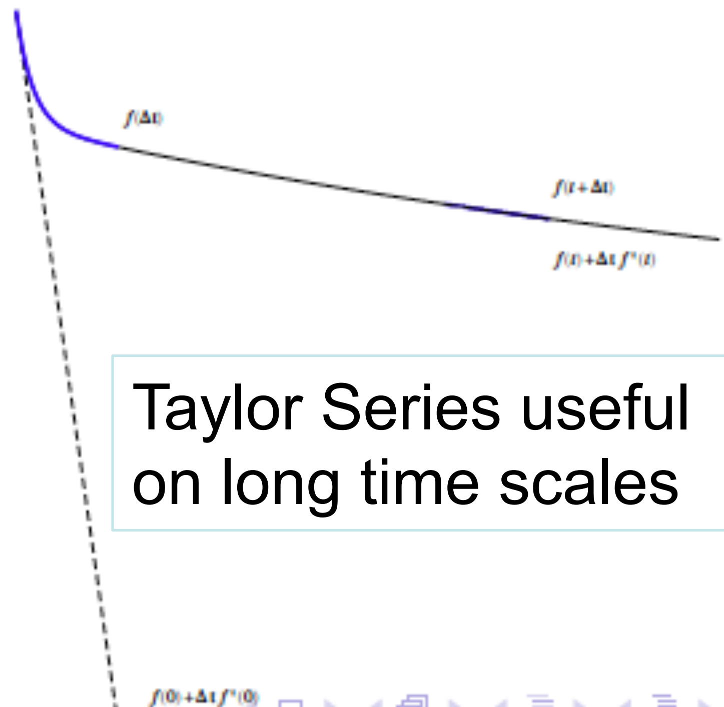
For many PDEs that govern physics there are two main types of limiting cases : slow singular limits and fast singular limits (multiple time scales)

FAST singular limit



Taylor Series only useful on small time scales

SLOW singular limit



Taylor Series useful on long time scales

The early days of numerical weather prediction

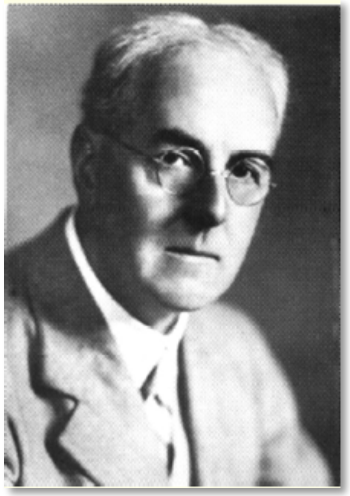
UNIVERSITY OF EXETER

Slow Dynamics and Asymptotic theory

L.F. Richardson in (1922) - using ‘computers’

Charney (1948) and Charney (1950) - derived ‘slow’ or Quasi-Geostrophic (QG) equations (important conceptual model)

Charney and Phillips (1953) – the first realistic numerical weather prediction using the QG eqs



L.F. Richardson

- **There is an important notion that the fast frequencies get ‘swept’.**

- Important counter example: some of the fast motions are ‘in resonance’ and collaborate to create ‘slow’ motions. Example, The **Stepwise Precession of the Resonant Swinging Spring**, “P. Lynch and D. Holm, 2002

J. von Neumann

J. Charney

Does the real atmosphere behave asymptotically?

Slow Manifolds (nonlinear normal mode initialization, center manifolds, dynamical systems, etc)

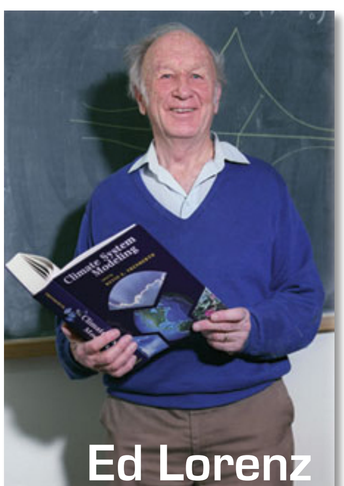
Machanauer (1977), Baer (1977), Tribbia (1979), etc

Leith, **Nonlinear Normal Mode Initialization and Quasi-Geostrophic Theory** (1980)

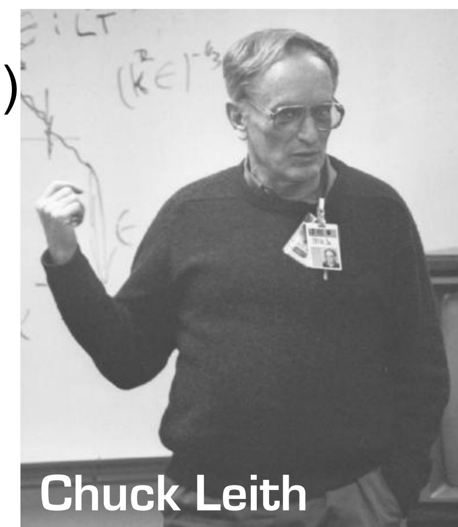
Lorenz, **On the Existence of a Slow Manifold** (1986)

Lorenz and Krishnamurthy, **On the non-Existence of the Slow Manifold** (1987)

Lorenz, **The Slow Manifold – what is it**? (1991)



The dynamics is not asymptotic and is not accurate for numerical weather prediction and climate simulations. Non-invariant manifold.



UNIVERSITY OF EXETER

A superlinear parareal method: complexity analysis and error bounds for APINT

For fixed k and decreasing epsilon superlinear convergence:

Banach space characterising the regularity

THEOREM 1. *Assuming that* $\mathbf{u}_0 = \mathbf{u}(T_0) \in B_{j+k+1}$, *the error,* $\mathbf{u}(T_n) - \mathbf{U}_n^k$, *after the kth parareal iteration is bounded by*

$$\left\| \mathbf{u}(T_n) - \mathbf{U}_n^k \right\|_{B_j} \leq C_{k,j} \left(\Delta T^p + \epsilon \right) \left(\Delta T^p + \frac{\epsilon}{\Delta T} \right)^k \left\| \mathbf{u}_0 \right\|_{B_{k+j+1}},$$

where $C_{k,j}$ *is a constant that depends only on the constants* C_m, $m = 0, 1, \ldots, k+j$.

