Extreme Events and Extreme Computing in Turbulence

P.K. Yeung	Georgia Tech (USA)	pk.yeung@ae.gatech.edu
K. Ravikumar	Georgia Tech (USA)	kiran.r@ae.gatech.edu
K.R. Sreenivasan	New York Univ. (USA)	krs3@nyu.edu

Euromech Colloquium 620
Extreme Dissipation and Intermittency in Turbulence
Delft, Netherlands (virtual); May 2021

Resources at US DOE Oak Ridge Nat. Lab & Texas Adv. Comp. Ctr (NSF)

Two Questions about Extreme

Extremely large fluctuations of energy dissipation rate, enstrophy, etc.

- 10 times the mean? 100 times? 1000 times? What is the criterion?
- Are data reliable, whether experiment or simulation? Too large or too small?
- Statistical theory of extrema is difficult (more samples may not $\not\Rightarrow$ convergence)
- Effects of the Reynolds number: need to separate the physics from the numerics.

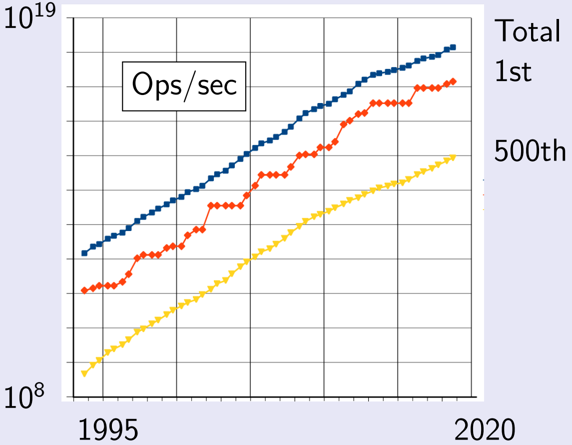
Extreme-scale computing: largest problem size possible at any time

- Time flies: In 2001, 1 Teraflop was extreme; In 2022, expect 1 Exaflop.
- DNS of turbulence: are we keeping pace? What are the challenges?
- Will ever-faster computers also become ever-more-difficult to use?
- Price of entry is high. How can others in the community benefit?
  (Working with Johns Hopkins Turbulence Database group)

"Extreme-Scale" Computing (Tera, Peta, Exa, ..)

"Top500" list shows nearly exponential increase in speed for some 30 years, factor of 10^6 from 1995 to 2019. Pre-Exascale in 2019 (200 PFlops/s)

- But, most user codes get only a small fraction of "theoretical peak"
- Massive distributed parallelism brings challenges in scalability
- Heterogeneous architectures, such as GPUs, are increasingly dominant. Next machine at ORNL will reach 1.5 Exaflops (10^{18} ops/sec) by 2022
- Adaptability to new programming models is crucial for best outcomes



Currently #2: "Summit" at US DOE ORNL

Outline of This Talk

- Extreme-scale computing: the landscape and our present capabilities
- "MRIS" approach for the largest simulations of stationary turbulence
- Results on intermittency obtained at grid resolutions up to $18{,}432^3$
- New results at higher Reynolds numbers for scalar dissipation
- Summary, ongoing work and some remarks

Pseudo-spectral DNS on a 3D Periodic Domain

Incompressible N-S eqs. for velocity fluctuations, with $\nabla \cdot \mathbf{u} = 0$

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

- No mean velocity, but stationary state achieved via low-wavenumber forcing ($\mathbf{f}$)
- Take Fourier transform, project onto plane $\perp$ to wavenumber vector $\mathbf{k}$

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

- 2nd or 4th order Runge Kutta in time; integrating factor for viscous term
- *Pseudo-spectral*: form nonlinear products in physical space, then transform. (No. of operations scales as $N^3 \ln_2 N$ instead of N^6 for convolution integral — although this does not determine overall speed of code...)

