

An asymptotic-parallel-in-time method for weather and climate prediction

Beth Wingate

University of Exeter

Terry Haut

Los Alamos National Laboratory

Jared Whitehead

Brigham Young University

Outline

- o A brief introduction to parallel-in-time
- o Describe a locally asymptotic slow solution
- o Show some results from the shallow water equations
- o Summarize and lead into highly parallel linear propagators

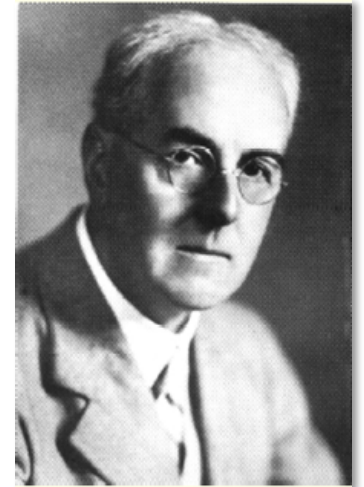
The early days of computational science

Slow Dynamics

L.F. Richardson in (1922) - using 'computers' (**us**)

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

And then our community proceeded to not use QG for weather prediction.

We have been dealing with strategies for overcoming or managing small time steps ever since.



J. von Neumann



J. Charney

Characteristics of emerging computer architectures

- o Unprecedented parallelism: Current High Performance computers can scale up to 250,000 processors. Efficient use of new architectures will require we know how to scale up to 1 Billion parallel processors.
- o Processors speeds will not be significantly faster than current processors.
- o Memory per processor limited.
- o We will have a greater need to have fault-tolerant algorithms
- o We will have to understand *asynchronous computing*

Algorithms and new computing architectures

- o **Fixed grid on N processors**

For a fixed grid you may have an optimal distribution of the grid on N processors, then the only dimension left available for parallelization is time.

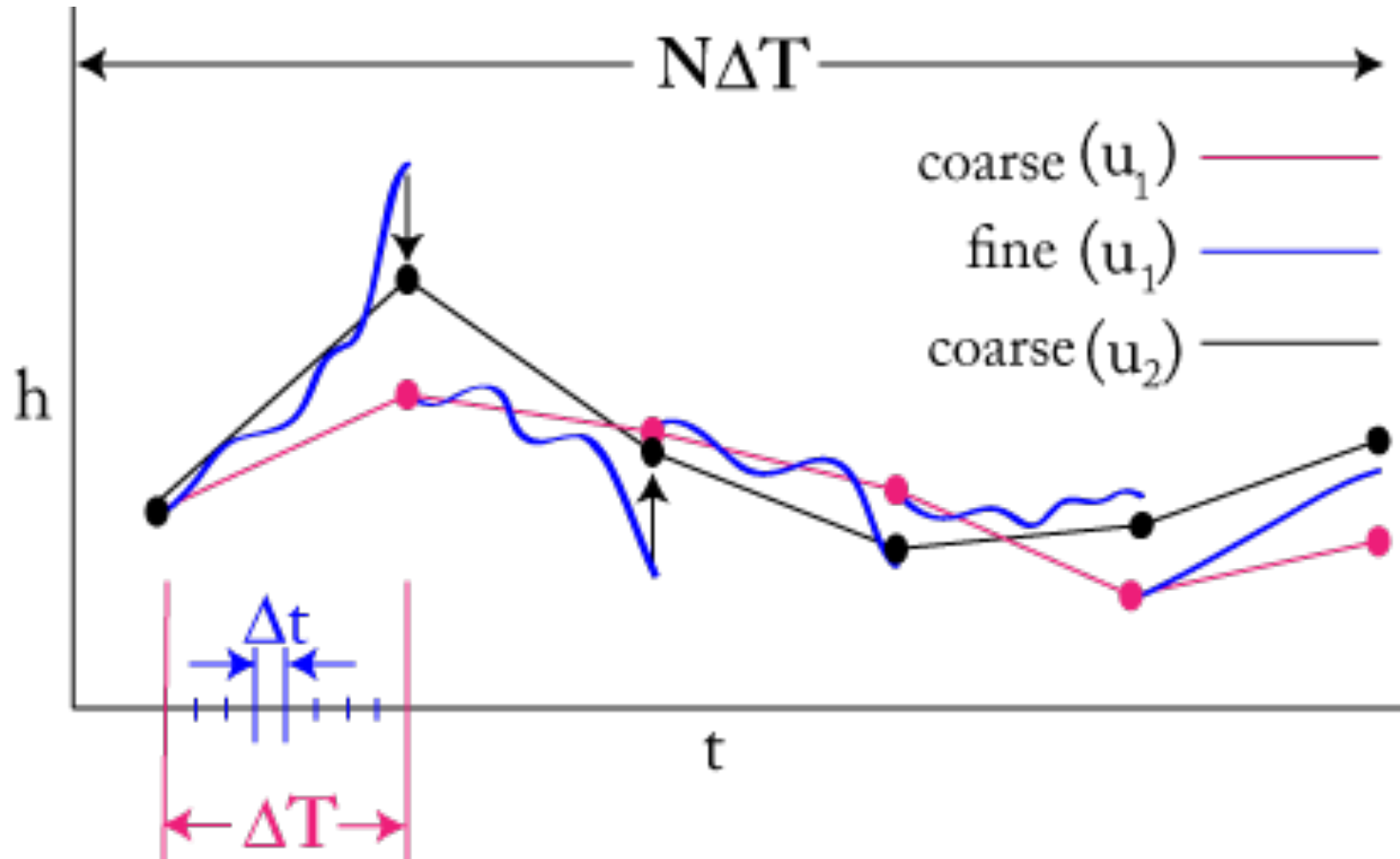
- o **Grid refinement** (we'll still have to wait for each time step)

