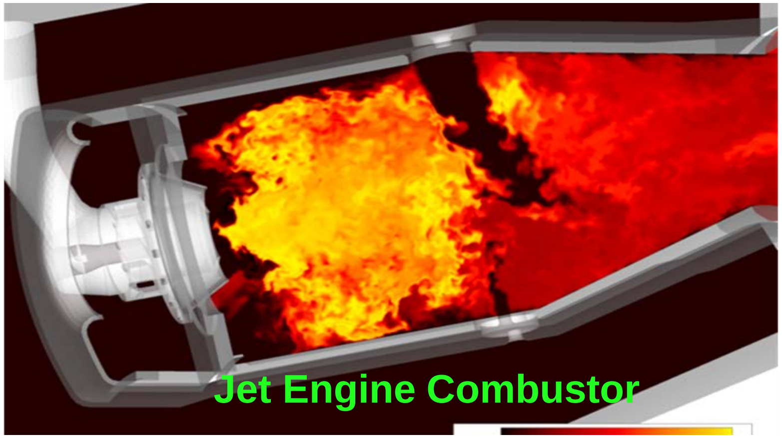
Towards Scalable Parallel Long Time Integration of Chaotic Dynamical Systems

Qiqi Wang

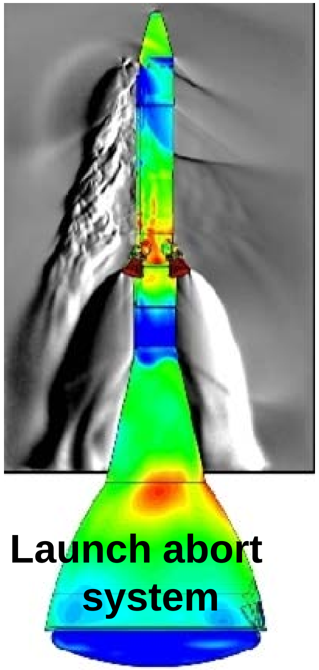
MIT AEROASTRO

Space-Time-Parallel Workshop
2013.6.19 Manchester UK

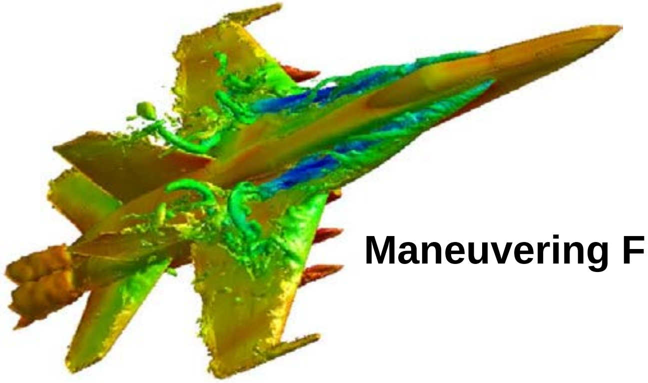
Motivation: Turbulent flow simulations often require millions of timesteps



Jet Engine Combustor

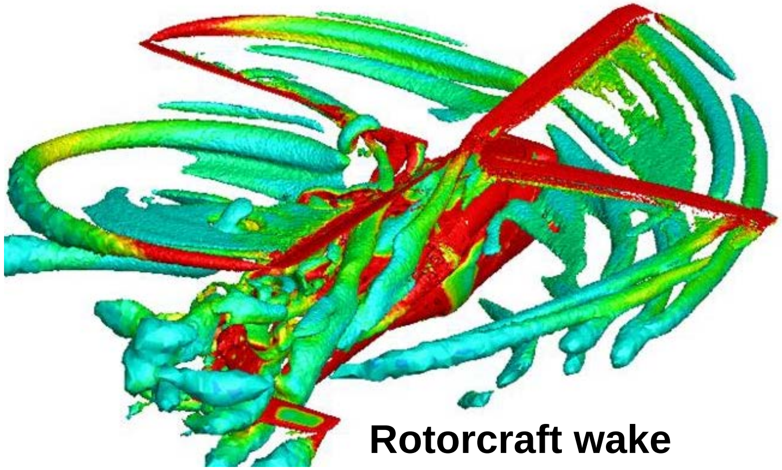


Launch abort system



Maneuvering F18

Many applications require computing long time averaged quantities.

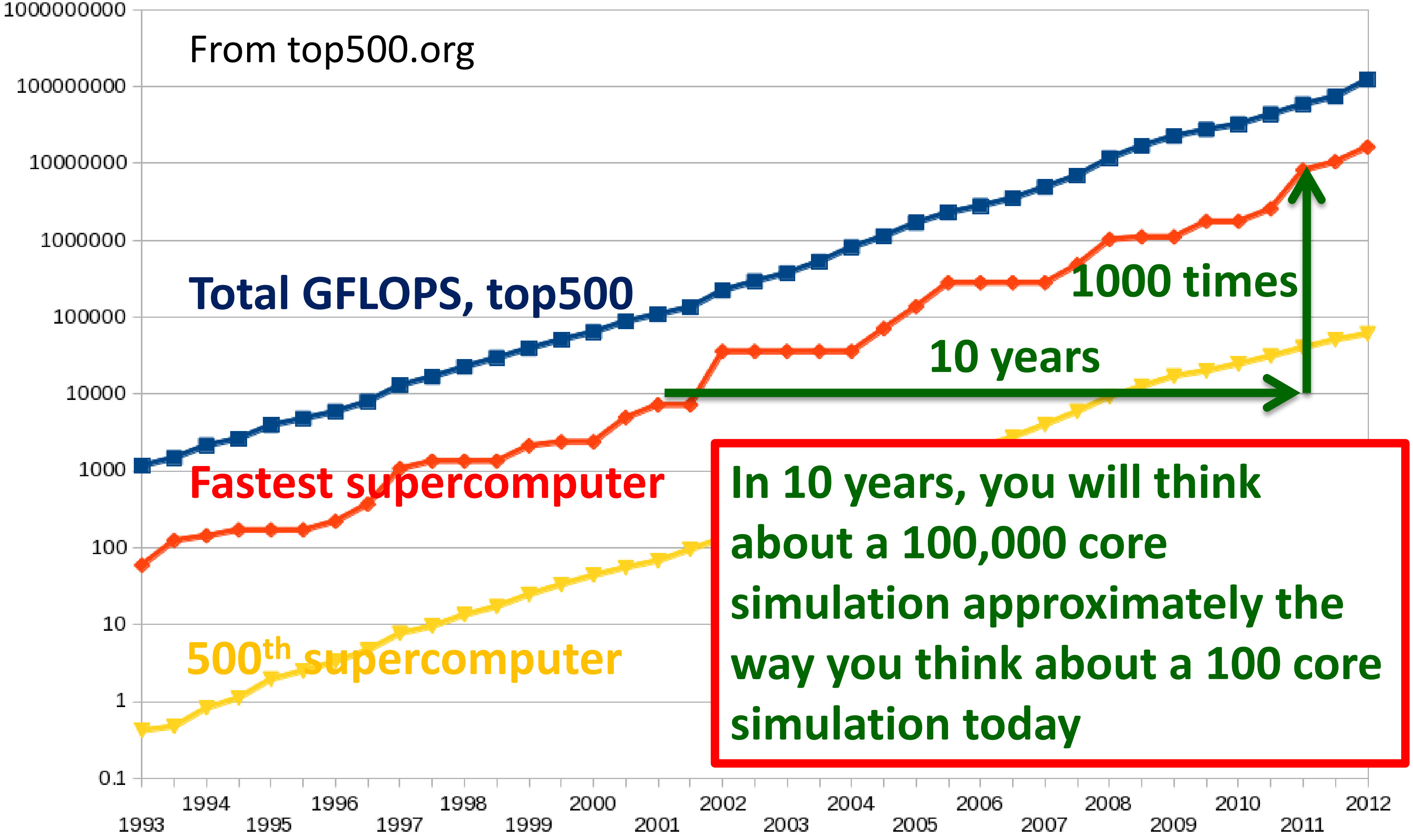


Rotorcraft wake

Multiscale chaotic dynamical system

Time step size constrained by fastest timescale, time integration length multiple of slowest timescale

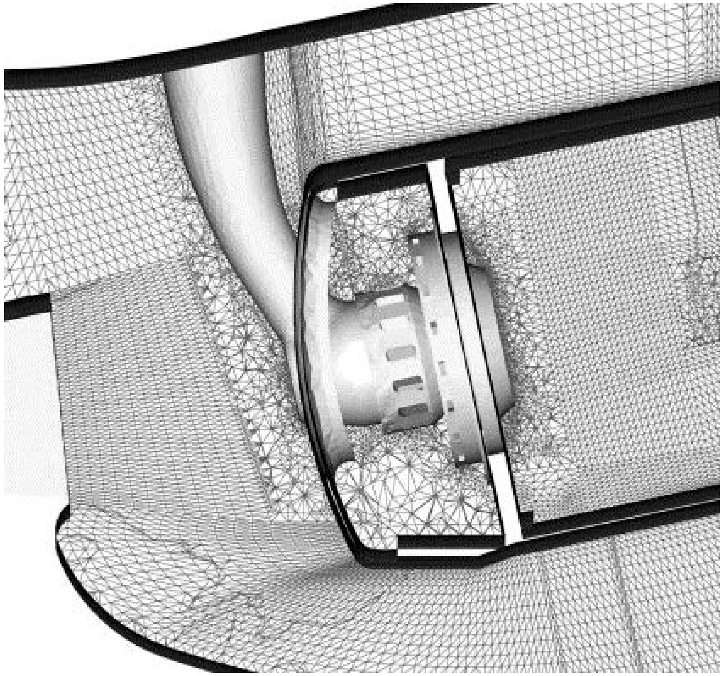
Motivation 2: Next generation parallel computers calls for space-time parallelism



