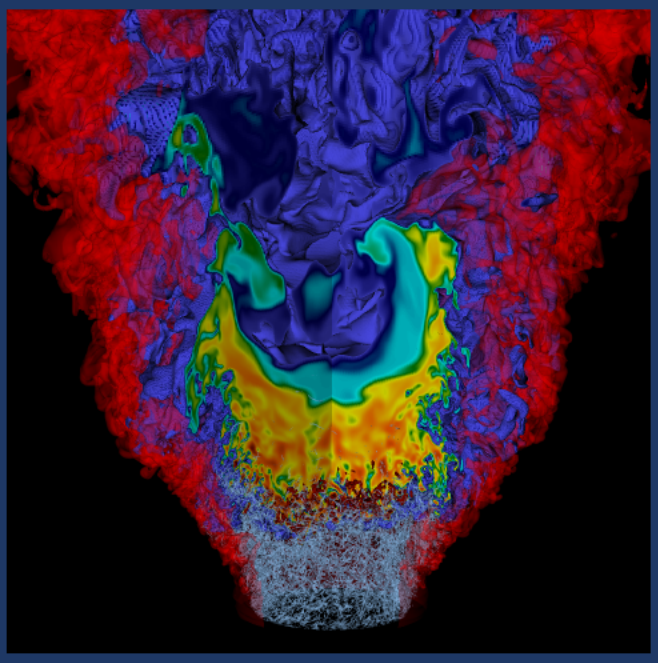
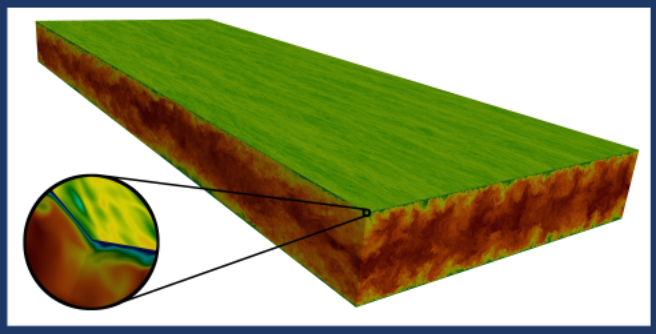
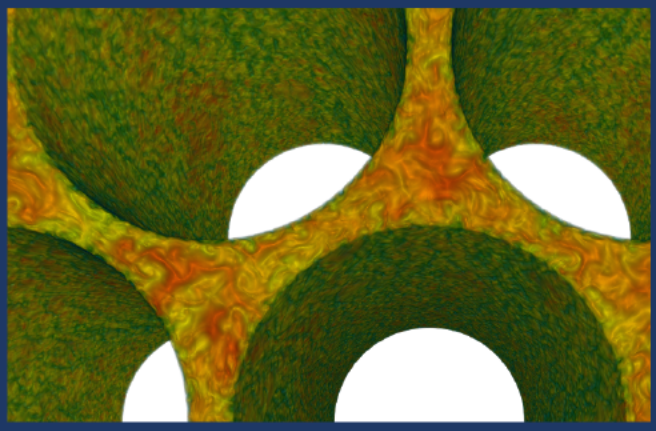
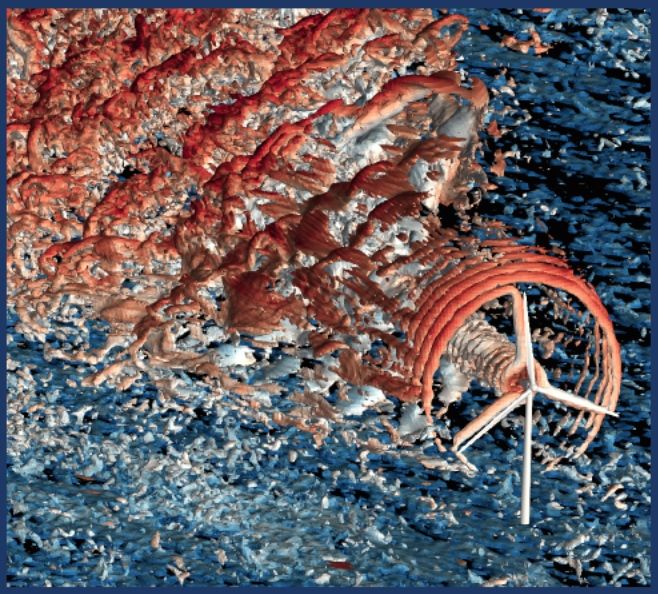
Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop

August 4–5, 2015, Washington, D.C.









Program Committee:
Michael A. Sprague
Stanislav Boldyrev
Paul Fischer
Ray Grout
William I. Gustafson Jr.
Robert Moser

DOE Points of Contact:
Michael Martin
Robinson Pino

Sponsored by the Office of Advanced Scientific Computing Research

U.S. DEPARTMENT OF ENERGY | Office of Science

Cover
Details for the cover images, going clockwise and starting from the upper left corner, are as follows: Simulation of a lean premixed hydrogen flame stabilized on a low-swirl burner (image courtesy of Dr. John Bell); visualization of the turbulent flow in a planar channel at friction Reynolds number 5200 (image courtesy of Dr. Myoungkyu Lee); volume rendering of the velocity magnitude in a nuclear reactor (image courtesy of Dr. Elia Merzari); isosurfaces of vorticity colored by the instantaneous streamwise velocity in the wake of a wind turbine (image courtesy of Dr. Matt Churchfield).

Availability of This Report
This report is available, at no cost, at http://www.osti.gov/bridge. It is also available on paper to the U.S. Department of Energy and its contractors, for a fee, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone (865) 576-8401
fax (865) 576-5728
reports@adonis.osti.gov

Auspices Statement
This work was performed under the auspices of the U.S. Department of Energy. Sprague and Grout's contributions were supported as employees of the Alliance for Sustainable Energy, LLC (Alliance) by the U.S. Department of Energy under Contract No. DE-AC36-08GO28308. A portion of Sprague's contribution was supported by the DOE Wind Energy Technologies Office. Gustafson's contribution was supported by Pacific Northwest National Laboratory (PNNL) Institutional Overhead funds; PNNL is operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830. Hittinger's contribution was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Chang's contribution was supported by U.S. DOE ASCR and FES through the SciDAC program and by OLCF for the computing resources. Fischer's contribution was supported by U.S. DOE ASCR under Contract DE-AC02-06CH11357. The U.S. Government and the Alliance retain a nonexclusive royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. NREL report number TP-2C00-67648.

Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop

Sponsored by
U.S. Department of Energy
Office of Science
Office of Advanced Scientific Computing Research

Program Committee
Michael Sprague (chair, National Renewable Energy Laboratory)
Stanislav Boldyrev (University of Wisconsin–Madison)
Paul Fischer (Argonne National Laboratory and University of Illinois at Urbana–Champaign)
Ray Grout (National Renewable Energy Laboratory)
William I. Gustafson Jr. (Pacific Northwest National Laboratory)
Robert Moser (University of Texas at Austin)

DOE/ASCR Points of Contact
Michael Martin and Robinson Pino

Abstract

This report details the findings and recommendations from the *Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop*, which was held August, 4–5, 2015, and was sponsored by the U.S. Department of Energy (DOE) Office of Advanced Scientific Computing Research (ASCR). The workshop objectives were to define and describe the challenges and opportunities that computing at the exascale will bring to turbulent-flow simulations in applied science and technology. The need for accurate simulation of turbulent flows is evident across the DOE applied-science and engineering portfolios, including combustion, plasma physics, nuclear-reactor physics, wind energy, and atmospheric science. The workshop brought together experts in turbulent-flow simulation, computational mathematics, and high-performance computing. Building upon previous ASCR workshops on exascale computing, participants defined a research agenda and path forward that will enable scientists and engineers to continually leverage, engage, and direct advances in computational systems on the path to exascale computing.

Contents

Executive Summary

This report details the findings and recommendations from the *Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop*, which was held August 4–5, 2015, and was sponsored by the U.S. Department of Energy (DOE) Office of Advanced Scientific Computing Research (ASCR). The workshop objectives were to define and describe the challenges and opportunities that computing at the exascale will bring to turbulent-flow simulations in applied science and technology. The need for accurate simulation of turbulent flows is evident across the DOE applied-science and engineering portfolios, including combustion, plasma physics, nuclear-reactor physics, wind energy, and atmospheric science. The workshop brought together experts in turbulent-flow simulation, computational mathematics, and high-performance computing. Building upon previous ASCR workshops on exascale computing, participants defined a research agenda and path forward that will enable scientists and engineers to continually leverage, engage, and direct advances in computational systems on the path to exascale computing.

The realization of exascale simulations of systems with turbulent flow will bring new opportunities in scientific discovery and applied research. Significant advances in application areas include:

- simulation of nonlinear multi-scale, self-organized turbulence in burning plasma, covering a whole fusion chamber,
- simulation of combustion with sufficient physical complexity and resolution in real geometries,
- nuclear reactor simulation to generate experimental-quality simulations of steady-state and transient reactor behavior with extreme-fidelity resolution,
- simulation of boundary-layer turbulence and clouds over large areas, such as continental scales,
- simulation of an entire wind farm under realistic atmospheric flow conditions and in complex terrain where turbine geometry is well resolved.

However, with the many opportunities that exascale computing will bring, there are abundant challenges. The computational simulation of turbulent flows, which are inherently multi-scale, presents a unique set of challenges including, e.g., the necessity for subgrid-turbulence modeling/filtering, global communication, and short simulated time spans due to prohibitively small time steps. We summarize below some of the important take-aways from the workshop:

- *The anticipated evolution of hardware architecture toward exascale will require that current turbulence simulation codes be re-tooled even to maintain current turbulence simulation capabilities.* As the DOE leadership computing facilities transition towards exascale computing, it is important that scientists and engineers engage in that transition and understand the implications to turbulent-flow simulations. First and foremost, power-usage restrictions will change dramatically the makeup of computing architecture, leading to decreases in processor clock rates; at the same time, core counts are increasing, memory hierarchies are getting more complex, and available memory bandwidth per processor core is decreasing. These changes will affect high-performance computing (HPC) on the largest leadership systems down to small clusters.
- *The introduction of exascale computing capabilities will enable only an incremental increase in the Reynolds number accessible through direct numerical simulation (DNS).* In regard to scientific discovery, exascale computational efforts should be directed at flows for which a modest increase in Reynolds number will expose transitions to new flow regimes.

- *High-order numerical methods show promise for next generation platforms, but their advantages over low-order methods need to be demonstrated.* Compared to low-order methods, high-order methods are appealing for two reasons. First, the computation intensity for data in cache is higher and fits well with the idea that "FLOPS are free" in the exascale landscape. Second, fewer degrees of freedom are required for a given accuracy level, which yields smaller model sizes and potentially less data movement. These ideas need rigorous vetting on next-generation platforms.

- *The global linear solver required for implicit and incompressible/low-Mach algorithms is a significant challenge to exposing more parallelism in many turbulence simulations.* This presents a risk in leveraging next-generation platforms, and readers are referred to the recommendations of the Exascale Mathematics Working Group[1] regarding scalable solvers.

- *Weak scaling behavior of large-eddy simulation (LES) models will permit much larger spatial domains at exascale, assuming data I/O can keep up.* An example of where this will bring scientific benefit is understanding cloud organization and propagation over the Central United States, which is an important area of research for improving climate models. However, an increase in spatial-domain size requires a commensurate increase in time-domain size.

- *Uncertainty quantification (UQ) techniques for estimating parameter uncertainty in models rely extensively on the ability to run many permutations of the model.* Exascale will permit running many of these permutations concurrently, which will significantly increase the permissible parameter space that can be evaluated. Importantly, running these simulations as a closely coupled ensemble promises significantly shorter time to solution. There are several approaches for UQ under coupled ensemble calculations that need to be explored and vetted for the different applications.

- *Exascale simulation of problems that are "weakly scaled up" from today's petascale problems will likely take longer to run.* This is because an increase in spatial-domain size requires a commensurate increase in time-domain size, and, for a fixed time step size, the number of time steps required for a simulation will increase. Unless the minimum number of degrees of freedom per thread (for scalability) decreases even faster as the number of available threads increases, there will be a longer time to solution.

- *Strong scaling must be dramatically improved on next generation platforms.* Otherwise, a simulation that runs today on a petascale system will have a longer time to solution on an exascale system. The minimum number of degrees of freedom per thread for scalable solutions must be reduced in order for shorter time to solution on exascale (and petascale) systems. Another avenue to shorter time to solution may lie in new parallel-in-time methods.

- *The transition to exascale computing will in no way reduce the need for improvements in turbulence modeling and filtering as required by LES, hybrid LES/DNS, and hybrid LES/RANS (Reynolds Averaged Navier Stokes).* The nature of turbulent flows makes DNS a wholly unrealistic approach for many applied and scientific problems, and therefore, the smallest turbulence scales will always need to be modeled or filtered. It is important that existing approaches are advanced to best leverage the advances in computing, and that new methods be explored.

[1] Applied Mathematics Research for Exascale Computing, March 2014, Exascale Mathematics Working Group, http://science.energy.gov/~/media/ascr/pdf/research/am/docs/EMWGreport.pdf

Acknowledgments

The Program Committee expresses gratitude to the many contributors to the processes that led to the successful workshop and to the composition of this report. First and foremost, we thank the attendees for taking the time to attend the workshop. We are thankful to Prof. C.-S. Chang, who played a major role in writing Section §3.4, and Dr. Elia Merzari, who wrote §3.5. We thank our plenary speakers Drs. John Bell, Jeff Hittinger, and Jeff Slotnick for giving excellent talks that helped to anchor discussions in breakout sessions, and whose slides provided material to this report. Very important was the support from our volunteer breakout-session leaders: Jon Baltzer, Ramanan Sankaran, Elia Merzari, David Womble, C.-S. Chang, Ramesh Balakrishna, Igor Bolotnov, Karen Pao, and Jeff Hittinger. We thank Kathryn Ruckman from NREL for her editing suggestions. Finally, we thank the efforts of Deneise Terry and Jody Crisp from ORAU for organizing so many of the administrative details necessary to hold the workshop.