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.

From a wall-clock-time point of view, current models may appear to dissipate into the machine.

See Jean Côté's talk from PDEs on the Sphere Cambridge at <http://www.newton.ac.uk/programmes/AMM/seminars/2012092814201.html>

Graphical representation of the parallel-in-time algorithm

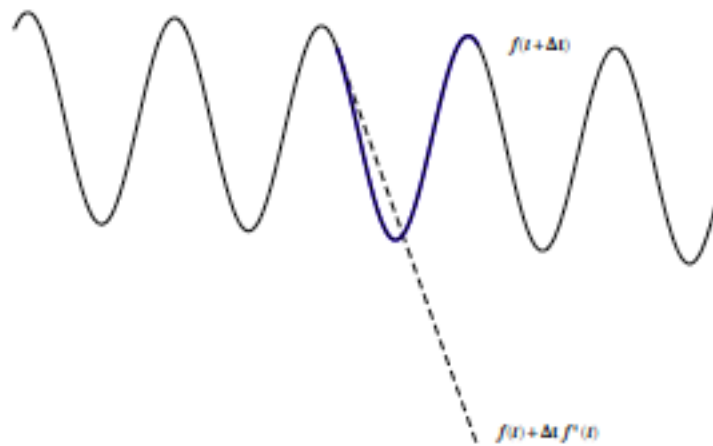


Successful for
dissipative stiffness

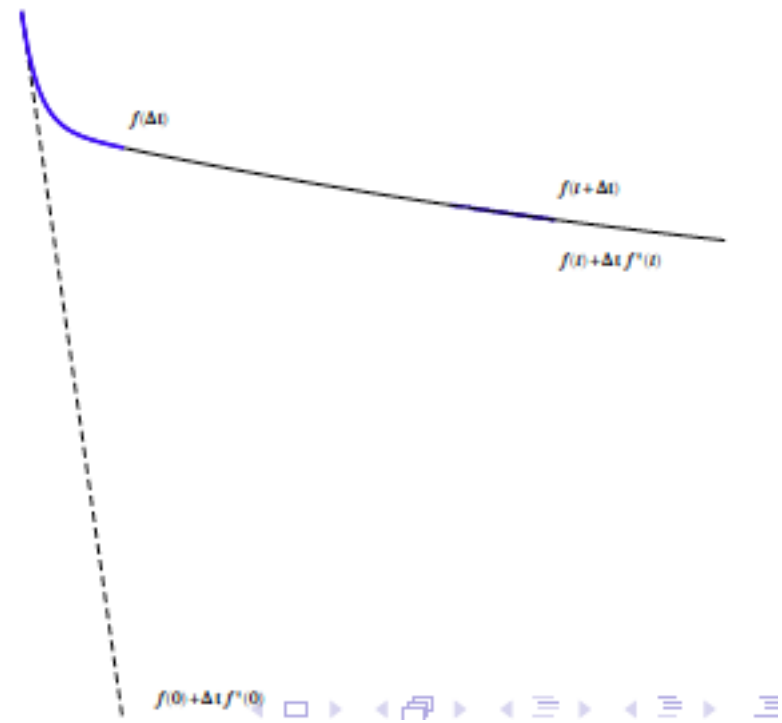
Nievergelt, 1964
Lions, Maday, Turinici, 2001