Constants

Order of the time stepping method

Example:

if you choose $\Delta t \approx \epsilon^{\frac{1}{2}}$ the error scales like $\epsilon^{k+\frac{1}{2}}$

Papers:

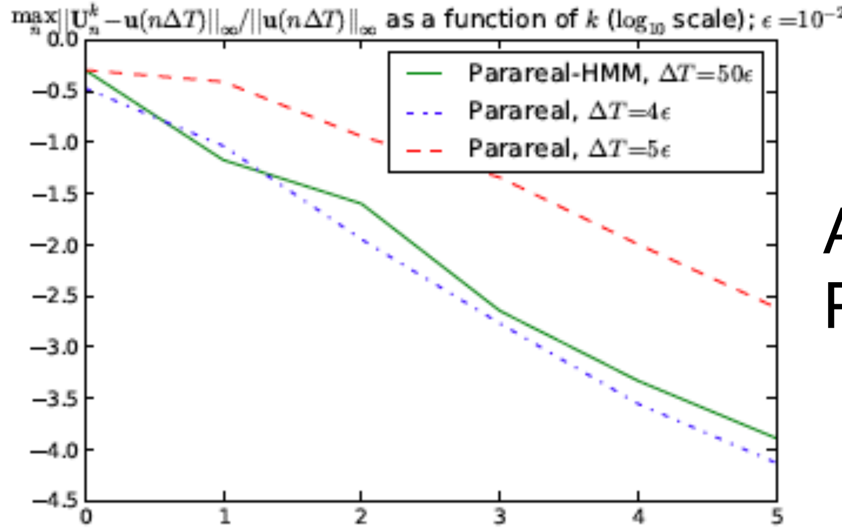
An Asymptotic Parallel-in-Time Method for Highly Oscillatory PDEs, T. Haut, B. Wingate, *SIAM Journal of Scientific Computing*, 2014

Key Proofs:

On the convergence and stability of the parareal algorithm to solve partial differential equations, Bal, In Domain Decomposition Methods in Science and Engineering, Springer, pp 425 2005 (stability too)

Nonlinear convergence analysis of the parareal…, Gander & Hairer, 2008 (superlinear convergence)

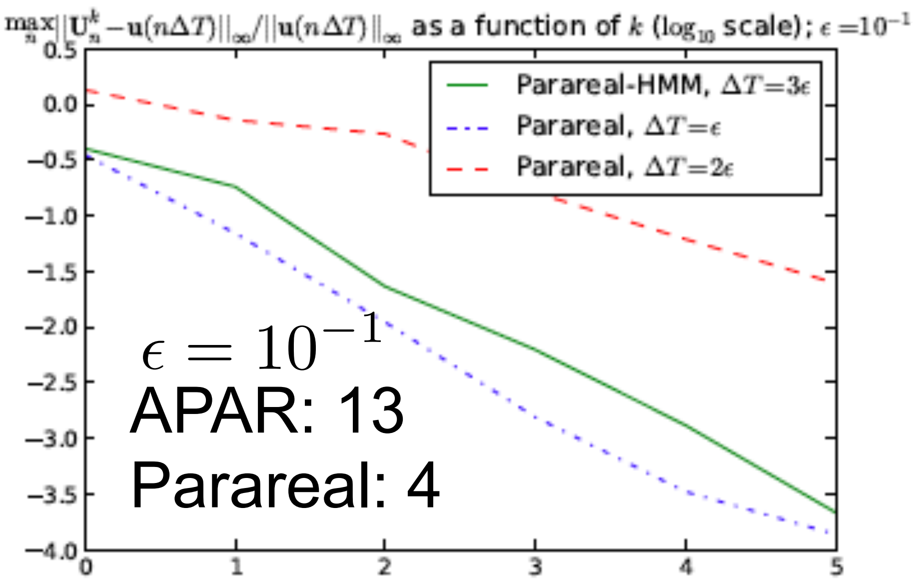
Maximum relative error versus number of iterations

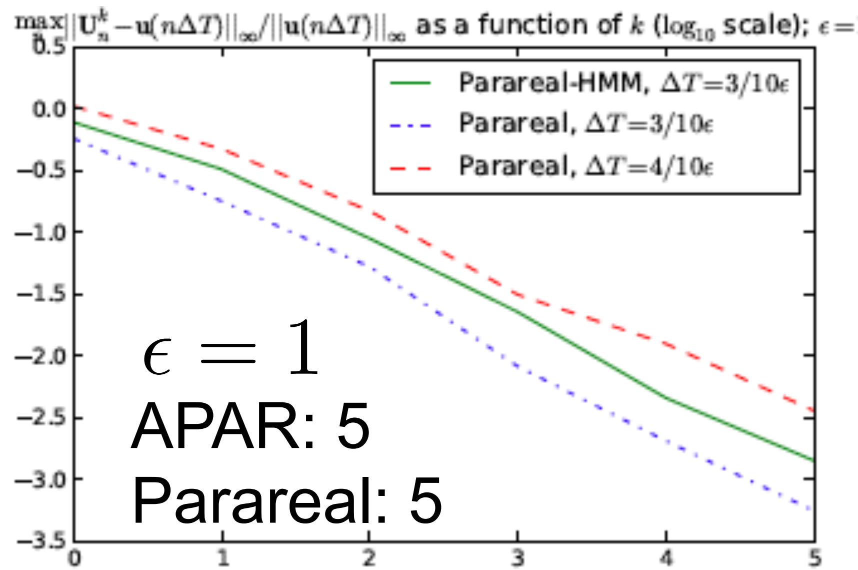


$$\epsilon = 10^{-2}$$

APAR : 100

Parareal: 10





Oscillatory Stiffness in the PDE

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{1}{\epsilon}\mathcal{L}\mathbf{u} + \mathcal{N}\left(\mathbf{u},\mathbf{u}\right) = \mathcal{D}\mathbf{u}, \quad \mathbf{u}\left(0\right) = \mathbf{u}_0,$$

