Petascale Computing Workshop
NCAR
7 May 2008

Structures and statistics:
Understanding turbulent transport using petascale resources

Mark Rast
Laboratory for Atmospheric and Space Physics
Department of Astrophysical and Planetary Sciences
University of Colorado, Boulder

"If you had access to a petascale computing system, what would you do with it?"

"If you had access to a petascale computing system, how would you use it?"

"If you had access to a petascale computing system, what problem would you solve?"

LASP

APS

Naïve assumption: direct numerical solution of the Navier-Stokes equations on unprecedentedly fine grids will be possible using these platforms

Greatest challenge: Data volumes

What problem?

Turbulence: the "perfect" problem for petascale

Formulate a statistical description of small-scale properties which is sensitive to large-scale driving and provides a model for transport

→ **Statistics of Turbulent Structures**

LASP

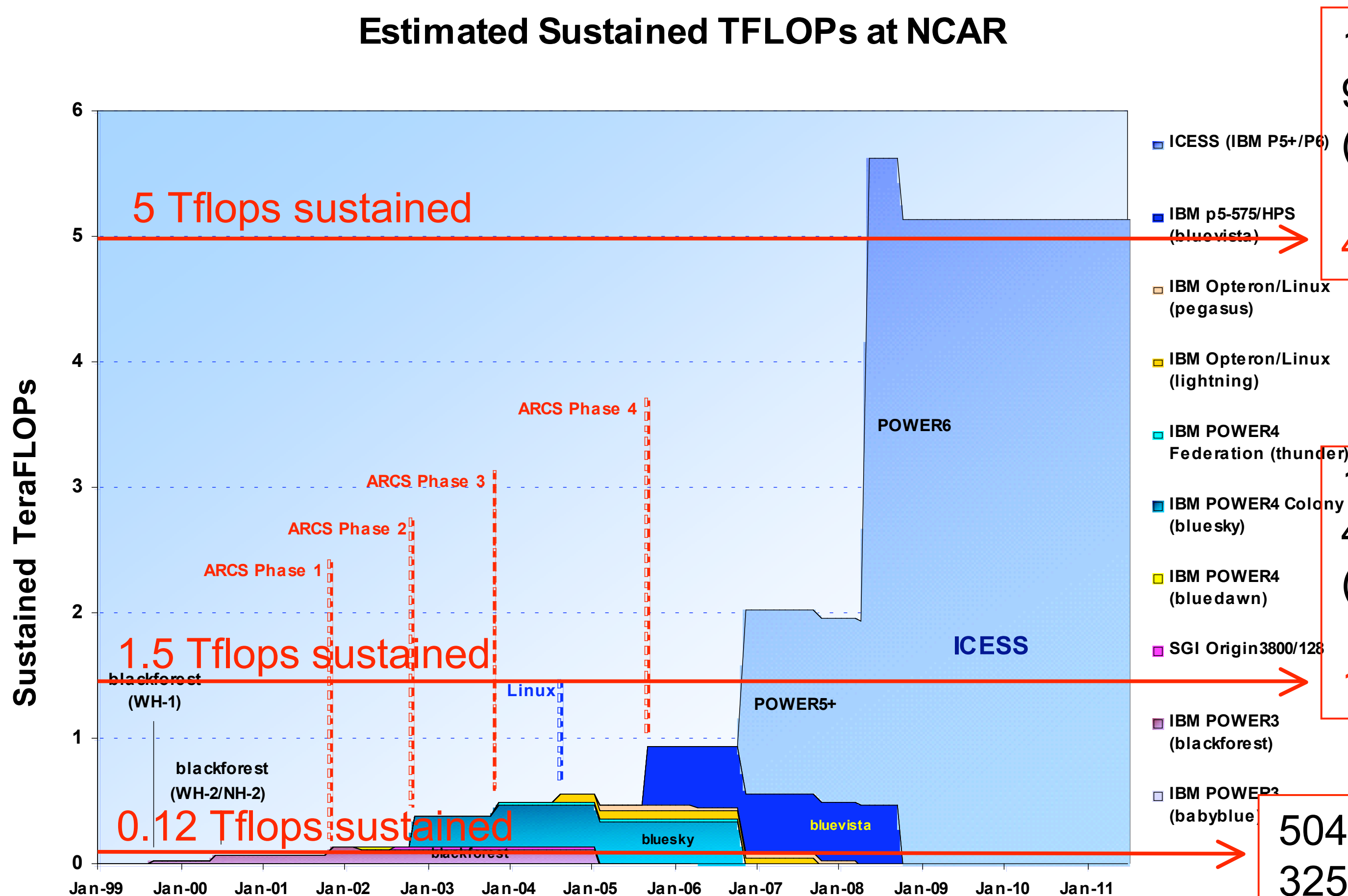
APS

The postdoc class problem (sustained effort on 1/10 of the machine):

1 Pflop sustained

8200^2 x 65536
145,000,000 pe hrs
(50,000/500,000 6.0GHz pe,
4 - 16 months)
85000TB data 16394^3

85 petabytes



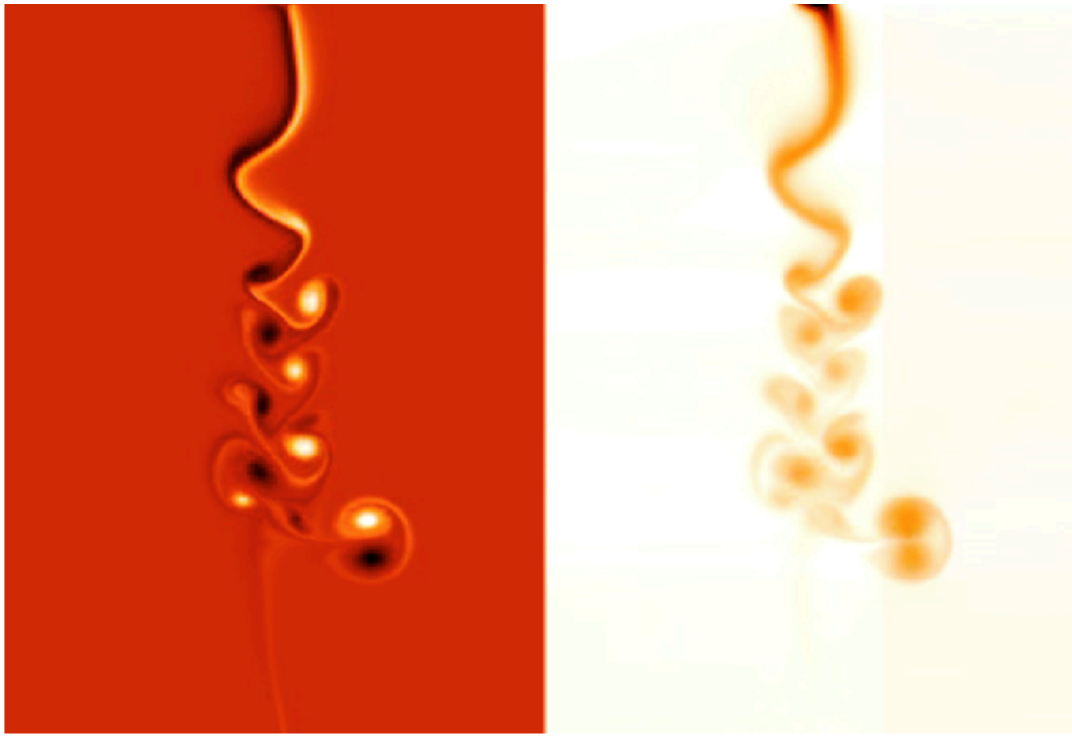
1650^2 x 8192
930,000 pe hrs
(320/3200 4.0GHz pe,
4 - 16 months)
430TB data 2814^3

1260^2 x 4096
460,000 pe hrs
(160/1600 1.9GHz pe,
4 - 16 months)
125TB data 1866^3

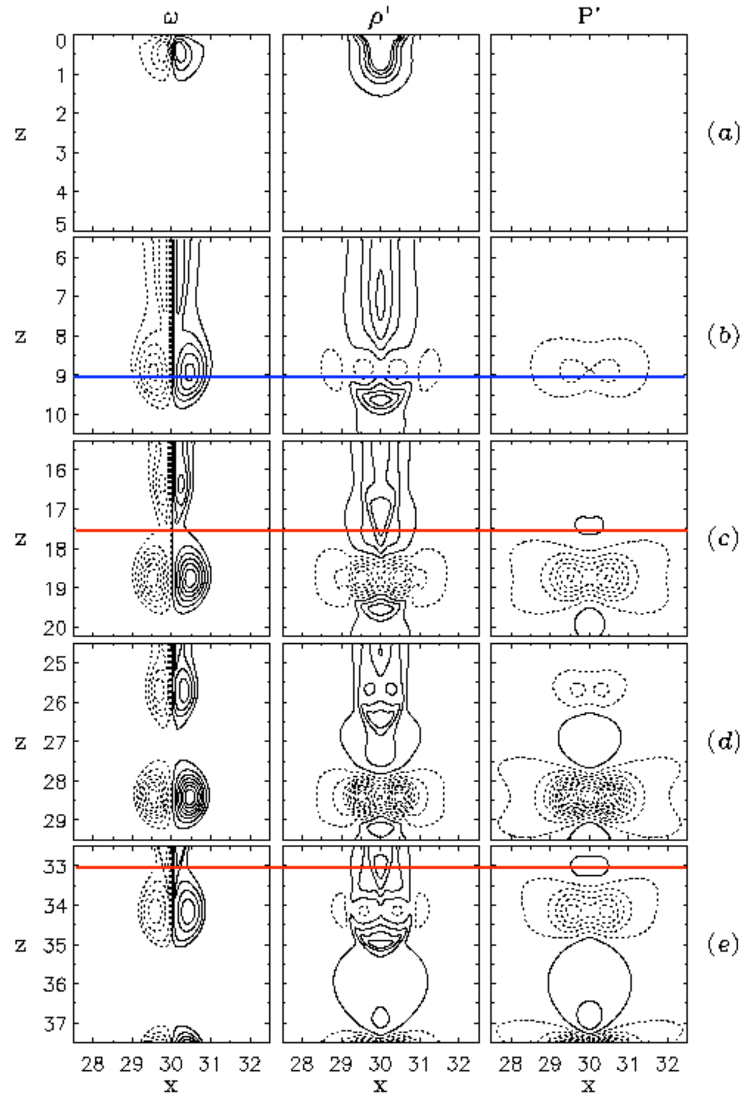
504^2 x 2048 (483 time steps)
325,000 pe hrs
(112/1160 375MHz pe,
4 - 16 months)
10TB data 804^3

LASP

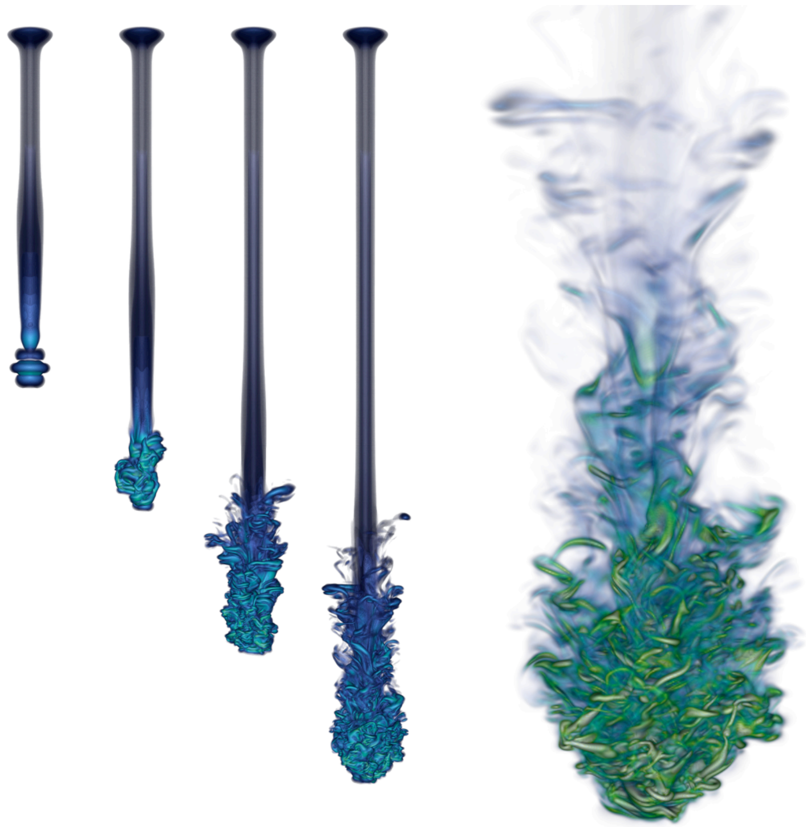
IDL era (2D slice and dice):



