Turbulence in the era of Big Data: Recent experiences with sharing large datasets

Charles Meneveau[1, a)] and Ivan Marusic[2, b)]

[1)] *Department of Mechanical Engineering, and Institute for Data Intensive Engineering and Science (IDIES), The Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218, USA*

[2)] *Department of Mechanical Engineering, University of Melbourne, Melbourne, Victoria 3010, Australia*

(Dated: 8 December 2015)

Placing turbulence research in the context of the contemporary push for "big data" in many fields, we review recent experiences building large databases for turbulence research. We consider data from Direct Numerical Simulations (DNS) of various canonical flows and from experimental studies and related numerical simulations of wall-bounded turbulence, where the data storage needs are particularly challenging due to the very large range of length and time scales that exists in these flows at high Reynolds numbers. The focus is on a move from the traditional approach of data-handling and analysis where datasets are moved to individual computers, to one where much of the analysis is moved to the hosting system that stores the data. In this context we give a summary of a unique open numerical laboratory that archives over 200 Terabytes of DNS data, including full spatio-temporal structure of various canonical flows. Particular attention is given to the unique access requirements for large datasets to become open to the research community and the success the system has had in democratizing access to large datasets.

[a)]Electronic mail: meneveau@jhu.edu

[b)]Electronic mail: imarusic@unimelb.edu.au

I. INTRODUCTION

The notion of "Big Data" has in recent years become ubiquitous in many fields, ranging from the natural to social and political sciences. Big Data means different things to different people. In the area of turbulence, its meaning is clear: the ability of both numerical simulations as well as experiments to generate a huge amount of data has been outpacing our ability to efficiently analyze it. A recognized visionary of the big data revolution, Jim Gray writes in his "Fourth Paradigm of Science"[1], that data analysis in general leads to unforeseen insights and generates hypotheses that guide modeling, new theory, and further experimentation. Turbulence research, especially when based on large numerical simulations, has been challenged to fully realize the true power of data-driven discovery. The largest simulations of the fundamental laws governing fluid flows use tens of millions of hours of CPU time,[2–6] yet most of the analysis must be performed on the memory of compute nodes while the simulation is running because the simulation state is too large to be transferred over networks or stored on traditional file systems. Even transferring and analysis of a few representative snapshots at select time-steps is challenging. It is difficult, if not impossible, to confront model predictions and observations with the exponentially increasing amounts of such simulation data. The push has been to run ever larger DNS of turbulent flows at ever increasing Reynolds numbers. Based on Moore's law continued progress in computing technologies means that within 10-20 years DNS of wall-bounded turbulence will reach Reynolds numbers obtained in the largest laboratory facilities. However, at $Re_\tau = 10^4$ (such as in a large wind tunnel), the DNS would generate 23 Terabytes of data at each time-step, while for $Re_\tau = 10^5$ (such as in the Princeton superpipe) the DNS file size grows to nominally 23 Petabytes per time step, and the full simulation would require an estimated 10^7 time steps[7]. While it is not feasible to keep data from all time steps (for $Re_\tau = 10^5$ that would be 0.2 yottabytes (10^{24})!), it is clear that although these simulations may be realizable in the coming decades, the processing and data-handling tools required to deal with these large datasets will need to keep up. Deciding what and when it is stored requires a high level appreciation of the physics, and invariably this process is an iterative and evolving one. Continuing future efforts will likely benefit greatly from collaborations with computer scientists in devising efficient and tailored data-handling tools.[8]

Big Data is also generated from physical simulations of turbulence that involve laboratory and field study experiments, where very high Reynolds numbers are accessible,[7] and rapid developments in camera, laser, and computer processing technologies have enabled major advances in planar and volumetric particle image velocimetry (PIV) techniques.[9] However, in this case as in DNS, since the dynamical range of length and time scales of the turbulent motions increases with Reynolds number, acquiring fully time-resolved and spatially resolved information requires extremely large datasets.[10,11] Modern time-resolved tomographic PIV experiments of turbulence routinely generate multiple Terabytes of data per minute.

Meaningful progress in the field requires not only that we run sophisticated and massive direct numerical simulations and large-scale experiments, but that the results of these simulations persist for further discovery. Once the relevant and large datasets are stored it is increasingly important to create usable science products from numerical and physical simulations accessible to a broader pool of users. Data arising from the largest simulations must be released publicly, shared, reanalyzed, and archived over extended periods of time. But simply hosting the simulation output files for download, as is done by most projects today, is not good enough once the data volumes exceed a few Terabytes. Once the data volume is too large, one has to move much of the analysis to the data rather than the traditional approach, which moved the data to our computers.