Oscillatory Stiffness in Parareal

(a) Stiffness from oscillations



(b) Stiffness from dissipation



The roots of the locally asymptotic slow/coarse solution

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

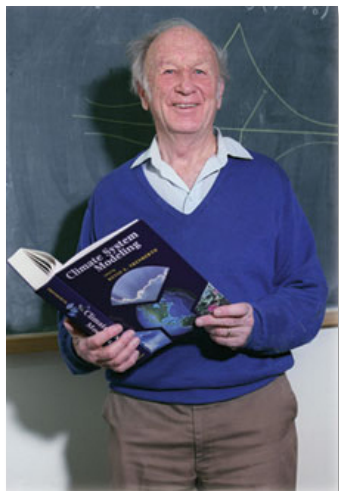
Machanauer (1977), Baer (1877), Tribbia (1979), etc

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

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

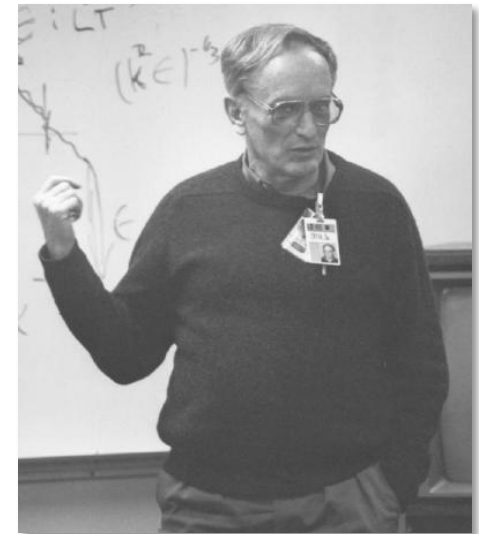
Lorenz and Krishnamurthy, **On the non-Existence of the Slow Manifold** (1987)

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



Ed Lorenz

(....so far the answer is 'it is a fuzzy manifold'.)



Chuck Leith

PDEs with oscillatory stiffness – the equations of climate and weather

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

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

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

One way to view the problem (also used for exponential integrators)

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

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

$$\frac{\partial\mathbf{v}}{\partial t} = \mathcal{O}(1)$$

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

\mathbf{v} varies slower than \mathbf{u} , but \mathbf{v} still has some oscillations

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

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

Asymptotic solution looks like:

$$\mathbf{u}(t) = \exp(-t/\epsilon \mathcal{L}) \bar{\mathbf{u}}(t) + \mathcal{O}(\epsilon)$$

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

$$\bar{\mathcal{D}} \bar{\mathbf{u}}(t) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (e^{s\mathcal{L}} \mathcal{D} e^{-s\mathcal{L}}) \bar{\mathbf{u}}(t) ds.$$

$$\bar{\mathcal{N}}(\bar{\mathbf{u}}(t)) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T e^{s\mathcal{L}} \mathcal{N}(e^{-s\mathcal{L}} \bar{\mathbf{u}}(t)) ds.$$