1 Introduction

1.1 Background and motivation

To better understand the opportunities and challenges in computational fluid dynamics (CFD) and turbulent flow simulation that will come with exascale computing, the U.S. Department of Energy (DOE) Office of Advance Scientific Computing Research (ASCR) held a workshop in August 2015 that brought together experts in computational fluid dynamics, computational mathematics, and high-performance computing—*Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop* [6]. Turbulent flows are a critical component for many science and applied programs, broadly and for programs important to the DOE mission. The presence of turbulence increases mixing rates, so in any field where the design is sensitive to mixing in a fluid, the design is likewise sensitive to the smallest scales of turbulent motion.

In this workshop, participants were drawn to consider the relationship between analysis of turbulence and computing in key areas that represent the broad space where turbulence is of critical importance to the mission of the DOE: combustion, atmospheric science, wind energy, fusion, and nuclear energy. These application areas all have performance-limiting aspects that require simulation of turbulent flows, as well as commonalities in terms of the underlying character of the mathematical equations. These areas also have distinguishing features, and a one-size-fits-all description is prone to missing subtleties that are of paramount importance. Combustion science underpins the efficiency of conversion of energy from chemical storage in fuels to thermal and mechanical energy; in this area, the turbulent mixing limits not only the device performance in concert with the limits of material science, but also the resulting pollution emissions. In atmospheric science, improvements in the fundamental understanding of the planetary boundary layer are necessary to understand its interaction with, and contribution to, the formation of clouds as well as transport from the surface up into the free troposphere. Optimizing the performance of wind farms requires understanding and optimizing the turbulent flow within and around wind farms, including turbine-turbine wake interaction, siting to account for complex terrain effects, and wind-plant-wind-plant interaction. Realizing the vision of safe and abundant energy from fusion energy has current scientific attention focused on two exemplar problems, both of which involve understanding and controlling the onset of turbulence arising from instabilities. On one hand, controlling hydrodynamic instabilities driven by density differences are critical to successful laser-induced plasma ignition, and on the other hand understanding the roles of temperature gradients and plasma density in triggering instabilities that lead to confinement degradation appear to be a key concern in magnetic fusion reactors. Finally, all traditional and advanced nuclear reactor concepts rely on fluid flow to cool and transport heat from the reactor core to energy conversion systems that ultimately deliver the useful energy, so analysis of the heat transfer rates in these flows is critical to safe and efficient designs. A commonality to be explored in the following sections is that for an incremental increase in computational power, all of these research areas must choose between increasing dynamic range, increasing the fidelity with which the largest or smallest scales of motion are represented, or increasing the fidelity of the models for the physical processes at play. The distinctions between the applications arise from differences between which facets of the problem limit accuracy and relevance of the tractable computational problem to the ultimate problem of interest.

The computational simulation of turbulent flows is a difficult thing. While the Navier-Stokes (NS) equations constitute a broadly accepted mathematical model, the computational solution of those equations, i.e., CFD, is especially challenging for turbulent flows due to the chaotic and inherently multi-scale nature of turbulence. In a chaotic system such as turbulence, although the system remains deterministic, the smallest scales impact the largest scales, and small changes to

boundary conditions, initial conditions, or grid resolution, for example, can have a dramatic impact on large-scale responses. Analytical solutions to the NS equations have only been found for the simplest of problems in the absence of turbulence, and proving the existence and smoothness of NS solutions remains one of the greatest unsolved problems in physics [46]. Figure 1 shows example simulation results that hint at that multi-scale nature in energy-related applications. A particular challenge is the nonlinear cascade of turbulent energy from large eddy scales to the small dissipation scales (i.e., the Kolmogorov microscales), at which turbulent energy is converted to heat. These scales are typically separated by many orders of magnitude. Consider Figure 2, which shows an example energy spectrum as a function of wave number for a turbulent flow. While energy is being injected on the left side of the spectrum (at large spatial scales), that energy is converted to smaller scales (via the breakup of larger eddies into smaller ones) down to the Kolmogorov scale. A simulation that captures/resolves all of these motion scales (e.g., as described by the NS equations) is called direct numerical simulation (DNS). Except for the simplest of problems, CFD of turbulent flows require high-performance computing (HPC). However, even with today's state-of-the-art algorithms, codes, and petascale computing systems, DNS is feasible for only a small class of problems, namely those at moderate Reynolds number (Re, a ratio of inertial forces to viscous forces) and in simple geometries. For example, the largest to-date isotropic-turbulence simulation in a cube domain with periodic boundary conditions employed over 68 billion grid points, but reached only $Re \approx 5 \times 10^4$ [45], whereas turbulent flow problems of applied and scientific interest can have Re values greater by many orders of magnitude.

Given that DNS is limited in the range of scales that it can capture and that in many cases the large-scale flow features are of significant interest, engineers often turn to turbulence models. These approximations, which account for the effects of turbulent motions not directly resolved in the computation, allow for a more accurate representation of the problem geometry at the large scale and focus available dynamic range on the larger scales. In a design setting, the shift in modeling approximations from the large scales to the small may result in an increased relevance of the solution obtained despite the increased errors due to approximating the smallest scales. The most common turbulence modeling approaches are Reynolds Averaged Navier Stokes (RANS), large-eddy simulation (LES), and hybrid-RANS/LES approaches like detached-eddy simulation (DES). In a "classical" RANS approach, the flow is decomposed to solve for the stationary mean flow. In an unsteady RANS (URANS) approach, "slow" variations in the mean flow are captured. In LES, the equations are filtered to solve for the resolved part of the flow. For flow around bodies, boundary-layer eddies can become so small that LES with sufficient resolution near the boundary (i.e., wall-resolved LES) can be impractically expensive. DES addresses that issue by using a RANS flow model near the boundary that transitions to an LES model away from the boundary. Figure 3 illustrates RANS, URANS, and DES simulation results for flow around a cylinder, where one can see the increasing fidelity-capturing capabilities. A general argument in favor of moving modeling from the large to the small scales is that the small-scale flow features are more universal, resulting in a broader applicability in efforts to develop effective models than large scales, which tend to be highly problem dependent. As we move towards exascale computing, it is important to consider the role of turbulence modeling, and to understand where new opportunities might exist for DNS, LES, and DES.