In the following section we describe a unique BigData Open Numerical Laboratory, the Johns Hopkins Turbulence Databases (JHTDB), which presents persistent storage and public access to a select set of DNS data. The system preserves the significant computation effort of simulation and enables further experiments leveraging the data to accelerate discovery.

II. AN OPEN NUMERICAL LABORATORY FOR TURBULENCE: JHTDB

A. System description

The system is part of a set of big data prototypes for open numerical laboratories including open laboratories in other areas of science such as astronomy[12,13], computational cosmology[14,15] and neuroscience[16]. The Johns Hopkins Turbulence Databases (JHTDB, see `http://turbulence.pha.jhu.edu`) is an open numerical laboratory that addresses the increasingly untenable situation of the large HPC-generated data being inaccessible to the vast majority of researchers in turbulence. Its primary goal is to expose large-scale turbulent data to the research community while at the same time providing easy-to-use client interfaces for retrieving and interacting with the data.

Fundamental to JHTDB's approach is the ability of the user community to interact easily and flexibly with these massive amounts of data. A Database Web Service has thus been built that handles requests over the web for velocities, pressure, various space-derivatives of velocity and pressure, interpolation functions, particle tracking for trajectories, etc. We index the data using a space-filling fractal curve (Z-index) and apply the "move the program to the data" philosophy[12], a fundamental tenet in the design of large-scale scientific databases. A first dataset was placed in the open laboratory in 2008[17], and consisted of a DNS of forced isotropic turbulence on $1,024^3$ grid points. A total of $1,024$ time steps were captured, encompassing the time evolution during a single large-eddy turnover time.

In Figure 1, a schematic of the Johns Hopkins Turbulence Database (JHTDB) system is shown, with remote users indicated at the top of the figure. They are separated from the databases by the Web server which processes incoming requests and returns the requested data. The first component of JHTDB is the Web server (front-end) that provides the layer with which remote clients can interact. The client interfaces remove the need for the end user to know details of the data storage, such as the distribution of data across database nodes and how they are indexed. The database cluster (backend), composed of a networked database system running Microsoft SQL Server, is the second component. The cluster contains the datasets and provides a scalable infrastructure that supports data-intensive analyses. This Web service model is modeled after the successful SkyServer approach[18] in which multi-terabyte astronomy data archives have been available to researchers[12] for some time.

The data have been generated by a large-scale simulation performed using a HPC facility. Within the database cluster, data are partitioned across several nodes. The Web server issues queries to the database cluster asynchronously using multiple connections per node in order to leverage the multicore architecture of each node. Data from a simulation are stored in the database as binary large objects (BLOBs) indexed by the (Morton) Z-curve[20,21]. The Z-curve maps the 3D data to the one dimensional index space. It passes through each location in space once and only once (i.e. space filling). It is easy to compute and provides good spatial locality which is important in order to support contiguous data access for typical usage patterns of the databases.

There are four types of flows presently available (over 200 TBytes) in the open turbulence laboratory: (1) A forced isotropic turbulence case at a Taylor-scale Reynolds number of $R_\lambda = 433$ from a pseudo-spectral simulation on 1024^4 space-time data points[17]. (2) A 56 TB database of steady-state magnetohydrdynamic (MHD) turbulence, obtained from a $1,024^3$ pseudo-spectral simulation of unit magnetic-Prandtl-number, incompressible MHD, forced with a Taylor-Green flow. Taylor-scale Reynolds numbers are $Re_{\lambda,u} = 186$ and $R_{\lambda,b} = 144$ for the velocity and magnetic fields, respectively. The velocity, pressure, magnetic field,

