# Simulating Turbulence: Resolution, Extreme Events and Towards Extreme Computing

P.K. Yeung

Georgia Inst of Tech, USA
*E-mail: pk.yeung@ae.gatech.edu*

*Turbulence from Angstroms to Light Years*

*Discussion Meeting in honor of Prof. Katepalli R. Sreenivasan*

International Center for Theoretical Sciences

Bangalore, INDIA, January 2018

**Sreeni as a visionary, and catalyst**

# KRS70@Denver, Nov 17-18, 2017

# Simulation Database: a brief overview

"Deep Science" philosophy, to advance fundamental understanding

- Intermittency, inertial range and dissipative range
- Turbulent mixing, Schmidt numbers $O(0.001)$ to $O(1000)$
- Turbulent dispersion, backward tracking (w/ B.L. Sawford)

Forced isotropic turbulence, $R_\lambda$ up to 1300; various resolutions

- Largest production run at $8192^3$, using 262,144 parallel processes
- Some shorter (but quite arduous) runs at $12288^3$ (D. Buaria, poster at this conference) and $16384^3$ (X.M. Zhai, M.P. Clay)
- Some 3 Petabytes stored at supercomputer centers

First paper with Sreeni published in 2002. His tremendous and benevolent influence has endured since mid 1990's. Visited over 10 times.

# Outline of This Talk

- Assessing the accuracy of large simulations
- Resolution in space and time for "extreme events"
- Towards extreme-scale computing: GPU algorithms

# As ever-larger simulations become possible...

Within the realm of 3D homogeneous turbulence:

- What do we want to use the computing power for?
  (higher Reynolds no., Mach no., other physical processes..)
- How do we recognize the limitations?
  (deterministic as well as statistical sources of errors...)

Can a simulation at higher resolution resolve any doubts?

- Higher moments, extreme fluctuations inherently more sensitive
- An expensive proposition, subject to statistical sampling
- Resolution in space (especially critical for work on intermittency)
- Resolution in time (seems to be discussed less in literature)

# Spatial Resolution in Pseudo-Spectral DNS

$$k_{max} = \sqrt{2}N/3; \qquad \Delta x/\eta \approx 2.96/(k_{max}\eta)$$

- $k_{max}\eta \approx 1.5$ may be fine for low-order velocity statistics
- But better resolution needed for small-scale statistics, increasingly so at higher Reynolds numbers (Yakhot & Sreenivasan 2005).
- Refine the grid, run again at same *Re*, and compare (e.g. acceleration statistics, Yeung *et al.* PoF 2006)
- For a given snapshot, what features may be subject to errors?
  - Take best-resolved velocity field, truncate at some wavenumber $k_c$. Compute various statistics again, compare, and repeat
  - Large differences would indicate insufficient accuracy
  - A post-processing task, not a large new simulation
  - No information on global error after N-S time evolution

# Temporal Resolution

- Courant number constraint for numerical stability:

$$C = \left[\frac{|u|\Delta t}{\Delta x} + \frac{|v|\Delta t}{\Delta y} + \frac{|w|\Delta t}{\Delta z}\right]_{max} \sim \alpha \frac{u'\Delta t}{\Delta x}$$

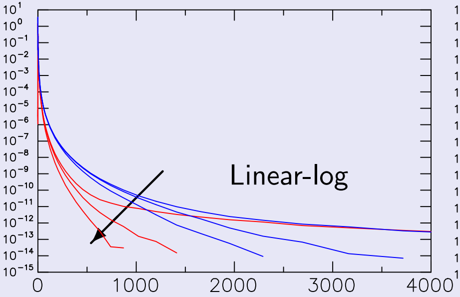
- Classical scaling and our experience in the simulations (take $\alpha = 12$)

