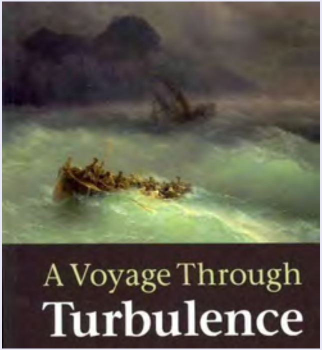
Computing Petascale Turbulence on Blue Waters: Advances Achieved and Lessons Learned

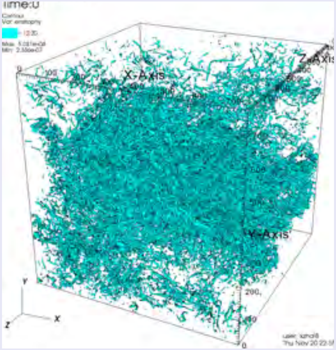
P.K. Yeung (PI)
Schools of AE and ME, Georgia Tech
E-mail: pk.yeung@ae.gatech.edu

NSF:	PRAC (0832634, 1036170, 1640771) and Fluid Dynamics Programs
BW Team, Cray:	Scaling, Reservations, Help Requests, Storage, Visualization
Collaborators:	T. Gotoh, S.B. Pope, B.L. Sawford, K.R. Sreenivasan
PhD Students:	K.P. Iyer (2014), D. Buaria (2016), M.P. Clay (2017), X.M. Zhai (2019); K. Ravikmar (current)
Postdocs:	K.P. Iyer (w/ KRS at NYU, 2017 –)

Blue Waters Symposium, June 3-6, 2019

Altogether, one decade of Blue Waters





A rewarding ride, nervy at times, but many thanks to BW staff:

- First PRAC grant from NSF in 2009; Access to machine since 2012
- High-resolution simulations allowed us to address difficult questions
- Learned some lessons, but perhaps that is how science is done (?)

Turbulence and High-Performance Computing

Disorderly fluctuations over a wide range of scales

- Pervasive in many branches of science and engineering
- Reynolds number: a measure of the range of scales
- Numerical simulation often best source for detailed information

A Grand Challenge problem in computing

- Flow is 3D: domain decomposition, and communication-intensive
- Every step-up in problem size: 8X in number of grid points

Some notable references in the field:

- Kaneda *et al.* PoF 2003: 4096^3, on Earth Simulator
- Yeung, Zhai & Sreenivasan PNAS 2015: 8192^3, on Blue Waters
- Ishihara *et al.* PRF 2016: 12288^3, on K Computer

What Blue Waters Has Enabled (Not Over Yet!)

Forced isotropic turbulence, R_λ up to 1300; various resolutions

- Largest production run at 8192^3, using 262,144 parallel processes
- Some shorter (yet arduous) runs at 12288^3 and 16384^3 (4 trillion)
- Hundreds of millions of core hours, 2.5 PB Nearline storage

Topics and Publications (to date):

- Extreme events (Y, Zhai & Sreenivasan PNAS 2015)
- Velocity increments and similarity (Iyer, S & Y, PRE 2015, 2017)
- Nested OpenMP for low-diffusivity mixing (Clay, *et al.* CPC 2017)
- Highly scalable particle tracking (Buaria & Y, CPC 2017)
- Resolution and extreme events (Y, S & Pope, PRF 2018)
- A few more since after BW Symposium of 2018

Turbulence and Pseudo-Spectral Methods

- 3D Navier-Stokes eqs. (conservation of mass and momentum)

$$\partial \mathbf{u}/\partial t + (\mathbf{u}\cdot\nabla)\mathbf{u} = -\nabla(p/\rho) + \nu\nabla^2\mathbf{u} + \mathbf{f} \tag{1}$$

- Periodic domain: $\mathbf{u}(\mathbf{x},t) = \sum_{\mathbf{k}} \hat{\mathbf{u}}(\mathbf{k},t)\exp(\iota\mathbf{k}\cdot\mathbf{x})$ in a discrete Fourier representation. In wavenumber space, $\hat{\mathbf{u}} \perp \mathbf{k}$ and evolves by

$$\partial\hat{\mathbf{u}}/\partial t = -\widehat{\nabla\cdot(\mathbf{uu})}_{\perp\mathbf{k}} - \nu k^2\hat{\mathbf{u}} + \hat{\mathbf{f}} \tag{2}$$