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

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

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

Test of the method on the shallow water equations

- Test problem: shallow water equations,

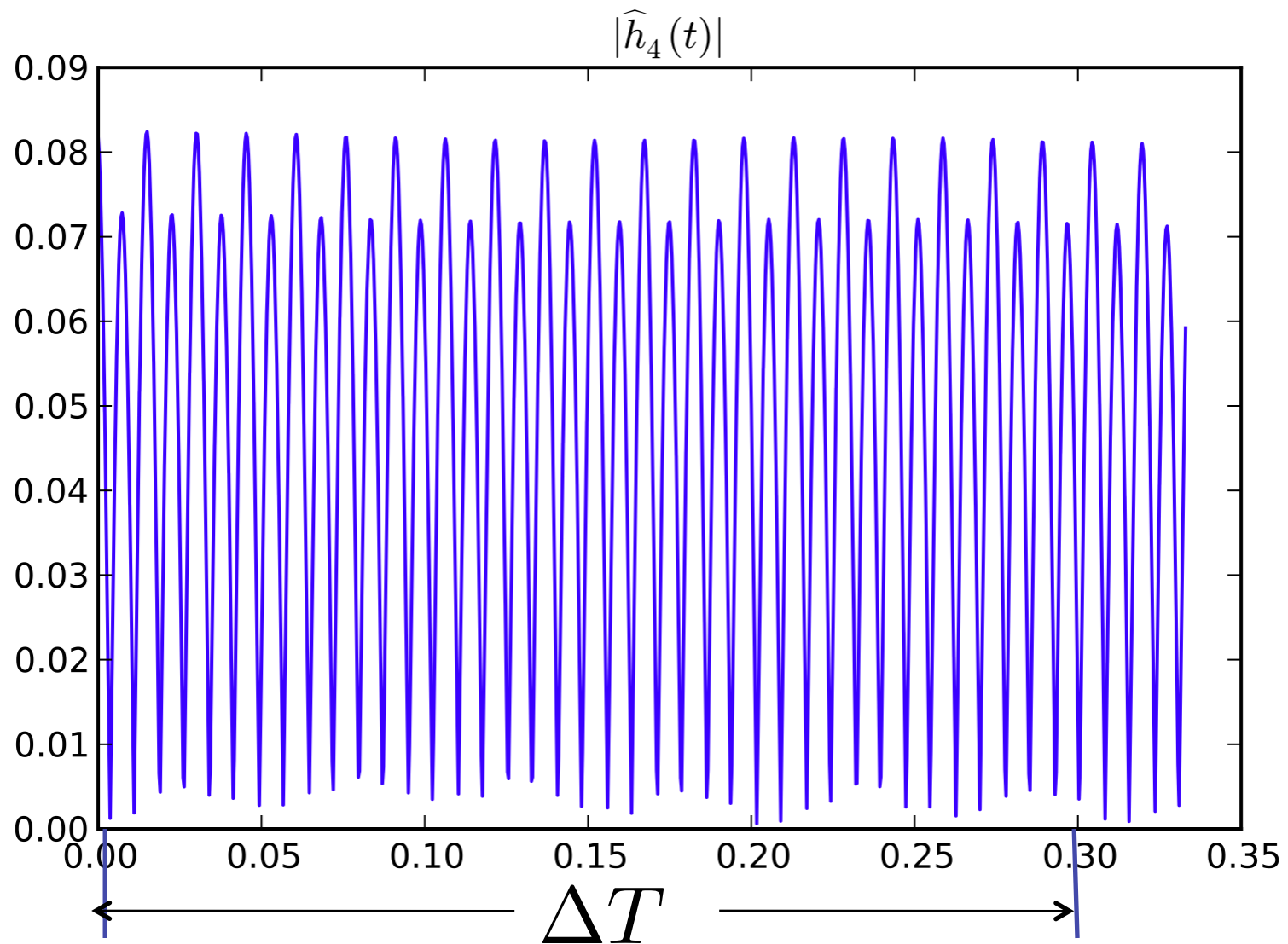
$$\begin{aligned} \frac{\partial v_1}{\partial t} + \frac{1}{\epsilon} \left(-v_2 + F^{-1/2} \frac{\partial h}{\partial x} \right) + v_1 \frac{\partial v_1}{\partial x} &= \mu \partial_x^{2p} v_1, \\ \frac{\partial v_2}{\partial t} + \frac{1}{\epsilon} v_1 + v_1 \frac{\partial v_2}{\partial x} &= \mu \partial_x^{2p} v_2, \\ \frac{\partial h}{\partial t} + \frac{F^{-1/2}}{\epsilon} \frac{\partial v_1}{\partial x} + \frac{\partial}{\partial x} (h v_1) &= \mu \partial_x^{2p} h. \end{aligned}$$

Here ϵ denotes the Rossby number.