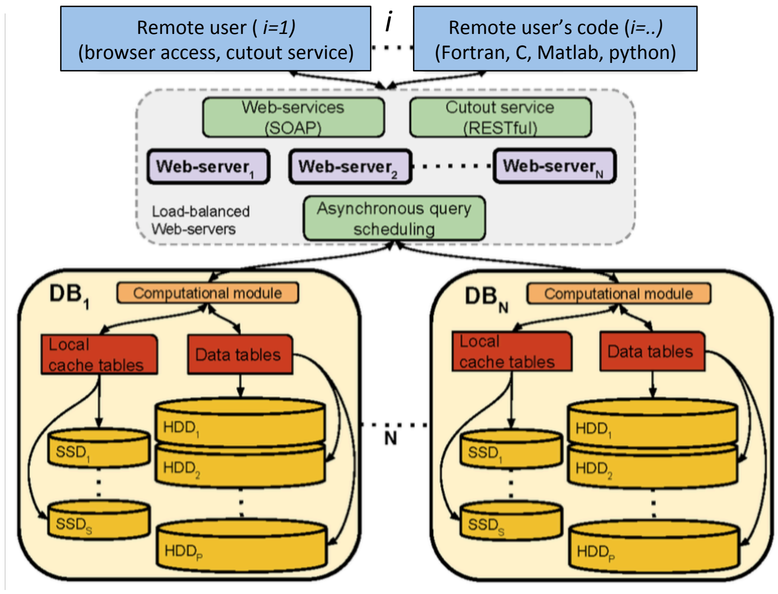

FIG. 1. Schematic of the existing architecture of the JHTDB Open Numerical Turbulence Laboratory, showing logical layout of the remote users (clients), Web server, and the database cluster (sketch adapted from Kanov *et al.*[19]).

and Coulomb-gauge vector potential are stored on 1024^4 space-time grid points, spanning about one large-scale eddy turnover time. (3) A channel flow dataset at $Re_\tau = 1,000$ on $2,048 \times 1,536 \times 512$ spatial data points and 4,000 time steps, simulated in collaboration with researchers at Univ. Texas at Austin,[22–24] yielding over 110 TBytes, and (4) a variable-density mixing flow on $1,024^3$ spatial and 1,015 temporal data points simulated by Los Alamos (LANL) researchers.[25,26]

B. Data access

A hallmark of the system is the data access philosophy, predicated on the notion of *virtual sensors*. We base our virtual sensor data access philosophy on an experimental analogy: an experimenter would like to place sensors at specified positions and times (x, y, z, t) and "measure" velocities, pressure and other derived quantities there. This can be a one-time measurement, or as function of time at a physical (Eulerian) location or the points can move with the flow as fluid particle trackers. Derived quantities can be based on a variety of operators, like the gradient or Laplacian of a field, or applying various filters[27] and thresholds.[28] Such data access patterns also enable the users to run time backwards[29], impossible in a direct simulation of a dissipative system. Snapshots are saved frequently enough so that one can interpolate field values smoothly enough. Sensors can back-track their original trajectory and one can see where they came from, all the way back to the initial conditions. Furthermore, virtual sensors placed anywhere in the 4D flow domain typically require interpolations. For spatial interpolation, Lagrange polynomial[17] or spline interpolation[?] methods are used. For Lagrange polynomials the client has the option between no interpolation, 4th order, 6th order, or 8th order interpolation. For splines, the client can select between 3rd and 5th order methods. In the case of the isotropic, MHD and variable density mixing databases, interpolation with uniform grid spacing is employed. For cases with non-equal grids, a generalized barycentric Lagrange interpolation method is used (e.g. non-uniform grid spacing in the wall normal direction for the channel flow database). Spatial differentiation is based on various methods such as finite difference or differentiation of the interpolation splines/polynomials. Temporal interpolation is performed using piecewise cubic Hermite interpolating polynomial.

The existing Web services provide a convenient mechanism for remote clients to interact with the databases via the immersive approach. The Web services use Simple Object Access Protocol (SOAP) which provides a standard protocol for sending and receiving messages over the internet. Client functions are provided by the Web services for spatial and temporal interpolation, differentiation, fluid particle tracking and other secondary calculations performed within the database. These calculations are in addition to client functions to request primary fields. Packaged libraries which allow easy use of the Web service functions are provided for C, Fortran, Matlab and Python. Fig. 2 shows a snippet of matlab code to read a subset of the channel flow velocity and generate a contour plot.

At this stage, what is noteworthy about the availability of a 4D dataset "at your fingertips" and the ability to make "casual" queries from anywhere at any time, is that it is beginning to change how we think about the data. Researchers can come back to the same place in space and time and be sure to encounter the same values. They can follow and observe phenomena forward and backwards in time. The reader is invited to use any web-browser and visit *http://turbulence.pha.jhu.edu/webquery/query.aspx*, select (e.g.) the "channel" dataset and the "*GetPressureGradient*" function, and enter a particular time between 0-25 along with x, y, z positions in the range $(0 : 8\pi) \times (-1 : 1) \times (0 : 3\pi)$. The database will return the three components of pressure gradient at that position in space-time. Velocity, pressure, the velocity gradient tensor, and other quantities of interest are available by

```
%  Chose a random time step
time = 0.002 * randi(1024,1);
spacing = 3.0*pi/1536; % z-spacing = 1/2 x-spaci
% Set domain size and position
nx = 256;
nz = nx;
xoff = 2*spacing*randi([0,2048-nx]);
yoff = -0.99; % y+ = 10
zoff = spacing*randi([0,1536-nz]);
npoints = nx*nz;
% Create surface
x = linspace(0, (nx-1)*spacing, nx) + xoff;
z = linspace(0, (nz-1)*spacing, nz) + zoff;
[X Z] = meshgrid(x, z);
points(1,:) = X(:)';
points(3,:) = Z(:)';
points(2,:) = yoff;

% Get the velocity at each point
result3 = getVelocity(authkey, dataset, time, Lag4, NoTInt, npoints, points);

% Assign x-velocity to Y
y = result3(1,:);
Y = reshape(y, nx, nz);

% Plot x-velocity contours
contourf(X, Z, Y, 30, 'LineStyle', 'none');
```