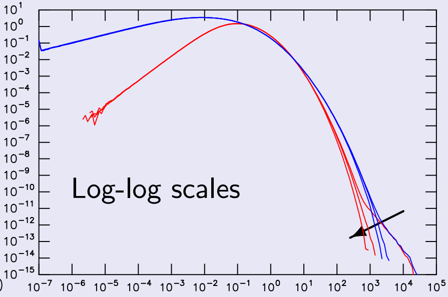
$$\Delta t/\tau_\eta \approx (C/12)(15)^{1/4}(\Delta x/\eta)\ {R_\lambda}^{-1/2}$$

- In large simulations, $\Delta t/\tau_\eta$ frequently well under 1%
- But small scales evolve fast (at fixed position, swept by large scale motions of speed $\sim u'$, with a time scale of order $\eta/u'$)
- Temporal intermittency is stronger than spatial intermittency, as comparisons between Eulerian and Lagrangian statistics suggest
- Short tests (approx 10 $\tau_\eta$, say) at $C = 0.6$, 0.3, 0.15 can help.

# Dissipation and Enstrophy: Spatial Filtering

- Enstrophy is more intermittent. Is it more sensitive to resolution?
- PDFs of $\epsilon/\langle\epsilon\rangle$ and $\Omega/\langle\Omega\rangle$, derived from velocity fields at $R_\lambda$ 650, $k_{max}\eta = 2.8$ ($4096^3$); apply cutoff at $k_c/k_{max} = 1, 0.75, 0.5$



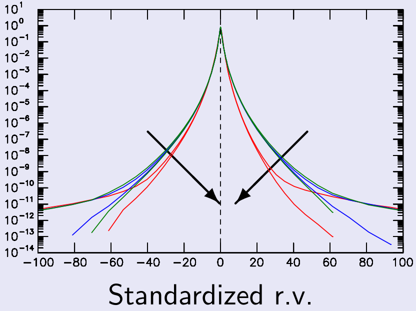




Convergence between PDFs of $\epsilon$ and $\Omega$ at far (power-law like) tails appears to be due to high-$k$ modes, contaminated by aliasing errors

# Why is dissipation more sensitive to resolution?

$$\epsilon \equiv 2\nu s_{ij} s_{ij} \ ; \quad \Omega \equiv \omega_i \omega_i$$

- For incompressible isotropic turbulence: $\nabla_{\parallel}\mathbf{u}$ is negatively skewed and contributes only to $\epsilon$; while $\nabla_{\perp}\mathbf{u}$ is more intermittent




- Narrower PDF tails when some high $k$ modes removed by filtering.
- $\nabla_{\parallel}\mathbf{u}$ more sensitive to resolution than $\nabla_{\perp}$ or $\nabla \times \mathbf{u}$

# Incompressibility and 1-D Spectra

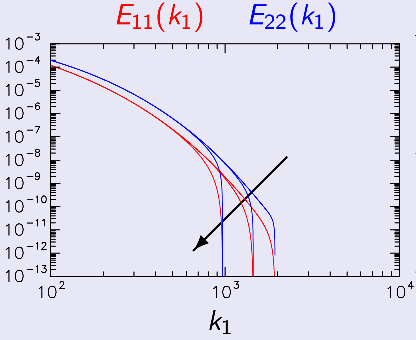
Incompressibility: $\hat{\mathbf{u}}(\mathbf{k}) \perp \mathbf{k}$ in wavenumber space

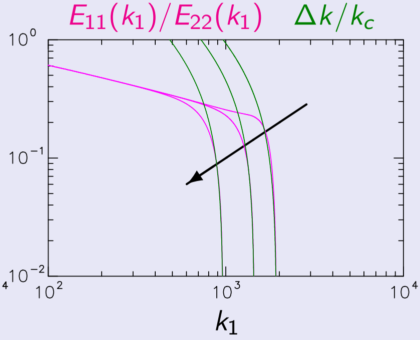
- At high $k \equiv |\mathbf{k}|$, spectral content of $\hat{u}_1$ is mostly in Fourier modes with large $k_2$ and/or $k_3$ (i.e., $k_\perp$)