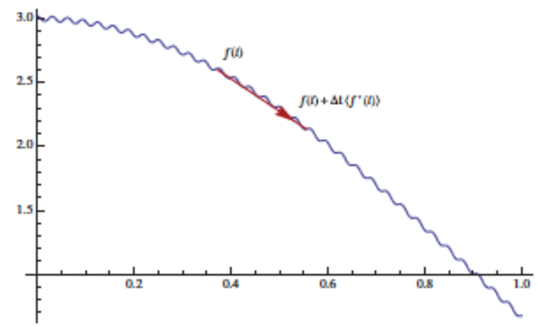


FIG. 3. *Schematic depiction of the moving time average.*

Setting the dissipation to zero,

$$\mathbf{u}\left(t\right) = e^{-t/\epsilon\mathcal{L}}\mathbf{v}\left(t\right)$$

$$\frac{\partial \mathbf{v}}{\partial t} + e^{t/\epsilon\mathcal{L}}\mathcal{N}\left(e^{-t/\epsilon\mathcal{L}}\mathbf{v}(t), e^{-t/\epsilon\mathcal{L}}\mathbf{v}\left(t\right)\right) = 0$$

$$\boxed{\frac{\partial \mathbf{v}}{\partial t} = \mathcal{O}\left(1\right)} \qquad \boxed{\frac{\partial^2 \mathbf{v}}{\partial t^2} = \mathcal{O}\left(\frac{1}{\epsilon}\right)}$$

Klainerman & Majda, Schochet, Embid, & others

....but goes all the way back to Bolgoliubov & Mitropolskiy 1961

$$\frac{\partial \mathbf{u^0}}{\partial \tau} + \mathcal{L}\left(\mathbf{u^0}\right) = 0$$

$$\boxed{\mathbf{u}^0(\mathbf{x}, t, \tau) = e^{-\tau\mathcal{L}}\overline{\mathbf{u}}(\mathbf{x}, t) + \ O(\epsilon)}$$

Where $\overline{\mathbf{u}}$ solves:

$$\boxed{\frac{\partial \overline{\mathbf{u}}}{\partial t}(\mathbf{x}, t) = -\lim_{\tau\to\infty}\frac{1}{\tau}\int_0^\tau e^{s\mathcal{L}}\left(\mathcal{N}(e^{-s\mathcal{L}}\overline{\mathbf{u}}, e^{-s\mathcal{L}}\overline{\mathbf{u}})\right)\,\mathrm{ds}}$$

These ideas are the foundation for the locally asymptotic parallel-in-time numerical method

An asymptotic method-of-multiple scales in time (another way to derive Quasi Geostrophy is a singular perturbation in time):

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{1}{\epsilon}\mathcal{L}\mathbf{u} + \mathcal{N}(\mathbf{u},\mathbf{u}) = \mathcal{D}\mathbf{u}, \quad \mathbf{u}(0) = \mathbf{u}_0,$$

There exits a finite [0,T], T independent of ϵ :

$$\mathbf{u}(t) = e^{-t/\epsilon\mathcal{L}}\ \overline{\mathbf{u}}(t) + \mathcal{O}(\epsilon)$$

$$\frac{\partial \overline{\mathbf{u}}}{\partial t} + \overline{\mathcal{N}}(\overline{\mathbf{u}},\overline{\mathbf{u}}) = \overline{\mathcal{D}}\overline{\mathbf{u}}, \quad \overline{\mathbf{u}}(0) = \mathbf{u}_0,$$

where

$$\overline{\mathcal{D}}\ \overline{\mathbf{u}}(t) = \lim_{T\to\infty}\frac{1}{T}\int_0^T \left(e^{s\mathcal{L}}\mathcal{D}\ e^{-s\mathcal{L}}\right)\overline{\mathbf{u}}(t)\ \mathrm{ds}$$

$$\overline{\mathcal{N}}(\overline{\mathbf{u}}(t),\overline{\mathbf{u}}(t)) = \lim_{T\to\infty}\frac{1}{T}\int_0^T e^{s\mathcal{L}}\left(\mathcal{N}(e^{-s\mathcal{L}}\overline{\mathbf{u}}, e^{-s\mathcal{L}}\overline{\mathbf{u}})\right)\mathrm{ds}$$

Embid and Majda, 1996, 1998, Majda and Embid, 1998, Schochet, 1994, Klainerman and Majda 1981, Wingate, Embid,, Cerfon-Holme Taylor, 2011

Compare coordinate transformation to the asymptotic solution

- Coordinate transformation

$$\mathbf{u}(t) = e^{-t/\epsilon\mathcal{L}}\mathbf{v}(t)$$

$$\frac{\partial \mathbf{v}}{\partial t} + e^{t/\epsilon\mathcal{L}}\mathcal{N}\left(e^{-t/\epsilon\mathcal{L}}\mathbf{v}(t), e^{-t/\epsilon\mathcal{L}}\mathbf{v}(t)\right) = 0$$

Step back from the limit
As tau goes to infinity.

- Asymptotic Solution

$$\mathbf{u}^0(\mathbf{x}, t, \tau) = e^{-\tau\mathcal{L}}\overline{\mathbf{u}}(\mathbf{x}, t) \quad + \quad O(\epsilon)$$

$$\frac{\partial \overline{\mathbf{u}}}{\partial t}(\mathbf{x}, t) = -\lim_{\tau\to\infty}\frac{1}{\tau}\int_0^\tau e^{s\mathcal{L}}\left(\mathcal{N}(e^{-s\mathcal{L}}\overline{\mathbf{u}}, e^{-s\mathcal{L}}\overline{\mathbf{u}})\right)\mathrm{ds}$$