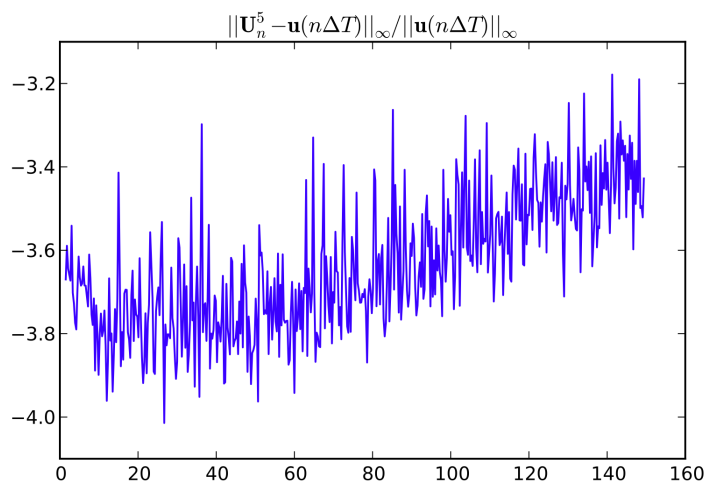
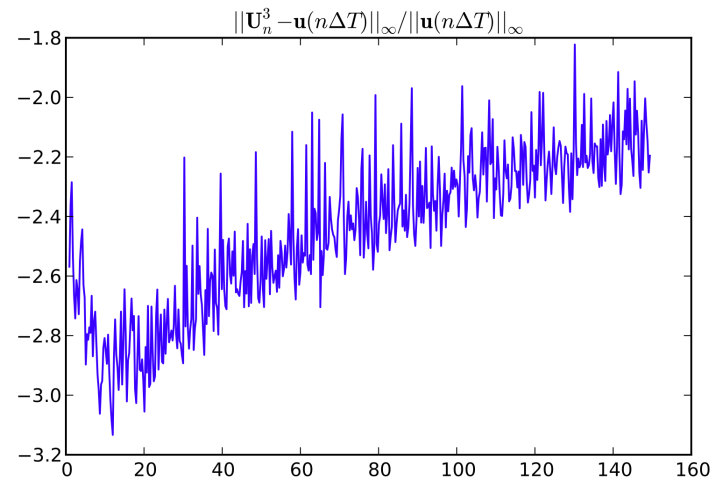
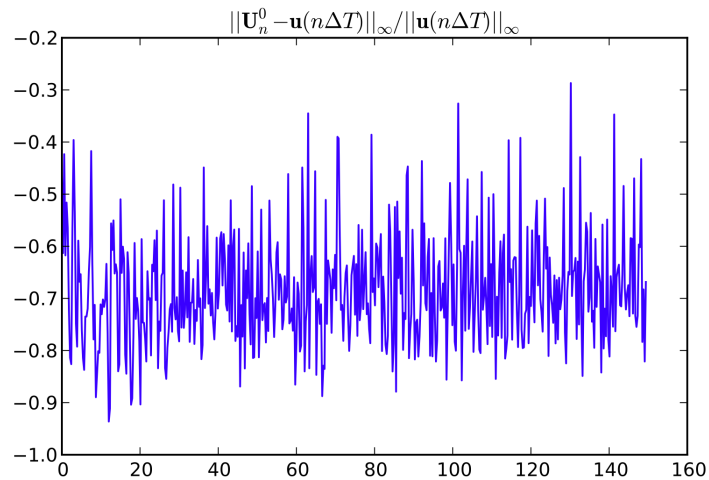
- Example: solve with $\epsilon = 10^{-2}$, $\mu = 0$, and

$$h(0, x) = C_h \left(e^{-2(x-\pi)^2} + h_0 \right), \quad v_1(0, x) = v_2(0, x) = 0.$$