The U.S. DOE is committed to advancing high-performance computing (HPC) simulation capabilities to the exascale regime in order to solve the grand-challenge science problems important to U.S. energy and national security. Exascale computing systems, i.e., those capable of at least 10^{18} floating-point operations per second (FLOPS) or an exaFLOPS, are planned to come online in 2023 (see, e.g., Ref. [2]). The anticipated hardware changes in next-generation platforms (NGPs), will, in many cases, require changes in the algorithms embodied in today's HPC simulation soft-

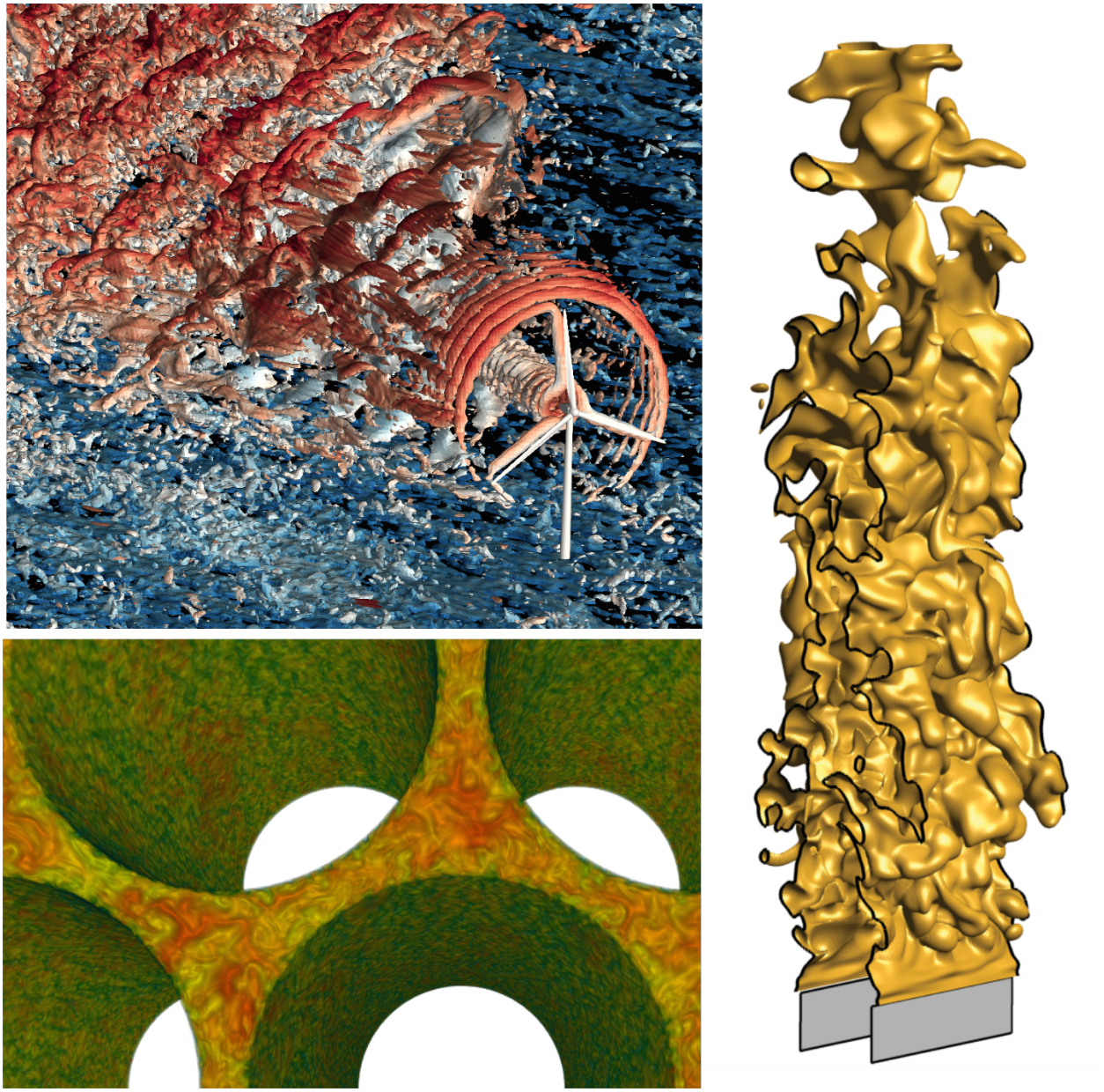


Figure 1: Representative CFD-simulation results emphasizing the complex, multi-scale nature of turbulent flows that makes predictive computational simulations so challenging: (top left) isosurfaces of vorticity colored by the instantaneous streamwise velocity in the wake of a wind turbine [19], (lower left) volume rendering of the velocity magnitude in a nuclear reactor, (right) visualization of the flame surface in the DNS of a lean premixed turbulent methane-air flame in a slot burner by Chen et al. [17].

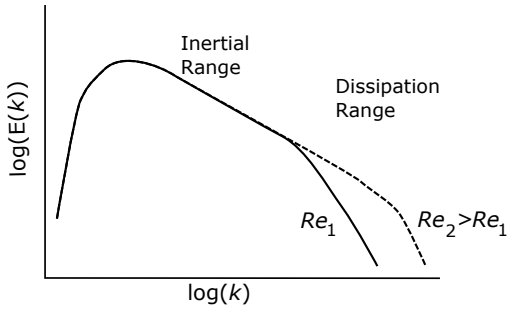


Figure 2: A symbolic description of a turbulent kinetic energy spectrum as a function of wave number. The spectrum can span many orders of magnitude, and, as Re increases, that span increases.

ware. Building upon previous ASCR workshops on exascale computing, workshop participants were tasked with defining a path forward that enables scientists and engineers to continually leverage, engage, and direct advances in computational systems on the path to exascale computing. Workshop priorities were to understand the new scientific and applied problems that will be solved on exascale systems, how today's codes will need to adapt to scale to exaFLOPS, and what new algorithms and software might be enabled by exascale systems. With respect to those priorities, we describe in this report the most important outcomes of the turbulent flow simulation workshop.

1.2 Previous exascale workshops and the NSCI

The U.S. commitment to exascale computing is perhaps best evidenced by the number of workshops and reports directed at exascale computing, including the DOE Scientific Grand Challenges Workshops that were convened in the 2008-2010 time frame [23–30], the Advanced Scientific Computing Advisory Committee (ASCAC) report [22] on "The Opportunities and Challenges of Exascale Computing," and the 2014 Workshop on Exascale Applied Mathematics [31]. The importance of exascale computing is also called out by the Executive Order establishing the National Strategic Computing Initiative [4] (NSCI), which is a "whole-of-government effort designed to create a cohesive, multi-agency strategic vision and Federal investment strategy executed in collaboration with industry and academia, to maximize the benefits of HPC for the United States." The NSCI Executive Order calls out five specific objectives [4]:

- Accelerating delivery of a capable exascale computing system that integrates hardware and software capability to deliver approximately 100 times the performance of current 10 petaFLOPS systems across a range of applications representing government needs.
- Increasing coherence between the technology base used for modeling and simulation and that used for data analytic computing.
- Establishing, over the next 15 years, a viable path forward for future HPC systems even after the limits of current semiconductor technology are reached (the "post-Moore's Law era").
- Increasing the capacity and capability of an enduring national HPC ecosystem by employing a holistic approach that addresses relevant factors such as networking technology, workflow, downward scaling, foundational algorithms and software, accessibility, and workforce development.