Considerations in coding for extreme parallelism

Distributed-memory parallelism: across "nodes", or across CPU cores

- Distribute data, and work, among multiple parallel execution processes
- These processes will need to communicate, by passing (sending/receiving) messages
- Messages usually have to be contiguous: nontrivial pack before and unpack after
- Communication overhead is worst if process count is high and message size is small

Shared-memory on "many-core" nodes and GPUs

- Each parallel process can spawn a number of execution *threads*
- Can some threads be computing, while others communicate? (e.g Clay *et al.* 2017)

Costs of data movement (within and across the parallel processes) are often the greatest limiting factor at large problem sizes

Towards Exascale: Heterogeneous Computing

Heterogeneous: multiple processor types, dissimilar co-processors or accelerators (such as GPUs), usually with specialized capabilities

- GPUs can compute very fast, using optimized libraries
- Data movement between CPU and GPU: minimize it and make it fast
- Memory of the GPU (smaller than CPU) may be an issue at large problem sizes
- New opportunities for asynchronism: e.g. CPU, GPU, and data transfer can be operating on different portions of the data simultaneously

Challenge: We have a communication-intensive problem. How to best benefit from hardware whose prime strength is fast computation?

$18{,}432^3$ Pseudo-spectral DNS on 200 PF/s "Summit"

- DOE Oak Ridge National Lab., USA
- IBM Power9 CPUs + NVIDIA Volta GPUs
- Large CPU memory allows large problem sizes
- Special software for optimizing data copies
- Batched asynchronism:
  - divide "slabs" into sub-units ("pencils")
  - *overlap* copy, compute and copy for different pencils; while optimizing communication

- Details given in Ravikumar, Appelhans & Y, Supercomputing Conference 2019
- Emphasize large problem size: 3X GPU speedup at $18{,}432^3$ (over 6 trillion points)
- Also handles passive scalars and tracks fluid particles, at modest extra cost.

Plenty of Reasons to Keep Pushing the Envelope

Such as ...

- Higher Reynolds no. and wider range of scales: $\ell/\eta \propto R_\ell^{3/4}$ (or ${R_\lambda}^{3/2}$)
- Resolving small scales well, especially for "extreme events"
- Capturing large scales well, minimizing finite domain size effects
- Smaller time steps, for physical or numerical reasons
- Longer simulation in time, depending on the flow physics
- Tracking larger populations of fluid/inertial/Brownian particles
- Mixing at high Schmidt number: smaller grid spacing (Batchelor scale)
- Mixing at low Schmidt number: larger domain, smaller time step

A Challenge: the length of large simulations

Each halving of Δx causes at least 16X increase in resource requirements

- 8X in grid points, 2X in time steps, plus imperfect scalability (inevitable) — but each new machine usually within 10X and in high demand
- Seems ironic: as computing power grows, even going into Exascale era: it becomes harder to simulate long enough, at the largest problem size?

Not surprising: Most record-size simulations were relatively short

Studies of small-scale intermittency in stationary isotropic turbulence

- Ideally, sample statistics over several (5–10) large-eddy time scales (T_E)
- This will become harder at ever-larger problem sizes in the future
- But "extreme events" have short time scales ($O(\tau_\eta)$), $\ll T_E$ at high Re

MRIS: Multiple-Resolution Independent Simulations

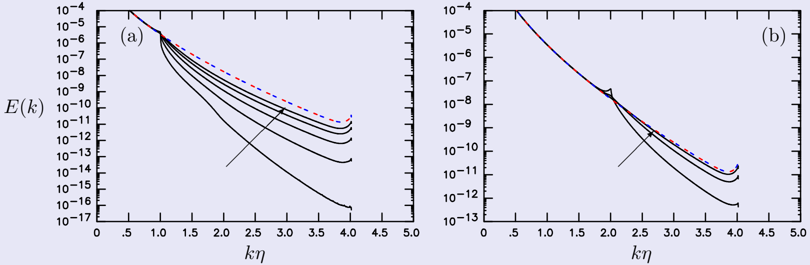
For a given Re, suppose well-sampled results from long-running DNS at desired resolution N^3 is unfeasibile, but feasible at some N_1^3 (with $N_1 < N$).