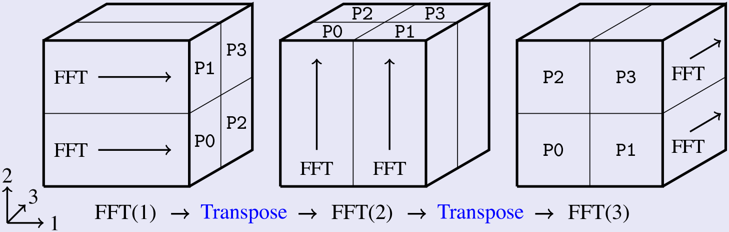
- Pseudo-spectral: nonlinear terms formed in physical space, transformed back and forth in $O(N^3 \ln_2 N)$ operations on N^3 grid (avoiding convolution integral, whose cost would be $\propto N^6$)

- 3D FFT: wide relevance spanning many domain science specialties

- Parallel computing: first decision is how to divide up the domain.

Massive (Distributed) Parallelism for 3D FFTs

- 2D domain decomposition allows up to N^2 MPI processes
- row and column communicators: $P_r \times P_c$ 2D processor grid



- FFTs taken 1 direction at a time (complete lines of data needed)
- Transpose (re-distribution of data) via all-to-all communication
- Local packing and unpacking needed for non-contiguous messages

Communication-intensive nature is main barrier to scalability, especially at large core counts

Communication and Contention

How to make the code communicate more efficiently?

- Reduce communication overhead via fewer MPI processes. (May not necessarily lead to reduction in overall wall time.)
- Non-blocking all-to-all, overlap w/ OpenMP computation (May not be effective if communication-to-computation ratio is high)
- Remote memory addressing (Fortran Co-Arrays, Cray Compiler)
  - declare major buffers as co-arrays, accessible to other processes
  - one-sided "get" operation for pairwise exchange
  - copy of data between regular and co-arrays

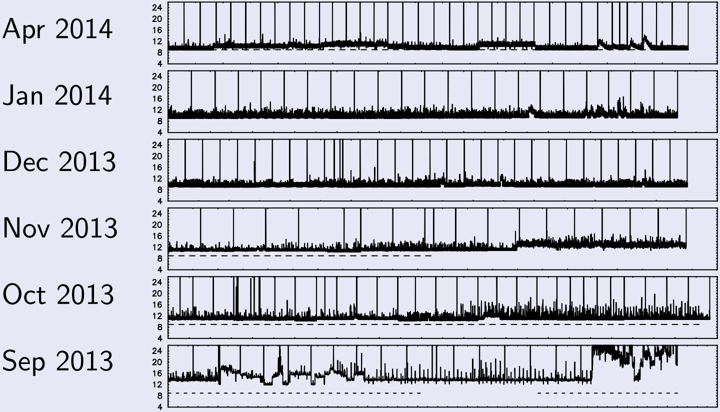
  (Thanks to R.A. Fiedler for co-array all-to-all implementation)

Performance degradation due to contention with other jobs

- Best performance was obtained when running on a reserved partition designed to minimize contention from network traffic
- Likewise, much helped by Topologically Aware Scheduling (TAS)

Impact of Network Topology / Reservation

- 262144 MPI tasks, Fortran co-arrays, single-prec, RK2



- Best timing was 8.897 secs/step; with other traffic minimized
- I/O on Blue Waters is good: 40 secs to write 8192^3 checkpoint

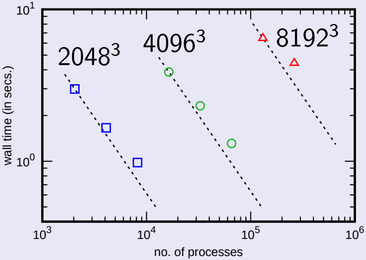
Particle tracking

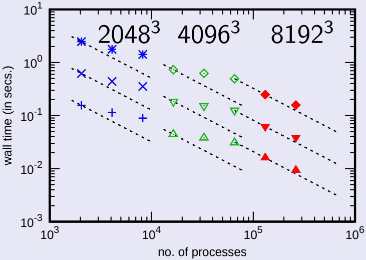
- Study of turbulent dispersion (pollutants, soot, bioterrorism, etc)
- Fluid particles (w/o inertia, diffusion): $\mathbf{u}^+(t) = \mathbf{u}(\mathbf{x}^+, t)$
  — interpolate for particle velocity based on instantaneous position
- Cubic spline interpolation (Yeung & Pope, JCP 1988): $(N+3)^3$ spline coefficients computed in manner analogous to 3D FFT, also distributed among the MPI processes.
- Hundreds of millions of fluid particles (Buaria & Yeung, CPC 2017):
  - A given MPI task always tracking the same particles, or
  - Dynamic mapping between MPI tasks and particles determined by instantaneous positions, minimizing communication cost
  - Communication of spline coefficients for particles close to sub-domain boundaries implemented efficiently using Fortran Co-Arrays