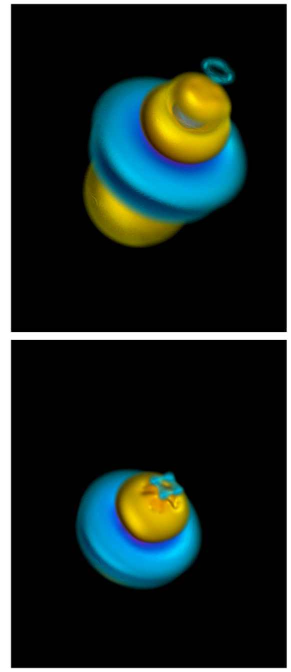
23plumevideo.mpg

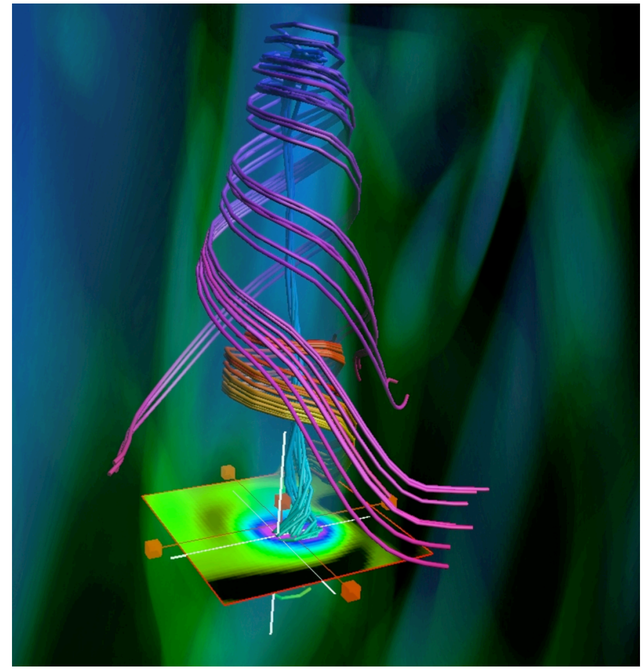


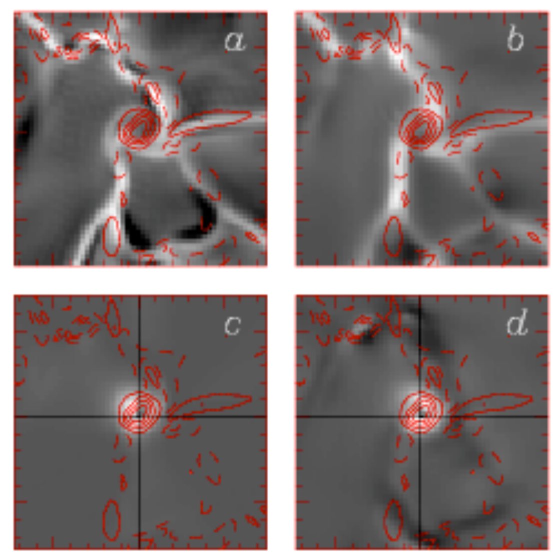
VAPOR era
(3D multi-resolution and sub-domain selection for interactive analysis):



504^2 x 2048







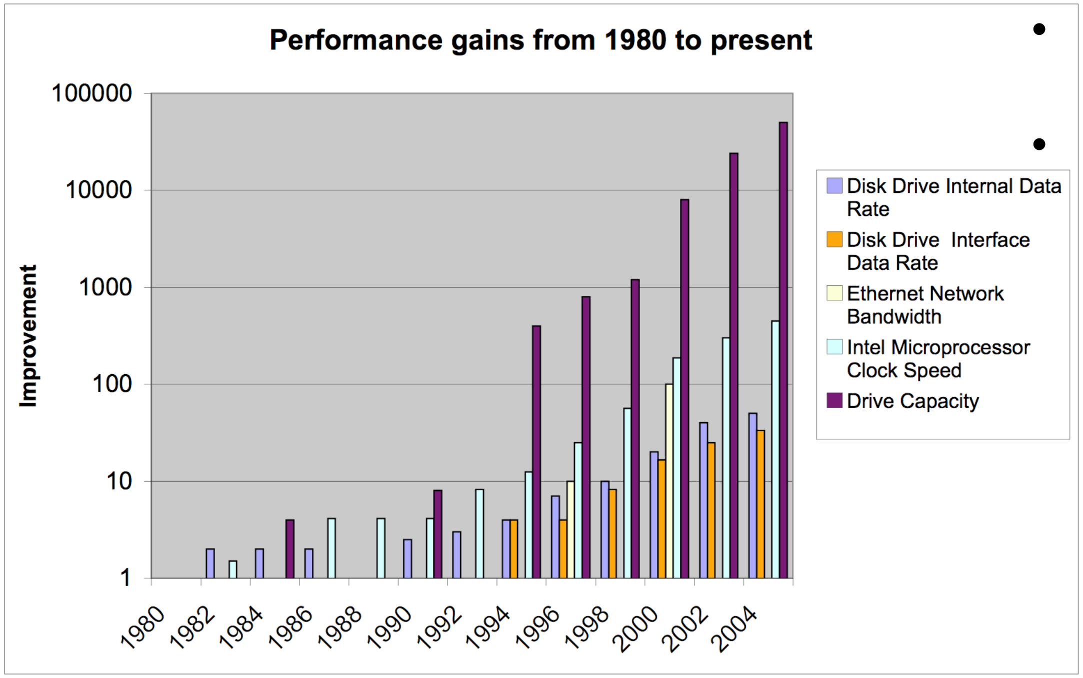
$-\nabla_H \cdot \rho \mathrm{u}$

$-\frac{\partial p}{\partial z} + \rho g$

$\frac{u_\theta^2}{r}$

$\frac{1}{\rho}\frac{\partial p}{\partial r}$

A posteriori analysis and visualization of the data volumes can not keep up with batch capabilities:



- Not all technologies advance at the same rate
- Multi/*insanely-large-number* processor simulation vs. single/dual/quad/*small-number* processor analysis and visualization

16384^3 simulation will require decimation by factors of about 32^3 for interactivity

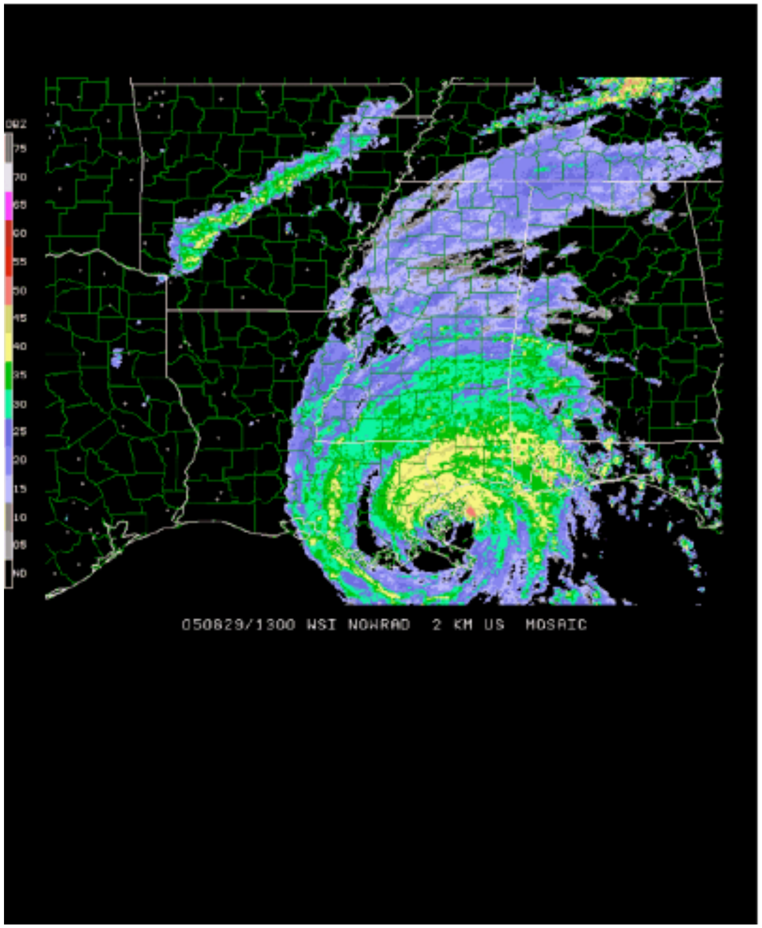
THE HOPELESS SITUATION THEOREM:
Doubling the resources available to a batch execution will increasingly overload a corresponding doubling of the resources available for interactive *analysis* and visualization.

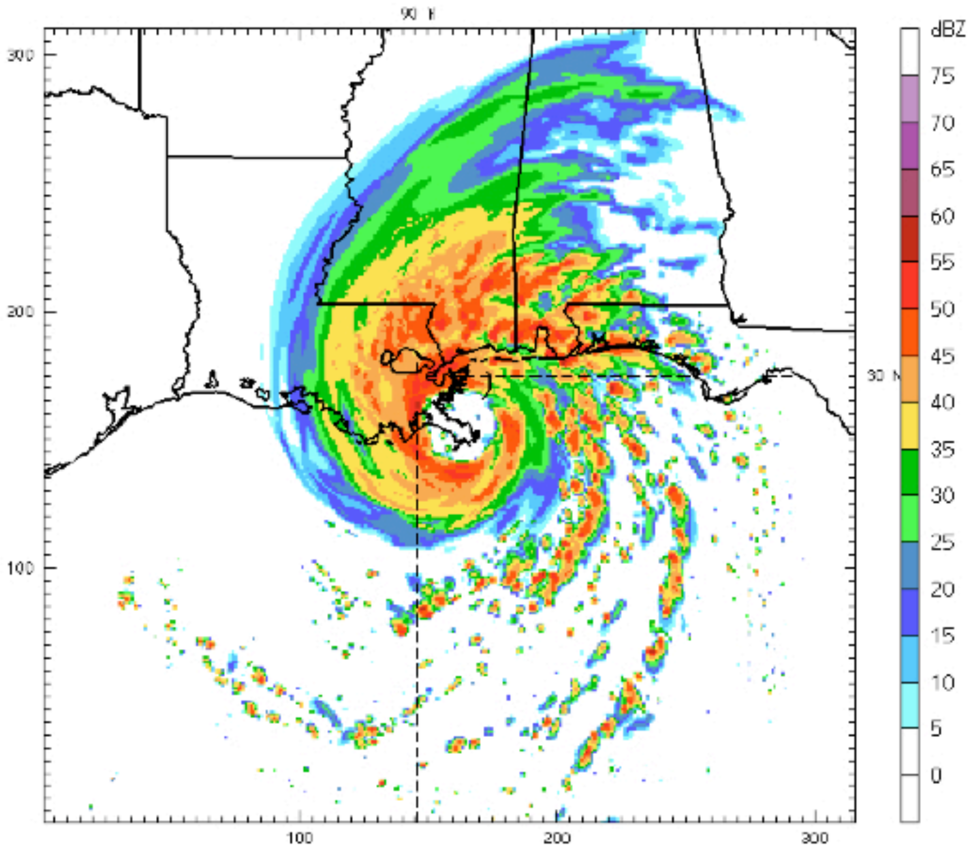
Data decimation BEFORE batch output will be essential.

LASP

APS

Forecast models – reduced output on reduced grid:

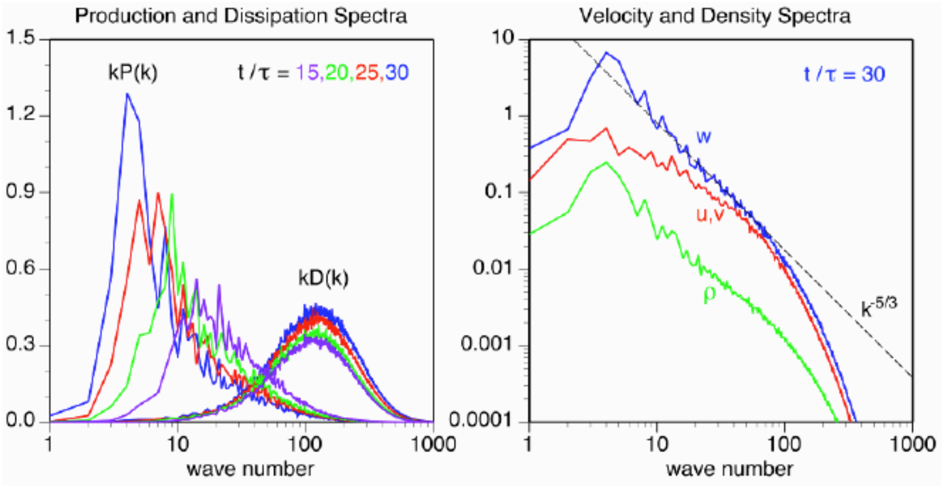


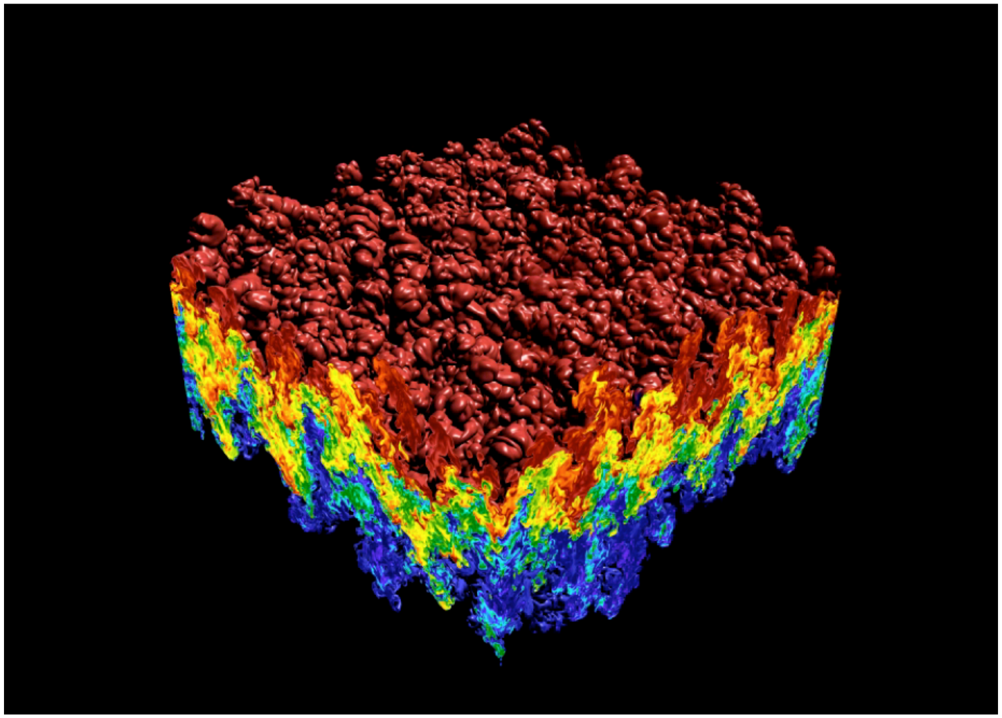


NCAR's Advanced Research version of the
Weather Research and Forecasting model (WRF)