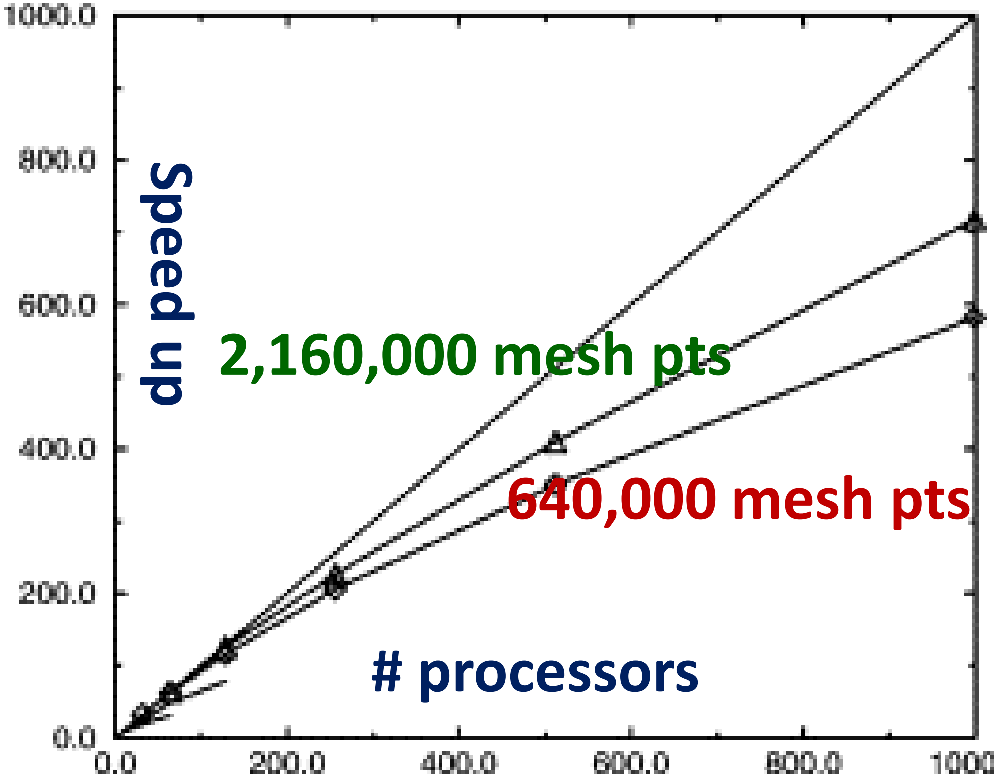
From top500.org

In 10 years, you will think about a 100,000 core simulation approximately the way you think about a 100 core simulation today

Bottleneck of LES turnaround time



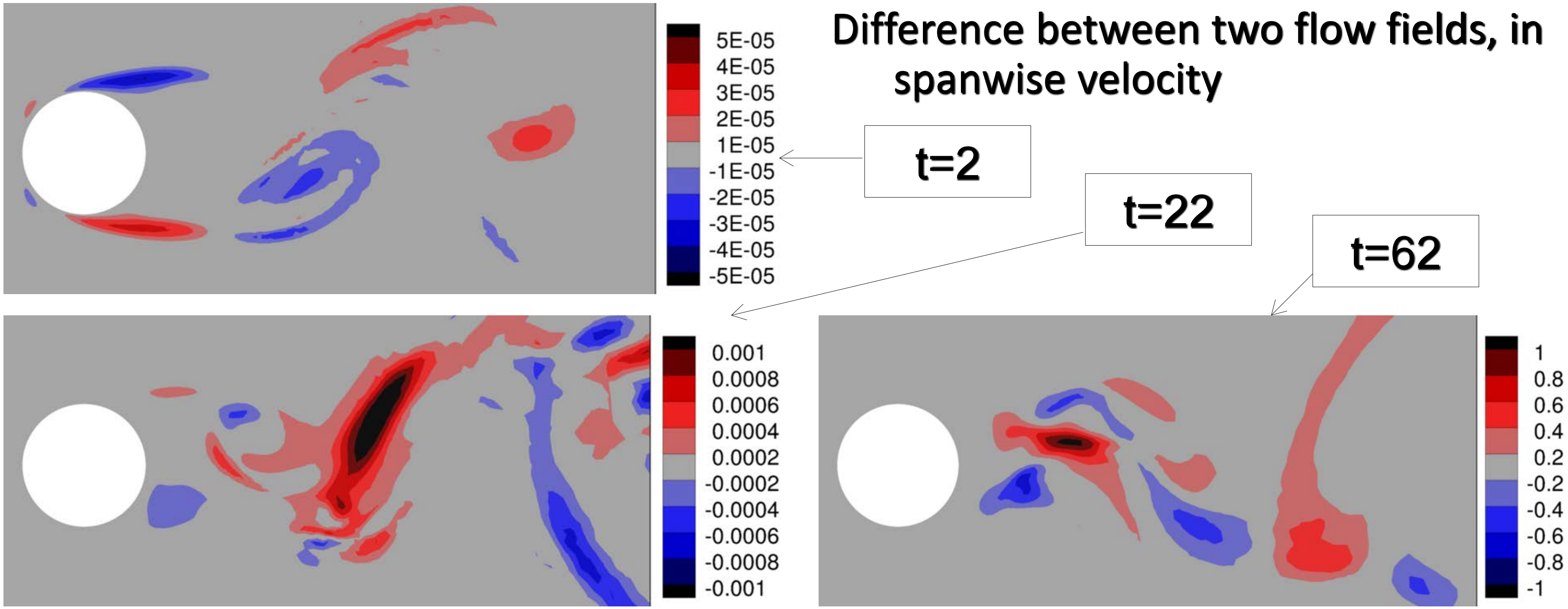
Moin 2002



- **Good scaling is difficult** below 5000 gridpoint per core.
- Next generation, **massively parallel HPC** on the horizon.
- **Paralle-in-time is necessary for applications requiring fast turnaround time.**

A Critical barrier to efficient time parallelism

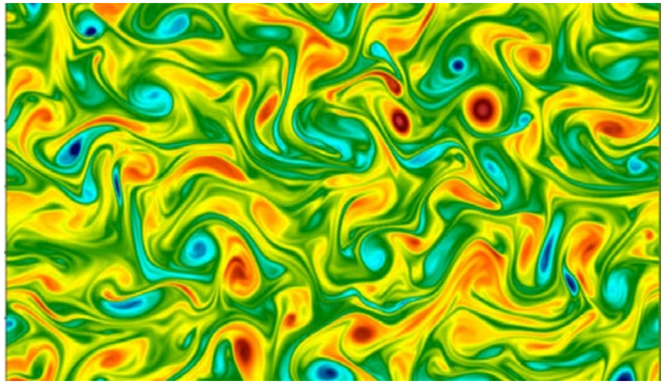
- Turbulent flows can be **chaotic – sensitive to perturbations**.
- **Small error** in the coarse solver at **earlier timesteps** can cause **large error** in the estimate at **later timesteps**
- Causing **slow convergence** of time parallel solvers.
- Flow across circular cylinder at Re=500:



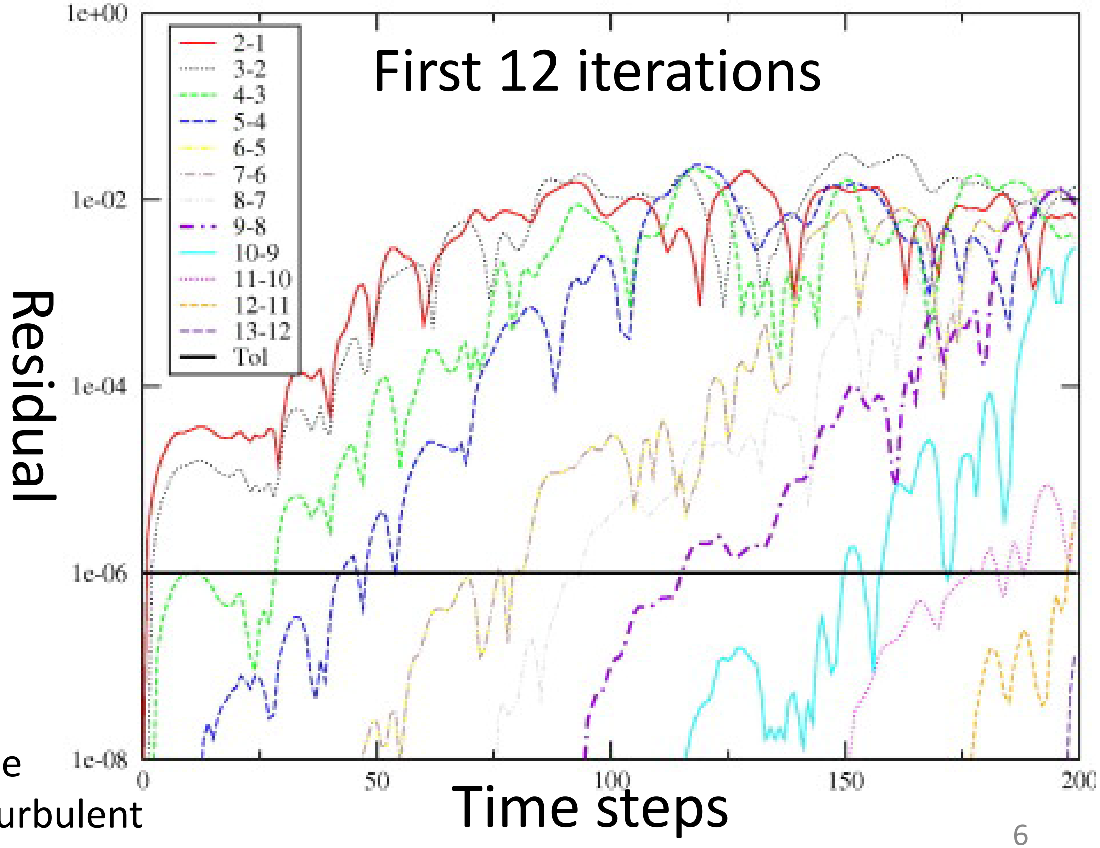
Difference between two flow fields, in spanwise velocity

A Critical barrier to efficient time parallelism

- Turbulent flows can be **chaotic – sensitive to perturbations**.
- **Small error** in the coarse solver at **earlier timesteps** can cause **large error** in the estimate at **later timesteps**
- Causing **slow convergence** of time parallel solvers.
- **Iterations required by Parareal is proportional to the length of time integration for chaos.**

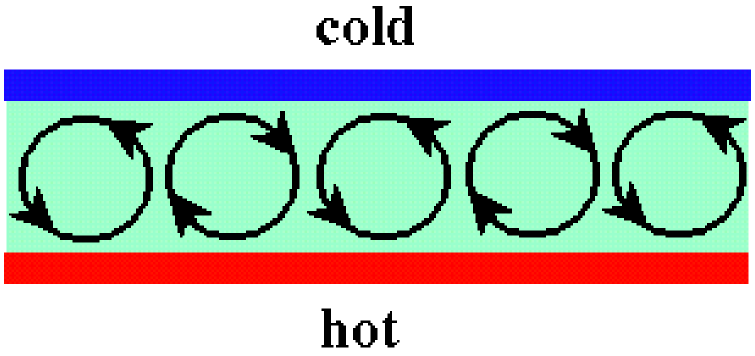


Reynolds-Barredoa et al. Mechanisms for the convergence of time-parallelized, parareal turbulent plasma simulations. J. Comp. Phys. 231-23, 2012



Why slower convergence at later timesteps?

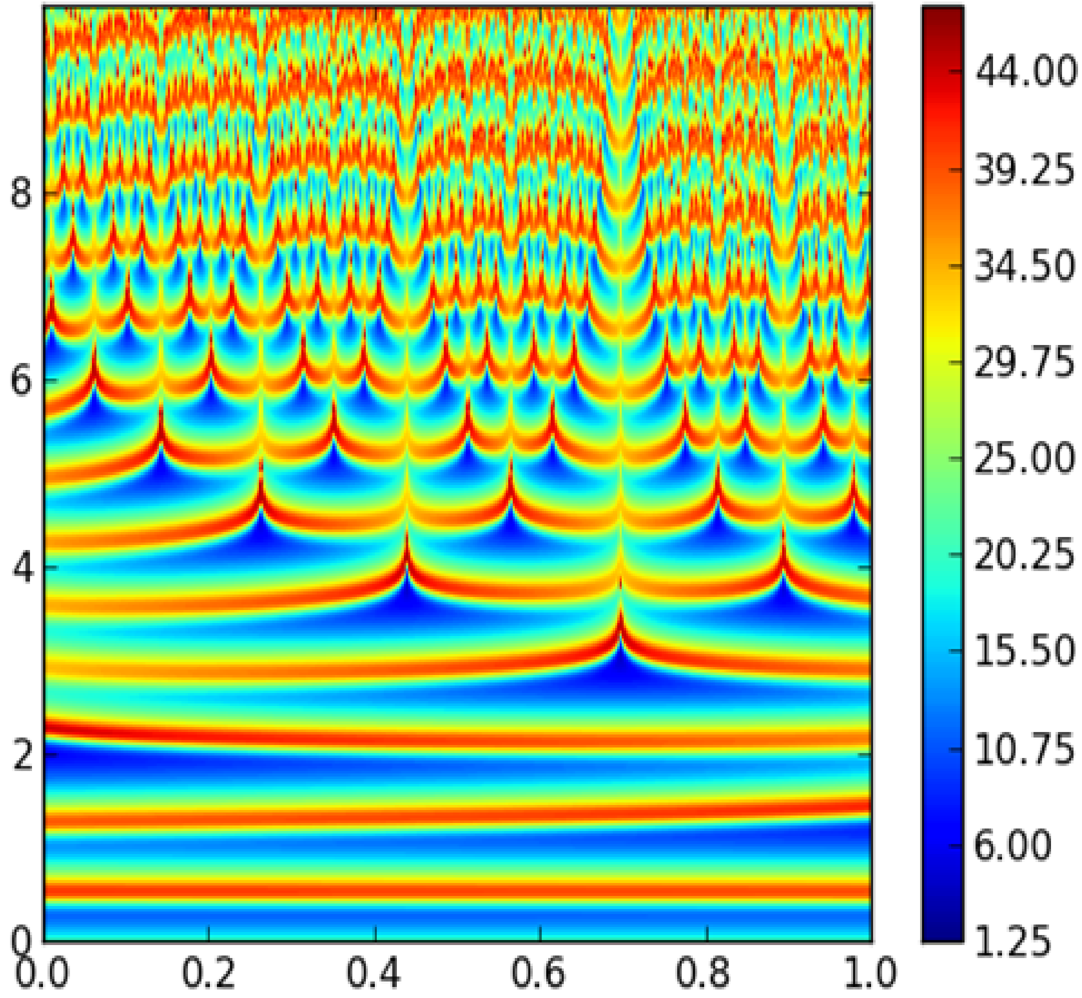
Initial value problems of chaotis are ill-posed.



Example:

Lorenz system

Rate of heat transfer



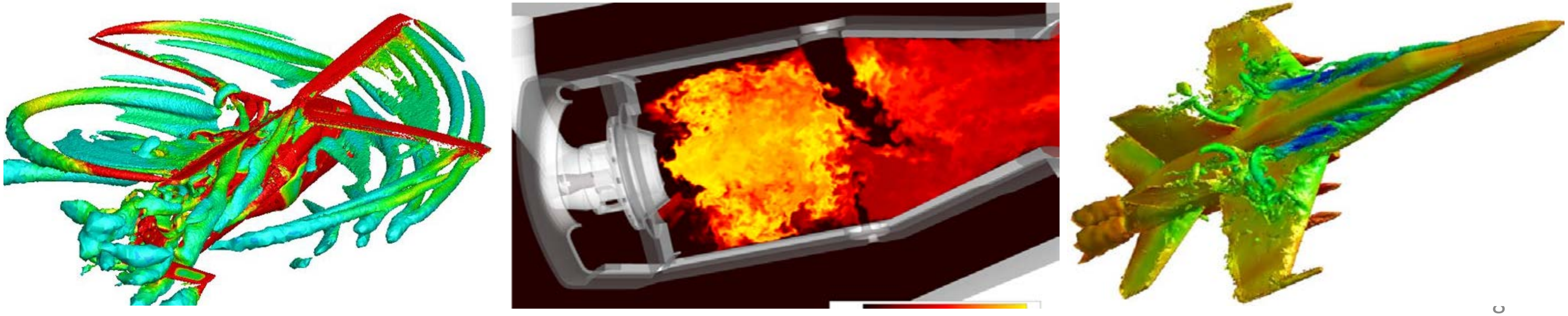
Fine solver (Lorenz RK4) ⟷ Coarse solver (Lorenz FE)

Target quantity of interest: time averaged quantity

$$\langle J \rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T J(u)\,dt$$

where $$\frac{du}{dt} = f(u, s)$$

For a large class of chaotic systems, ‹J› is independent of the initial condition (ergodic).



Reformule turbulent flow simulation for efficient time parallelism

- **Ergodic assumption: time averaged quantities of interest are independent of the initial condition.**
- **Replace** the traditional initial value problem

$$\dot{u} = \mathcal{L}(u), \qquad u(0) = u_0$$

with one **without the initial condition**.

- Aim for solution of the flow equation near a reference solution (coarse solver, a nearby parameter value).
- Chaotic systems **with initial condition** is **ill-conditioned**; **without initial condition** it can be **well-conditioned**.