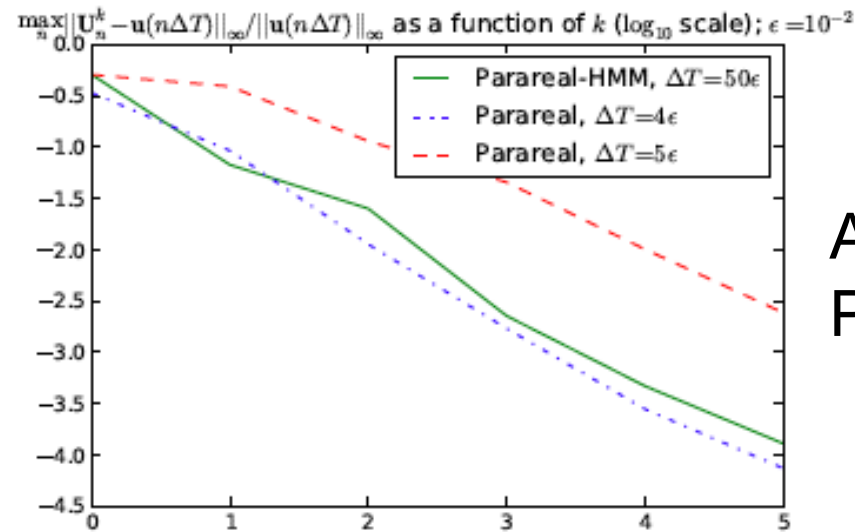
Example of the solution in time: epsilon=.01, Delta T = .3



Relative L^∞ errors for $\epsilon=10^{-1}$, using the asymptotic based parallel-in-time integrator and taking $\Delta T=3/10$ and $\Delta t=1/500$. Each graph is for fixed a iteration k (1,3,5), and the errors are plotted as a function of the time $n \Delta T$, $n=1..N$. The errors are shown on a \log_{10} scale.



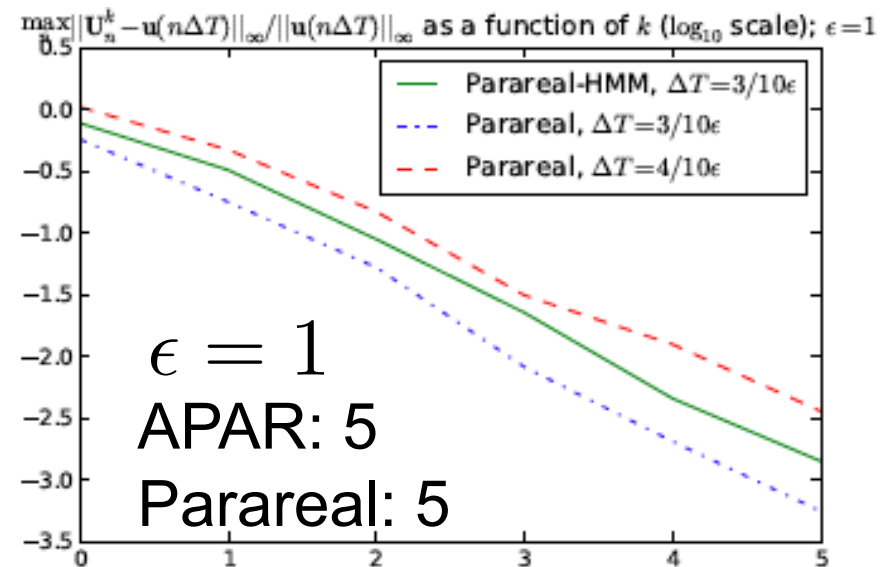
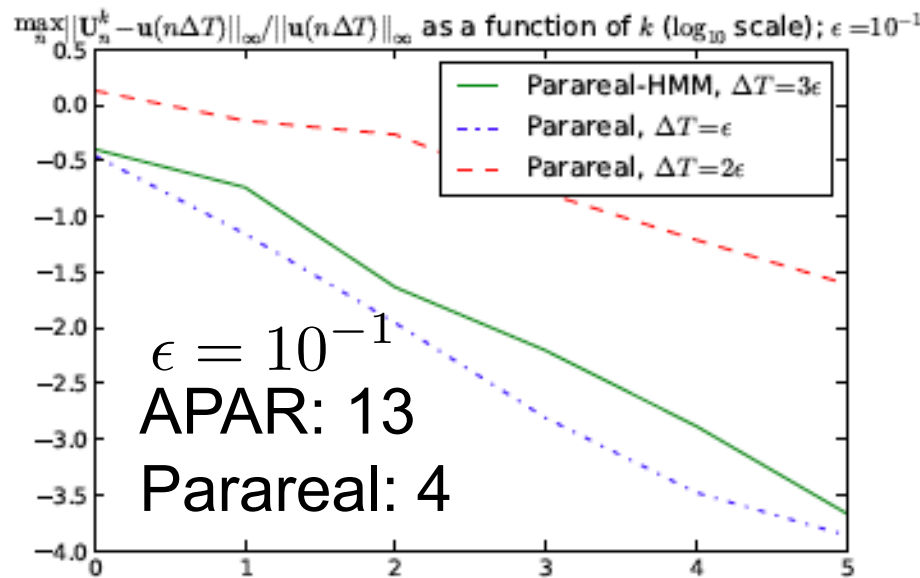
Maximum relative error versus number of iterations