Scalability of new particle tracking algorithm

Time to compute $(N+3)^3$ spline coeffs. from velocity field on N^3 grid

Time to interpolate for velocity of $N_p = 16\mathrm{M}$, 64M and 256M particles





- Splines scale like 3D FFTs, despite some load imbalance due to $N+3$
- Interpolation time actually scales *better* at larger N
  - computation scales as N_p/P (particles evenly distributed in domain)
  - communication depends on no. of particles located within 2 grid spacings of a sub-domain boundary. For 8192^3 with 32×8192 domain decomposition this also scales as N_p/P

Multi-particle clusters and post-processing

Some physical questions (beyond the simplest):

- How is a particle trajectory affected by local flow conditions in space?
- Relative dispersion: How quickly can a pair of particles move apart?
- Mixing: How quickly can a pair of particles come together?
- Shape distortion: What happens to a collection of 3 or 4 particles as they move? Is there a preferred shape, even if size keeps growing?

"Backward tracking" via post-processing

- N-S equations are irreversible in time. To learn about past history, need to have stored a lot of data at earlier times
- N_p particles, and $O(N_p^2)$ possible pairs: trace back their trajectories, mostly on pairs close together at "final time" of simulation
- Four-particle tetrads: careful, selective sampling even more important: cannot deal with N_p^4 when N_p is many millions!

The Study of Extreme Events

Local deformation of a fluid element involves changes in shape and orientiation, due to intense velocity gradients

- Fluctuations of dissipation rate (strain rates squared) also pivotal to intermittency in turbulence theory
- Extreme events: samples of $> O(10^3)$ times mean value seen in DNS. But sensitive to resolution in both space and time (and statistics)

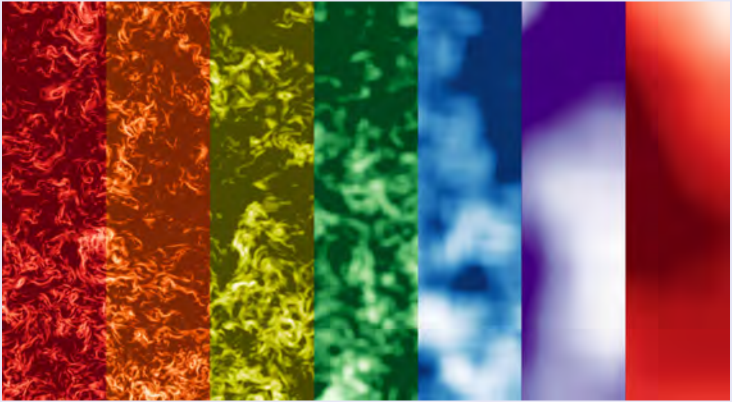
Local averages (in 3D) of dissipation rate

$$\epsilon_r(\mathbf{x}, t) = \frac{1}{Vol} \int_{Vol} \epsilon(\mathbf{x} + \mathbf{r}', t)\ d\mathbf{r}'$$

- Rarely reported in the past; 1D averages can be misleading
- Intermediate range of r is most important
  — and less sensitive to numerics

Local Averaging of a Highly Intermittent Signal

[K.P. Iyer *et al*, APS-DFD 2018, with help from R. Sisneros (NCSA)]



Locally averaged slices of dissipation at $r/\Delta x = 1, 2, 4, 8, ...$, taken from a single 16384^3 DNS snapshot. Left to Right: from wrinkled to smooth.

A Summary of our Blue Waters Experience

Advances in domain science (turbulence) using up to 8192 BW nodes

- First full-length 8192^3 DNS (and much shorter 16384^3), w/ attention to extreme events and spatial resolution
- Highest Reynolds number DNS for turbulent dispersion
- Dual-resolution simulations of high Schmidt number mixing

Algorithmic challenges faced and innovations achieved

- Fortran co-arrays for 256K MPI tasks alltoall (further helped by TAS)
- Ideas applied to massive particle tracking (CPC 2017)
- Nested OpenMP on Cray XE6; OMP 4.5 on XK7 (CPC 2017, 2018)

Data Management (on NCSA Nearline system)

- Learned lessons about handling of a large number of "small" files
- Some 2.5 PB. Off-site transfer in progress. Data compression desired

Future Goals: Still Thirsty for More Computing Power

Increase in grid resolution: 12288^3, 16384^3, dreaming of 32768^3

- Need exascale, but also constantly adapt to new architectures
- Communicate faster, and/or overlap with other operations?