Initial value problem of chaos is ILL-conditioned

New approach: replace the initial condition with least squares

$$\min \int_{T_1}^{T_2} \|u(\tau) - \underset{\uparrow}{u_r}(\tau)\|^2 + \alpha^2 \underset{\uparrow}{\left(\frac{dt}{d\tau} - 1\right)^2} d\tau$$

Reference solution

Time dilation

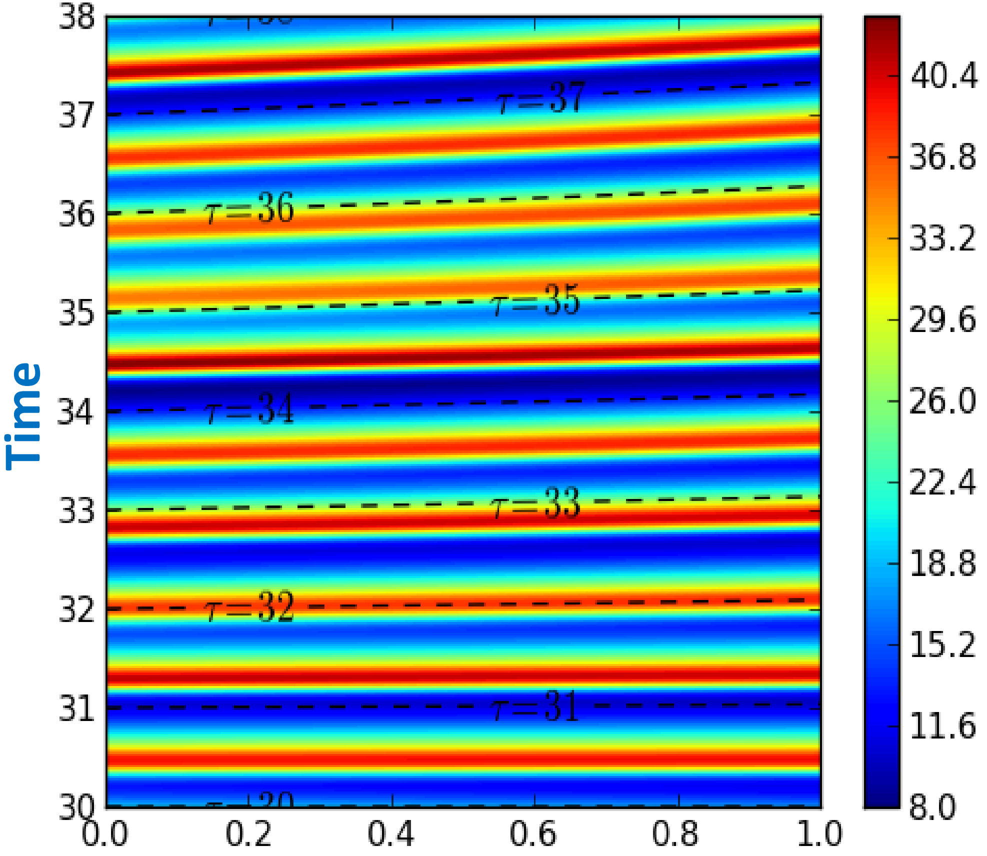
$$\text{s.t.} \quad \frac{du}{dt} = f(u, s)$$

- “Integral phase condition” for finding homoclinic cycles in dynamical systems
- **Does using least squares instead of an initial condition remove the ill-conditioning of chaos?**