Over a few oscillations we approximate the time integral using HMM:

slow time scale

$$
\begin{aligned}
\overline{\mathcal{N}}\left(\overline{\mathbf{u}}(t),\overline{\mathbf{u}}(t)\right) &= \lim_{T\to\infty}\frac{1}{T}\int_0^T e^{s\mathcal{L}}\mathcal{N}\left(e^{-s\mathcal{L}}\overline{\mathbf{u}}(t),\, e^{-s\mathcal{L}}\overline{\mathbf{u}}(t)\right)ds \\
&\approx \frac{1}{T_0}\int_0^{T_0}\rho\left(\frac{s}{T_0}\right)e^{s\mathcal{L}}\mathcal{N}\left(e^{-s\mathcal{L}}\overline{\mathbf{u}}(t),\, e^{-s\mathcal{L}}\overline{\mathbf{u}}(t)\right)ds \\
&\approx \frac{1}{M}\sum_{m=0}^{M-1}\rho\left(\frac{s_m}{T_0}\right)e^{s_m\mathcal{L}}\mathcal{N}\left(e^{-s_m\mathcal{L}}\overline{\mathbf{u}}(t),\, e^{-s_m\mathcal{L}}\overline{\mathbf{u}}(t)\right)
\end{aligned}
$$

fast time scale

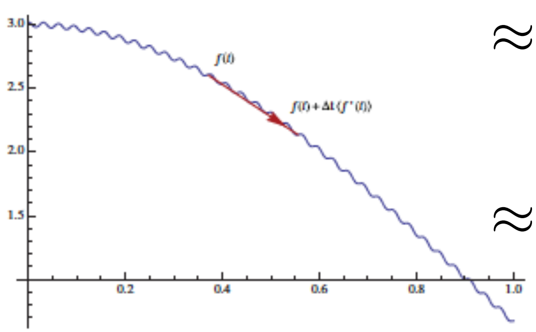


FIG. 3. *Schematic depiction of the moving time average.*

- The sum is fully parallelisable.
- The sum is over the nonlinear **operator**, not the solution itself!
- Resolving the near-resonant frequencies appears to be important for accuracy

Interacting wave frequencies

Near resonances in nonlinearity of the PDE

If we look at the nonlinear term expanded in terms of the eigenfunctions of the linear operator :

$$v(\mathbf{x},t) = \sum_{\mathbf{k}\in\mathbb{Z}} \sum_{\alpha=-1}^{1} \sigma_{\mathbf{k}}^{\alpha}(t) e^{i\mathbf{k}\cdot\mathbf{x}} \mathbf{r}_{\mathbf{k}}^{\alpha}$$

Look at the nonlinear term:

$$e^{t/\epsilon\mathcal{L}} \mathcal{N}(e^{-t/\epsilon\mathcal{L}} v(x,t), e^{-t/\epsilon\mathcal{L}} v(x,t)) =$$

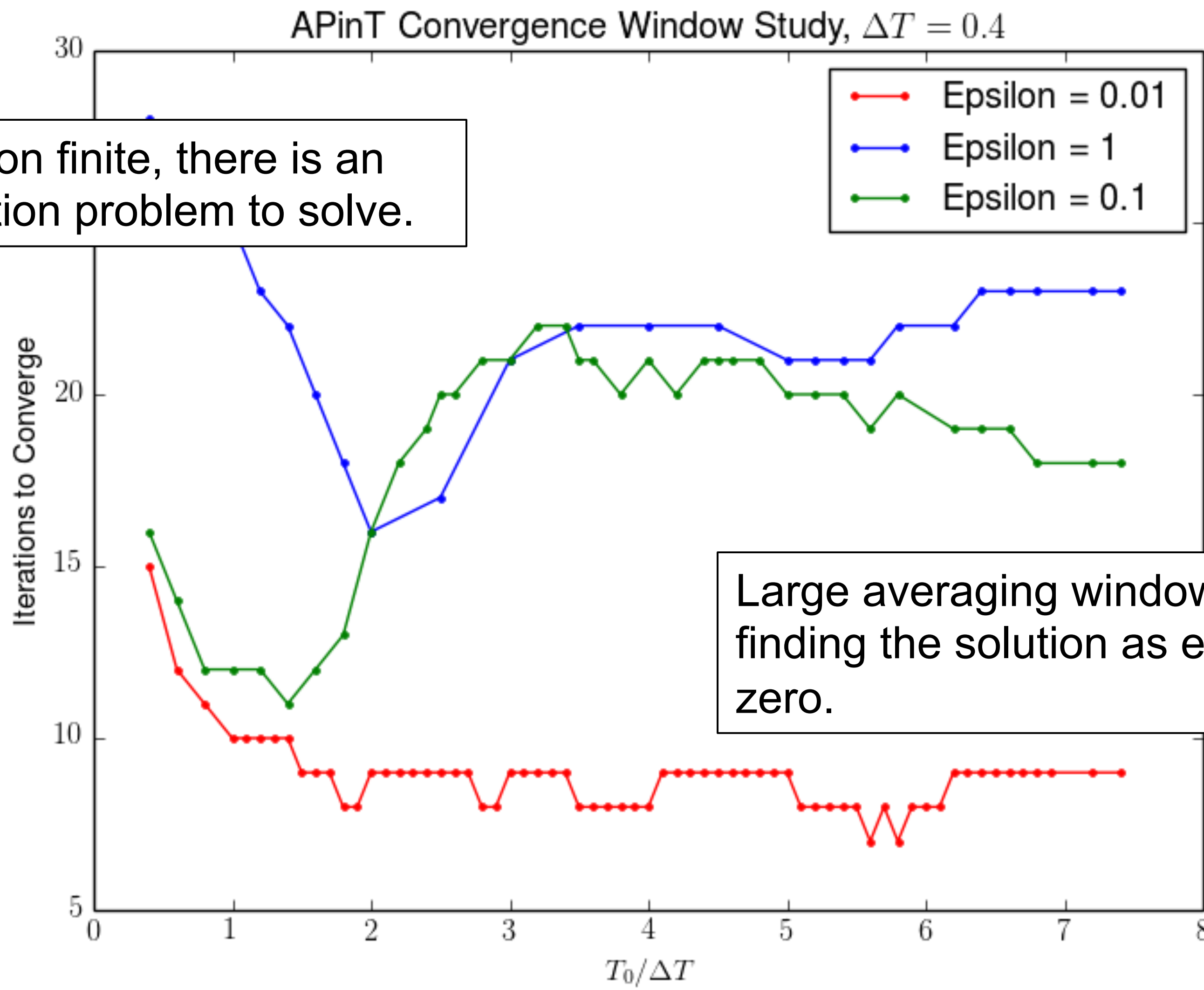
$$\sum_{\mathbf{k}\in\mathbb{Z}} \sum_{\alpha=-1}^{1} \left(\sum_{\mathbf{k_1}+\mathbf{k_2}=\mathbf{k}} \sum_{\alpha_1,\alpha_2=-1}^{1} C_{\mathbf{k_1},\mathbf{k_2},\mathbf{k}}^{\alpha_1,\alpha_2,\alpha_3} \sigma_{\mathbf{k_1}}^{\alpha_1}(t)\ \sigma_{\mathbf{k_2}}^{\alpha_2}(t)\ e^{i\left(\mathbf{k}\cdot\mathbf{x} - (\omega_{\mathbf{k_1}}^{\alpha_1} + \omega_{\mathbf{k_2}}^{\alpha_2} - \omega_{\mathbf{k}}^{\alpha})t/\epsilon\right)} \right) \mathbf{r}_{\mathbf{k}}^{\alpha}$$

Something interesting happens when there are near resonances:

$$|\omega_{\mathbf{k_1}}^{\alpha_1} + \omega_{\mathbf{k_2}}^{\alpha_2} - \omega_{\mathbf{k}}^{\alpha}| \leq \epsilon$$

Near-resonance when epsilon not small

Adam Peddle's thesis at University of Exeter



For epsilon finite, there is an optimisation problem to solve.