Relations for 1D spectra with $k_1$ near cutoff $k_c$ ($k \approx k_1$)

$$\begin{aligned} E_{11}(k_1) &= \int_{k_1}^{\infty} \frac{E(k)}{k}(1 - k_1^2/k^2)\, dk \\ &\approx \frac{E(k_1)}{k_1} \int_{k_1}^{k_c} (1 - k_1^2/k^2)\, dk \approx E(k_1)(\Delta k/k_1)^2 \\ E_{22}(k_1) &= \frac{1}{2} \int_{k_1}^{\infty} \frac{E(k)}{k}(1 + k_1^2/k^2)\, dk \approx E(k_1)(\Delta k/k_1) \end{aligned}$$

- A spherical cutoff filter based on $k$ removes more of high wavenumber content in longitudinal spectra than transverse (hence affects $\nabla_{\parallel}\mathbf{u}$ more than $\nabla_{\perp}\mathbf{u}$)
- For small $\Delta k = k_c - k_1$: $E_{11}(k_1)/E_{22}(k_1) \approx \Delta k/k_c$
- $R_\lambda \sim 650$, $k_{max}\eta \approx 2.8$, $4096^3$, $k_c/k_{max} = 0.75, 0.5$:



# Outline of This Talk

- Assessing the accuracy of large simulations
- Resolution in space and time for "extreme events"
- Towards extreme-scale computing: GPU algorithms

# Tests of Spatial and Temporal Resolution

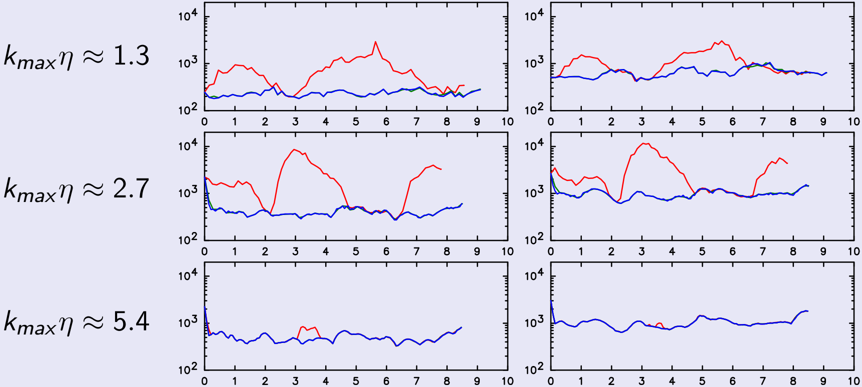
A wide parameter space: easier if not at the highest Reynolds number:

| | | $R_\lambda \sim 390$ | | |
|---|---|---|---|---|
| $k_{max}\eta$ | $N$ | $C = 0.6$ | $C = 0.3$ | $C = 0.15$ |
| 1.33 | 1024 | 1.31 | 0.699 | 0.349 |
| 2.67 | 2048 | 0.609 | 0.333 | 0.165 |
| 5.38 | 4096 | 0.327 | 0.164 | 0.082 |
| | | $R_\lambda \sim 650$ | | |
| 1.33 | 2048 | 0.646 | 0.484 | 0.238 |
| 2.67 | 4096 | 0.411 | 0.236 | 0.116 |
| 5.38 | 8192 | 0.205 | 0.102 | 0.062 |

Resolution parameters and percentage $\Delta t/\tau_\eta$ [Note: RK2 in time] for simulation datasets (roughly 10 $\tau_\eta$) at Reynolds numbers corresponding (within statistical error) to past simulations at $R_\lambda$ 390 and 650.