Least squares problem

Condition number $O(T^2)$

Time dependent output

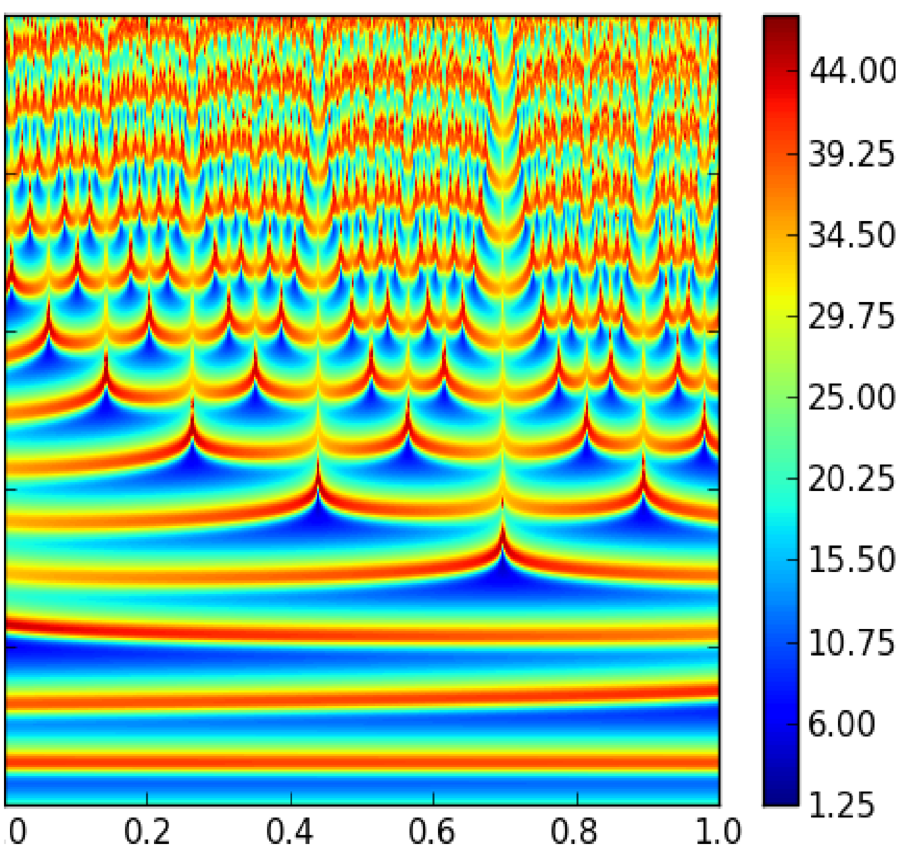


Fine solver (Lorenz RK4) ↔ **Coarse solver (Lorenz FE)**

Initial value problem

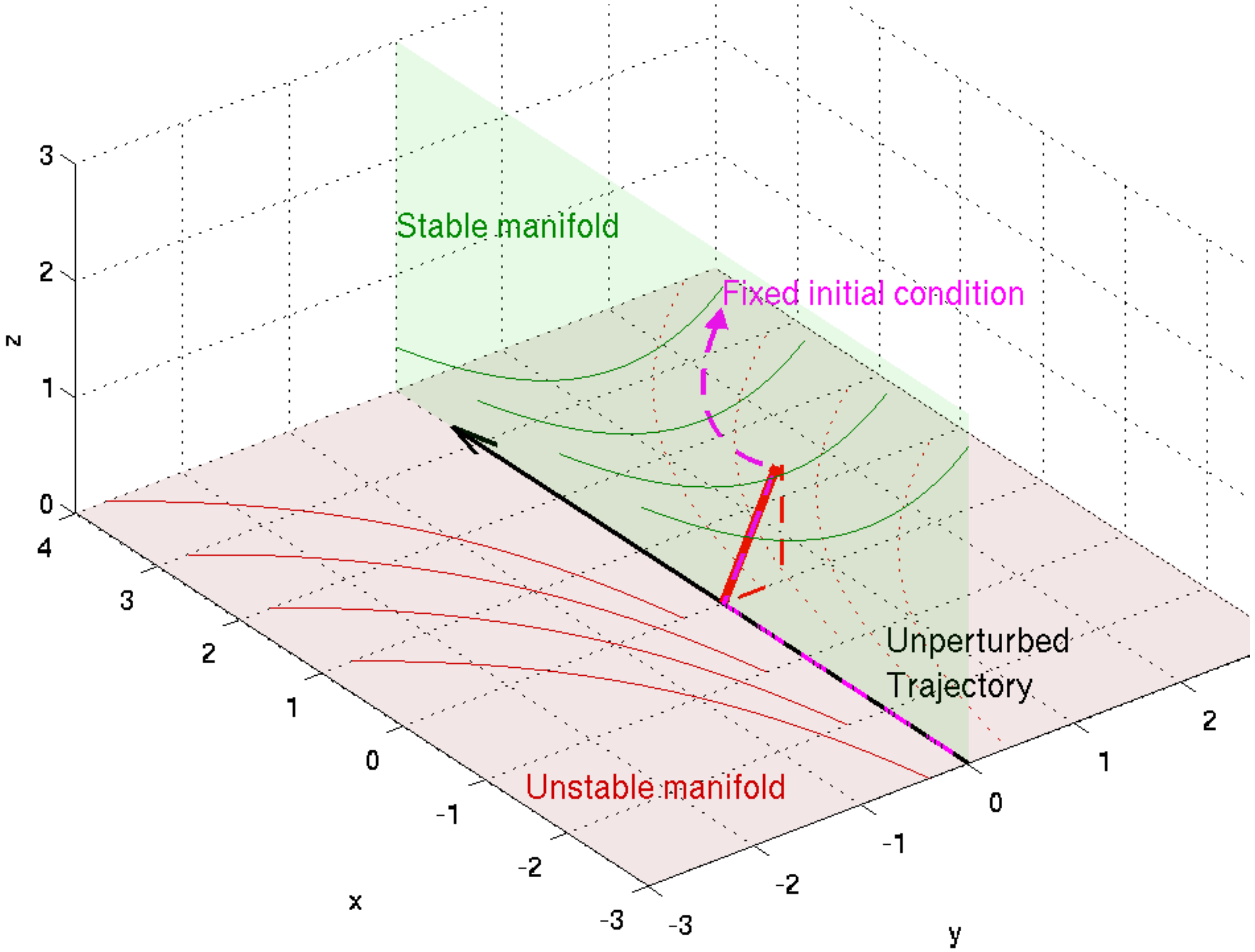
Condition number $O(e^{\lambda T})$

Time dependent output

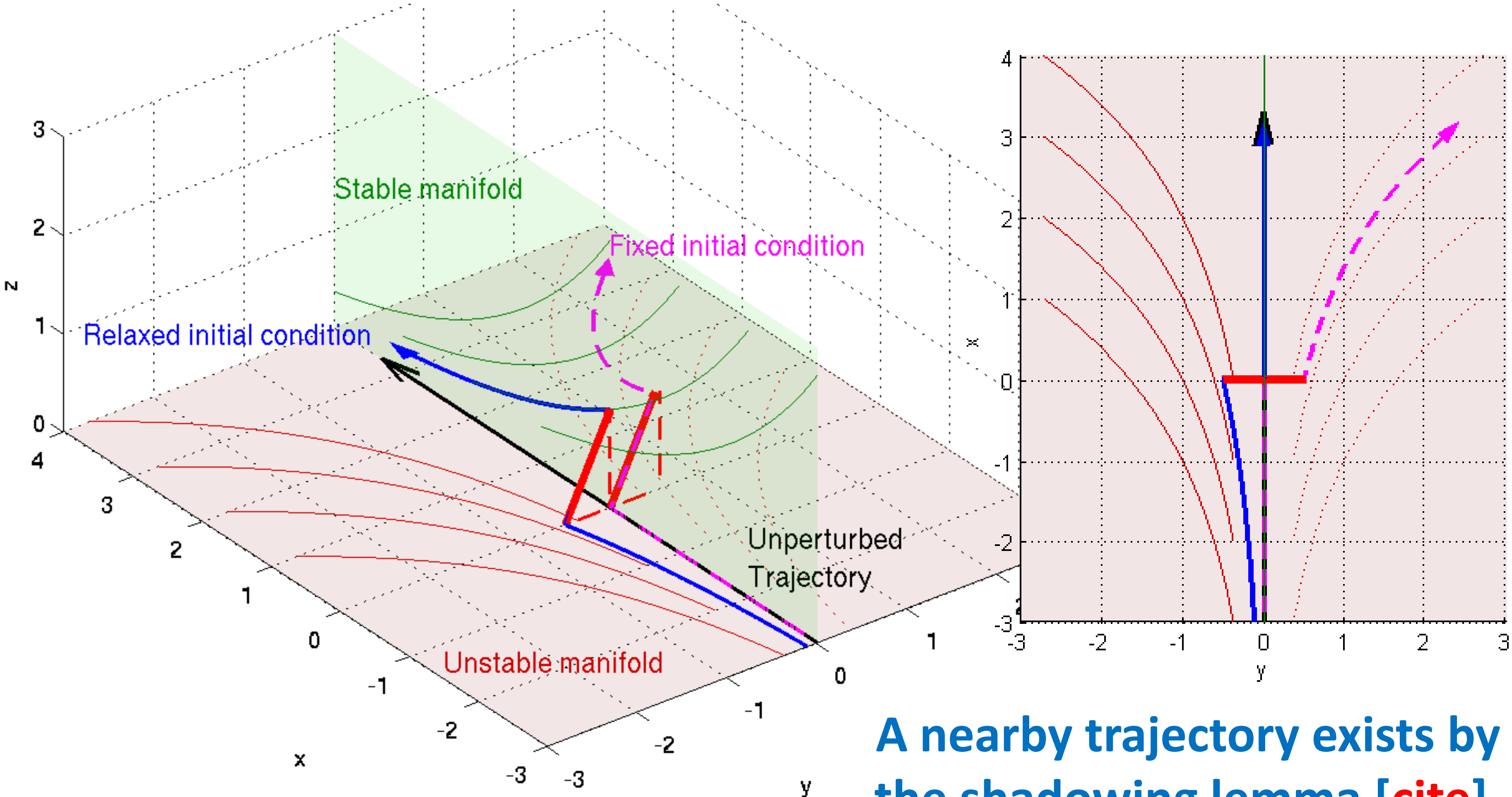


Fine solver (Lorenz RK4) ↔ **Coarse solver (Lorenz FE)**

Phase space cartoon of a perturbed initial value problem



Phase space cartoon of a perturbed least squares problem

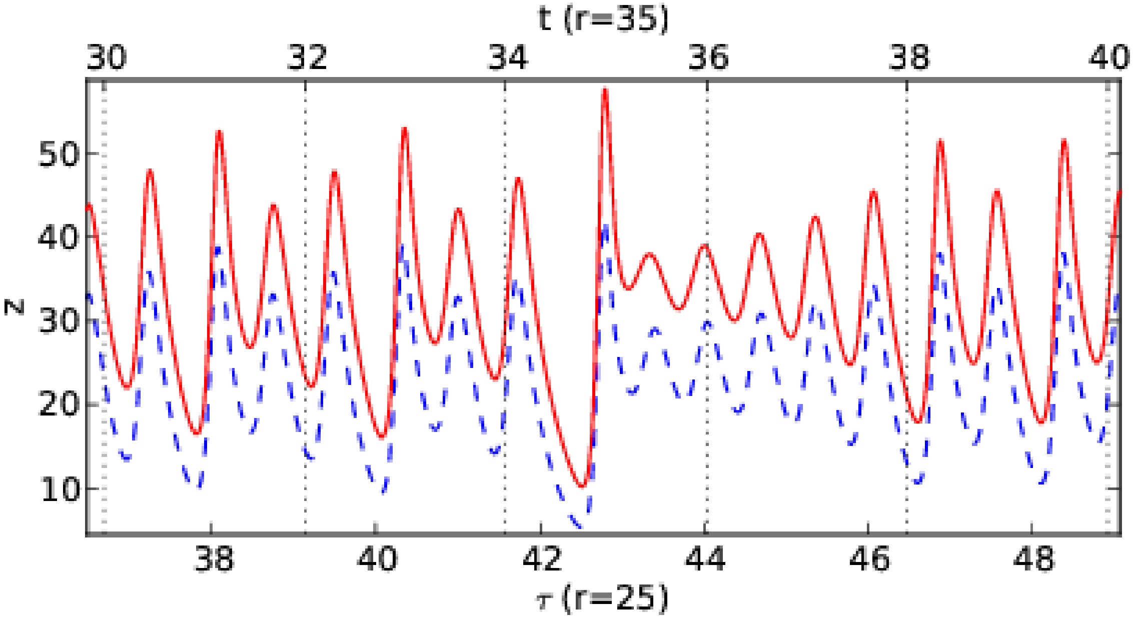


A nearby trajectory exists by the shadowing lemma [cite].

Shadowing Lemma for Hyperbolic Chaos

For any δ>0 there exists ε>0, such that for every "ε-pseudo-solution" u satisfying $\|du/d\tau - f(u)\| < \varepsilon$, there exists a true solution **u** satisfying $d\mathbf{u}/dt - f(\mathbf{u}) = 0$ under a time transformation t(τ), such that