Large averaging windows are like finding the solution as epsilon goes to zero.

We finally have a convergence proof for epsilon finite – a few slides from now

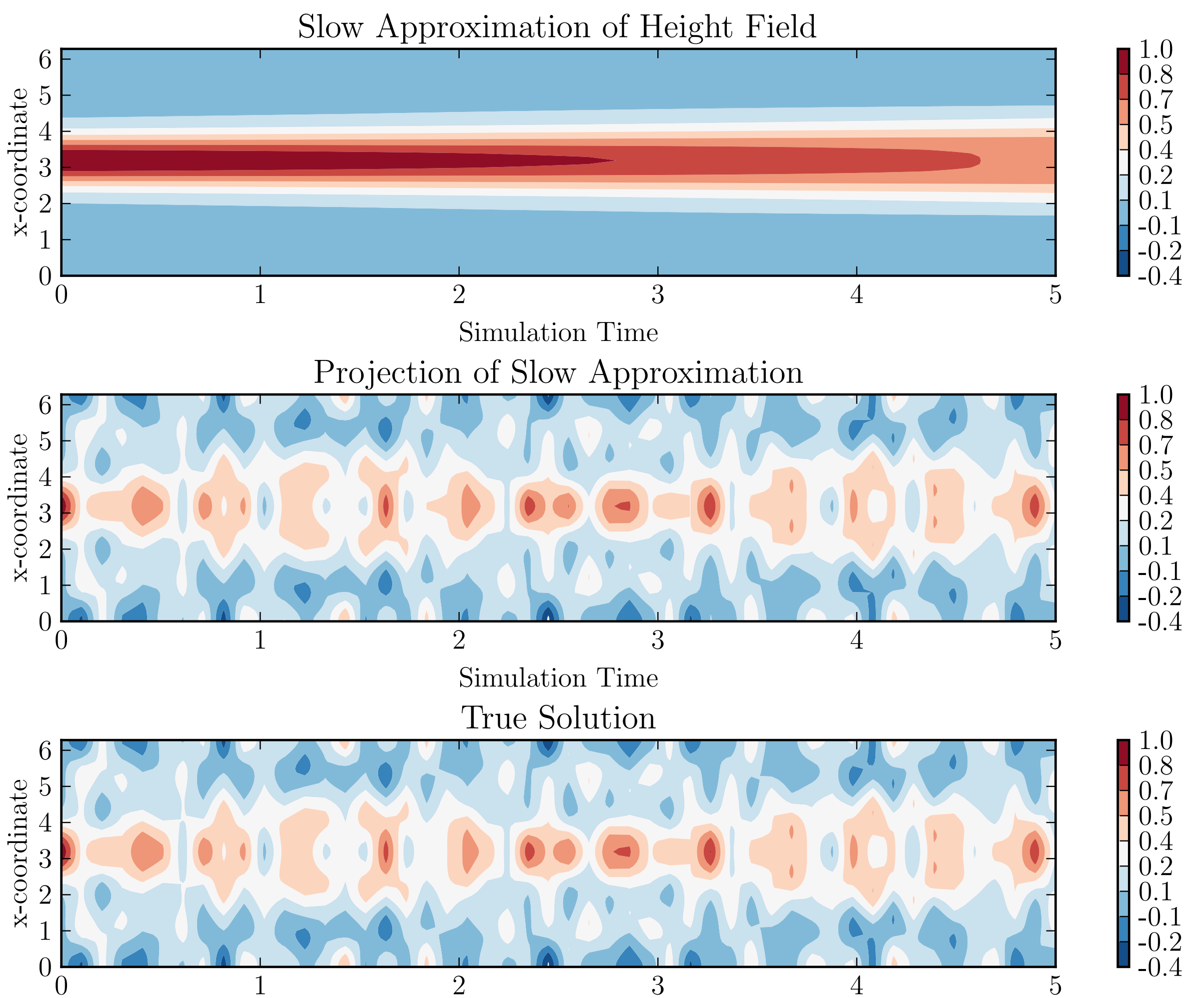
What happens if the dynamics isn't asymptotic

The coarse propagator is:

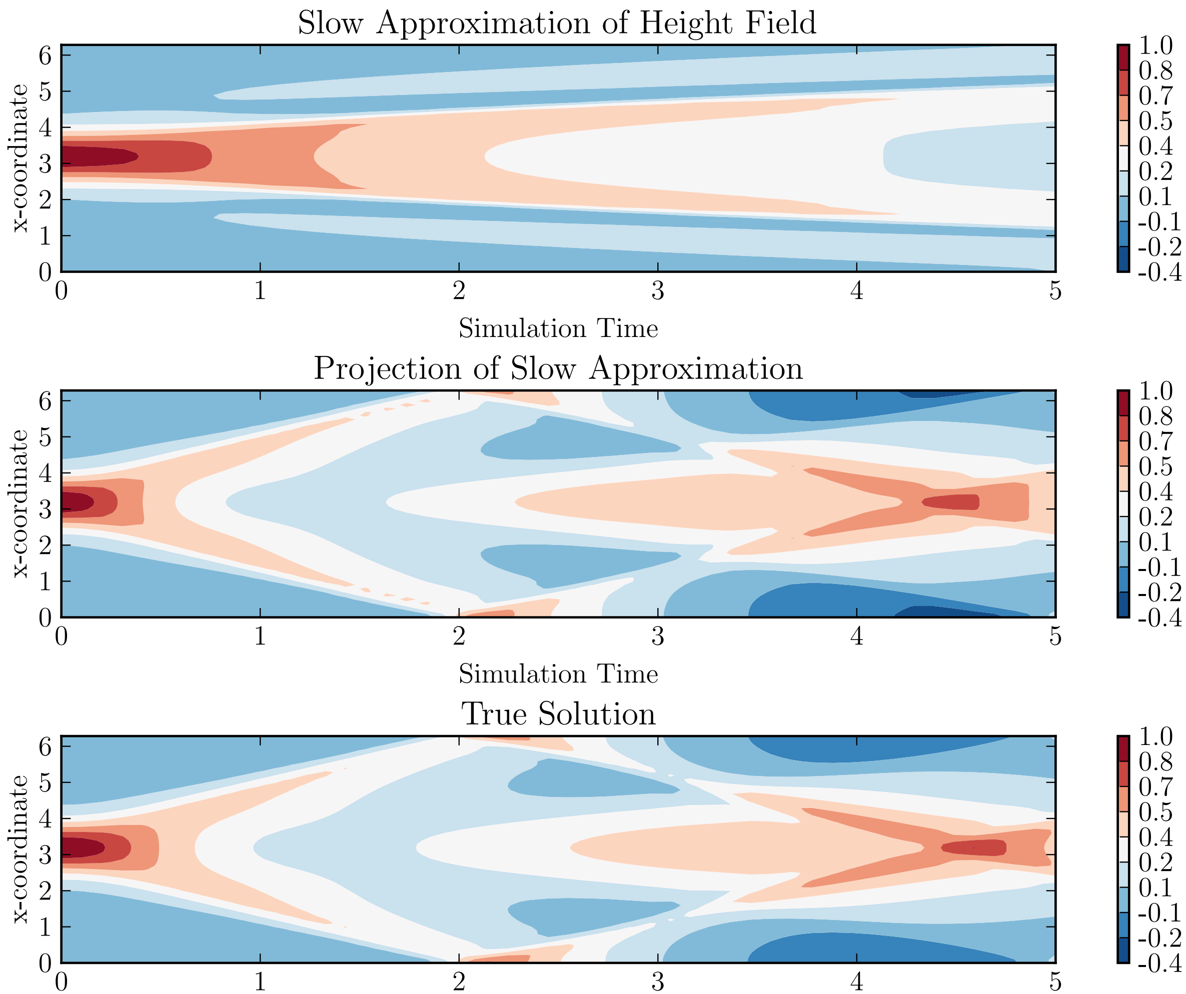
$$\mathbf{u} = e^{-t\mathcal{L}}\overline{\mathbf{u}}$$

$$\frac{\partial \overline{\mathbf{u}}}{\partial t} \approx \frac{1}{T_0}\int_0^{T_0} \rho\left(\frac{s}{T_0}\right) e^{s\mathcal{L}} \mathcal{N}\left(e^{-s\mathcal{L}}\overline{\mathbf{u}}\,(t)\,, e^{-s\mathcal{L}}\overline{\mathbf{u}}\,(t)\right) ds$$

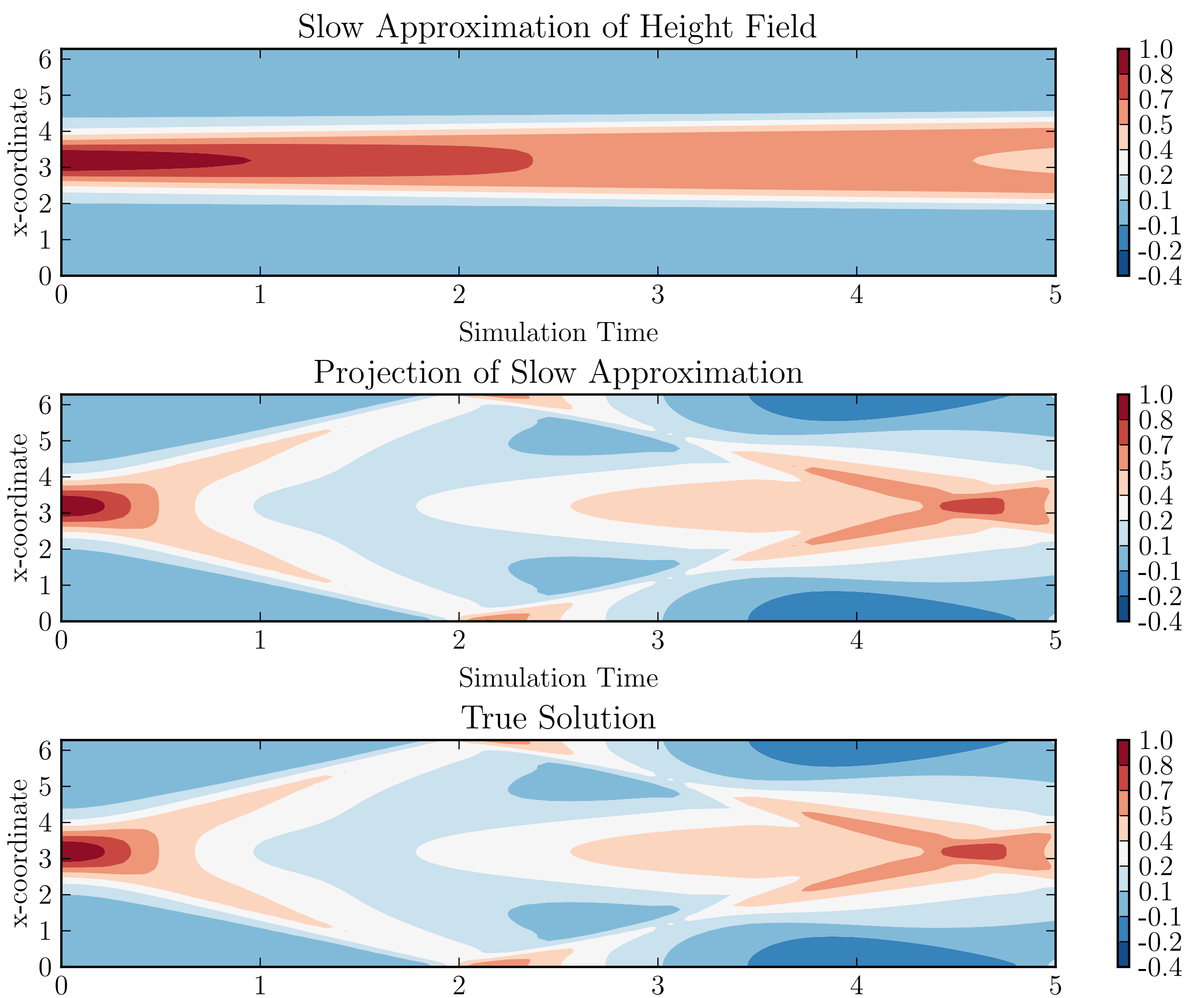
ϵ = .1, superlinear convergence for parareal