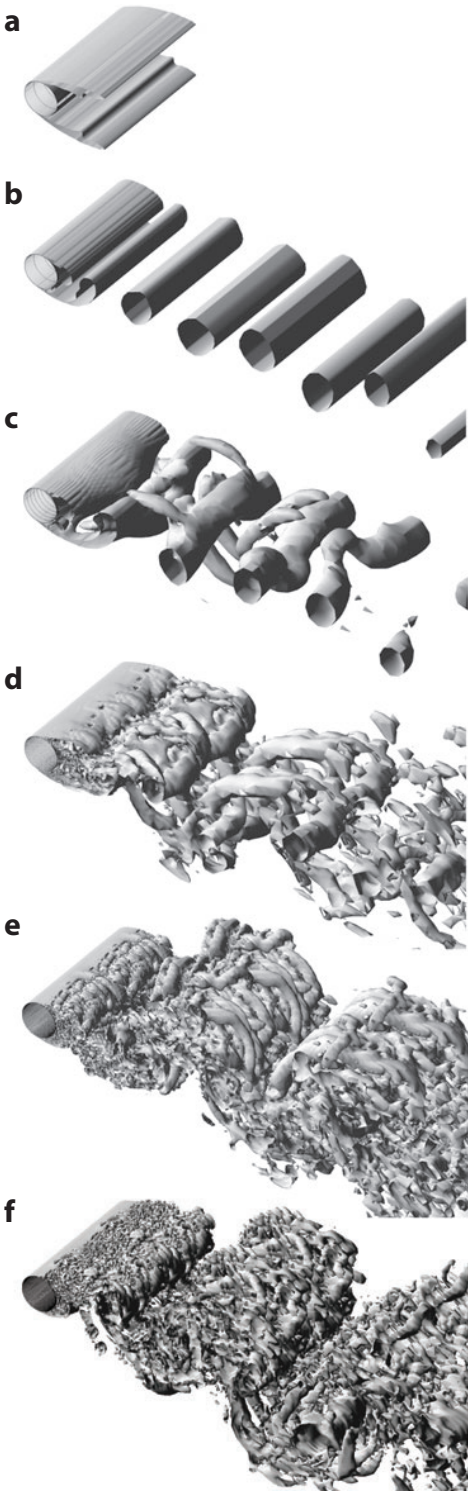


Figure 3: Simulation results (vorticity isosurfaces) for flow over a cylinder, where turbulence is modeled as (a) shear-stress-transport (SST) steady RANS, (b) 2D SST URANS, (c) 3D SST URANS, Spalart-Allmaras DES on a (d) coarse grid and a (e) fine grid, and (f) SST DES on a fine grid (figure courtesy of A. Travin, Boeing Company; also published in [83]).

- Developing an enduring public-private collaboration to ensure that the benefits of the research and development advances are, to the greatest extent, shared between the United States government and industrial and academic sectors.

1.3 Expectations around exascale systems

The transition from today's petascale-class leadership systems (e.g., Edison, Titan, Mira) to tomorrow's exascale systems is expected to be disruptive to the scientific computing community. Perhaps most significant are the changes to hardware architecture that will be, necessarily, dramatically different due to power constraints. For example, as described in the 2010 ASCAC report [22],

> Based on current technology, scaling today's systems to an exaFLOPS level would consume more than a *gigawatt* of power, roughly the output of Hoover Dam.

In order to prepare for exascale systems, it is important to understand as best we can how those systems will be designed and the challenges those systems will present. Table 1 shows how the characteristics of leadership-class machines have evolved since 2000 and provides a notional exascale system. Note that, compared to Sequoia, an exascale system will have about an order of magnitude more processors, but each processor will have about 1,000 cores. While there will be huge gains in processing power, memory gains will lag, which implies a shift to a "FLOPS for free" paradigm, where memory issues are the limiting factors to scalability. The following points are adapted from the report for the 2014 Workshop on Exascale Applied Mathematics [31]; these points describe the key architectural changes expected in order to build an exascale machine:

> **Electrical power:** Power is the driving force behind the changes in supercomputer architecture. In some sense, exascale computing should really be thought of more as "low-power, high-performance computing." To continue to design supercomputers using standard commodity technologies is not sustainable; the power requirements of such a machine rapidly become prohibitive [55]. The goal has therefore been set to achieve exaFLOPS performance with a power limit of 20 MW [84]. This restriction has direct implications for the structure and organization of the hardware components as well as algorithms. It is conceivable that the energy used by a simulation may replace the CPU time as the cost metric for supercomputer use; hence, numerical algorithms may need to become more power aware.
>
> **Extreme concurrency:** From hand-held devices to supercomputers, processor clock speeds have stagnated because of power-density limitations. Instead, increased node performance is being obtained by increasing the number of processing elements on a chip (multiple cores) and supporting threading. It is estimated that exascale machines will have two to three orders of magnitude of parallelism over petascale-computer levels, with much greater parallelism on nodes than is available today. Intra-node concurrency, which is being used to cover memory latency, and performance variability, arising from hardware thermal fluctuations and elsewhere, will challenge the bulk-synchronous execution models that dominate today's parallel applications. In such an environment, strategies that reduce synchronization and communication without sacrificing algorithmic optimality will be advantageous. Dynamically scheduled task parallelism may help, but will introduce a new challenge, reproducibility, that will make determination of code correctness more difficult.
>
> **Limited memory:** Without a significant change in technology, memory density is not expected to increase at the same rate as the number of processing units. Again, power

is a limiting factor; current volatile RAM technology, for example, consumes a great deal of power to maintain its state. Therefore, while the amount of memory per node will increase, the amount of memory per core will decrease. Many current algorithms will thus be memory constrained and will need to be redesigned to minimize memory usage.

Data locality: Similarly, memory bandwidth is not expected to increase at the same rate as the number of processing units. Consequently, on-node memory bandwidth will increase, but the bandwidth per core will actually decrease. Interconnect transfer rates are also not expected to increase at the same rate as the number of cores. In addition, the energy used for a double-precision floating-point operation is expected to decrease by roughly an order of magnitude, which will expose differences in the energy cost not just of off-chip data motion but of on-chip transfers as well. Future systems may use a variety of different memory technologies including nonvolatile memory, stacked memory, scratchpad memory, processor-in-memory, and deep cache hierarchies to try to ameliorate some of these challenges. Algorithms will need to be more aware of data locality and seek to minimize data motion, since this will be a more significant energy cost than computation.

Resilience: Because of the shear number of components, hardware failures are expected to increase on exascale computers. Traditional checkpoint-restart recovery mechanisms are too expensive in terms of both the time and energy with bulk synchronization and I/O with the file system. Such global recoveries could conceivably take more time than the mean time between failures. Local recovery mechanisms are required that leverage the mathematical properties of the algorithms in the application. In addition, efforts to reduce power by computing with lower threshold voltages and other environmental disturbances may lead to more soft errors that may not be caught by the hardware.

These characteristics will force changes throughout the software stack in ways that cannot be completely hidden from the application and its associated numerical algorithms.

Table 1: Evolution of leadership-class systems since 2000 and a prediction of an exascale system (adapted from a workshop plenary talk [43]).