# Resolution in time: Peak dissipation and enstrophy

$R_\lambda \sim 390$, $C = 0.6, 0.3, 0.15$; $\epsilon/\langle\epsilon\rangle$ and $\Omega/\langle\Omega\rangle$ vs. $t/\tau_\eta$




- $C = 0.6$ gives spuriously large peaks; 0.3 and 0.15 almost the same
- Sensitivity greater for $\epsilon/\langle\epsilon\rangle$, weaker at higher $k_{max}\eta$ (such as 5.4)

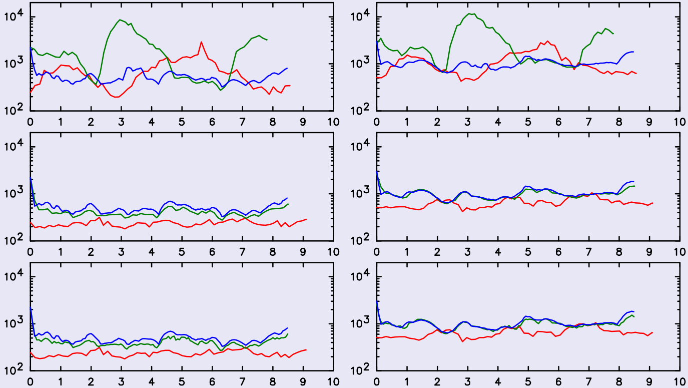
# Resolution in space: Peak dissipation and enstrophy

$R_\lambda \sim 390$, $k_{max}\eta = 1.3, 2.7, 5.4$; $\epsilon/\langle\epsilon\rangle$ and $\Omega/\langle\Omega\rangle$ vs. $t/\tau_\eta$

$C = 0.6$

$C = 0.3$

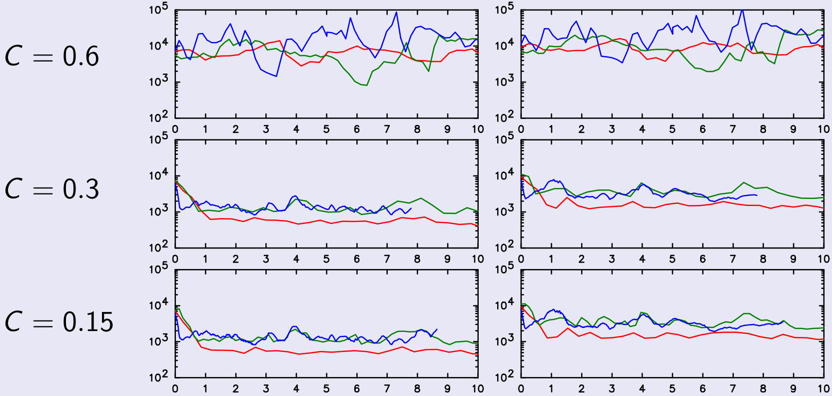
$C = 0.15$




- Impact of using higher $k_{max}\eta$ not clear if $C$ is too high
- Low $C$: higher $k_{max}\eta$ does allow larger gradients, $\epsilon$ and $\Omega$

# Would a lower Courant no. be needed at higher $R_\lambda$?

$R_\lambda \sim 650$, $k_{max}\eta =$ 1.3, 2.7, 5.4; $\epsilon/\langle\epsilon\rangle$ and $\Omega/\langle\Omega\rangle$ vs. $t/\tau_\eta$



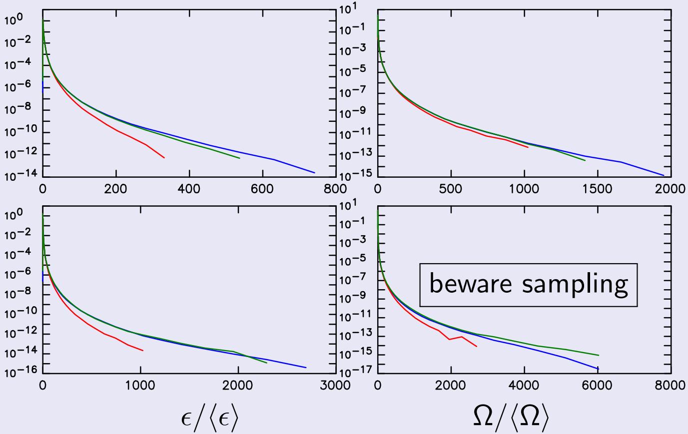

Not perfectly clear, but sensitivity to $C$ at $k_{max}\eta \sim 5.4$ seems stronger than at $R_\lambda \sim 390$, even as $\Delta t/\tau_\eta$ drops