Larger simulation can be used for many different purposes

- A wider range of scales (higher Reynolds & Schmidt numbers)
- Resolving small scales better, or a larger domain size
- Longer simulations, more time steps

Interest in other phenomena (generalize eqs of motion), such as:

- Buoyancy effects due to temperature and salinity in the ocean
- Magnetic fields: one-way coupling (liquid metal applications) or two-way coupling (Maxwell equations, astrophysics)
- Couplings among body forces: rotation, buoyancy, electromagnetic

Publications based on use of Blue Waters

- Yeung, P.K., Zhai, X.M. and Sreenivasan, K.R. (2015) Extreme events in computational turbulence. *Proc. Nat. Acad. Sci*, **112**, 12633-12638.
- Buaria D., Sawford, B.L. and Yeung, P.K. (2015) Characteristics of two-particle backward dispersion in turbulence at different Reynolds numbers. *Phys. Fluids*, **27**, 105101.
- Iyer, K.P., Sreenivasan K.R. and Yeung, P.K. (2015) Refined similarity hypothesis using 3D local averages. *Phys. Rev. E*, **92**, 063024.
- Buaria, D., Yeung, P.K. and Sawford, B.L. (2016) A Lagrangian study of turbulent mixing: forward and backward dispersion of molecular trajectories in isotropic turbulence. *J. Fluid Mech.*, **799**, 235-382.
- Iyer, K.P., Sreenivasan K.R. and Yeung, P.K. (2017) Reynolds number scaling of velocity increments in isotropic turbulence. *Phys. Rev. E*, **95**, 021101(R).
- Clay, M.P., Buaria, D., Gotoh, T. and Yeung, P.K (2017) A dual communicator, dual grid-resolution algorithm for Petascale simulations of turbulent mixing at high Schmidt number. *Comput. Phys. Comm*, **219**, 313-328.

Publications (cont'd)

- Buaria, D. and Yeung, P.K (2017) A highly scalable particle tracking algorithm using partitioned global address space (PGAS) programming for extreme-scale turbulence simulations. *Comput. Phys. Comm*, **221**, 246-258.
- Clay, M.P., Buaria, D., Yeung, P.K and Gotoh, T. (2018) A dual communicator, dual grid-resolution algorithm for Petascale simulations of turbulent mixing at high Schmidt number. *Comput. Phys. Comm*, **219**, 313-328.
- Yeung, P.K., Sreenivasan, K.R. and Pope, S.B. (2018) Effects of finite spatial and temporal resolution on extreme events in direct numerical simulations of incompressible isotropic turbulence. *Phys. Rev. Fluids* **3**, 064603.
- Iyer, K.P., Schumacher, J., Sreenivasan, K.R. and Yeung, P.K. (2018) Steep cliffs and saturated exponents in three-dimensional scalar turbulence. *Phys Rev. Lett.* **121**,. 264501.
- Zhai, X.M., Sreenivasan, K.R. and Yeung, P.K. (2019) Cancellation exponents in isotropic turbulence and magnetohydrodynamic turbulence. *Phys. Rev. E* **99**, 021302.

Publications (cont'd)

- Iyer, K.P., Schumacher, J., Sreenivasan K.R. and Yeung, P.K. (2019) Scaling of locally averaged energy dissipation and enstrophy density in isotropic turbulence. *New J. Phys.* **21**, 033016.
- Buaria, D., Pumir, A., Bodenschatz, E. and Yeung, P.K. (2019) Extreme velocity gradients in turbulent flows. *New J. Phys.* **21**, 043004.
- Iyer, K.P., Sreenivasan K.R. and Yeung, P.K. (2019) Circulation in high Reynolds number isotropic turbulence is a bifractal. Revised version under review at *Phys. Rev. X*.

THANK YOU AGAIN TO THE BLUE WATERS PROJECT TEAM

ESPECIALLY THE FOLLOWING CURRENT AND FORMER MEMBERS

G. Bauer, B. Bode, R.A. Fiedler, S. Islam, J. Li (POC), R. Sisneros