1. Pick an N_1^3 snapshot as IC, refine the grid (fill extra wavenumber modes by zeroes). The small scales will adjust rapidly, in $O(\tau_\eta)$, while the large scales change little.
2. After $t \sim O(\tau_\eta)$ we run short "simulation segments" at N^3, for say $\beta\tau_\eta$, where β =1–2, which may suffice to capture events with very short time scales.
3. Repeat 1 and 2 for M of N_1^3 snapshots, sufficiently far apart in time for small scales being statistically independent. Average over resulting M segments of size N^3.
4. If N_1/N is deemed too small, go through intermediate resolution: $N_1 \rightarrow N_2 \rightarrow N$.

Overall cost, measured at N^3 resolution, is similar to cost of $M\beta\tau_\eta$ instead of MT_E. Savings especially great if Re is high, such that T_E/τ_η is large.

Validation of MRIS

Start with "reference simulation" at say $N = 3072$, R_λ 390, $k_{max}\eta = 4.2$.
Truncate down (in Fourier space) to $N_1 = 768$, perform the MRIS.
Do we get the "good" results back? And quickly, as hoped?

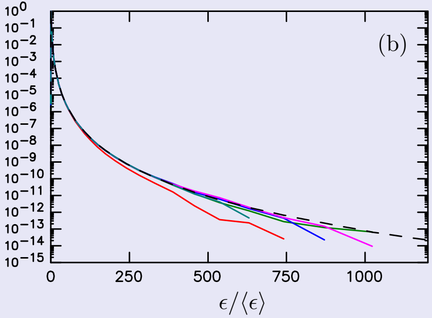


- Energy spectrum shows quick adjustment at small scales (black lines: 0.1 τ_η apart; converging to reference simulation result at end of MRIS segment)
- $N_1 \to N_2 \to N$ more economical than $N_1 \to N$: less run time needed at N^3

Validation: statistics of energy dissipation

PDF of $\epsilon/\langle\epsilon\rangle$ has wide tails
(as in talk by A. Pumir, this morning)

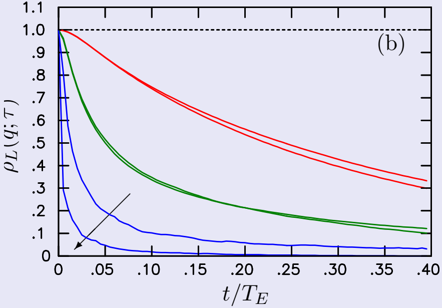
More resolution-sensitive than enstrophy
(Y, Sreenivasan, Pope, PRF 2018)



Rapid convergence as resolution improves

How quickly do measurements in time become approximately independent?

Lagrangian autocorrelations for u^2, ϵ, ϵ^2 at two different resolutions for each



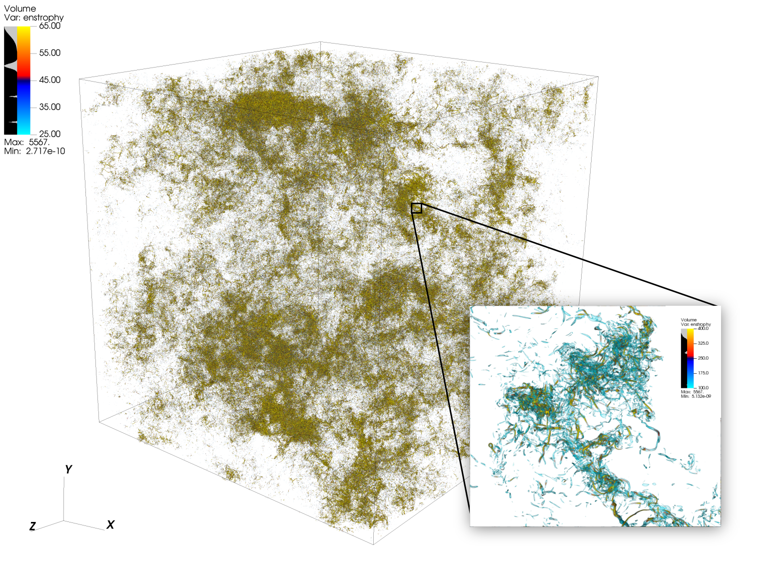
0.4 T_E seems reasonable. Even less may work for finer-scale quantities