ϵ = 1, poor guess for To – takes longer to converge

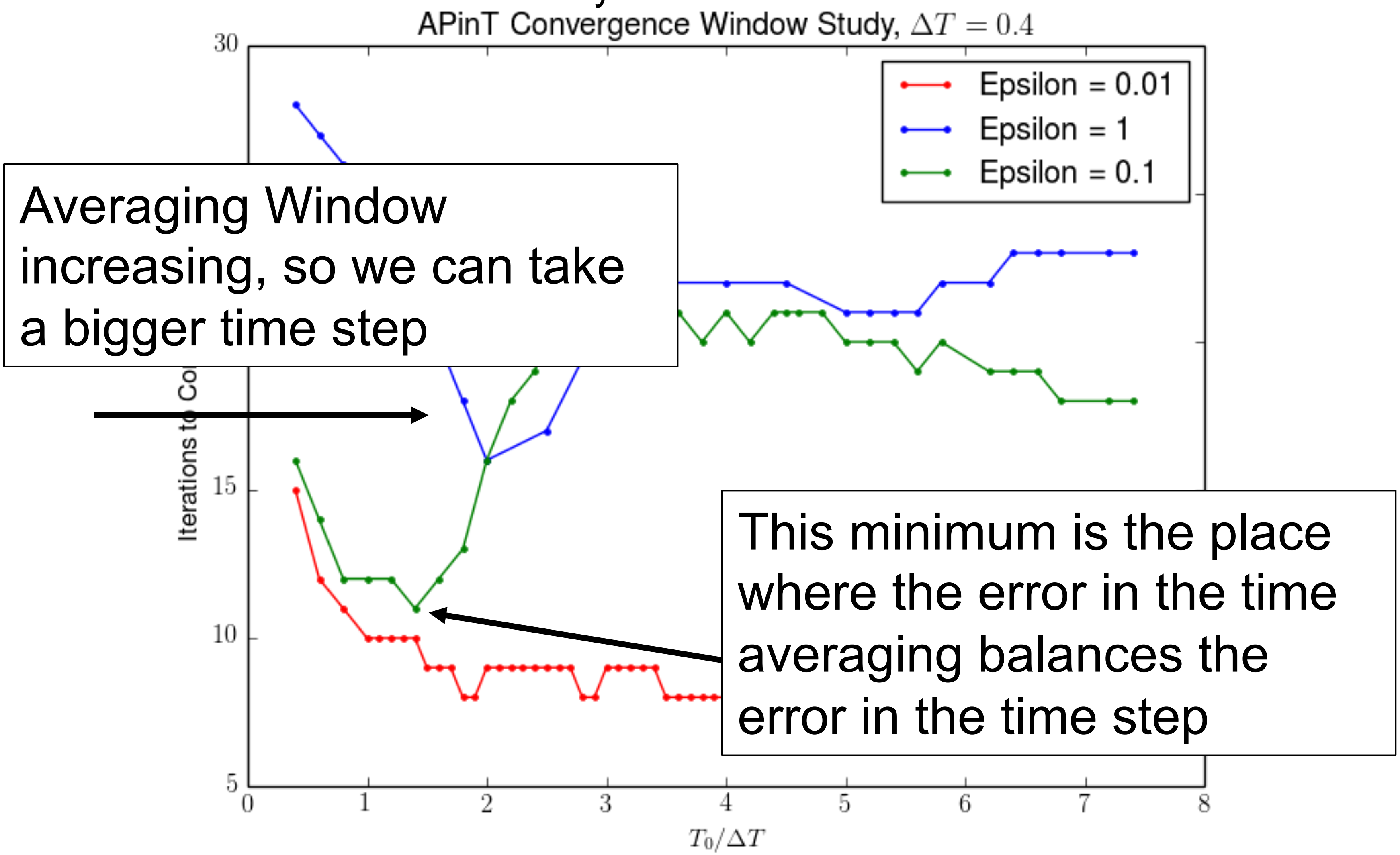


ϵ = 1, good estimate for To (faster parareal convergence)



New theorem for when epsilon finite

Adam Peddle's thesis at University of Exeter



A new theorem and optimization problem for epsilon finite

Adam Peddle's thesis at University of Exeter
In collaboration also with Terry Haut from LLNL

For a pth order time-stepping method, and η is To

$$||\mathbf{u}(T_n) - \mathbf{U}_n^k||_{B_j} \leq MC_g \left(C_1 \Delta T^{p+1} \epsilon \Lambda(\eta) + (C_2 + C_3 \epsilon)\epsilon\eta\right)^{k+1} ||\mathbf{u}_0||$$

- The 3 waves near-resonances play a key role
- **But they are not only to do with the scale separation, they are the near-resonant set relevant to the time step ΔT**
- This minimum is where the equations become locally regularized (less stiff!) over some interval η such that over ΔT

$$\eta|\lambda_n| < \delta$$

Convergence for any ε

$$C_1 \Delta T^{p+1} \epsilon \Lambda(\eta) + (C_2 + C_3 \epsilon)\epsilon\eta \leq 1$$

- η is the averaging window
- ΔT is the coarse time step
- ε is the time scale separation
- Λ is

$$\Lambda(\eta) = \max_{x \in \mathbb{R}} \lambda_n^p \int_0^1 \rho(s) e^{i\lambda_n \eta \Delta T s} \mathrm{d}s$$

The role of average of the nonlinear operator averaging

$$\Lambda(\eta) = \max_{x \in \mathbb{R}} \lambda_n^p \int_0^1 \rho(s) e^{i\lambda_n \eta \Delta T s} \mathrm{d}s$$

- This is a measure of the degree to which the averaging can mitigate the stiffness from oscillations.
- When λ_n is large (for highly oscillatory problems) it creates large gradients in the fluid that require a small timestep.
- In contrast, the integral tends to zero with $\rho(s)$ the ‘smooth kernel’ for the average.
- In summary, this term tells us how the averaging of the nonlinear operator regularises the solution – it achieves a lower magnitude than λ_n itself.