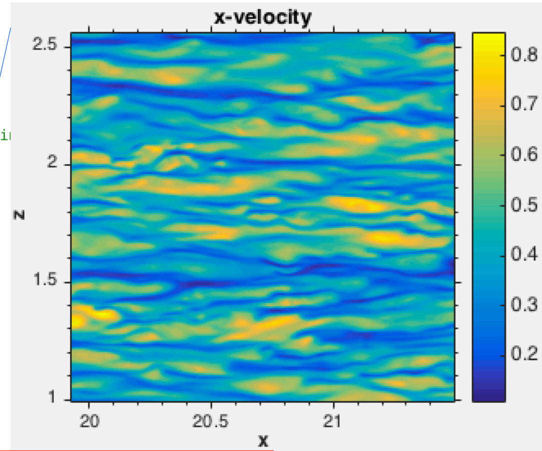

FIG. 2. Snippet of Matlab code for extraction of velocity data (using *getVelocity*) from JHTDB channel flow database at $Re_\tau = 1000$ and visualization of streamwise velocity in plane near the wall (at $y^+ = 10$).

similar simple queries. There are tremendous benefits to being able to re-visit repeatedly the complex flow phenomena at leisure, as physical intuition about the complicated dynamics begins to develop.

C. Community response and impact

While researchers who themselves perform very large simulations could often (but not always) analyze their own data more effectively within original HPC environments, the rest of the community could not do so effectively or not at all. Thus, many researchers started to access the JHTDB data and carry out their research in the open Numerical Laboratory. A recent analysis of JHTDB's usage patterns has been presented in Kanov *et al.*[19], as part of a special journal issue on open numerical laboratories.[30] Figure 2 in Kanov *et al.*[19] shows the worldwide distribution of requests to the turbulence database, measured by total number of points queried. There is heavy usage and it has become a global resource. Figure 3 displays the cumulative number of points queried from the system.[31] In 2015, the number of points has exceeded 12 trillion.

The availability of massive, high-resolution turbulence datasets in web-accessible databases

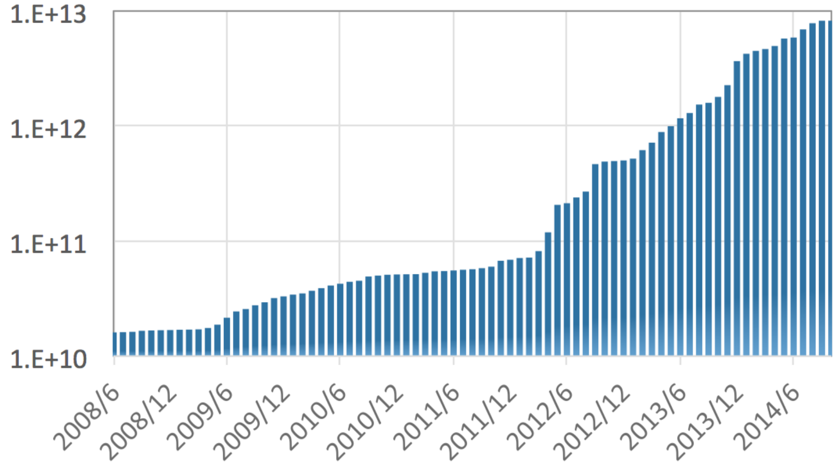

FIG. 3. Cumulative number of points queried as a function of time. As of Jan. 2015 the number of points has exceeded 12 trillion.[31]

has already proven to have major impact on turbulence research and education. The system has been used to address many research questions in turbulence, both from the internal group that has generated and is curating the DNS data, as well as from external users. From the internal set of users, a number of publications can be mentioned, such as Refs.[29,32–47]. External users have used the databases in studies of extreme events in turbulence,[48], shape evolution[49] characterization of particles-turbulence interactions,[50] calibration of experimental measurement tools,[51–54] the study of velocity gradient tensor properties,[55–57] subgrid-scale model assessments,[58,59] and many others.[49,60–72] Typically there are now about 1-2 individualized tokens requested and assigned to new outside users every week (a token is needed if a user wishes to download more than 4096 points in a single call).