	ASCI Red	Road Runner	K Computer	Sequoia	Exascale System
Year	2000	2008	2011	2012	2023
Peak (FLOPS)	1.3e12	1.7e15	11.3e15	20.1e15	1.2e18
Linpack (FLOPS)	1.0e12	1.0e15	10.5e15	16.3e15	1.0e18
Total Cores	9,298	130,464	705,024	1,572,864	1e9
Processors	9,298	12,960(6,912)	88,128	98,304	1e6
Cores/Proc	1	9(2)	8	16	1e3
Power (MW)	0.85	2.35	9.89	7.9	20

1.4 Workshop structure

In order to engage the community in defining the vision for turbulent flow simulations at the exascale, the *Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop* was

held on August 4–5, 2015, at the Autograph Mayflower Hotel, Washington, D.C. About sixty people attended from national labs, academia, and industry (see the Workshop Participants appendix). The workshop consisted of four plenary talks (talk titles and abstracts are in the Plenary Talks appendix) and four breakout sessions. The breakout sessions and plenary talks were structured to address five framing questions:

1. What is the potential impact of exascale simulations of turbulent flow on our fundamental understanding of turbulence?

   Most of the recent progress in fundamental understanding of turbulence has come from simulations that push down to the smallest length scales in the flow. In some cases, molecular-level effects have been incorporated. This is an extremely computationally intensive approach. The potential for discovery science in turbulence using increased computing power should be examined in detail.

2. What are the potential impacts on DOE Applied Technology programs (Wind Energy, Nuclear Energy, Stockpile Stewardship)?

   While simulations aimed at discovery science in turbulence are generally done using approaches such as direct numerical simulation, engineering simulations in turbulent flows are performed using LES and RANS simulations. Therefore, the benefits, and potential impacts, of improved simulation capabilities in these areas should be considered separately.

3. What are the potential impacts of exascale simulations that include improved turbulent flow simulations on problems of scientific interest to the Department of Energy?

   The potential impacts of increased simulation capability in turbulent flows for climate, fusion, and other DOE Office of Science problems is likely to differ from those of applied programs, and should be considered separately.

4. What are the challenges in implementing existing turbulent flow simulations at the exascale?

   The architecture changes created by exascale computing change the relative cost of operations. Memory and chip-to-chip communication capabilities are limited relative to floating point operations. This will create new challenges for implementing turbulent flow solvers that may be particularly relevant in a multi-scale problem such as turbulent flow.

5. What are the opportunities for new turbulence simulation approaches that may be enabled by exascale capabilities and architectures?

   In some areas, the FLOPS are free paradigm created by exascale may lead to new computational approaches, and new capabilities. The correct solution approach will not always be a modified version of a petascale code. We propose examining this question for turbulent fluid mechanics.

Given the broad range of DOE science and technology areas that require effective simulation of turbulent flows, a workshop on turbulent fluid mechanics at the exascale can be expected to have a far-ranging impact. The potential impacts are particularly large in applied programs such as wind energy that are making increasingly ambitious use of high-performance computing. These programs will want to move to exascale platforms as they become available, and they will feel the impacts of exascale-like architectures in their high-performance codes. However, it is not clear that lessons from the two existing exascale centers will readily translate to resolving these issues, which may require simulation across a broad range of length scales using a range of modeling approaches.

1.5 White papers

For all individuals invited to the workshop, white papers were solicited that address one or more of the workshop's framing questions listed in §1.4. Nineteen white papers were received, which are listed in the White Papers appendix. Electronic copies of the white papers were made available at the workshop website.

1.6 Report goals and outline

The goal of this workshop report is to ensure that the relevant stakeholders, e.g., DOE program managers, and scientists and engineers from national laboratories and universities, understand how the transition to exascale computing will impact turbulent flow simulations. Impacts as *challenges* could include, e.g., rendering existing codes obsolete without major restructuring. Impacts as *opportunities* could include predictive simulations for grand challenge problems now inaccessible with petascale machines, for example. It is important that the stakeholders understand that the transition to exascale computing will be disruptive, and we must now prepare our codes and algorithms for that transition. Further, it is important to establish realistic expectations regarding what we will be able to accomplish with exascale computing. It is our hope that this report paints an accurate picture of the exascale-computing landscape for opportunities in turbulent flow simulations, and that it brings to the forefront the particular challenges that will be faced.

The report follows the workshop structure. Chapter 2 focuses on the impacts of exascale computing on turbulent flows of scientific interest, whereas Chapter 3 focuses on five applied areas: combustion, atmospheric turbulence, wind energy, fusion, and nuclear energy. Chapter 4 describes the opportunities and challenges in adapting today's codes (that run at the petascale) to future exascale systems, and Chapter 5 describes new algorithms and codes that may be enabled by exascale systems. Chapter 6 summarizes conclusions and recommendations from the workshop.

2 Exascale computing for scientific problems of DOE interest and fundamental understanding of turbulence

2.1 Background and history

Since its inception in the early 70s [35, 75], the numerical simulation of turbulence, using direct and large-eddy simulation, has been used as a powerful tool for the fundamental study of turbulence and transition. Turbulence simulation has provided the information required to make many advances in our understanding of turbulence. Examples of these advances include: identification of intense vortex tubes with diameters that scale with the Kolmogorov scale as the dominant small-scale vortical structure [50]; discovery of an autonomous dynamical mechanism that sustains near-wall turbulence [49] and characterization of the scaling and non-uniqueness of turbulent wakes [69]. In addition, turbulence simulations have been used to validate and improve our theoretical understanding of turbulence [32], and to develop and evaluate turbulence control techniques. For summaries of many other contributions of turbulence simulation to the science of turbulence, see reviews such as [66]. The utility of turbulence simulation in fundamental turbulence research is due to two unique characteristics [66]. First, turbulence simulations provide access to the complete three-dimensional time-dependent fluctuating fields (velocity, pressure, temperature, etc.), which allows any quantity of interest to be determined from the simulation, including those that would be difficult or impossible to determine experimentally. Through the analysis of the solutions, turbulence simulations provide access to any mathematically well-defined diagnostic quantity without the limitations im-

posed by the capabilities of experimental instruments. With such simulations, researchers probing the nature and dynamics of turbulence are limited only by the insight and ingenuity they bring to the problem. Second, turbulence simulation gives the researcher complete control over the turbulence, through initial and boundary conditions, and manipulation of the equations being solved. This allows non-physical numerical "experiments" to be performed to test scientific hypotheses or proposed techniques for manipulating turbulence. Such numerical experiments have been crucial to the advances in near-wall dynamics, wake scaling and non-uniqueness, and turbulence control discussed above. In short, these two uses of turbulence simulations allow researchers to precisely determine *what* the characteristics of turbulence are, and *why* they are that way.