$$\|\mathbf{u}(\tau) - u(t)\| < \delta, \; |1 - dt/d\tau| < \delta$$



Space-time Parallel nonlinear Solver
Iterative Least Squares Algorithm

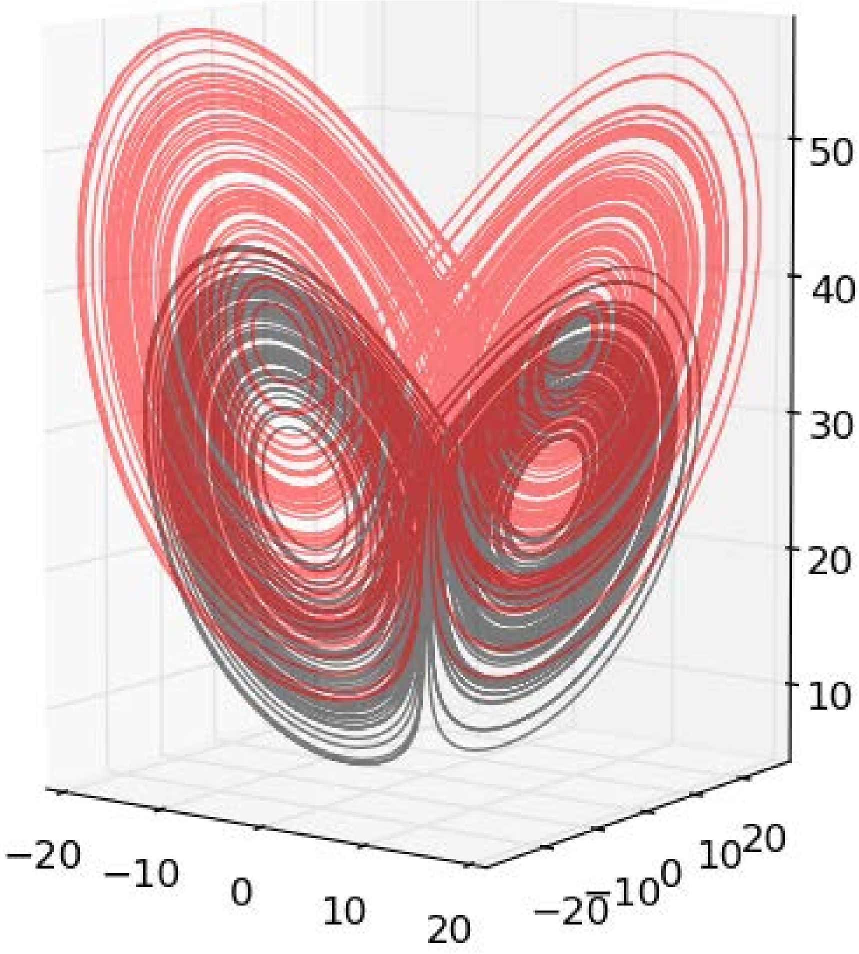
1. Start at step k=0 with a reference solution $u_0(t)$
2. Evaluate residual $R_k = du_k/dt - \mathcal{R}(u_k)$
3. Compute the **L**east **S**quares **S**olution of the Linearized Equation with **space-time-parallel solver**

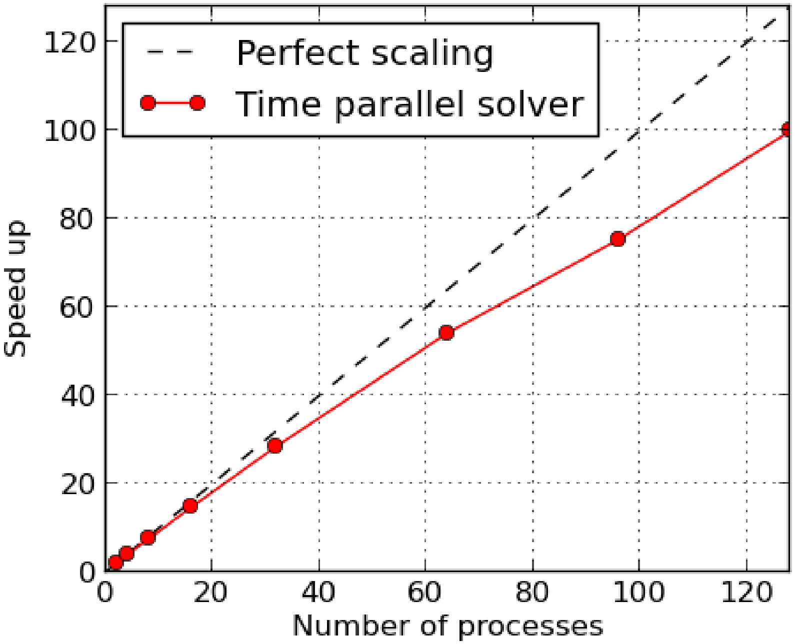
$$\min \int_0^T v^T v \, dt \quad \text{s.t.} \quad \frac{dv}{dt} = \mathcal{L}_{u_k} v + R_k$$

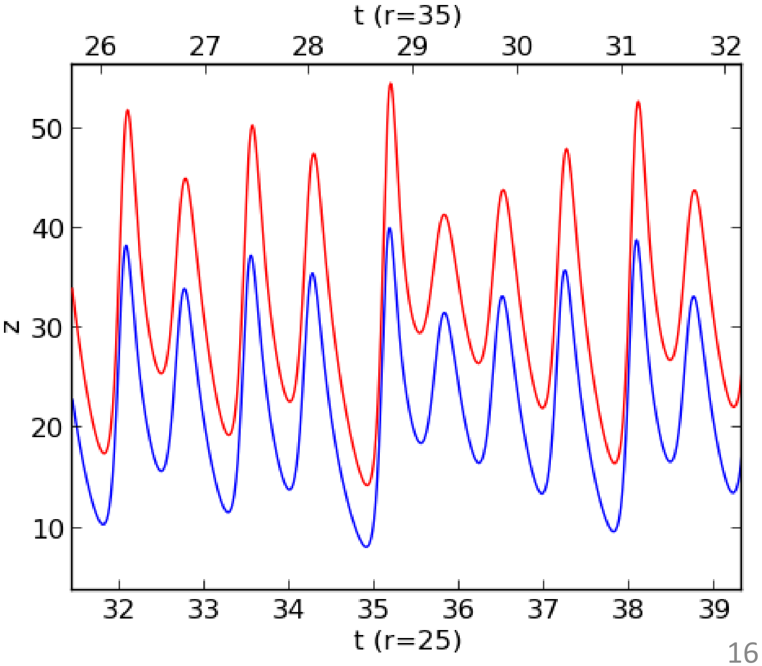
4. Newton step $u_{k+1}(t) = u_k(t) - v(t)$
5. Go to Step 2 if convergence not achieved.

Validation on the Lorenz System

Ra=25 **Ra=35**

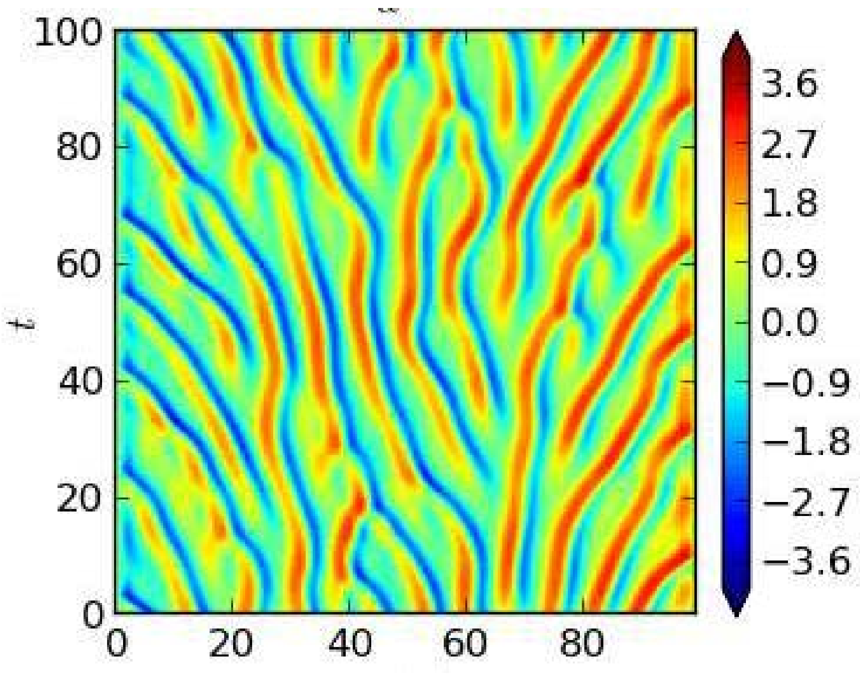




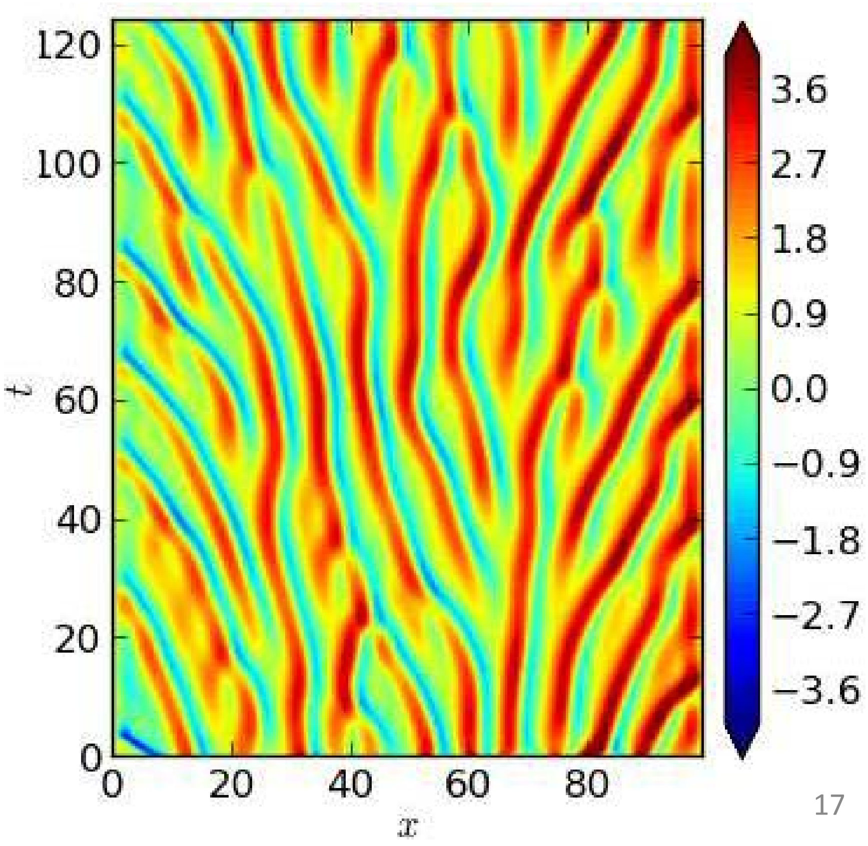


The Kuramoto-Sivashinsky equation

$$\frac{\partial u}{\partial t} = -(u+c)\frac{\partial u}{\partial x} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^4 u}{\partial x^4}$$



Initial condition c=-1



Converged solution c=-0.1

What does it takes to do Least Squares?