In terms of education, JHDTB has been used in various classes around the world, as well as in student workshops (e.g. the "Tutorial School on Fluid Dynamics: Topics in Turbulence" held at the UMD College Park - see `http://www2.cscamm.umd.edu/programs/trb10/`, that has been repeated in 2015 `http://burgers.umd.edu/school/`). Especially the Matlab interface has been very popular to facilitate student access to turbulence data.

III. CONCLUSIONS

Turbulence research has arguably been at the forefront of "big data" for a long time, as it has been generating very large datasets both through numerical simulations and experiments.

It thus corresponds to turbulence researchers to seek out and establish interdisciplinary collaborations especially with computer data scientists, and with them to propose and adapt new tools that enable wider access to turbulence data sets. Here we have presented details of a unique 'open laboratory for turbulence' at the Johns Hopkins University as a example of the required paradigm shift where one has to move much of the analysis to the data rather than the traditional approach, which moved the data to our computers. It is hoped that broader access to data, from simulations and experiments, will further accelerate turbulence research in coming years.

Acknowledgements: The authors congratulate Prof. William George, for whom this Festschrift is intended, on occasion of his 70th birthday and salute him for his many accomplishments in turbulence research. CM gratefully acknowledges the large collaborative and interdisciplinary effort that has made the JHTDB open numerical turbulence laboratory possible, specifically the following colleagues, scholars, students and staff: A. Szalay, R. Burns, G. Eyink, S. Chen, E. Vishniac, T. Zaki, E. Perlman, Y. Li, T. Budavari, M. Wan, H. Yu, E. Frederix, K. Buerger, H. Aluie, J. Graham, C. Lalescu, K. Kanov, S. Hamilton, P. Johnson, D. Livescu, R.D. Moser, M. Lee and N. Malaya, J. Vandenberg, S. Werner, V. Paul and G. Lemson. The funding from the National Science Foundation (grants CMMI-0941530 and CBET-1507469) is gratefully acknowledged. IM thanks the support of the Australian Research Council.

REFERENCES

[1] A. Hey, S. Tansley, K. Tolle, *et al.*, "The Fourth Paradigm: Data-intensive scientific discovery," Microsoft Research Redmond, WA (2009).

[2] D. Donzis, P. Yeung, and K. Sreenivasan, "Dissipation and enstrophy in isotropic turbulence: Resolution effects and scaling in direct numerical simulations," Physics of Fluids **20**, 045108 (2008).

[3] T. Ishihara, T. Gotoh, and Y. Kaneda, "Study of high–Reynolds number isotropic turbulence by direct numerical simulation," Annu. Rev. Fluid Mech. **41**, 165–180 (2009).

[4] X. Wu and P. Moin, "Direct numerical simulation of turbulence in a nominally zero-pressure-gradient flat-plate boundary layer," Journal of Fluid Mechanics **630**, 5–41 (2009).

[5]Y. Khalighi, J. Nichols, S. Lele, F. Ham, and P. Moin, "Unstructured large eddy simulation for prediction of noise issued from turbulent jets in various configurations," AIAA paper **2886** (2011).

[6]M. Lee and R. Moser, "Direct numerical simulation of turbulent channel flow up to Re_τ = 5200," J. Fluid Mech. **774**, 395 – 415 (2015).

[7]A. Smits and I. Marusic, "Wall-bounded turbulence," Physics Today **66**, 25–30 (2013).

[8]I. Marusic, G. V. Candler, V. Interrante, P. K. Subbareddy, and A. Moss, "Data mining tools for the analysis of turbulent flows: Needs, issues and potential," in *Data Mining for Scientific and Engineering Applications,* R. Grossman, C. Kamath, P. Kegelmeyer, V. Kumar and R. Namburu (Editors), Kluwer Acad. Pub. (2001).

[9]R. J. Adrian and J. Westerweel, *Particle Image Velocimetry* (Cambridge University Press, 2011).

[10]A. Schröder, R. Geisler, K. Staack, G. Elsinga, F. Scarano, B. Wieneke, A. Henning, C. Poelma, and J. Westerweel, "Eulerian and lagrangian views of a turbulent boundary layer flow using time-resolved tomographic PIV," Experiments in fluids **50**, 1071–1091 (2011).