# Best results available for dissipation and enstrophy PDFs

$C = 0.15$, $k_{max}\eta \approx$ 1.3 2.7 5.4: Convergence apparently achieved

$R_\lambda \sim 390$

$R_\lambda \sim 650$



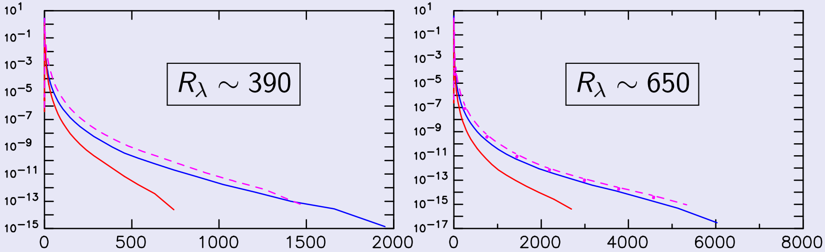

Tails stretch out further: as $R_\lambda$ increases, and $\Omega$ relative to $\epsilon$

# Re-visiting the issue of dissipation vs enstrophy PDFs

- Both are measures of the small scales: should they scale similarly at sufficiently high Reynolds no.? (Nelkin 1999)
- Most data sources, including DNS at finite Reynolds no., show that enstrophy is more intermittent
- Visualizations of spatial structure: vortex filaments in contrast to more diffuse, sheet-like dissipative structures
- Recent DNS at higher resolution suggest both $\epsilon$ and $\Omega$ attain amplitudes more extreme than previously reported, scaling similarly unless averaged over significant scale sizes (Yeung *et al.* 2015)
    - yet contaminated by the effect of time-stepping errors, which affect dissipation more than enstrophy
    - although $\Delta t \ll \tau_\eta$, need to have $\Delta t \ll \eta/u'$, to resolve rapid rate of small scales advected by the large scales (Tennekes JFM 1975)

# Compare dissipation and enstrophy PDFs again

Best data at $C = 0.15$ (RK2), $k_{max}\eta \approx 5.4$: tails do not coincide, but seem to have the same shapes (Magenta dotted line is PDF of $2\epsilon/\langle\epsilon\rangle$)




$\epsilon/\langle\epsilon\rangle$, $\Omega/\langle\Omega\rangle$ $\qquad$ $\epsilon/\langle\epsilon\rangle$, $\Omega/\langle\Omega\rangle$

Possible explanation from multi-fractal theory

- Assume $\langle\epsilon^n\rangle \propto {R_\lambda}^{f(n)}$, $\langle\Omega^n\rangle \propto {R_\lambda}^{g(n)}$, for $n \geq 1$. [with $f(1) = g(1)$]
- If $f(n)$ and $g(n)$ have the same functional *forms*: at a given $R_\lambda$, $\exists\ \beta$ s.t. $\beta\epsilon/\langle\epsilon\rangle$ and $\Omega/\langle\Omega\rangle$ have nearly the same PDF at large amplitudes

# Outline of This Talk

- Assessing the accuracy of large simulations
- Resolution in space and time for “extreme events”
- Towards extreme-scale computing: GPU algorithms

# How will we do larger simulations?

How do we do it now...

- Massive distributed-memory parallelism, with 2D domain decomposition (3D domain as "pencils")
- Usually, all-to-all communication for transposes of data is dominant
- On "Blue Waters" (Cray XE6): remote memory addressing via Co-Array Fortran, also helped by favorable machine topology

Changes in HPC landscape, towards Exascale (by 2021 in US)

- Many-core platforms with fewer but fatter nodes
- Heterogeneous computing: especially multi-threaded GPUs for fast arithmetic, but data movements are (even more) the key

GPUs: recent success for turbulent mixing at high Schmidt number using a combination of FFTs and compact finite differences (Gotoh *et al.* 2012)