Statistics – reduction in dimensionality:

BlueGene/L with a sustained speed of 280.6 teraFLOPS

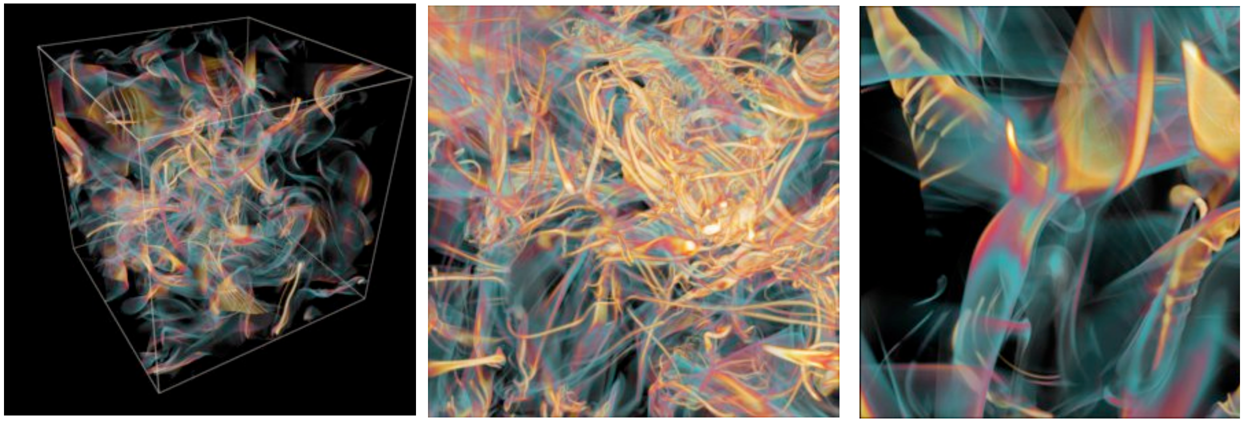




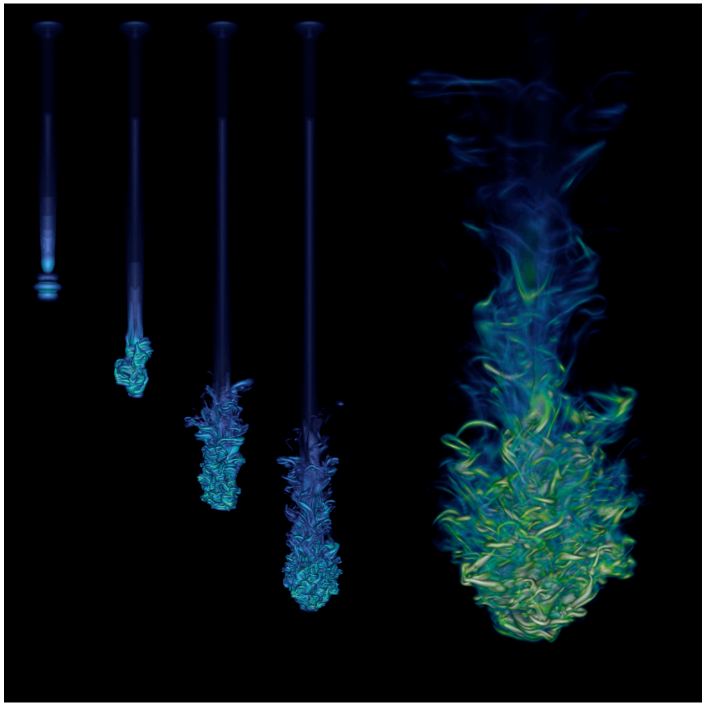
Resolved Rayleigh-Taylor instability
3072^3

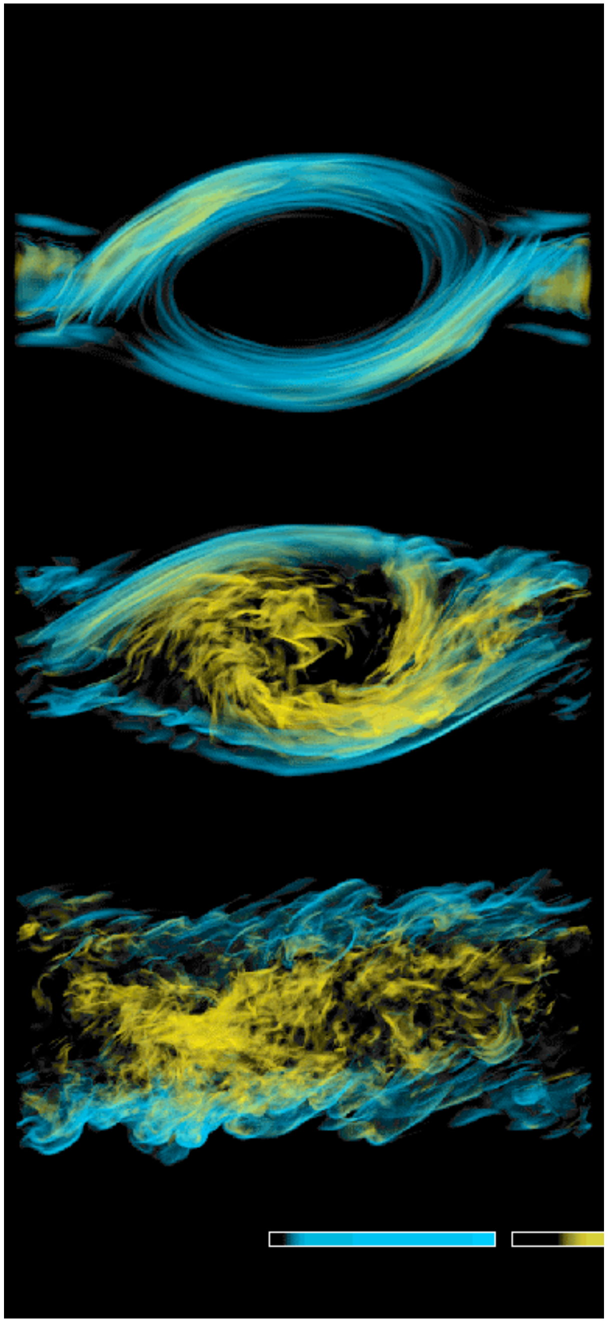
Lawrence Livermore National Laboratory (2006)

Structures in turbulent flows:



Compressible turbulence – Porter, Woodward, Winkler & Hodson

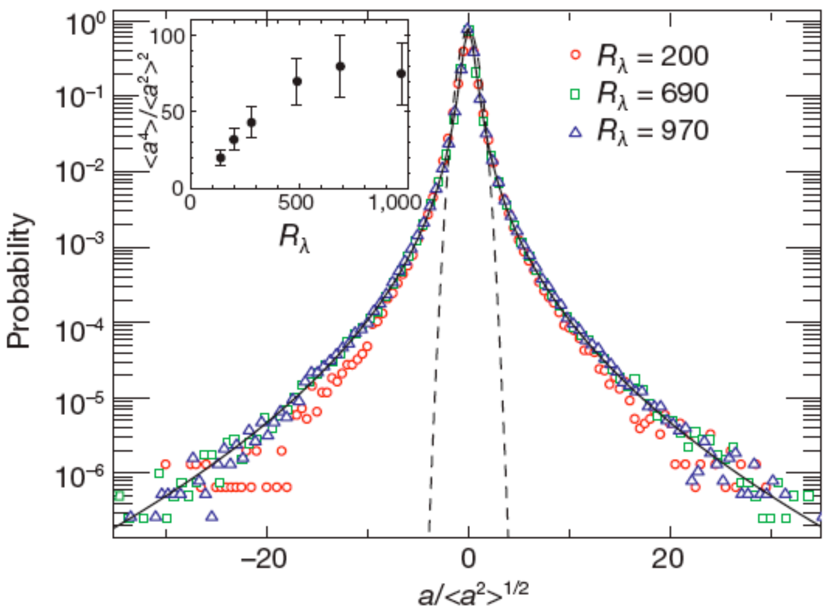


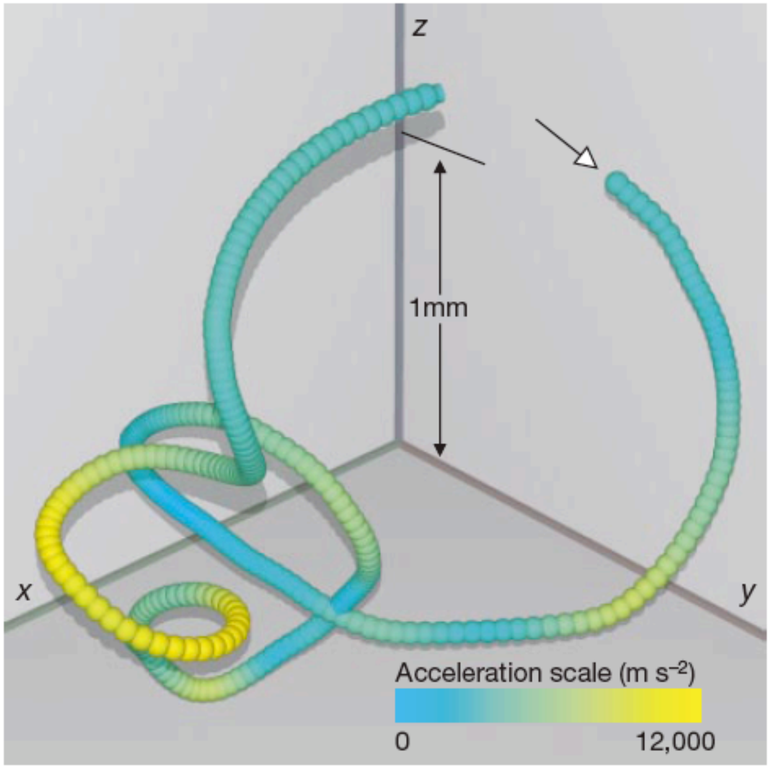


Viscous (yellow) and thermal (blue) dissipation in stratified shear turbulence – Werne & Fritts

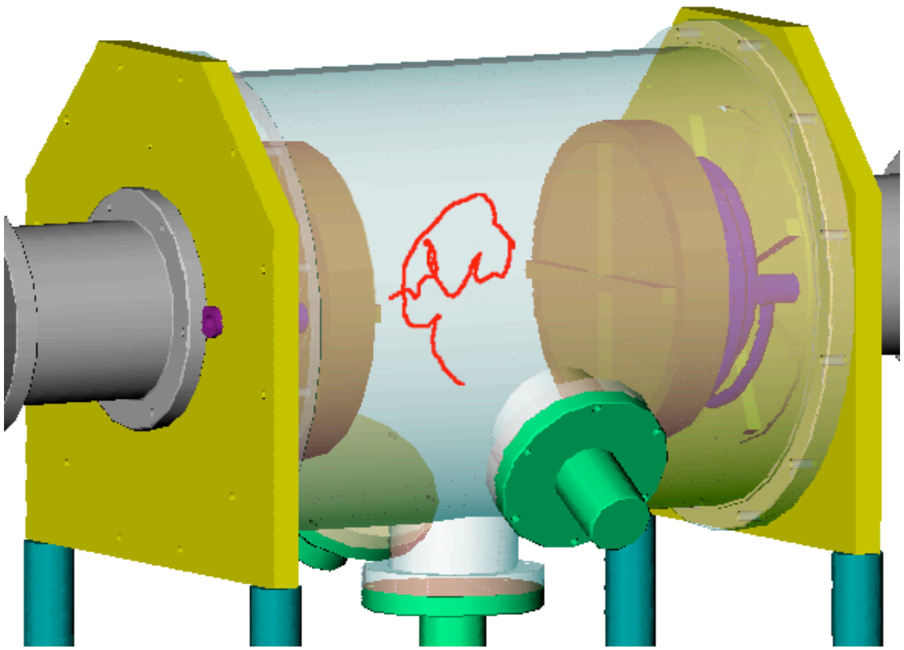
Lagrangian statistics:

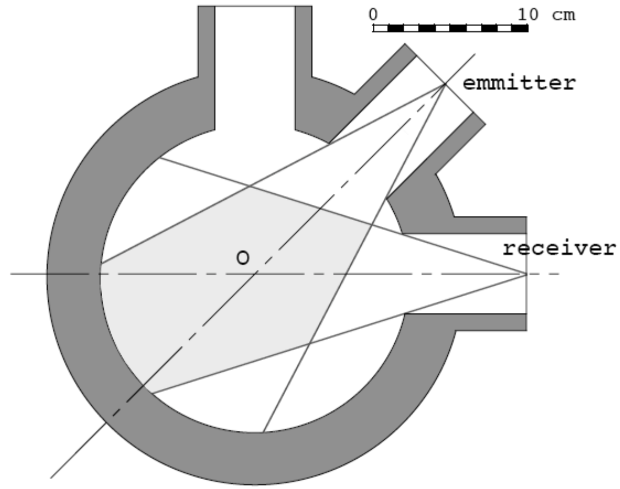
La Porta et al. 2001, Nature, 409, 1017

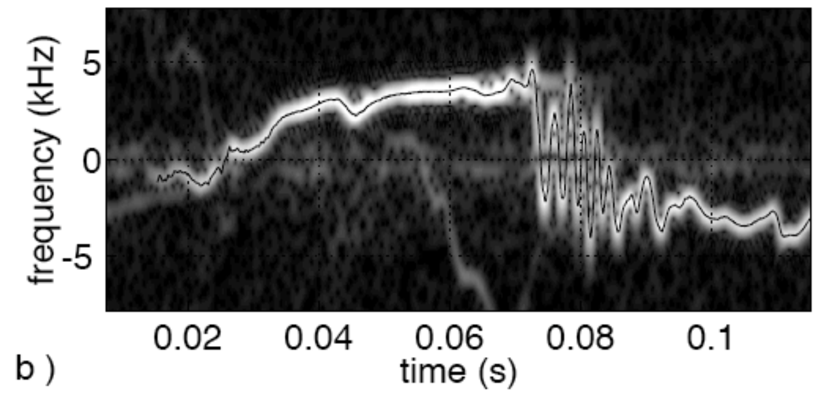




Mordant, N., Lévèque, E., & Pinton, J.-F.
Phys Fluids 2006

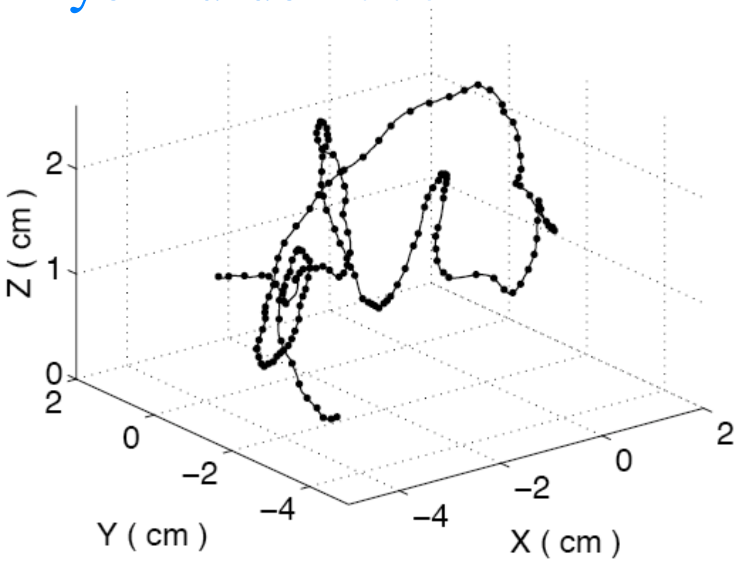


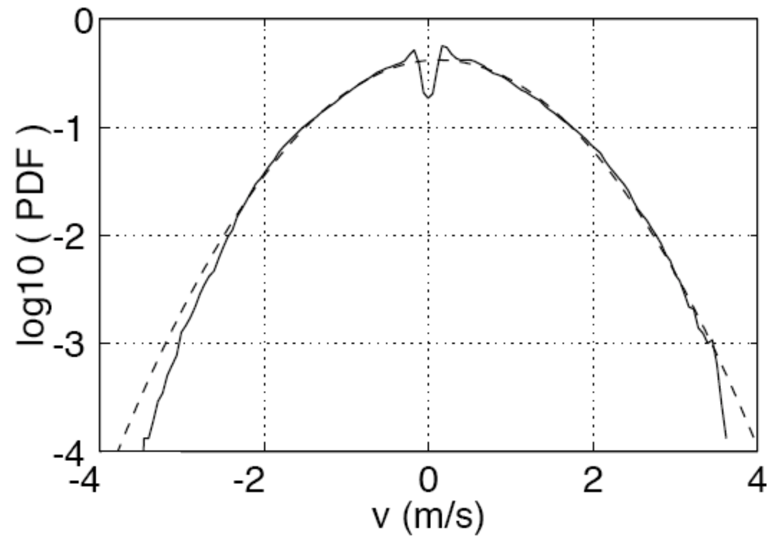




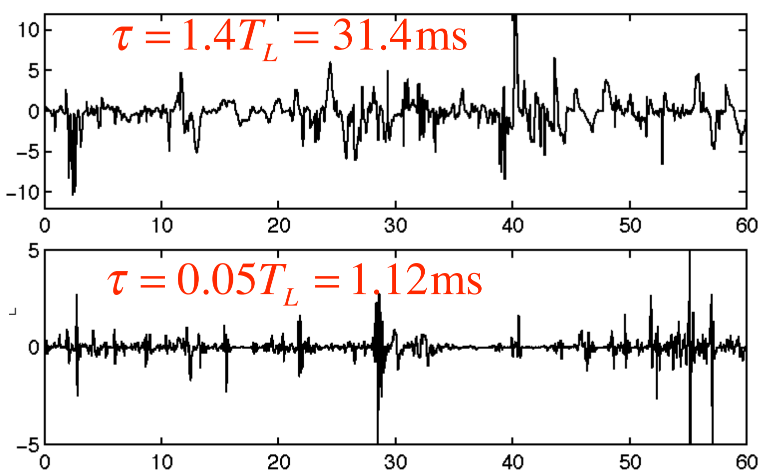
- Tank radius 10cm (9 liter volume) filled with water
- Counter rotating disks (9.5cm diameter, 18cm separation)
- 250μm diameter 1.06 g/cm^3 tracer particles (smaller than Taylor microscale)
- Beam width at center of volume (no mean flow) 10 cm
  (larger than the integral scale -- sample full range of Lagrangian motions)
- Total number of tracers small (less than two in sample volume at one time)

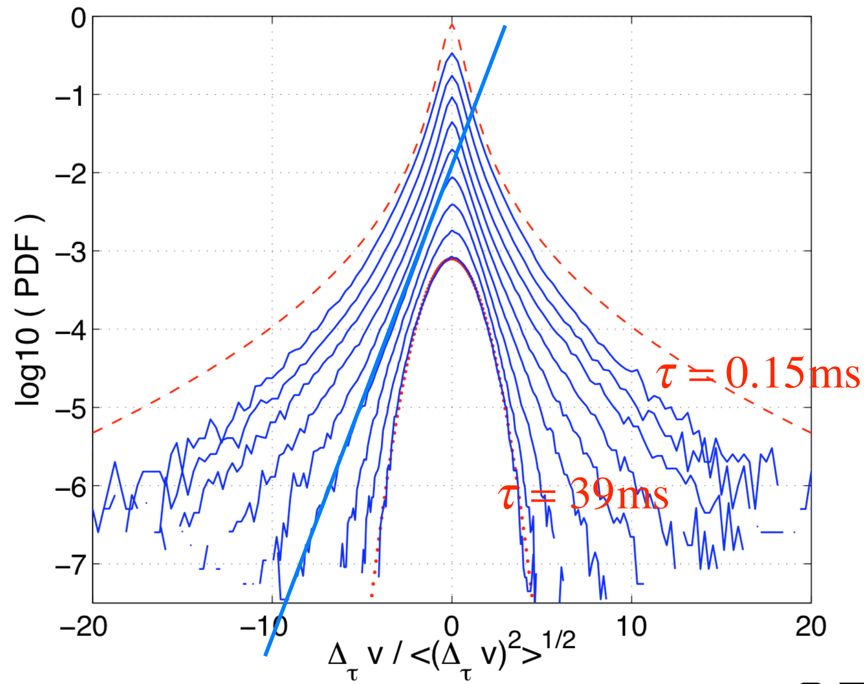
Mordant, N., Lévèque, E., & Pinton, J.-F.
Phys Fluids 2006





$\Delta_\tau v(t) = v(t+\tau) - v(t)$



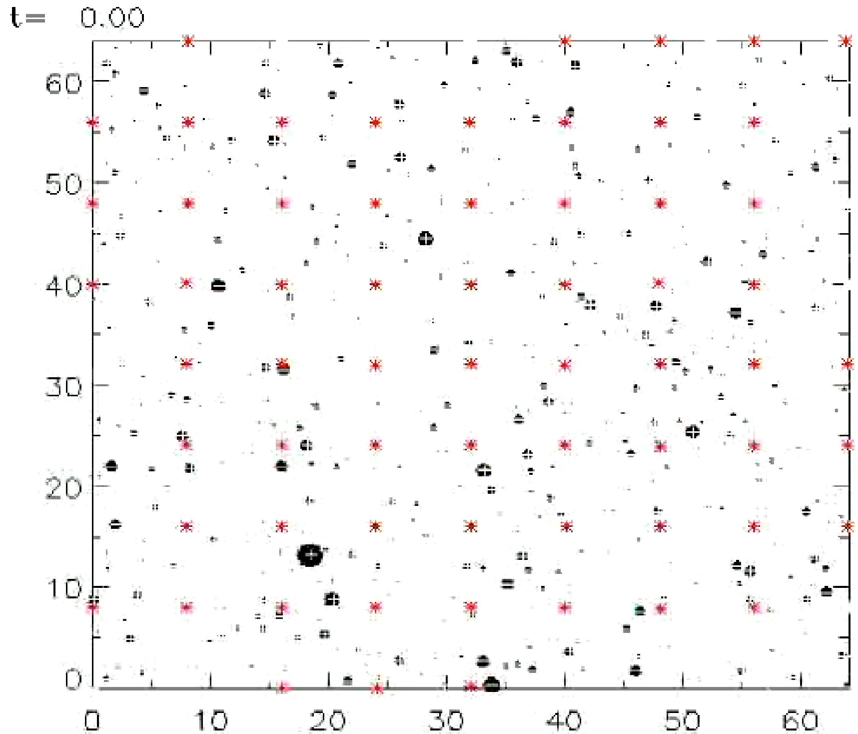


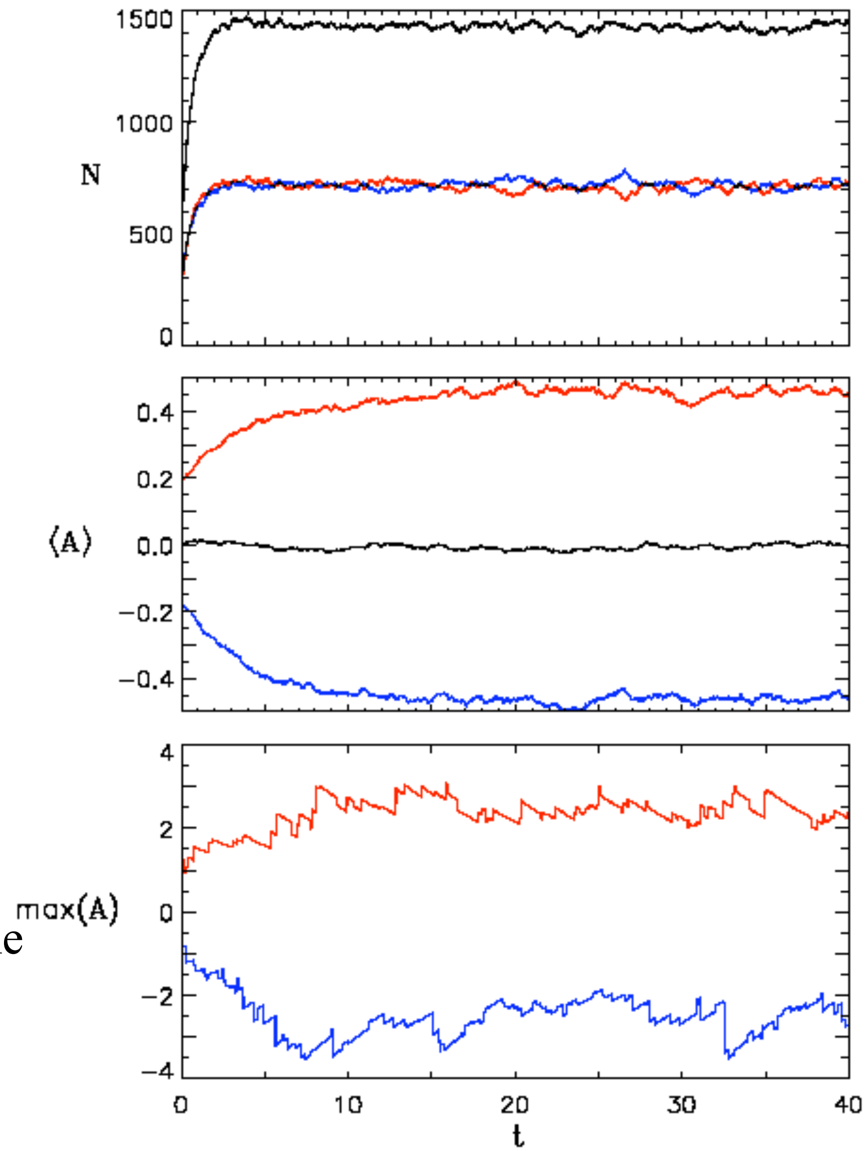
$T_L \simeq 22.4\text{ms}$ $T_L \equiv$ Lagrangian velocity autocorrelation time

Point vortex simulations:

$u_\theta \sim 1/r$

$u_r \sim 0$





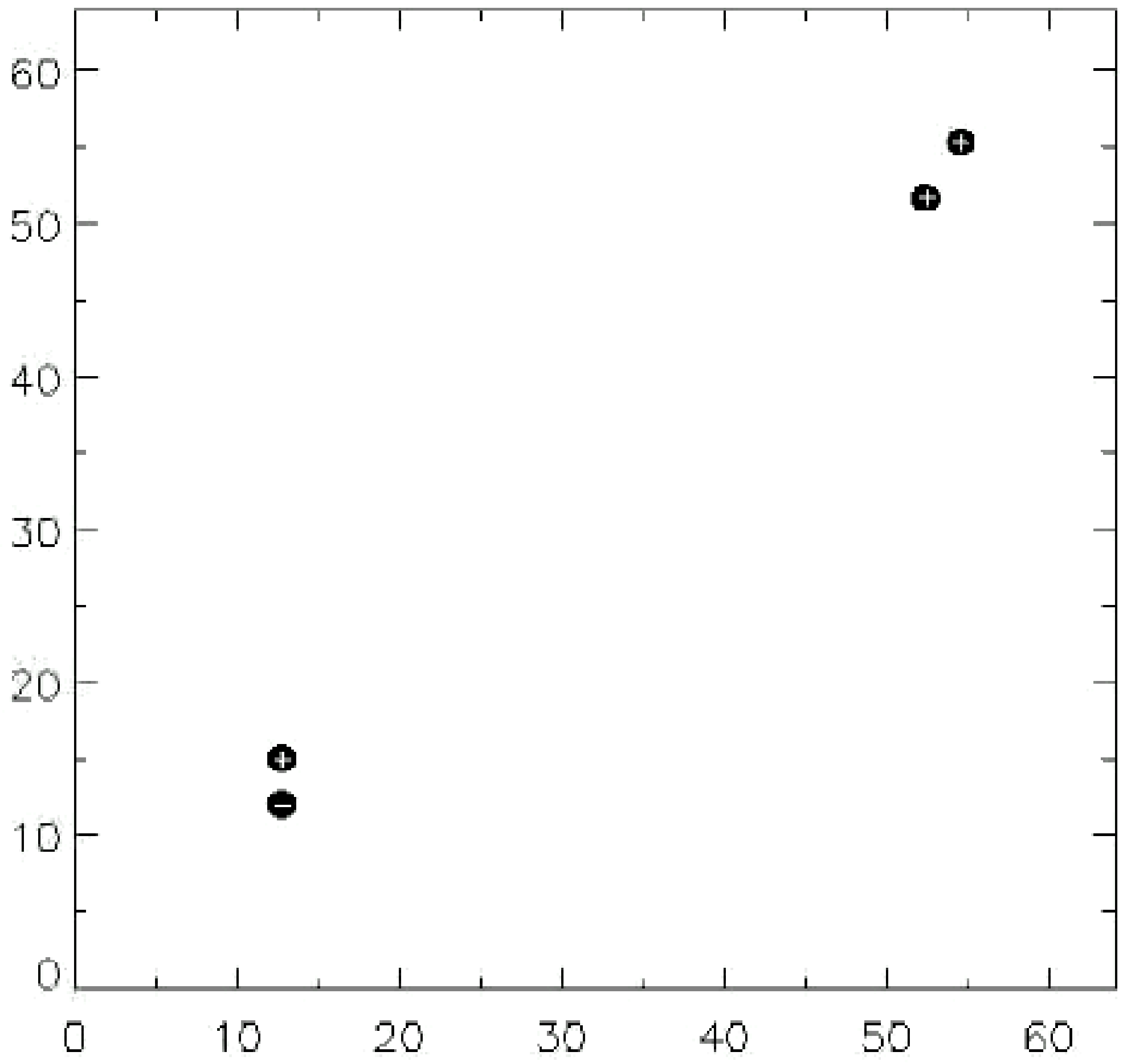
6.25 new vortex sites per unit area per unit lifetime

2.56 new vortex sites in domain per time step

~~$\tau = 10$~~

~~$dt = 10^{-4}\,\tau$~~

Every 100 time steps shown

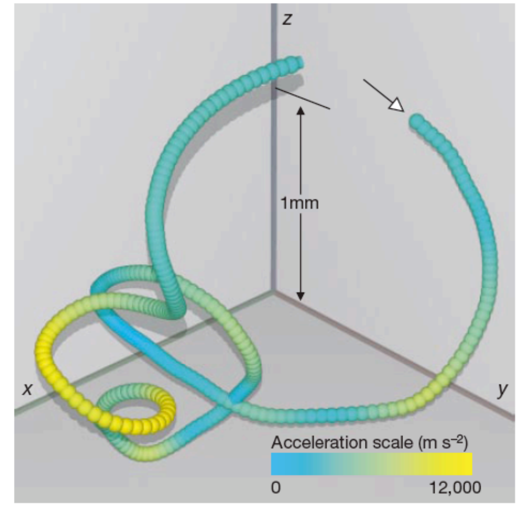


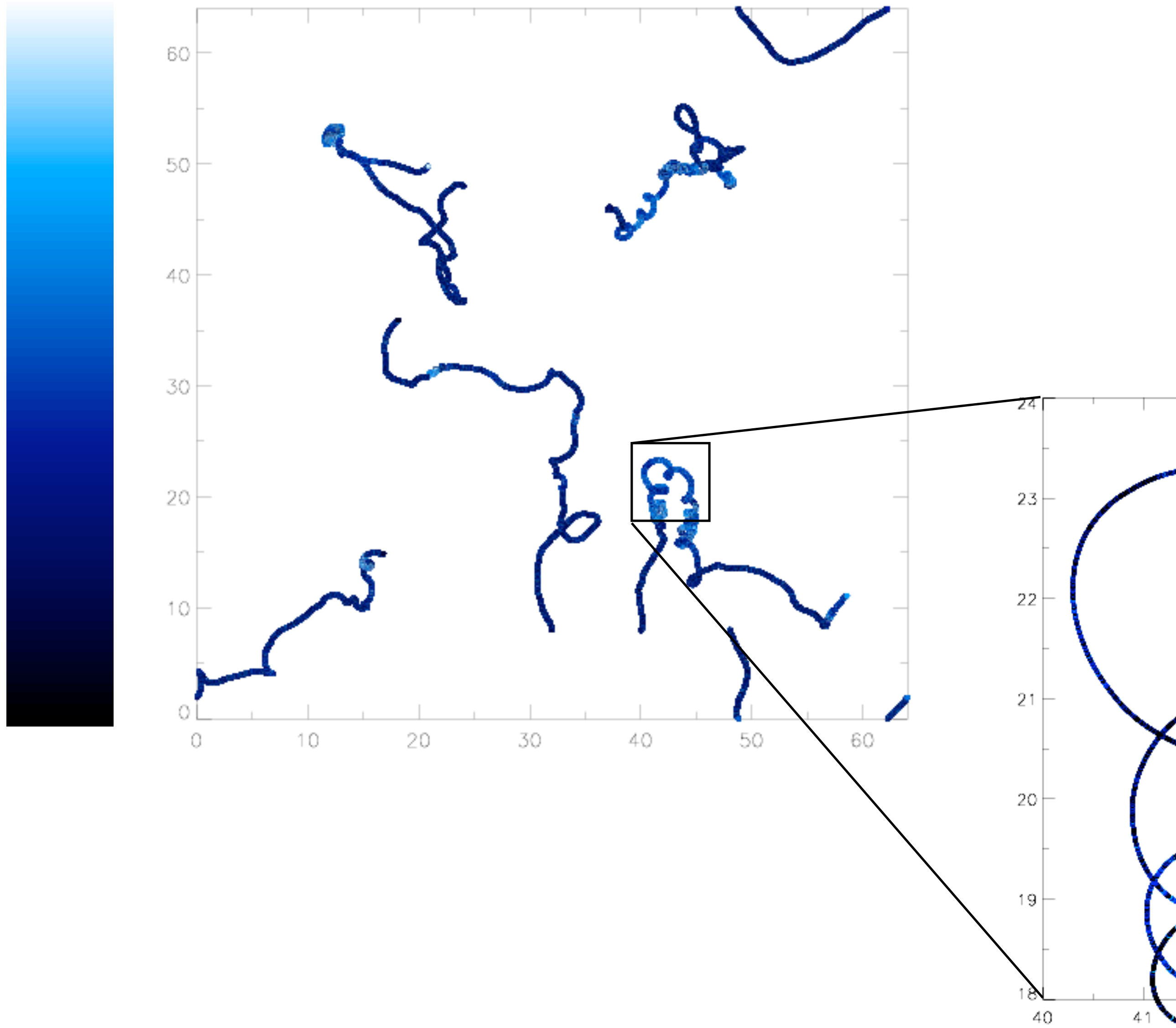
- Like sign vortices orbit
- Oppositely signed vortices translate
- Scattering leads to preferential merger of oppositely signed pairs

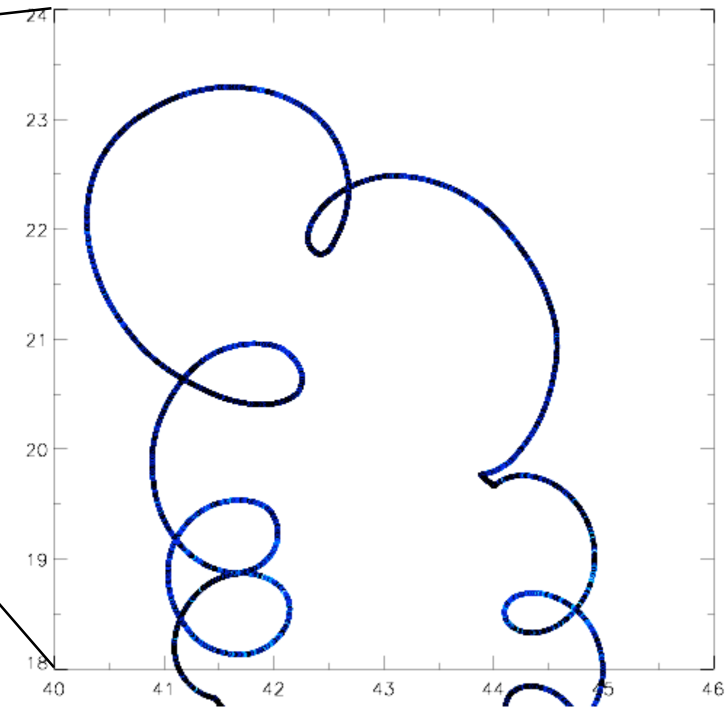
LASP

APS

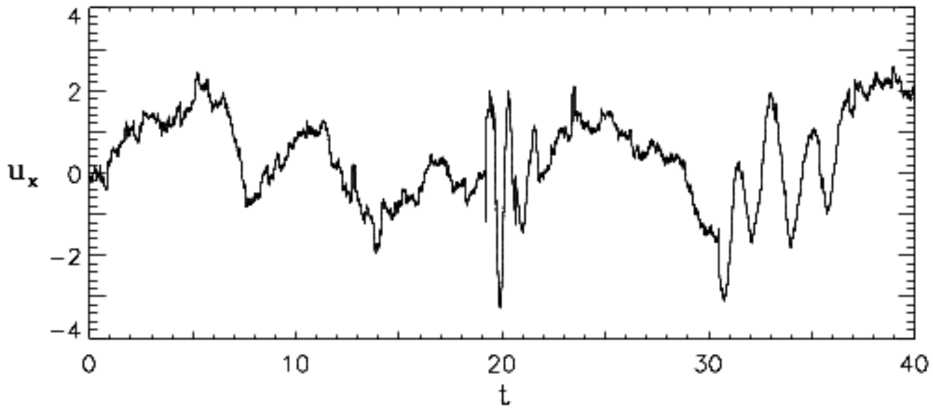
Particle trajectories in point – vortex model:

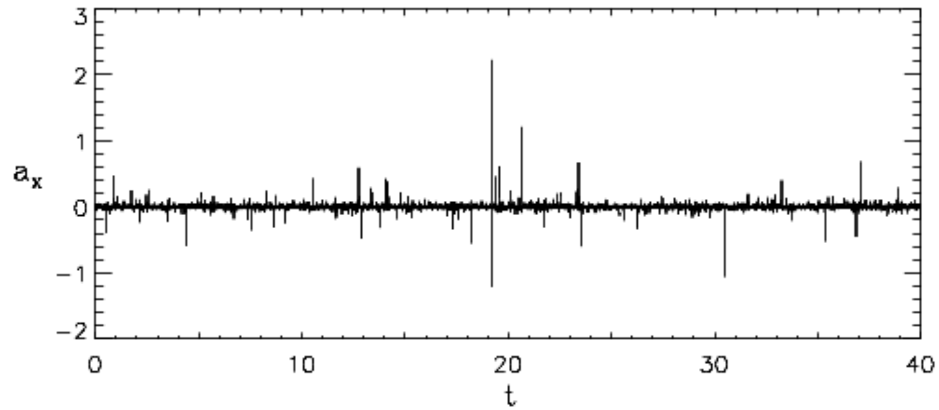


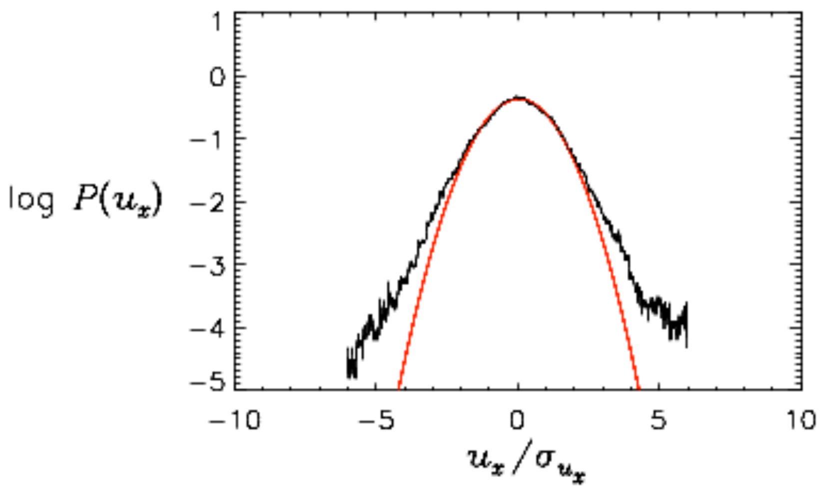


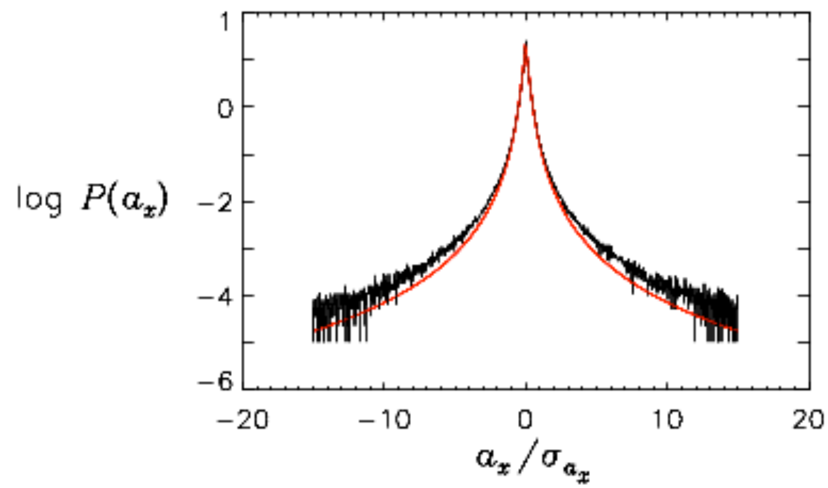


Velocity and acceleration distributions in point – vortex model:

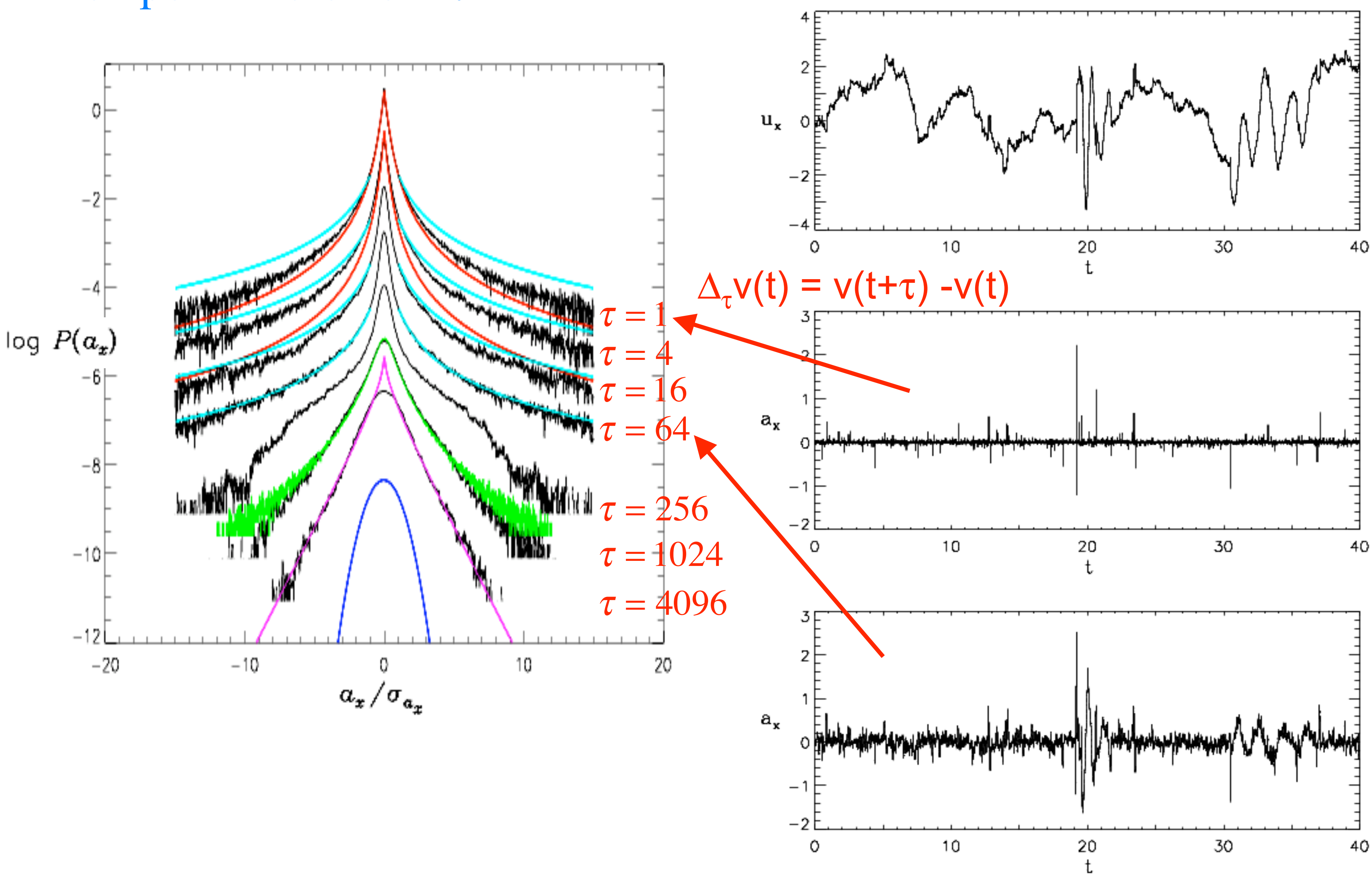








Temporal increment τ



t in units of 1000 time steps

Bivariate transformation of random variables:

Let x and y be independent random variables with probability densities $P(x)$ and $P(y)$ and joint probability density $P_{xy}(x,y) = P(x)P(y)$

Let $u = f(x,y)$ and $v = g(x,y)$ be functions of the random variables with inverse functions $x = h_1(u,v)$ and $y = h_2(u,v)$

Then the joint probability density of u and v is $P_{uv}(u,v) = P_{xy}(h_1,h_2)\begin{vmatrix} \dfrac{\partial h_1}{\partial u} & \dfrac{\partial h_1}{\partial v} \\ \dfrac{\partial h_2}{\partial u} & \dfrac{\partial h_2}{\partial v} \end{vmatrix}$

and $P(u) = \int P_{uv}(u,v)\, dv$ and $P(v) = \int P_{uv}(u,v)\, du$

Example:

Consider two Gaussianly distributed independent random variables each with a Mean value of zero and variance equal to one:

$$P_{xy}(x,y) = P(x)P(y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

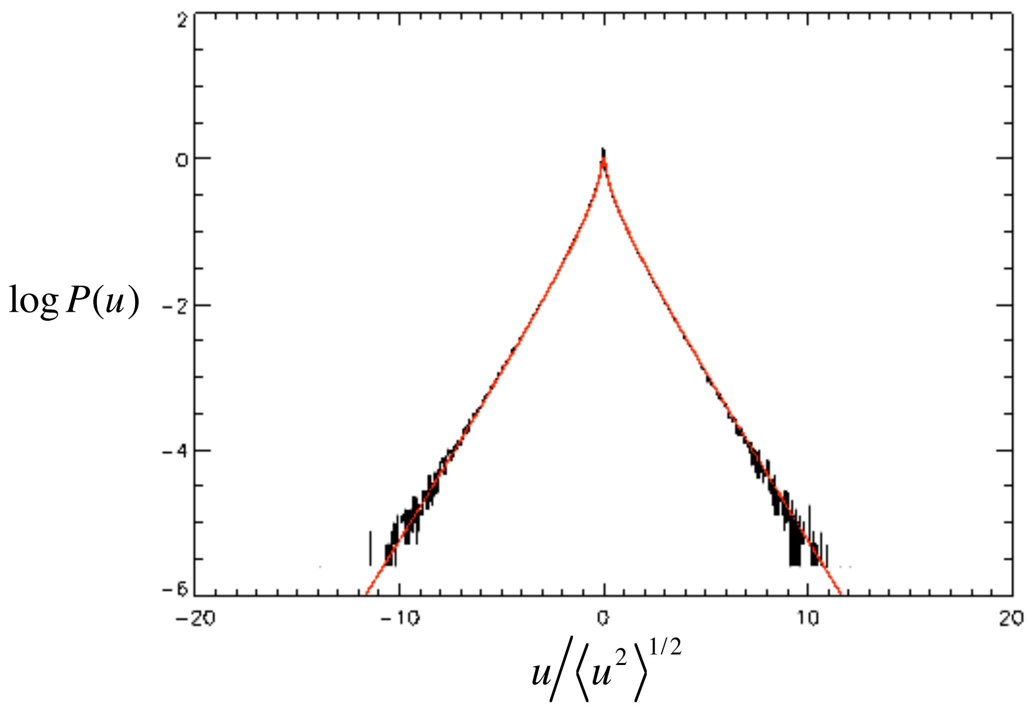
To derive the probability density of their product, let $u = xy$ and $v = y$ with inverses $x = u/v$ and $y = v$

The joint probability density of u and v is then

$$P_{uv}(u,v) = \frac{1}{2\pi} e^{-(u^2/v^2+v^2)/2} \begin{vmatrix} \frac{1}{v} & -\frac{u}{v^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{2\pi v} e^{-(u^2/v^2+v^2)/2}$$

and integrating over "dummy" function v yields $P(u) = \frac{1}{\pi} K_0\left(\sqrt{u^2}\right)$ $u = xy$

K_0 is the lowest order modified Bessel function of the second kind



Monte Carlo vs. analytic probability density for the Gaussian product $N_1 N_2$

Velocity around a single point – vortex:

When radial distance to vortex is sampled randomly in the plane:

$$\mathbf{u} = U_0 / r \ \hat{\theta} \qquad u_x = \frac{U_0}{r}\sin\theta \qquad U_0 = \text{constant}$$

$$P(r) \propto r \qquad P(\theta) \text{ uniformly distributed}$$

Bivariate transformation of random variables:

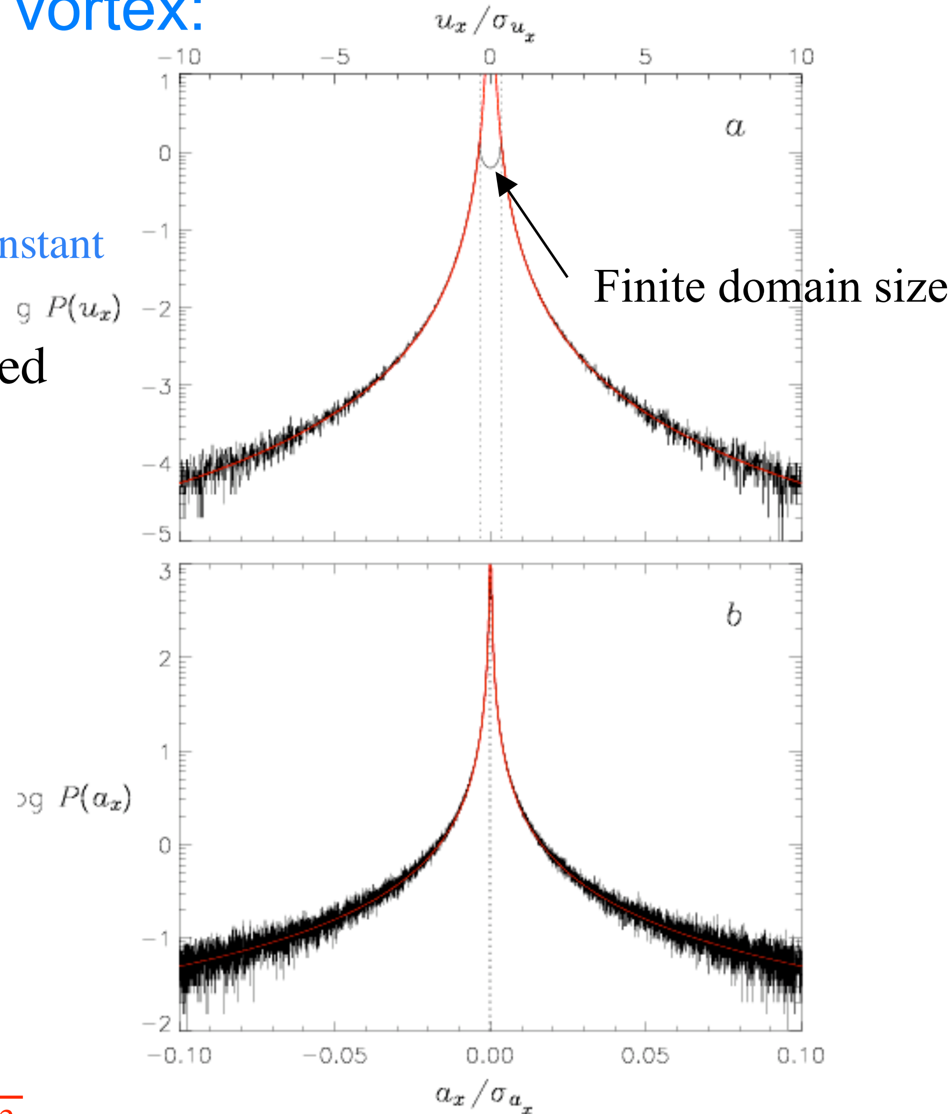
$$x = \sin\theta \qquad y = r$$

$$P(x) = \frac{1}{\cos\theta} = \frac{1}{\cos(\sin^{-1}x)} \qquad P(y) = r = y$$

$$P_{xy}(x,y) = P(x)P(y)$$

$$u = x/y \qquad v = \sin^{-1}x$$

$$\Rightarrow \quad P(u_x) \propto \frac{1}{u_x^3} \qquad P(a_x) \propto \frac{1}{a_x^{5/3}}$$