[11]C. M. de Silva, E. Gnanamanickam, C. Atkinson, N. A. Buchmann, N. Hutchins, J. Soria, and I. Marusic, "High spatial range velocity measurements in a high Reynolds number turbulent boundary layer," Phys. Fluids **26**, 025117 (2014).

[12]A. S. Szalay, P. Kunszt, A. Thakar, J. Gray, D. Slutz, and R. J. Brunner, "Designing and mining mult-terabyte astronomy archives: the Sloan digital sky survey," Proceedings of the 2000 ACM SIGMOD international conference on Management of data **.**, 451–462 (2000).

[13]W. O'Mullane, J. Gray, N. Li, T. Budavari, M. Nieto-Santisteban, and A. Szalay, "Batch query system with interactive local storage for SDSS and the VO," in *Proc. ADASS XIII, ASP Conference Series, eds: F. Ochsenbein, M. Allen and D. Egret* (ASP Conference Series, 2004) p. 314:372.

[14]G. Lemson and the Virgo Consortium, "Halo and galaxy formation histories from the millennium simulation: Public release of a VO-oriented and SQL-queryable database for studying the evolution of galaxies in the lambdacdm cosmogony," arXiv:astro-ph **.**, 0608019 (2006).

[15]R. Overzier, G. Lemson, R. Angulo, and et-al., "The millennium run observatory: first light," MNRAS **428**, 778 (2013).

[16]R. Burns, J. Vogelstein, and A. Szalay, "From cosmos to connectomes: the evolution of data-intensive science," Neuron **83**, 1249–1252 (2014).

[17]Y. Li, E. Perlman, M. Wan, Y. Yang, C. Meneveau, R. Burns, S. Chen, A. Szalay, and G. Eyink, "A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence," J. Turbulence **9**, No. 31 (2008).

[18]Y. Gu, R. Grossman, A. Szalay, and A. Thakar, "Distributing the Sloan Digital Sky Survey using UDT and sector," in *e-Science* (2006) p. 56.

[19]K. Kanov, C. Burns, R. Lalescu, and G. Eyink, "The johns hopkins turbulence databases: An open simulation laboratory for turbulence research," Computing in Science & Engineering **17:5**, 10–17 (2015).

[20]H. Samet, *Foundations of Multidimensional and Metric Data Structures* (Morgan Kaufmann Publishers, 2006).

[21]E. Perlman, R. Burns, Y. Li, and C. Meneveau, "Data exploration of turbulence simulations using a database cluster," SC07 **07** (2007).

[22]J. Graham, K. Kanov, E. Givelberg, R. Burns, G. Eyink, A. Szalay, C. Meneveau, M. Lee, N. Malaya, and R. Moser, "A web-services accessible database for channel flow turbulence at Re-tau= 1000," Bulletin of the American Physical Society **58** (2013).

[23]M. Lee, N. Malaya, and R. Moser, "Petascale direct numerical simulation of turbulent channel flow on up to 786k cores," in *Proceedings of SC13: International Conference for High Performance Computing, Networking, Storage and Analysis* (ACM, 2013) p. 61.

[24]J. Graham, K. Kanov, X. Yang, M. Lee, N. Malaya, C. Lalescu, R. Burns, G. Eyink, A. Szalay, R. Moser, and C. Meneveau, "A web services-accessible database of turbulent channel flow and its use for testing a new integral wall model for les," J. Turbulence **17:2**, 181–215 (2016).

[25]D. Livescu and J. Ristorcelli, "Variable-density mixing in buoyancy-driven turbulence," Journal of Fluid Mechanics **605**, 145–180 (2008).

[26]D. Livescu, "Numerical simulations of two-fluid turbulent mixing at large density ratios and applications to the rayleigh–taylor instability," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **371**, 20120185 (2013).

[27]K. Kanov, R. Burns, G. Eyink, C. Meneveau, and A. Szalay, "Data-intensive spatial filtering in large numerical simulation datasets," International Conference for High Performance Computing, Networking, Storage and Analysis (SC) **(IEEE)** (2012).

[28]K. Kanov, R. Burns, and C. Lalescu, "Efficient evaluation of threshold queries of derived fields in a numerical simulation database," Proc. 18th International Conference on

Extending Database Technology (EDBT) (2015).

[29]G. Eyink, E. Vishniac, C. Lalescu, K. Aluie, H.and Kanov, K. Bürger, R. Burns, C. Meneveau, and A. Szalay, "Flux-freezing breakdown in high-conductivity magnetohydrodynamic turbulence," Nature **497**, 466–469 (2013).