# Turbulent mixing at high Schmidt number

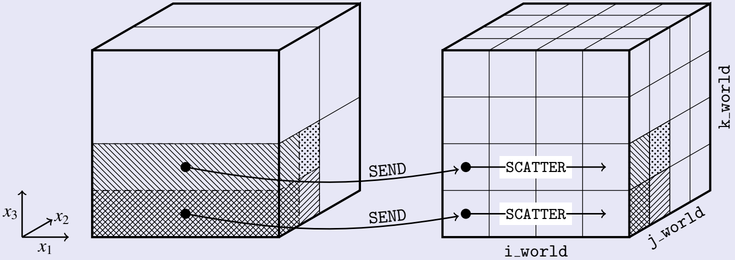
- $Sc = \nu/D$ varies: 7 for heat in water, 700 for salinity in ocean
- $Sc \gg 1$: smallest scale is Batchelor scale $\eta_B = \eta/\sqrt{Sc}$. Different scaling laws, and very difficult to resolve
- Disparate resolution requirements:
  - velocity on coarser grid, scalar on finer grid
  - scalar: use numerical method with lesser communication needs
  - some overlap of operations on velocity and scalar field?
  - CPU host to work on velocity, GPU for scalar
- Dual grid, dual communicator, dual scheme algorithm
  - CPU-based: Clay *et al.* 2017 (*Comput. Phys. Commun., published*)
  - GPU-based: Clay *et al.* 2018 (*Comput. Phys. Commun., in revision*)

  GPU: Performed well on 27-Petaflop TITAN (Cray XK7) at Oak Ridge

# Dual-Communicator Algorithm

Clay *et al.* *Comput. Phys. Commun.* 2017

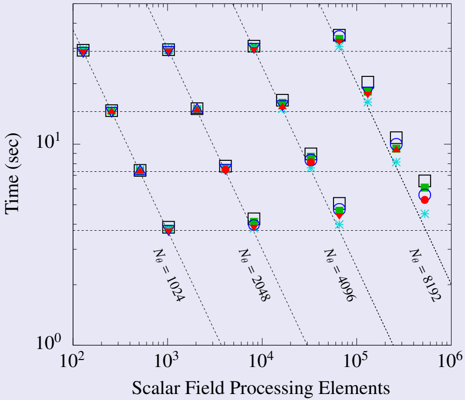
- Say $N_\theta / N_u = 8$: workload on finer grid is much heavier
- Define 2 groups of MPI processes, proportional to $N_u^3$ and $N_\theta^3$
- One-way intercommunicator transfer: need **u** on scalar side




- Overlap velocity field computations with scalar
- Overlap between operations on different velocity components

# CPUs: DNS Code Performance Comparisons

- Problem sizes $1024^3$ to $8192^3$, different implementations
  - black squares: blocking and single-threaded
  - cyan stars: overlapping, using OpenMP nested parallelism




- Also very high scalability
- About 65% of time is spent on CCD3D
- Interpolation approx 25%
- Performance gain less than for CCD3D, due to reduced overlapping

# GPU Programming Considerations

[Mainly based on experiences on TITAN (Cray XK7, 1 GPU/Node)]

- Directives-based, instead of CUDA, use OpenACC or (better yet) OpenMP 4.5 standard (facilitates asynchronous execution)
- FFTs using CUDA mathematical libraries extremely fast
- Memory of GPU only 1/5 of host: needs careful management
- Minimize data movement: almost all scalar field operations on GPUs
- Keep GPU(s) as busy as possible: e.g. by using 2 MPI processes on 16-core TITAN node sharing 1 GPU
- NVIDIA visual profiler can provide detailed diagnostics
- Currently, all communication go through CPUs, thus extra copying between CPU and GPU. Hope to avoid this on future machines