- Solve L2 regularized problem wo. initial condition

$$\min \int_{T_1}^{T_2} \|v\|^2 dt \quad \text{s.t.} \quad \frac{dv}{dt} = f_u v + f_s$$

What does it takes to do Least Squares?

- Solve L2 regularized problem wo. initial condition

$$\min \int_{T_1}^{T_2} \|v\|^2 dt \quad \text{s.t.} \quad \frac{dv}{dt} - f_u v = f_s$$

- Optimality (KKT) condition

$$\frac{dw}{dt} + f_u^T w = v$$

$$\frac{dv}{dt} - f_u v = f_s$$

$$w(0) = w(T) = 0$$

Substitute: a second order in time system

$$\frac{d^2 w}{dt^2} + \frac{d}{dt} f_u^T w - f_u \frac{dw}{dt} - f_u f_u^T w = f_s$$

$$w(0) = w(T) = 0$$

A second order in time, SPD system

$$\frac{d^2 w}{dt^2} + \frac{d}{dt} f_u^T w - f_u \frac{dw}{dt} - f_u f_u^T w = f_s$$

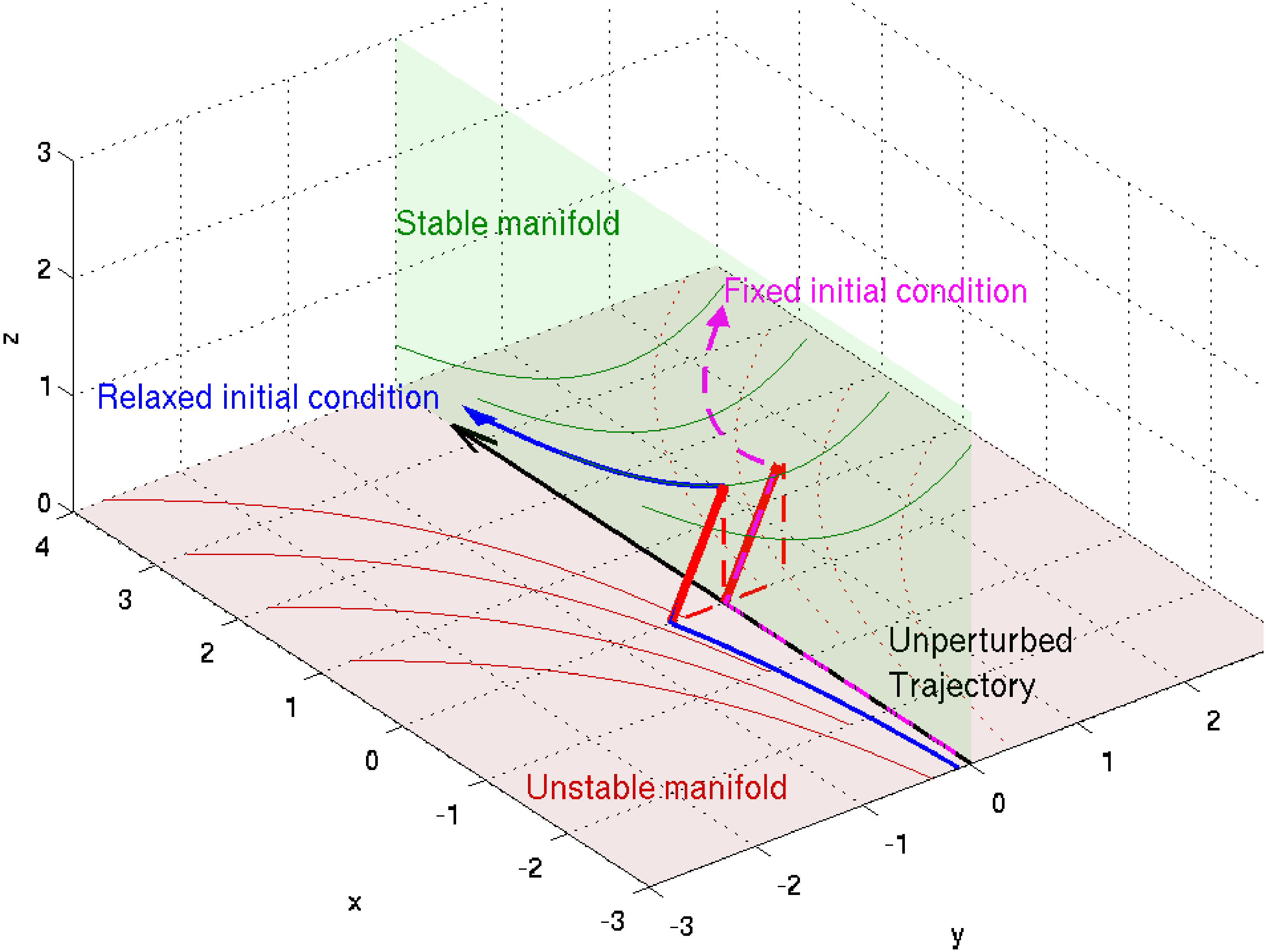
$$w(0) = w(T) = 0$$

Discrete version:

$$\begin{pmatrix} F_1 & G_1 & & & \\ & F_2 & G_2 & & \\ & & \ddots & \ddots & \\ & & & F_n & G_n \end{pmatrix} \begin{pmatrix} F_1^T & & & \\ G_1^T & F_2^T & & \\ & G_2^T & \ddots & \\ & & \ddots & F_n^T \\ & & & G_n^T \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = - \begin{pmatrix} f_{s1} \\ f_{s2} \\ \vdots \\ f_{sn} \end{pmatrix}$$

where $F_i = -\frac{I}{\Delta t} - \frac{f_{u,i}}{2}$, $G_i = \frac{I}{\Delta t} - \frac{f_{u,i}}{2}$

Phase space cartoon of a perturbed least squares problem



What does it takes to do Least Squares?

- Solve L2 regularized problem wo. initial condition

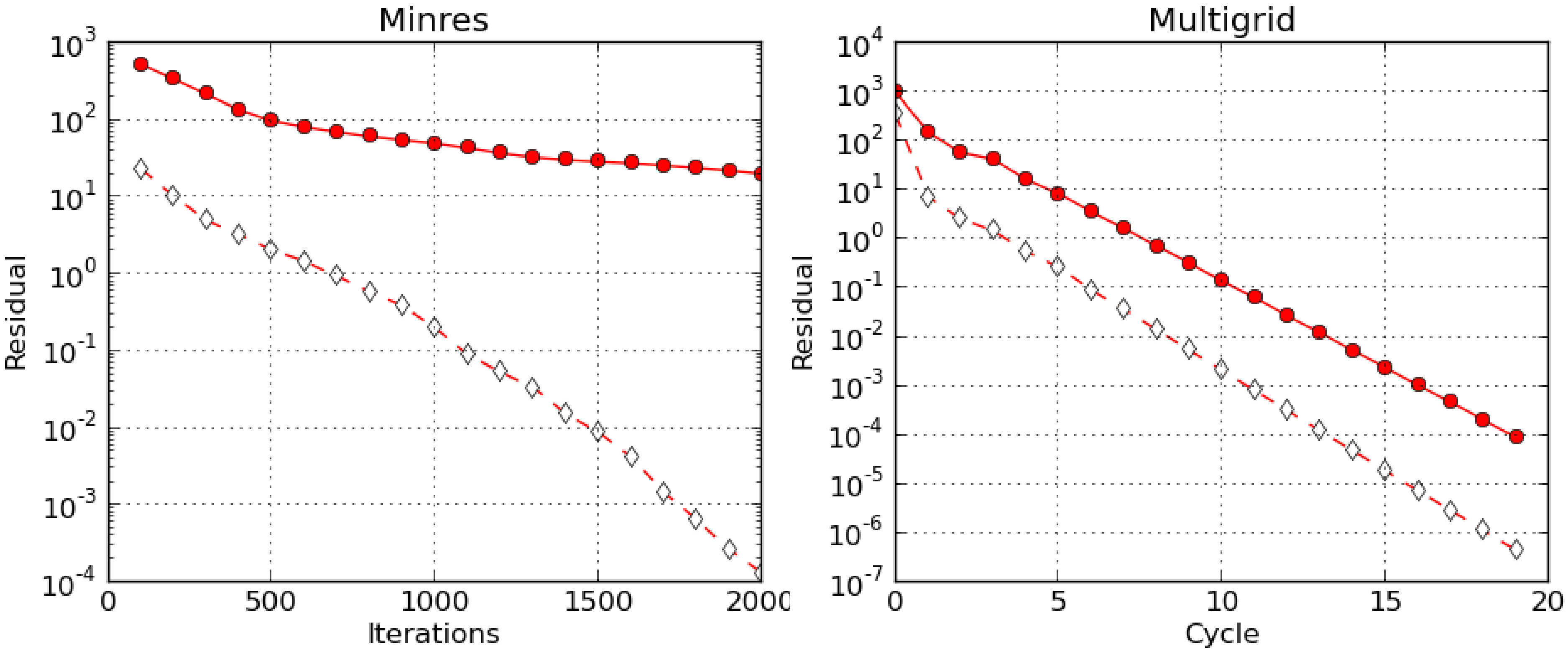
$$\min \int_{T_1}^{T_2} \|v\|^2 dt \quad \text{s.t.} \quad \frac{dv}{dt} = f_u v + f_s$$