List of MRIS Production Simulations

R_λ	N	$k_{max}\eta$	β	M	$\langle\epsilon^2\rangle/\langle\epsilon\rangle^2$	$\langle\Omega^2\rangle/\langle\Omega\rangle^2$
390	1024	1.4	2	22	3.869	7.665
390	1536	2.1	2	22	4.034	7.938
390	3072	4.2	2	22	4.074	7.969
650	2048	1.4	2	15	4.357	8.718
650	3072	2.1	2	15	4.575	9.133
650	6144	4.2	2	15	4.664	9.214
1000	4096	1.4	2	10	4.949	9.901
1000	6144	2.1	2	10	5.250	10.556
1000	12288	4.2	2	10	5.381	10.745
1300	12288	3.0	1	10	6.103	12.238
1300	18432	4.5	1	10	6.142	12.288

Table of MRIS parameters in Y & Ravikumar, PRF Nov 2020

Volume rendering of a 3D $12{,}288^3$ enstrophy field



Zoom-in shows high intensity region, with $100 < \Omega/\langle\Omega\rangle < 400$ in the colored areas
Courtesy of David Pugmire, Michael Matheson (Oak Ridge Natl. Lab.)

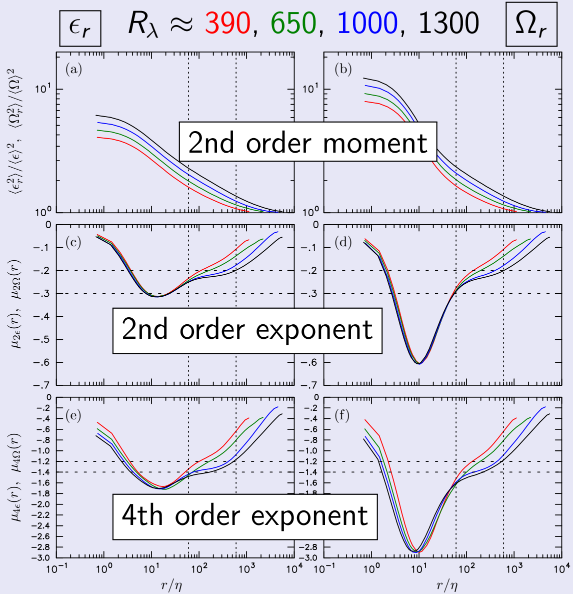
Refined Similarity Theory (started with Kolmogorov 1962)

- Dissipation averaged locally over a 3D volume of linear size r:

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

- Moments of ϵ_r used for intermittency corrections in inertial range
- Long-standing questions: e.g. Frisch (1995), Sreenivasan & Antonia (1997)
- Past experiments: usually averaged along a line, often also used 1-D surrogate ($(\partial u/\partial x)^2$), which is more intermittent
- 3D averages from DNS available only recently (Iyer *et al.* 2015)

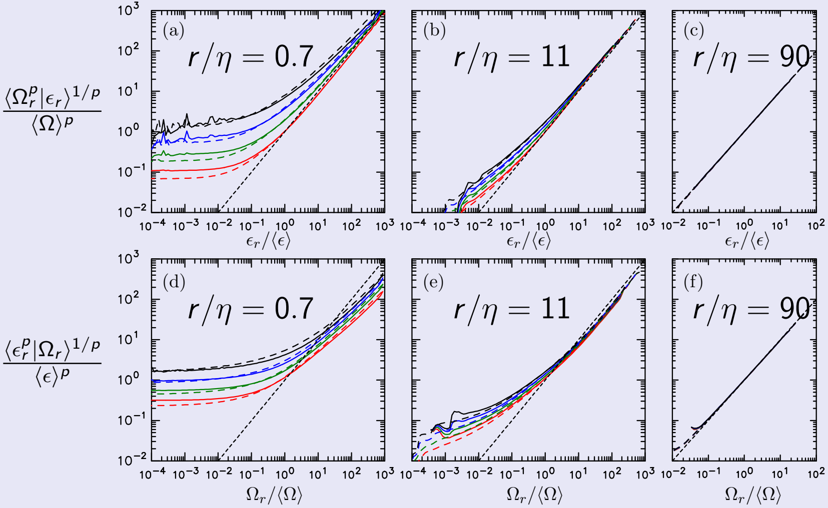
Moments and scaling exponents, orders 2 and 4