To demonstrate a convergent parareal algorithm for any epsilon

$$\eta = \frac{\Delta t}{\epsilon^s} \qquad \text{for} \qquad 0 < s < 1$$

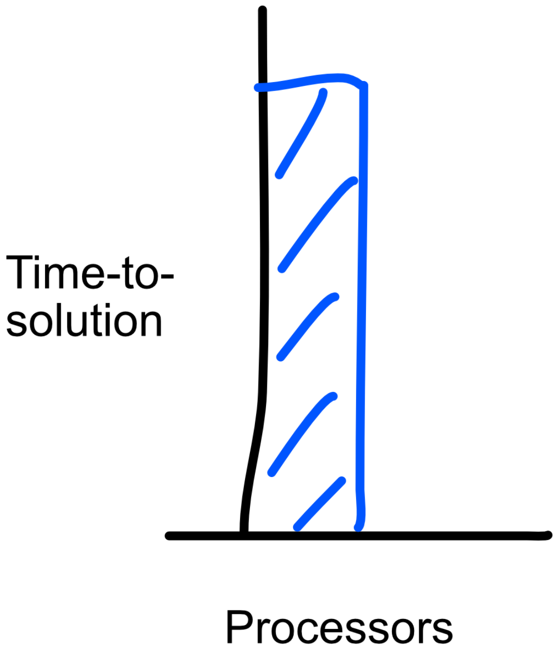
$$C_1 \Delta T^{p+1} \epsilon \Lambda(\frac{\Delta t}{\epsilon^s}) + C_2 \epsilon^{1-s} \Delta T + C_3 \epsilon^{2-s} \Delta T \leq 1$$

- For $\epsilon \to 0$, this also goes to zero for any s.
- For $\epsilon \to 1$, $\Lambda\left(\Delta T/\epsilon^s\right)$ is bounded.

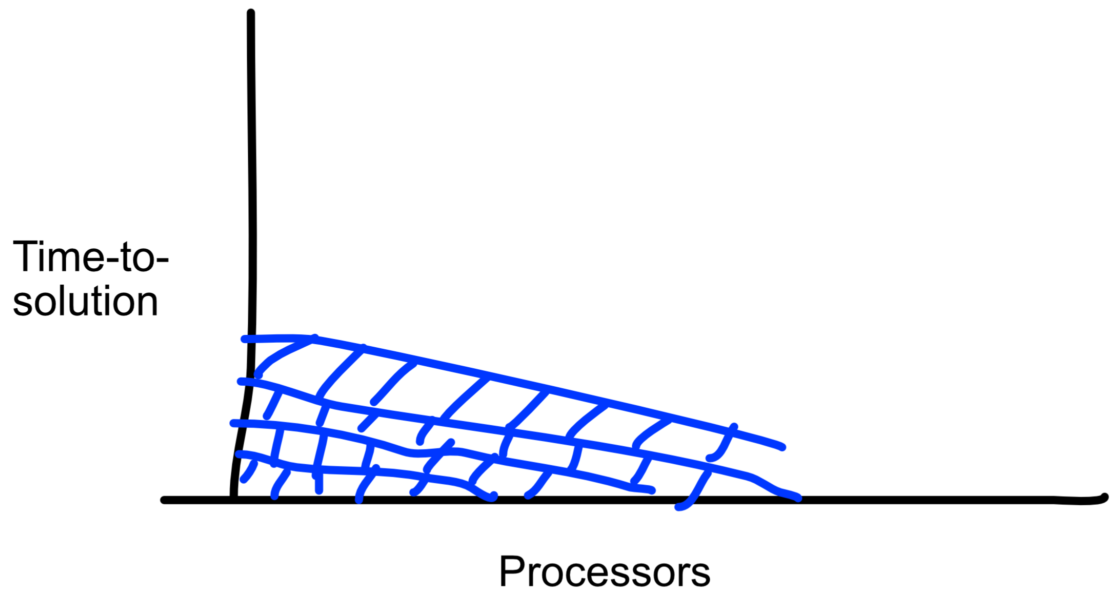
Therefore, we can then solve an optimisation problem to find a value of the averaging window that gives the minimum value and the parareal algorithm convergence for finite ε.

Adam Peddle will be presenting this at 10:00 in Room 211

How close are we to knowing this? Not very close!



"Monolithic serial"



"Sliding Window time-parallelism"

What about climate, weather, and exascale computing?

- Realisable exascale (next 5 year) climate and weather prediction will be ports (CS&E) of current models. Maybe some RDIC and exponential integrators?
- The ports will drastically underuse the available compute power of exascale machines, this will lead to more statistical scientific questions (ensemble science)
- Other science problems that can use the machines more efficiently will make enormous gains in understanding.
- While the above is happening, CS&E will be building new (but simpler) models from scratch that contain more ways of using compute resources, but they will be simple – spheres and boxes, with no land mass. Example: using firedrake
- Time to solution for climate scientists? The young people will adopt the new models for basic science, leading to their first use as sicentific models.

Related Minisymposia

Parallel in time

Part I will be on Tuesday, February 28 from 9:10 AM to 10:50 AM in Room 211
Part II will be on Tuesday, February 28 from 1:30 PM to 3:10 PM in Room 211

MS269 Advancing Cross-Cutting Ideas for Computational Climate Science

Will be on Tuesday, February 28 from 4:25 PM to 6:25 PM in Room 301

MS294 Finite Element Methods for Weather, Oceans and Climate

Part I wiil be on Friday, March 3 from 9:10 AM to 10:50 AM in Crystal AF - 1st Fl
Part Ii will be on Friday, March 3 from 11:20 AM to 1:00 PM in Crystal AF - 1st Fl