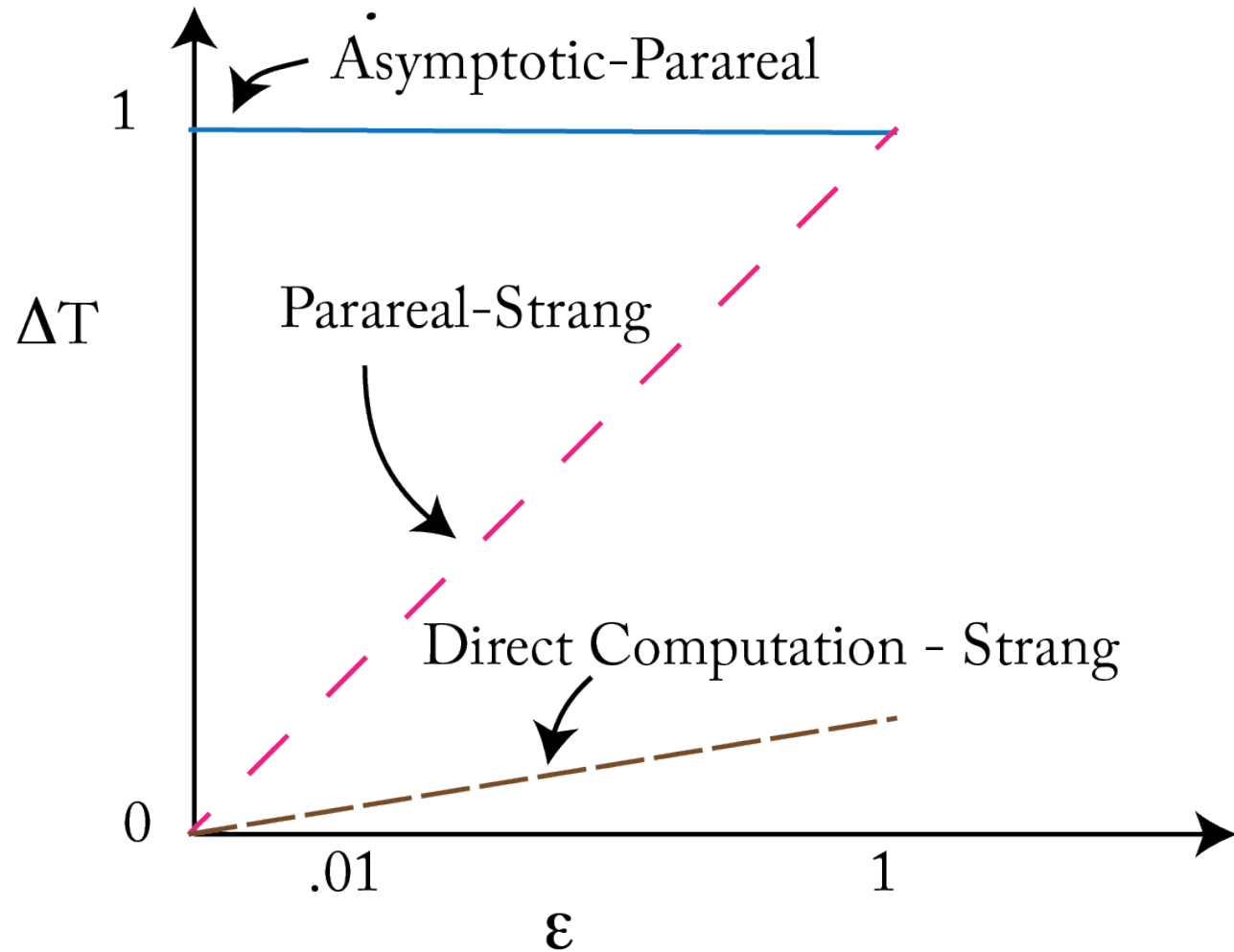
$$\epsilon = 10^{-2}$$

APAR : 100

Parareal: 10



Large time step dependence on epsilon



Summary (Haut & Wingate, SIAM J. Sci Comp 2014)