- Moments should be flat at low r if small scales sufficiently well-resolved
- Estimate exponents by logarithmic local slopes e.g $\mu_{p\epsilon} = d \ln \langle \epsilon_r^p \rangle / d \ln r$
- A trough around $r/\eta = 10$, deeper for Ω_r than ϵ_r and for higher orders
- $R_\lambda \sim 1300$: almost seeing a plateau in exponent; less well-defined at 4th order — almost same for ϵ_r as Ω_r

Do ϵ and Ω scale *together*? Look at conditional moments

Orders 1, 2, 3, 4; $R_\lambda \sim 390$ (solid); 1300 (dashed)



Ans.: No in dissipation range; but close in inertial range; Yes beyond.

Passive Scalars and Intermittency at high Reynolds no.

Canonical problem: isotropic turbulence, w/ uniform mean scalar gradient

- Production of scalar variance via turbulent scalar flux and mean gradient
- Schmidt number regimes: $Sc \ll 1, O(1), \gg 1$ all have important applications and different scaling properties (Gotoh & Y, 2013)
- Both $Sc \ll 1$ and $Sc \gg 1$ pose new numerical constraints (small Δt, small Δx)
- Focus here on $Sc = 1$, to compare intermittency in scalar vs velocity field

Fluctuations of the scalar dissipation rate: $\chi \equiv 2D|\nabla\phi|^2$

- Refined Similarity theory (e.g. Stolovitzky *et al.* JFM 1995) requires statistics of joint distribution of the local averages ϵ_r and χ_r
- More intermittent than energy dissipation, requires high resolution $\rightarrow$ use MRIS

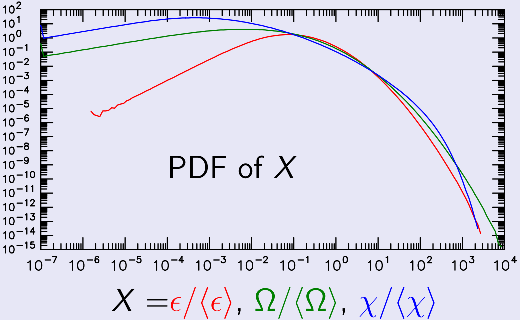
Passive scalars: parameters and one-point statistics

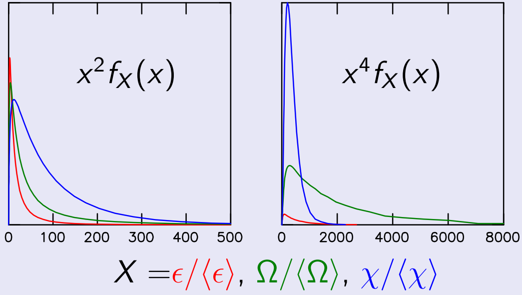
R_λ	N	$k_{max}\eta$	$\langle\epsilon^2\rangle/\langle\epsilon\rangle^2$	$\langle\chi^2\rangle/\langle\chi\rangle^2$
390	1024	1.4	4.19	13.42
390	1536	2.1	4.35	15.32
390	3072	4.2	4.39	15.77
650	2048	1.4	4.62	16.90
650	3072	2.1	4.88	19.70
650	6144	4.2	4.98	20.52
1000	4096	1.4	5.30	19.97
1000	6144	2.1	5.64	23.25
1000	12288	4.2	5.78	24.20

- For R_λ 390 and 650, averaged over scalars with mean gradients in x, y, z
- $k_{max}\eta \sim 4.2$ appears sufficient for second-order moments of both ϵ and χ