[30]F. Alexander and C. Meneveau, "Foreword to the "open simulation laboratories" special issue," Computing in Science & Engineering **17:5**, 7–9 (2015).

[31]K. Kanov, *Efficient Evaluation of Data-intensive Batch-queries in Open Simulation Laboratories*, Ph.D. thesis, Johns Hopkins University (2015).

[32]P. Johnson and C. Meneveau, "Large-deviation joint statistics of the finite-time Lyapunov spectrum in isotropic turbulence," Phys. Fluids **.**, under review (2015).

[33]G. Eyink, "Stochastic flux freezing and magnetic dynamo," Physical Review E **83**, 056405 (2011).

[34]H. Yu and C. Meneveau, "Lagrangian refined Kolmogorov similarity hypothesis for gradient time evolution and correlation in turbulent flows," Phys. Rev. Lett **104**, 084502 (2010).

[35]H. Yu and C. Meneveau, "Scaling of conditional Lagrangian time correlation functions of velocity and pressure gradient magnitudes in isotropic turbulence," Flow, Turbulence and Combustion **85**, 457–472 (2010).

[36]Y. Li, L. Chevillard, C. Meneveau, and G. Eyink, "Matrix exponential-based closures for the turbulent subgrid-scale stress tenso," Phys. Rev. E **79**, 016305 (2009).

[37]K. Kanov, E. Perlman, R. Burns, Y. Ahmad, and A. Szalay, "I/O streaming evaluation of batch queries for data-intensive computational turbulence," in *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis* (ACM, 2011) p. 29.

[38]X. Wang, E. Perlman, R. Burns, T. Malik, T. Budavári, C. Meneveau, and A. Szalay, "Jaws: Job-aware workload scheduling for the exploration of turbulence simulations," in *Proceedings of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis* (IEEE Computer Society, 2010) pp. 1–11.

[39]H. Yu, K. Kanov, E. Perlman, J. Graham, E. Frederix, R. Burns, A. Szalay, G. Eyink, and C. Meneveau, "Studying lagrangian dynamics of turbulence using on-demand fluid particle tracking in a public turbulence database," Journal of Turbulence **13** (2012).

[40]M. Treib, K. Burger, F. Reichl, C. Meneveau, A. Szalay, and R. Westermann, "Turbulence visualization at the Terascale on desktop PCs," Visualization and Computer Graphics, IEEE Transactions on **18**, 2169–2177 (2012).

[41]G. Eyink and D. Benveniste, "Diffusion approximation in turbulent two-particle dispersion," Physical Review E **88**, 041001 (2013).

[42]J. Graham, E. Givelberg, and K. Kanov, "Run-time creation of the turbulent channel flow database by an hpc simulation using mpi-db," in *Proceedings of the 20th European MPI Users' Group Meeting* (ACM, 2013) pp. 151–156.

[43]D. Benveniste and T. Drivas, "Asymptotic results for backwards two-particle dispersion in a turbulent flow," Physical Review E **89**, 041003 (2014).

[44]C. Meneveau, "Lagrangian dynamics and models of the velocity gradient tensor in turbulent flows," Annu. Rev. Fluid Mech. **43**, 219—245 (2011).

[45]C. Meneveau, "Germano identity-based subgrid-scale modeling: A brief survey of variations on a fertile theme," Physics of Fluids **24**, 121301 (2012).

[46]J. Boschung, P. Schaefer, N. Peters, and C. Meneveau, "The local topology of stream-and vortex lines in turbulent flows," Physics of Fluids **26**, 045107 (2014).

[47]Y. Ahmad, R. Burns, M. Kazhdan, C. Meneveau, A. Szalay, and A. Terzis, "Scientific data management at the Johns Hopkins Institute for Data Intensive Engineering and Science," ACM SIGMOD Record **39**, 18–23 (2011).

[48]H. Xu, A. Pumir, G. Falkovich, E. Bodenschatz, M. Shats, H. Xia, N. Francois, and G. Boffetta, "Flight–crash events in turbulence," Proceedings of the National Academy of Sciences **111**, 7558–7563 (2014).

[49]J. Jucha, H. Xu, A. Pumir, and E. Bodenschatz, "Time-reversal-symmetry breaking in turbulence," Phys. Rev. Lett. **113**, 054501 (2014).

[50]K. Gustavsson, J. Einarsson, and B. Mehlig, "Tumbling of small axisymmetric particles in random and turbulent flows," Physical Review Letters **112**, 014501 (2014).