It is possible to make reliable scientific inferences from the direct numerical simulation (DNS) of turbulence precisely because the mathematical model expressed by the Navier-Stokes (NS) equations is such a reliable description of the dynamics of most flows of simple fluids. The NS modeling assumptions of Newtonian viscous stress (compressible or incompressible NS), Fourier heat conduction and ideal gas thermodynamics (compressible NS) are known to be excellent characterizations of many real fluids, which are valid at all scales of turbulence, down to the Kolmogorov scale and smaller [68]. The primary limitations of DNS as a scientific instrument are that computational costs constrain both the Reynolds number and complexity of the turbulent flow that can be simulated. Advances in computational capacity over the past 40 years have allowed DNS to be performed for flows of ever increasing Reynolds number and ever increasing complexity, resulting in ever increasing scientific impact, and the advance to extreme-scale computing will continue this trend.

Many turbulent flow applications in the DOE portfolio will not be accessible to DNS even with exascale resources, or for the foreseeable future (see sections 3.2, 3.3 for examples). In such cases, the Reynolds number limitations of DNS can be relieved by using large-eddy simulation (LES), in which additional modeling assumptions are introduced to represent the smallest scales of the turbulence. These subgrid models are much less reliable than the Newtonian viscous stress model, calling into question the reliability of LES simulations. Nonetheless, LES can also be used to make valid scientific inferences, provided that great care is taken to ensure that the conclusions drawn using LES are not affected by the modeling assumptions. This will generally require sensitivity analysis and/or uncertainty quantification.

The simulation of more complicated flow phenomena, such as turbulent combustion, plasma turbulence, and multi-phase turbulence also involves additional modeling assumptions which may not have as reliable a pedigree as Newtonian viscous stress (e.g., chemical kinetics models in turbulent combustion). In this case, DNS is still of great scientific value because it allows the interaction of turbulence with other phenomena to be studied, without introducing the additional and often highly unreliable assumptions inherent to turbulence models. See section 3 for further discussion of simulation of such complex turbulent flows. As with LES, sensitivity analysis and/or uncertainty quantification will generally be needed to ensure the conclusions are not affected by the modeling assumptions.

2.2 Challenges to address in turbulence research

Turbulence is a ubiquitous fluid flow phenomenon that often has profound effects on the flows in which it occurs. Understanding and predicting these effects is the primary objective of scientific research in turbulence. However, the effects of turbulence are multitudinous, depending on the character of the fluid flows in which it is embedded, and what other physical phenomena (e.g., chemical reactions) it is interacting. Discussed below are a number of important challenges in turbulence research that were identified at the workshop, and the impact that extreme-scale computing could have on addressing these challenges.

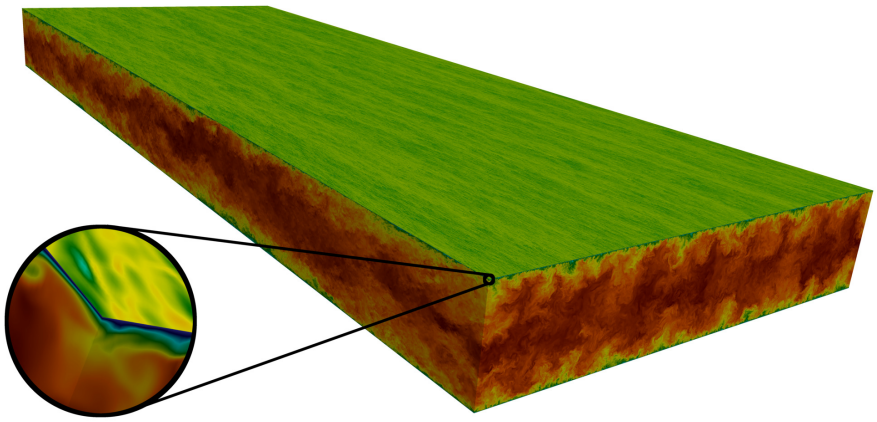


Figure 4: Visualization of the turbulent flow in a planar channel at friction Reynolds number 5200 [57]. Shown is the computation domain, with streamwise velocity visualized on the sides and the streamwise wall shear stress on the top. Apparent is the large difference in the length scales of the turbulent fluctuations at and away from the walls. (image courtesy of Dr. Myoungkyu Lee)

2.2.1 High Reynolds number turbulence

One of the defining challenges in turbulence research is the Reynolds number dependence of the turbulence and its effects. With increasing Reynolds number, the range of turbulence scales increases, with the ratio of the largest to the smallest eddies scaling like Reynolds number to a power between 3/4 and 1, depending on the flow (3/4 is from Kolmogorov scaling for isotropic turbulence). The cost of direct numerical simulation, in which eddies at all length and time scales are resolved, must therefore increase like Reynolds number to a power between three and four. This is important, because in many flows of scientific or technological interest, the Reynolds number is very high. Unfortunately, this cost scaling implies that a 100-fold increase in computational capability that could be achieved with an exascale machine would result in only a three- to five-fold increase in Reynolds number accessible by DNS (depending on the flow). This is an incremental increase in the Reynolds number, adding to the factor of 25 or so increase in DNS Reynolds number capability that has accumulated over the past 30 years.

Nonetheless, a three- to five-fold increase in Reynolds number could be of great significance in flows for which this increase takes the flow into a new regime. An example of this in wall-bounded shear flows is the simulation by Lee and Moser [57], which, with the increased computational capability offered by the Mira system at the Advanced Leadership Computing Facility (ALCF) at Argonne National Labs, was able to reach a high enough Reynolds number to exhibit characteristics of high Reynolds number flow that had not previously been observed in DNS. See figure 4 for an example result from this simulation. One example of a possible Reynolds number regime change that could be attained with exascale computing is the mixing transition in free shear flows. There may also be other possibilities. Attaining DNS data in high Reynolds number regimes is important to inform RANS and LES models applicable at high Reynolds number and for scientific study of high Reynolds number turbulence.

2.2.2 Turbulence with other phenomena

Often the impact of turbulence in a fluid is dominated by the interaction of the turbulence with other phenomena. The resulting turbulent flows can be very difficult to model, either by RANS or LES, because the interaction may occur at small scales and/or strongly affect the dynamics of the turbulence. For this reason, DNS is the preferred vehicle for scientific exploration in these cases. Often in these flows, DNS is the richest source of information about the interaction of turbulence and additional physical phenomena. Development of reliable LES and RANS models for these flows is an important active research topic. Examples that were discussed at the workshop of such complex turbulent phenomena are as follows:

- In turbulent combustion, the turbulence interacts with rapid chemical reactions. Turbulence is affected by the reactions due to large heat releases, and the turbulence impacts the transport of heat and chemical species. In addition to enabling a higher Reynolds number than is currently possible, exascale computing would allow the use of more complex chemistry models (more species) and therefore more complex fuels, simulation at higher pressure, and representation of more realistic flow configurations. This is important because important combustion applications (e.g., internal combustion and gas turbine engines) are at high pressure with complex fuels and in complex flow configurations. Further, new fuels such as bio-derived fuels and synthesis gas (syn-gas) will become increasingly important. DNS will be an important resource for the combustion research needed to support fuel flexibility. See the combustion discussion (section 3.1) for more details.