- o Proposed Locally Asymptotic slow integrator for the parareal algorithm. Has it's theoretical roots in the 'slow manifold'.
- o Separation of time scales not required for it to work, but it means you'll get a parallel-speed-up-in-time, allowing you to use things like asynchronous computing, fault tolerance, and large time steps.
- o Initial results from the shallow water equations are promising. A factor of 10 over the best parareal methods available and a factor of 100 over standard methods for $\epsilon=10^{-2}$
- o We can double the resolution and keep the same coarse time step.
- o We haven't even tried exploiting this parallelism on GPUs or getting more speed out of an asynchronous algorithm.
- o There is a crucial step to introducing even more parallelism is the representation of the **linear propagators**.

Pseudocode

Algorithm 4 Parallel-in-time integrator

Compute the initial guess using the slow solver:

$U_0^{\text{old}} \leftarrow u_0$

for $n = 1, \dots, N - 1$

$U_n^{\text{old}} \leftarrow \text{Coarse_Solver}(U_{n-1}^{\text{old}}, \Delta T)$

endfor

Now refine the solution until convergence:

$U_0^{\text{new}} \leftarrow u_0$

while ($\max_n \|U_n^{\text{new}} - U_n^{\text{old}}\| / \|U_n^{\text{new}}\| > \text{tol}$)

 parfor $n = 1, \dots, N - 1$:

$U_n^{\text{old}} \leftarrow U_n^{\text{new}}$

$V_n \leftarrow \text{Fine_Solver}(U_n^{\text{old}}, \Delta t, \Delta T)$

$V_n \leftarrow V_n - \text{Coarse_Solver}(U_n^{\text{old}}, \Delta T)$

 end parfor

 for $n = 1, \dots, N - 1$

$U_n^{\text{new}} \leftarrow \text{Coarse_Solver}(U_{n-1}^{\text{new}}, \Delta T) + V_{n-1}$

 endfor

end while

return $U_1^{\text{new}}, \dots, U_N^{\text{new}}$

Solve for the fast leading order solutions

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

$$\mathbf{u}^0(\mathbf{x}, t, \tau) = e^{-\tau \mathcal{L}_F} \bar{\mathbf{u}}(\mathbf{x}, t).$$

The order 1 solution is a function of the leading order solution:

$$\frac{\partial \mathbf{u}^1}{\partial \tau} + \mathcal{L}_F(\mathbf{u}^1) = -\left(\frac{\partial \mathbf{u}^0}{\partial t} + \mathcal{L}_S(\mathbf{u}^0) + B(\mathbf{u}^0, \mathbf{u}^0) - D(\mathbf{u}^0)\right)$$

Solve with Duhammel's formula and ensure the \mathbf{u}^1 solution is not large by using the concept of fast wave averaging.

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

$$\frac{\partial \bar{\mathbf{u}}}{\partial t}(\mathbf{x}, t) = -\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau e^{+s \mathcal{L}_F} \left[\mathcal{L}_S(e^{-s \mathcal{L}_F} \bar{\mathbf{u}}) + B(e^{-s \mathcal{L}_F} \bar{\mathbf{u}}, e^{-s \mathcal{L}_F} \bar{\mathbf{u}}) - D(e^{-s \mathcal{L}_F} \bar{\mathbf{u}}) \right] ds,$$

$$\bar{\mathbf{u}}(\mathbf{x}, t)|_{t=0} = \mathbf{u}_0(\mathbf{x}).$$

Boussinesq equations - nonlocal form in a Hilbert Space

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

Hilbert Space X of vector fields \mathbf{u} in L_2 that are divergence free and equipped with the L_2 norm.

2

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{1}{Ro} L_{Ro}(\mathbf{u}) + \frac{1}{Fr} L_{Fr}(\mathbf{u}) + B(\mathbf{u}, \mathbf{u}) = \frac{1}{Re} D(\mathbf{u})$$

$$\mathbf{u}|_{t=0} = \mathbf{u}_0(\mathbf{x})$$

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

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