[51]N. Buchmann, C. Willert, and J. Soria, "Pulsed, high-power LED illumination for tomographic particle image velocimetry," Experiments in Fluids **53**, 1545–1560 (2012).

[52]X. Liu and J. Katz, "Vortex-corner interactions in a cavity shear layer elucidated by time-resolved measurements of the pressure field," Journal of Fluid Mechanics **728**, 417–457 (2013).

[53]D. Xu and J. Chen, "Accurate estimate of turbulent dissipation rate using PIV data," Experimental Thermal and Fluid Science **44**, 662–672 (2013).

[54]D. Fiscaletti, J. Westerweel, and G. Elsinga, "Long-range μPIV to resolve the small scales in a jet at high Reynolds number," Experiments in Fluids **55**, 1–15 (2014).

[55]J. M. Lawson and J. R. Dawson, "On velocity gradient dynamics and turbulent structure," J. Fluid Mech. **780**, 60–98 (2015).

[56]B. Luethi, M. Holzer, and A. Tsinober, "Expanding the Q–R space to three dimensions," J. Fluid Mech. **641**, 497–507 (2010).

[57]J. Cardesa, D. Mistry, L. Gan, and J. Dawson, "Invariants of the reduced velocity gradient tensor in turbulent flows," Journal of Fluid Mechanics **716**, 597–615 (2013).

[58]H. Lu, "Assessment of the modulated gradient model in decaying isotropic turbulence," Theoretical and Applied Mechanics Letters **1**, 041004 (2011).

[59]G. A.G. and S. Menon, "A new two-scale model for large eddy simulation of wall-bounded flows," Progr. Aerospace Sci. **46**, 28–45 (2010).

[60]F. Grinstein, A. Gowardhan, and A. Wachtor, "Simulations of Richtmyer–Meshkov instabilities in planar shock-tube experiments," Physics of Fluids **23**, 034106 (2011).

[61]W. Liu and E. Ribeiro, "Scale and rotation invariant detection of singular patterns in vector flow fields," in *Structural, Syntactic, and Statistical Pattern Recognition* (Springer, 2010) pp. 522–531.

[62]P. Bhat and K. Subramanian, "Fluctuation dynamos and their Faraday rotation signatures," Monthly Notices of the Royal Astronomical Society **429**, 2469–2481 (2013).

[63]C. Keylock, T. Tokyay, and G. Constantinescu, "A method for characterising the sensitivity of turbulent flow fields to the structure of inlet turbulence," Journal of Turbulence **12** (2011).

[64]M. Holzner, M. Guala, B. Lüthi, A. Liberzon, N. Nikitin, W. Kinzelbach, and A. Tsinober, "Viscous tilting and production of vorticity in homogeneous turbulence," Physics of Fluids **22**, 061701 (2010).

[65]C. Wu and T. Chang, "Rank-ordered multifractal analysis (ROMA) of probability distributions in fluid turbulence," Nonlinear Processes in Geophysics **18**, 261–268 (2011).

[66]C. Keylock, K. Nishimura, and J. Peinke, "A classification scheme for turbulence based on the velocity-intermittency structure with an application to near-wall flow and with implications for bed load transport," Journal of Geophysical Research: Earth Surface (2003–2012) **117** (2012).

[67]W. Liu and E. Ribeiro, "Detecting singular patterns in 2D vector fields using weighted Laurent polynomial," Pattern Recognition **45**, 3912–3925 (2012).

[68]M. Mishra, X. Liu, M. Skote, and C.-W. Fu, "Kolmogorov spectrum consistent optimization for multi-scale flow decomposition," Physics of Fluids **26**, 055106 (2014).

[69]L. Moriconi and R. Pereira, "Vorticity statistics and the time scales of turbulent strain," Physical Review E **88**, 013005 (2013).

[70]F. Grinstein, A. Gowardhan, J. Ristorcelli, and A. Wachtor, "On coarse-grained simulations of turbulent material mixing," Physica Scripta **86**, 058203 (2012).

[71]T. Chang, C. Wu, M. Echim, H. Lamy, M. Vogelsberger, L. Hernquist, and D. Sijacki, "Complexity phenomena and ROMA of the earth's magnetospheric cusp, hydrodynamic turbulence, and the cosmic web," Pure and Applied Geophysics , 1–19 (2014).

[72]A. Pumir, H. Xu, G. Boffetta, G. Falkovich, and E. Bodenschatz, "Redistribution of kinetic energy in turbulent flows," Physical Review X **4**, 041006 (2014).