# GPU: Weak Scaling and GPU-to-CPU Speedup

| $N_\theta^3$ | $512^3$ | $1024^3$ | $2048^3$ | $4096^3$ | $8192^3$ |
|---|---|---|---|---|---|
| Nodes | 2 | 16 | 128 | 1024 | 8192 |
| CPU (s) | 14.80 | 14.94 | 15.33 | 15.71 | 16.13 |
| Weak (%) | — | 99.1 | 96.5 | 94.2 | 91.8 |
| Async. Off | 2.943 | 3.042 | 3.155 | 3.441 | 3.686 |
| Weak (%) | — | 96.7 | 93.3 | 85.5 | 79.8 |
| Speedup | 5.03 | 4.91 | 4.86 | 4.57 | 4.38 |
| Async. On | 2.929 | 2.900 | 2.938 | 3.024 | 3.259 |
| Weak (%) | — | 101 | 99.7 | 96.7 | 89.9 |
| Speedup | 5.05 | 5.15 | 5.22 | 5.20 | 4.95 |

# Conclusions and Next Steps: Science Aspects

Resolution effects and numerical challenges in time and space stronger than anticipated earlier (at RK2, needs Courant no. $C = 0.3$ or $0.15$)

- Although less intermittent than enstrophy, dissipation more sensitive to the numerics (role of incompressibility, vortex filaments, etc)
- Some past statements about extreme events in need of revision
- To examine/quantify resolution effects on various other statistics (perhaps inertial range statistics only weakly affected)

Successful development and execution of highly optimized algorithm for mixing at high $Sc$ on a 27-PF GPU machine

- Minimizing data movement and aggressive overlapping using OpenMP 4.5 on GPUs: factor of 5 speedup over CPU only
- To fully analyze passive scalar with mean gradient at $R_\lambda \sim 140$, $Sc = 512$; and extend to active scalars (stratified flow)

# Conclusions and Next Steps: Computing Aspects

Viability of $16384^3$ or higher on a production basis

- Adopt OpenMP 4.5 ideas in recent work on mixing at high *Sc*, towards new pseudo-spectral algorithm for velocity field
- Port codes to next big machine (150 PF SUMMIT at ORNL)
- Extension from 1 scalar to 2; passive to active

Obtaining, using, preserving resources

- Make science case in competitive proposals
- Large machines: problems are to be expected
- Where to put data and even hold on to them long term

More community and international cooperation would be very useful

# Acknowledgments (in addition to Sreeni)


Research Support

- U.S National Science Foundation (Fluid Dynamics & Cyber)

Supercomputing Resource Providers

- Blue Waters Project (University of Illinois, Urbana-Champaign)
- Texas Advanced Computation Center (XSEDE Consortium)
- Oak Ridge Leadership Computing Facility (US Dept of Energy)

Collaborators and Friends

- Toshiyuki Gotoh, Nagoya Inst. of Tech., Japan
- Stephen B. Pope, Cornell Univ., USA
- Alain Pumir, ENS-Lyon, France
- Dhawal Buaria (PhD 2016), Max Planck Inst, Germany
- Matthew P. Clay (PhD 2017), Air Force Research Lab, USA

Current PhD Students

- Shine X.M. Zhai, Kiran Ravikumar

# Some Publications


- Buaria, D., PKY, & Sawford, B.L. (2016) A Lagrangian study of turbulent mixing: forward and backward dispersion of molecular trajectories in isotropic turbulence. *JFM* **799**, 352-382.
- Clay, M.P., Buaria, D., Gotoh, T. & PKY (2017) A dual communicator dual grid-resolution algorithm for Petascale simulations of turbulent mixing at high Schmidt number. *Comput. Phys. Commun.*, **219**, 313-328.
- Donzis, D.A., Sreeni & PKY (2010) The Batchelor spectrum for mixing of passive scalars in isotropic turbulence. *Flow, Turb. & Combust.*, **85**, 549-566.
- Donzis, D.A., PKY & Sreeni (2008). Energy dissipation rate and enstrophy in isotropic turbulence: resolution effects and scaling in direct numerical simulations. *PoF*, **20**, 045108.
- PKY, Donzis, D.A. & Sreeni (2012). Dissipation, enstrophy and pressure statistics in turbulence simulations at high Reynolds numbers. *JFM*, **700**, 5-15.
- PKY, Zhai, X.M & Sreeni (2015). Extreme events in computational turbulence. *PNAS*, **112**, 12633-12638.