Velocity in field of randomly placed point – vortices:

$$u_x = \frac{U_0}{r}\sin\theta$$

$U_0 = \text{constant}$

$$P(\sin\theta) = \frac{1}{\cos\theta}$$

Keeping only nearest neighbor contribution:

$$P(r) = 2\pi n r e^{-\pi n r^2}$$

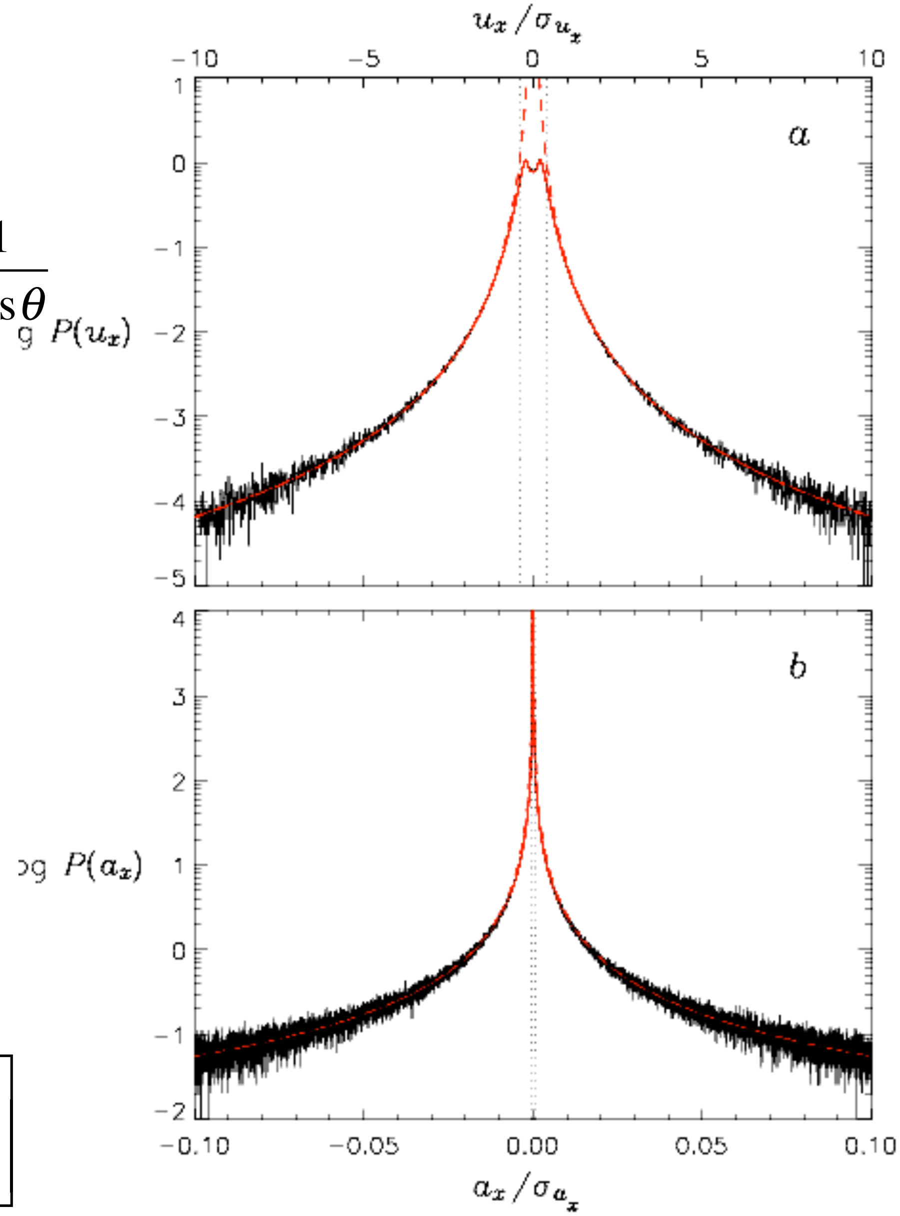
- nearest neighbor distance r
- n is vortex field number density

$$P(u_x) = \frac{n\pi}{u_x^3} e^{-\frac{\pi n}{2u_x^2}} \left[I_0\left(\frac{\pi n}{2u_x^2}\right) - I_1\left(\frac{\pi n}{2u_x^2}\right) \right]$$

modified Bessel functions of integer order

$$P(u_x) \propto \frac{1}{u_x^3}\left[u_0 + \frac{u_1 n}{u_x^2} + \frac{u_2 n^2}{u_x^4} + \cdots \right]$$

$$P(a_x) \propto \frac{\delta t^{2/3}}{a_x^{5/3}}\left[a_0 + a_1 n \left(\frac{\delta t}{a_x}\right)^{2/3} + a_2 n^2 \left(\frac{\delta t}{a_x}\right)^{4/3} + \cdots \right]$$



Velocity in field of randomly placed point – vortices of random amplitudes:

$$u_x = \frac{U_0}{r}\sin\theta$$

U_0 not constant

$$P(r) = 2\pi n r e^{-\pi n r^2}$$

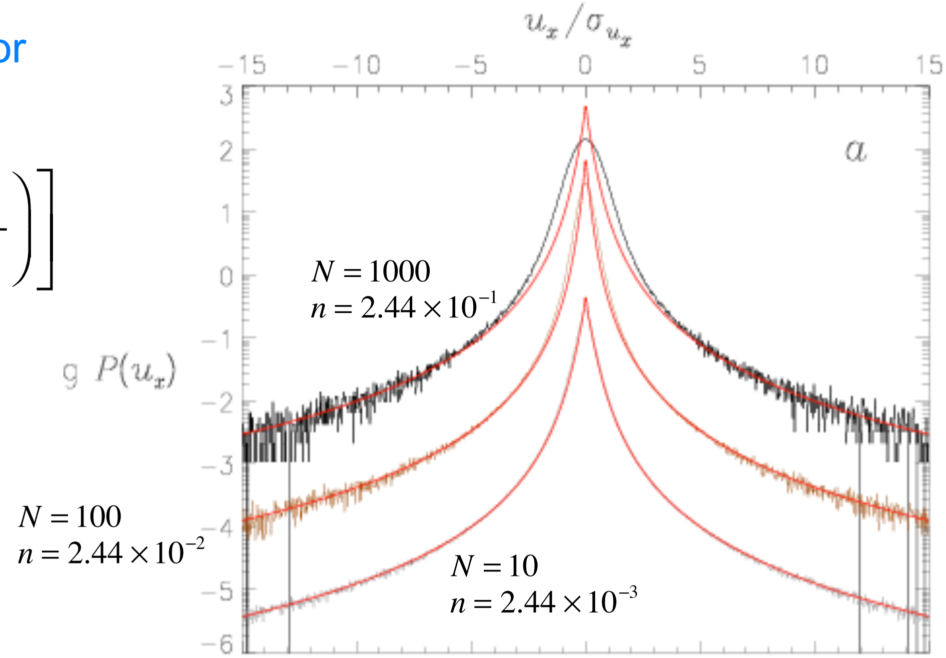
nearest neighbor as before

$$P\left(\frac{\sin\theta}{r}\right) = P(u) = \frac{n\pi}{u^3} e^{-\frac{\pi n}{2u^2}} \left[I_0\left(\frac{\pi n}{2u^2}\right) - I_1\left(\frac{\pi n}{2u^2}\right) \right]$$

$$P(U_0) = \frac{1}{\sigma\sqrt{2\pi}} e^{-U_0^2/2\sigma^2}$$

Gaussianly distributed amplitudes

$\Downarrow$



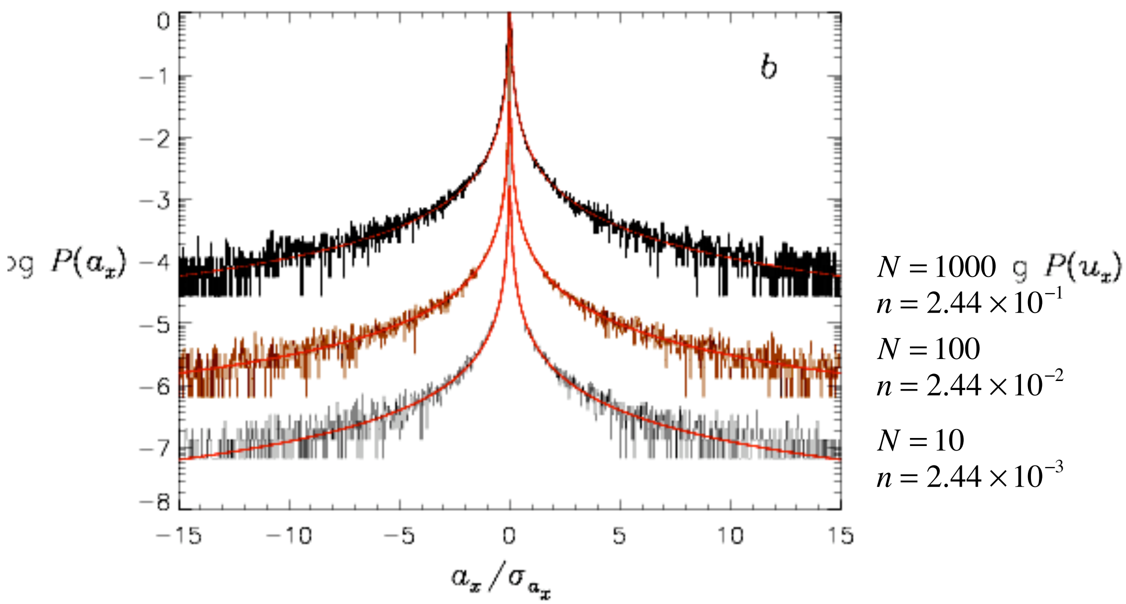
$$P(u_x) = \frac{2}{\pi^3} \frac{1}{\sqrt{u_x^2 + 2\pi n\sigma^2}} \left[K\left(\frac{2\pi n\sigma^2}{u_x^2 + 2\pi n\sigma^2}\right) - E\left(\frac{2\pi n\sigma^2}{u_x^2 + 2\pi n\sigma^2}\right) \right]$$

where K and E are the complete elliptic integrals of the first and second kind

LASP APS

Two important physical contributions to the velocity difference:

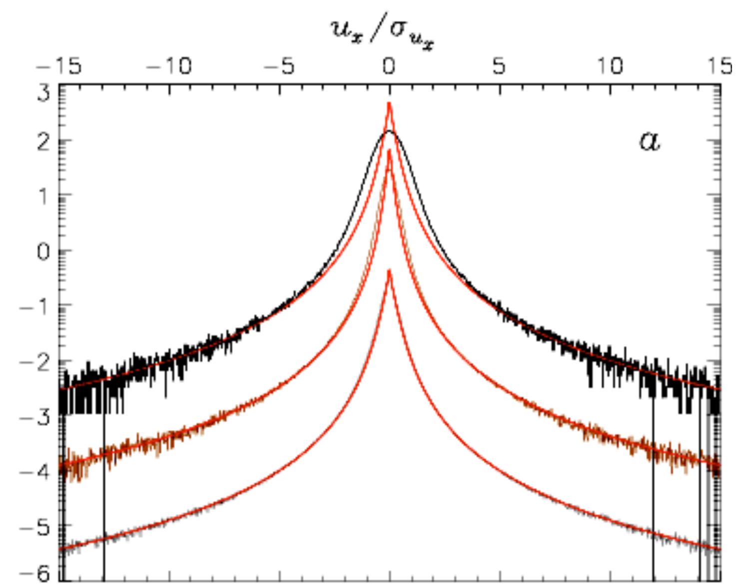
Advection over temporal increment τ by nearest neighbor:



$$P(a_x) \propto \frac{\delta t^{2/3}}{a_x^{5/3}} \left[a_0 + a_1 n \left(\frac{\delta t}{a_x} \right)^{2/3} + a_2 n^2 \left(\frac{\delta t}{a_x} \right)^{4/3} + \cdots \right]$$

Velocity difference in field of randomly placed point – vortices

Creation of new vortices in domain:

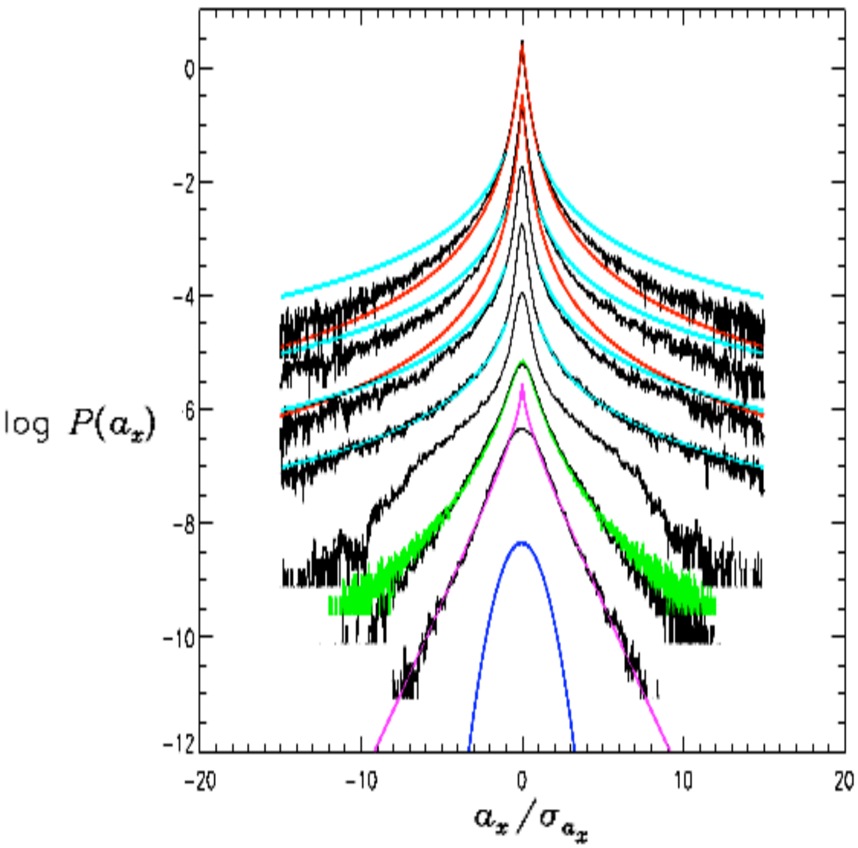


$$P(u_x) = \frac{2}{\pi^3} \frac{1}{\sqrt{u_x^2 + 2\pi n \sigma^2}} \left[K\left(\frac{2\pi n \sigma^2}{u_x^2 + 2\pi n \sigma^2} \right) - E\left(\frac{2\pi n \sigma^2}{u_x^2 + 2\pi n \sigma^2} \right) \right]$$

Velocity in field of randomly placed point – vortices of random amplitudes

LASP

APS



$$P(u_x) = \frac{2}{\pi^3} \frac{1}{\sqrt{u_x^2 + 2\pi n \sigma^2}} \left[K\left(\frac{2\pi n \sigma^2}{u_x^2 + 2\pi n \sigma^2} \right) - E\left(\frac{2\pi n \sigma^2}{u_x^2 + 2\pi n \sigma^2} \right) \right]$$

with NO fitting parameters!

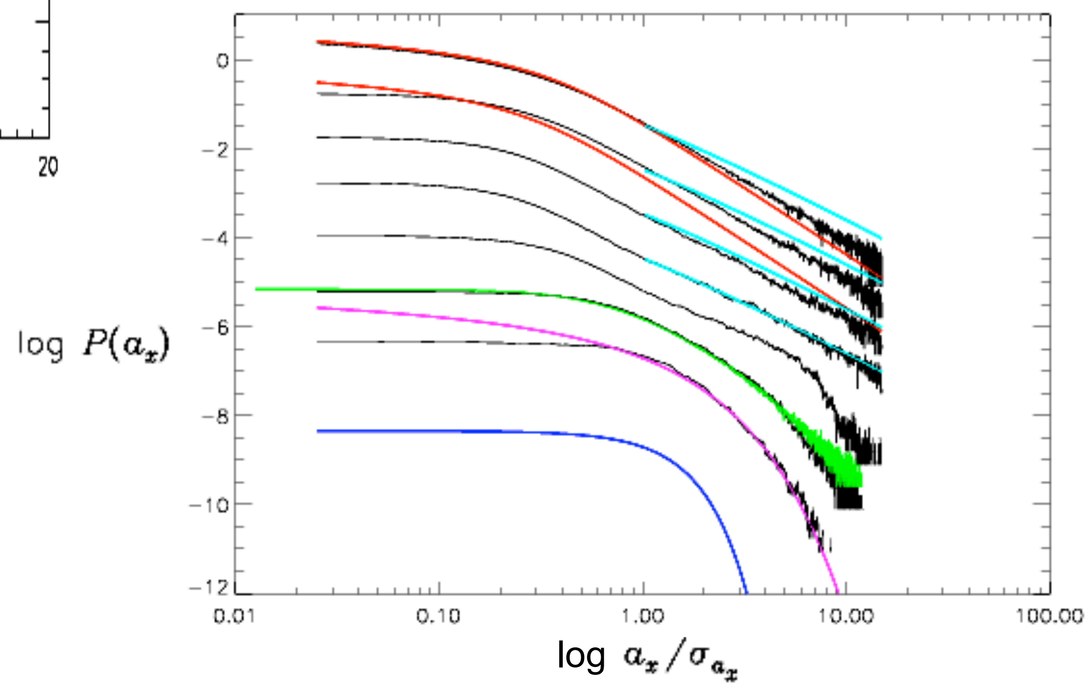
$$P(a_x) \propto \frac{\delta t^{2/3}}{a_x^{5/3}} \left[a_0 + a_1 n \left(\frac{\delta t}{a_x} \right)^{2/3} + a_2 n^2 \left(\frac{\delta t}{a_x} \right)^{4/3} + \cdots \right]$$

$$P(a_x) = \frac{1}{\pi \sigma_1 \sigma_2} K_0 \left(\frac{\sqrt{a_x^2}}{\sigma_1 \sigma_2} \right)$$

$$u(t_0) = u_{nn} + N_1$$

$$u(t_0 + \tau) = u_{nn} + N_1(1 + N_2)$$

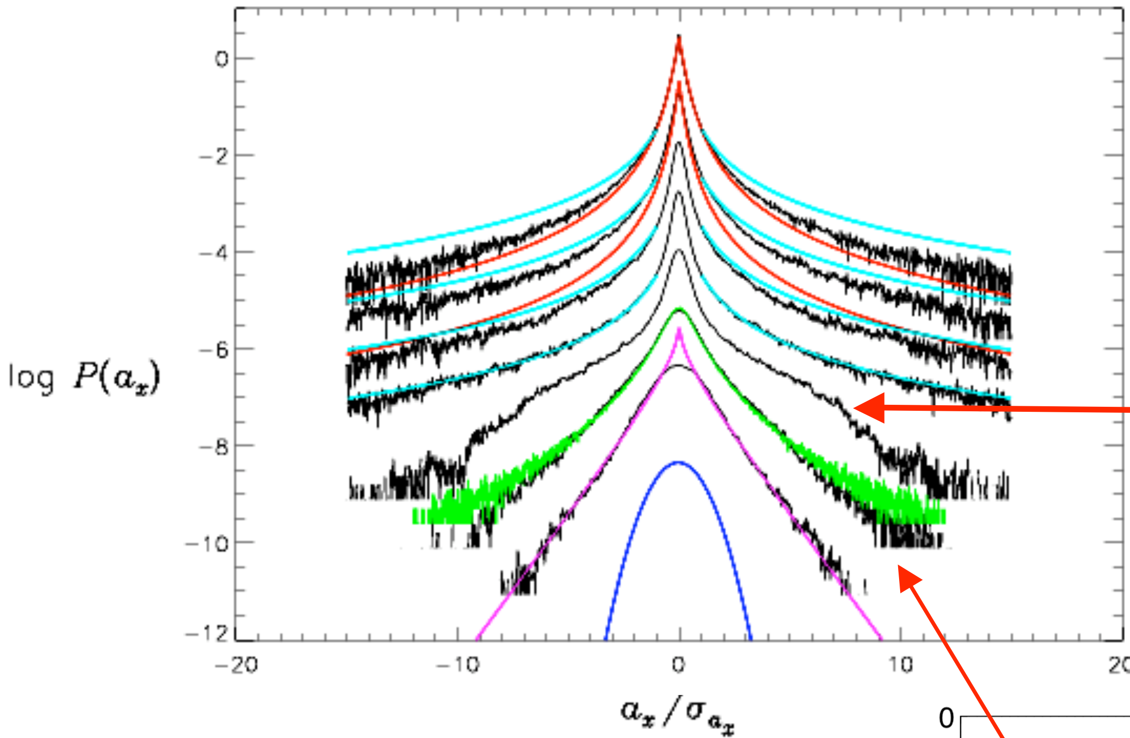
$$u(t_0 + \tau) - u(t_0) = N_1 N_2$$



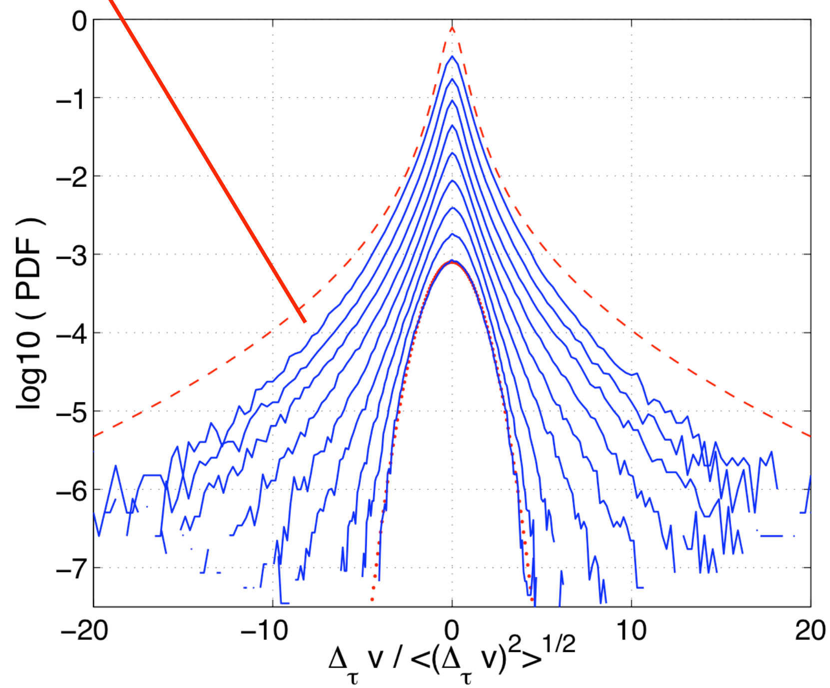
and finally Gaussian (uncorrelated)

LASP

APS



Recently observed in lab.



Implications and questions:

- The processes that dominate Lagrangian turbulent transport are dominated by nearest neighbor effects and are thus two-dimensional in the plane perpendicular to the closest vortex filament

- As the temporal increment $\tau \rightarrow 0$ the velocity difference probability density function approaches the new vortex nearest neighbor velocity pdf, because changes in the flow field resulting from new vortex creation overwhelm contributions from advection by existing filaments
  NEW vorticity changes do not have to be big (pdf normalized by rms)

- Lagrangian tracers randomly sample a random collection of vortices

- Random stirring mimics effects of vortex stretching

- As $\tau \rightarrow 0$, are velocity difference pdfs in driven turbulence significantly different from those in decaying turbulence?

- As $\tau \rightarrow 0$, are Eulerian and Lagrangian statistics different? If so, why? If not, why are we working so hard to measure Lagrangian motion at small τ?

LASP

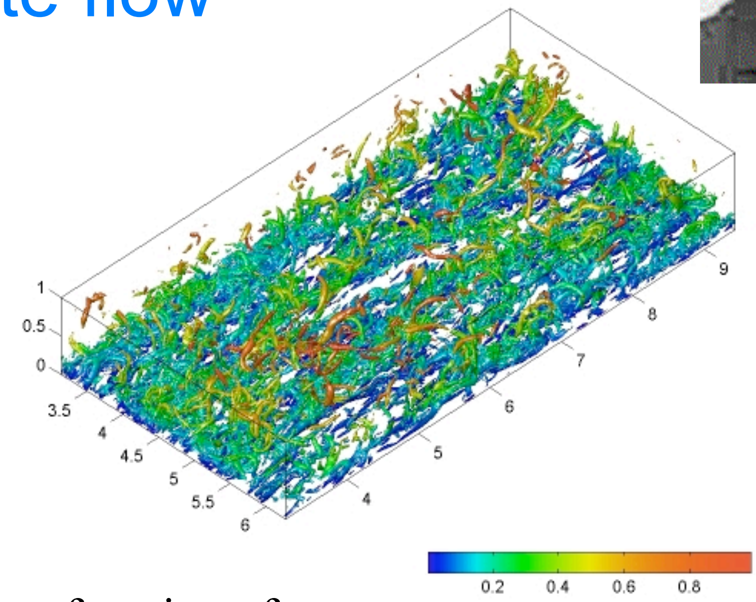
APS

Information rich data decimation based on turbulent structures:

e.g. petascale Plane Couette flow simulation:

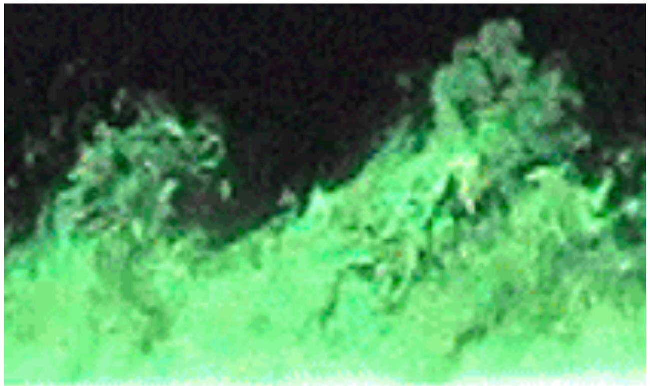
- large domain
- statistically steady state
- Output:
    - position
    - amplitude
    - orientation

  Of vortex filaments in domain



Assemble distributions of these quantities as function of imposed velocity and distance from boundary

Compute scalar and vector transport based on these distributions – develop a statistical mechanics of vortex structures on which to base transport coefficients



LASP

- Tera-scale computing offers sufficient data for robust point-wise statistics of the flow – and we can just handle the data volumes necessary to extract those

- Peta-scale computing will offer sufficient resources to develop a statistics of structures and a transport theory based on the statistical mechanics of these