Energy dissipation, enstrophy, scalar dissipation PDFs

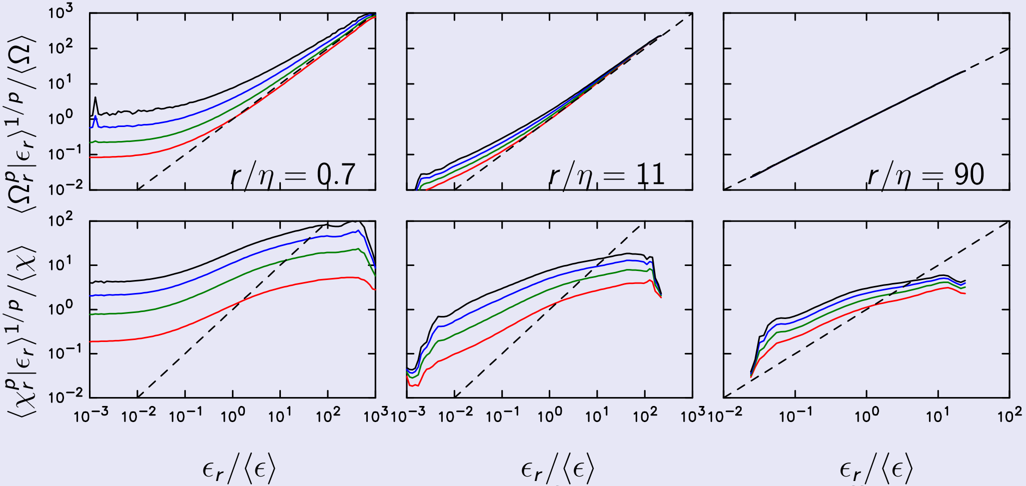
$R_\lambda \sim 1000$, $12{,}288^3$ data shown here ($Sc = 1$ only), normalized by mean





- Power-law for PDF at very small values (Y, Donzis & Sreenivasan JFM 2012)
- Characteristic crossover between PDFs of Ω and χ: suggests moments of $\chi/\langle\chi\rangle$ are largest at modest order but not higher orders (see plots of moment integrands)
- Tails of PDF of χ not similar in shape to those of ϵ and Ω
  — to investigate differences in nonlinear amplification mechanisms

Conditional moments of Ω_r, χ_r given ϵ_r: R_λ 650, $p = 1, 2, 3, 4$



- Extreme χ not tied to extreme ϵ; sites of peak ϵ and peak χ differ
- At large r, increasing homogeneity $\Rightarrow \epsilon_r \to \Omega_r$; but no such constraint for χ_r

Summary of Work Presented

Advancements in computing

- New GPU-optimized parallel algorithm on a pre-exascale machine has allowed us to reach higher grid resolution, $18{,}432^3$ for pseudo-spectral DNS

Multi-Resolution Independent Simulations

- Short yet well-sampled simulations at high resolutions provide a new paradigm necessary at the largest problem size(s) where long runs are prohibitive
- New results for statistics of local averages in isotropic turbulence, R_λ up to 1300

Passive scalar dissipation rate

- Simulations performed using MRIS approach, up to R_λ 1000, $Sc = 1$ (12288^3)
- Contrast with PDFs of ϵ and χ; conditional moments for local averages

Some Ongoing Investigations

Towards the next largest problem size:

- Portable heterogeneous computing crucial for the next few years
- In an Application Readiness program for "Frontier" at Oak Ridge

Multi-fractal analyses at high Reynolds number

- From local averages to local sums, 3D averages vs 1D in past literature
- Directly extendable to scalar dissipation (but Schmidt no. is also a parameter)

Fluid and Stokes particles

- Update of algorithm enables tracking at least 1 billion point particles
- Inertial particles, ultimately to include finite particle size

Anisotropic turbulent flows

- Active scalars, MHD turbulence, rotating turbulence, axisymmetric straining

Final Remarks on The Workshop Theme: "Extreme Dissipation"

1. Dependence on Reynolds number
2. Theories and comparisons with available data
3. Implications for combustion, atmospheric studies...etc.

Some remarks:

1. Computing can help, but eventually we need to extrapolate
2. Closer collaboration between data generators and data users
3. Needs more interactions across so-called disciplinary boundaries