- Gauss elimination of KKT system leads to

$$\begin{pmatrix} F_1 & G_1 & & & \\ & F_2 & G_2 & & \\ & & \ddots & \ddots & \\ & & & F_n & G_n \end{pmatrix} \begin{pmatrix} F_1^T & & & \\ G_1^T & F_2^T & & \\ & G_2^T & \ddots & \\ & & \ddots & F_n^T \\ & & & G_n^T \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = - \begin{pmatrix} f_{s1} \\ f_{s2} \\ \vdots \\ f_{sn} \end{pmatrix}$$

- S.P.D. but globally coupled system
- It takes efficient iterative solution method for size ndof x nstep sparse system

Multigrid is the method of choice for Least Squares Sensitivity



Time multigrid for Lorenz attractor

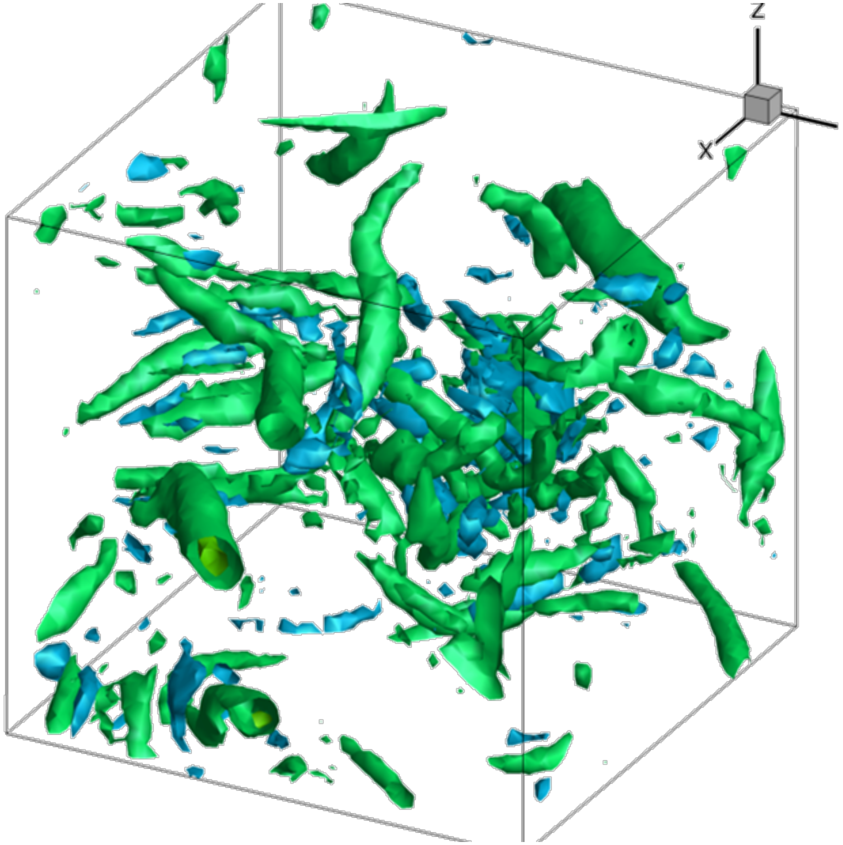
Solid dots: Dt = 0.001, Open dots: Dt = 0.008

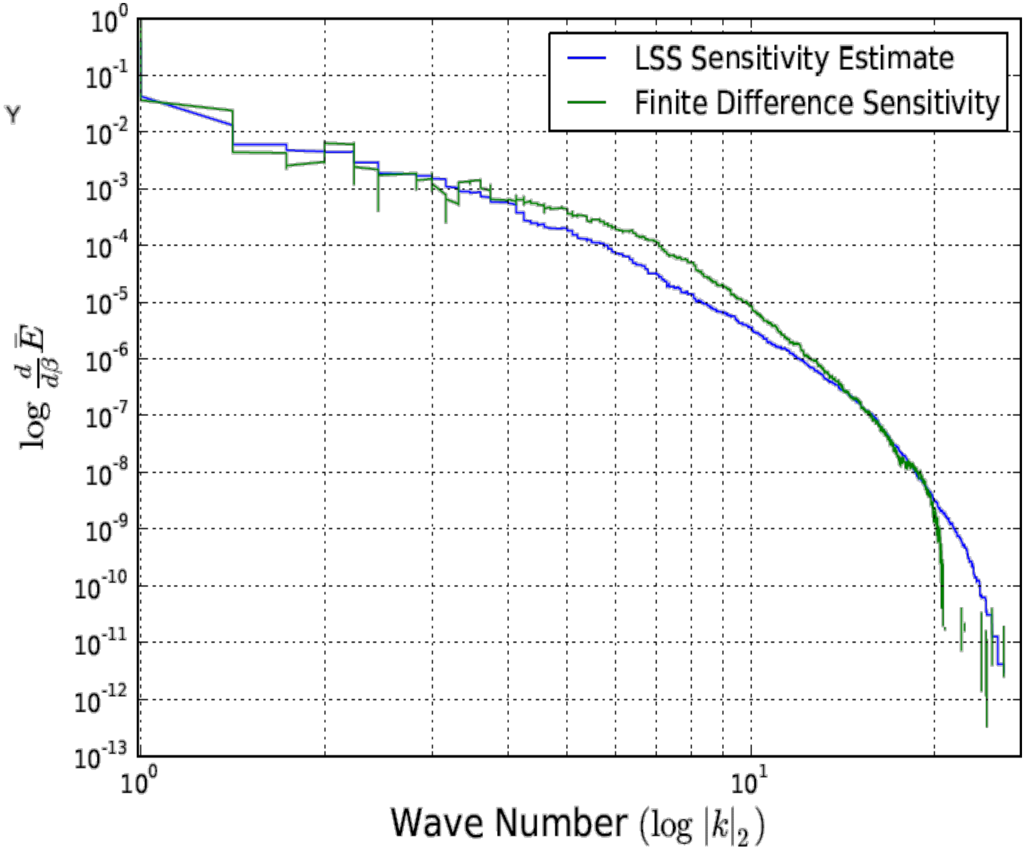
Multigrid for Least Squares Sensitivity: Lessons learned

$$\frac{d^2 w}{dt^2} + \frac{d}{dt} f_u^T w - f_u \frac{dw}{dt} - f_u f_u^T w = f_s$$

- f_u is a function of time.
  - Iterations on very coarse grids (larger Dt) useless.
  - Anti-aliasing in u(t) when coarsening is critical.
- Krylov iterations works better than fixed point smoother – why?
- Performance is sensitive to order of space and time coarsening.

Parallel space-time multigrid enables least squares computation for isotropic homogeneous turbulence





Taylor micro-scale Reynolds number = 33

A Potential Paradigm Shift in Simulating Chaos

Initial Value Formulation	Least Squares Formulation
Ill-conditioned	Well-conditioned
Suitable for time advancing Storing only a few time steps	Time advancing does not work Store all time steps at once

A Potential Paradigm Shift in Simulating Chaos

Initial Value Formulation	Least Squares Formulation
Ill-conditioned	Well-conditioned
Suitable for time advancing Storing only a few time steps	Time advancing does not work Store all time steps at once
Sequential in nature, do not parallelize well in time	Breaks causality, scalable parallelization in time

A Potential Paradigm Shift in Simulating Chaos

Initial Value Formulation	Least Squares Formulation
Ill-conditioned	Well-conditioned
Suitable for time advancing Storing only a few time steps	Time advancing does not work Store all time steps at once
Sequential in nature, do not parallelize well in time	Breaks causality, scalable parallelization in time
Suitable when computer size to problem size ratio is small **Current computing paradigm**	**Suitable when computer size to problem size ratio is large** **Potential "Exascale" computing paradigm**

Towards computational engineering of chaotic systems

- Solution of initial value problems of chaos is ill-conditioned and unsuitable for many computational engineering purposes.
- Our new formulation removes the initial condition, uses least squares instead
  - Stable trajectories for perturbed parameters
  - Computes correct, useful sensitivity
- Current work on scalability
  - Will enable design, control, characterization and UQ for an important class of aerospace applications on which current methods fail

Towards Scalable Long Time Integration of Chaotic Systems

- Many applications require **rapid simulation turnaround time** that is not achievable by spatial-only parallelism even on next generation computers.
- **Scalable parallel time integration** can be achieved by **replacing the initial condition** and with a **Least Squares problem**.