- High speed (high Mach number) turbulent flows often involve the interaction of turbulence with shocks. Strong shocks at high Reynolds number are generally much thinner than the smallest scales of turbulence, making DNS of such flows more expensive than comparable low speed flows. This is so even if the shocks are numerically thickened by shock-capturing schemes, as is common practice, since they must still be much thinner than the smallest turbulence scales in DNS. Exascale computing can enable such simulations in realistic situations, such as on reentry vehicles, scram jets, and flows arising from Richtmyer-Meshkov instabilities.

- Multi-phase turbulent flows are of two types: those with a dispersed second phase (e.g., particles, see [WP1]), and those with two extensive fluid phases. In the former, the particles interact with the turbulence through the drag forces on the particles. If the particles are much smaller than the smallest scales of turbulence they can be treated as points, and the flow around them need not be explicitly solved, a drag model is used instead, and the primary computational challenge is computing with the large number of particles generally required. If the particles are bigger, so that the details of the flow around them is important to the interaction, there is a much more difficult challenge to resolve these flows around large numbers of moving particles. With two extensive fluid phases, the computational challenge is to represent the interface between the phases, which interacts with the turbulence due to the density difference between phases and surface tension at the interface. Because of these challenges, current capability to perform DNS is greatly limited, and exascale computing can relieve some of these limitations.

- Turbulence in plasmas can be much more complicated than simple Navier-Stokes turbulence or the complications described above. For example, electromagnetic forces generally must be accounted for, plasmas are commonly not in thermodynamic equilibrium, and the plasma may be rarefied in parts of the flow so that a continuum representation is inappropriate (see

section 3.4 and [WP4]). In some cases, turbulence in plasmas is central to the phenomena of primary interest. For example, in magnetically confined plasmas (e.g., a Tokamak), the primary concern is with the confinement of the plasma away from the vessel wall, and turbulent transport at the edge of the plasma can limit the effectiveness of confinement. This interaction of turbulence with edge plasmas is not fully understood, and exascale simulation of the phenomenon may remedy this problem (see [WP4]).

2.2.3 Roughness effects

One of the big challenges associated with wall-bounded turbulent flows is the presence of surface roughness. Currently, roughness is modeled in terms of a single roughness parameter (called the "roughness height"), but there is no direct connection between roughness topography and this parameter, so that the way one determines the roughness height for a rough surface is to measure the effects of the roughness on a turbulent wall layer (see [48] for a recent review). With exascale computational resources, it will be possible to simulate via DNS turbulent flow over rough surfaces, with the roughness geometry resolved, to characterize the effects as a function of topography.

2.2.4 Complex turbulent flows

Turbulence simulation has yielded exceptionally valuable insights into the nature and dynamics of turbulent flows. Until recently such simulations have been confined to simple idealized flow scenarios. Simulations of turbulence, both DNS and LES, in complex flow geometries will be made possible by exascale computational resources. An example of such a complex flow of great technological interest is the turbulent flow in a turbine engine, where the wake turbulence of one row of vanes/blades impinges on those in the next row, while they are rotating relative to each other. Not only are such flows of great engineering importance, there are significant scientific issues that warrant investigation, such as the dynamical processes involved in the interaction of the wake and boundary layers. Reliable simulations of such flows is likely to be at least as valuable for investigating such complex flows as they have been in simple canonical flows.

2.2.5 Knowledge extraction

Another opportunity that exascale computational resources may provide is in advanced data analytics for knowledge extraction from turbulence simulations. It is widely appreciated that the cost of I/O at the exascale will make it imperative that more of the analysis of turbulence simulation results be done *in situ*. With the computational resources available at the exascale, there is an opportunity to do extensive *in situ* data analysis using advanced data analytics (e.g., machine learning or feature extraction) on the space-time structure of the turbulence. One of the major barriers to more effective use of advanced turbulence simulation data for turbulence research is the challenge of extracting knowledge from the huge amounts of data that are generated, and exascale systems may provide an opportunity to address this challenge in a novel way.

2.3 Challenges at exascale

As suggested above, the application of numerical simulation to scientific inquiry in turbulence is primarily through direct numerical simulation (DNS) and large-eddy simulation (LES) of turbulent flows. Several of the well-known issues associated with computational modeling of systems governed by partial-differential equations (PDEs) on anticipated exascale hardware have special consequences in DNS and LES. These are listed briefly below.

- **On-node memory bandwidth**: Limitations on memory bandwidth relative to available computational speed are generally unfavorable for PDE solution algorithms because in such algorithms there are often relatively few floating point operations that are performed with each data item while it is in cache. In DNS, it is generally desirable to use high-order, high-resolution numerical methods, which usually increase the number of operations that can be performed with data in cache, ameliorating to some extent the bandwidth limitation.

- **Inter-node communication**: When the particulars of the problem allow it, spectral numerical methods with global functional representations are usually used in DNS, and sometimes LES, because of their superior accuracy and resolution properties. Interconnect bandwidth limitations and power requirements will make the global transposes used in such algorithms particularly costly. It is often assumed that at exascale such algorithms will not be viable, but the cost of using lower-order numerics with only local communication will be an increase of a factor of 10 or more in the number of degrees of freedom in the simulation, for the same accuracy, with the resulting increased costs.

- **Adaptive numerics**: Adaptive grids are not commonly used in DNS of isotropic turbulence because in such cases resolving turbulent eddies requires fine numerical resolution throughout the turbulent domain. However, with the ability to treat problems of greater richness, including non-trivial large scales and localized spatial features (such as complex geometries, shocks, combustion), effective adaptive solvers for use on exascale hardware are essential for efficient treatment.

- **Resilience**: DNS and LES calculations require that the Navier-Stokes equations be advanced for many time steps, perhaps 10^5 or more, depending on the problem, and increasing with Reynolds number. A large exascale DNS or LES calculation could thus run (cumulatively) for many months on many millions of cores. Such long simulation time increases the probability of a silent hardware error, and because the solution is chaotic, such errors can be very hard to detect in the solution. Good fault detection techniques will thus be particularly important in exascale turbulence simulation.

- **I/O**: In exascale turbulence simulations the specification of the state of the computation will require a very large amount of data. For example the state in the large DNS simulation in [57] required 2 TB, and an analogous exascale calculation might have a 64 TB state. A months-long simulation would generally need to be divided up and performed in smaller pieces (say 12–24 hours a piece). Therefore, a large checkpoint and restart will be needed at least this often. An I/O system that can effectively handle this data volume will be needed.

- **Data analysis**: The cost of I/O and storage of exascale data is driving the use of *in situ* analysis in many problems, and the same will be true for DNS of turbulence and applied LES applications. However, the high computational cost of performing these simulations makes the ability to analyze check-pointed simulations after the fact extremely valuable, so that the same simulation can support many scientific inquiries over years. Fortunately, periodic checkpointing is required as discussed above. The scientific value of performing the DNS will be greatly increased by the ability to store these check-points, and to effectively analyze them.

3 Impacts of turbulent flow simulations at the exascale for DOE applied programs

3.1 Combustion

3.1.1 Background and motivation

Despite increasing deployment and research investment in alternative energy technologies, combustion continues to be of critical importance to our energy infrastructure, and the sheer volume energy converted from chemical to thermal / mechanical forms through combustion makes the aggregate impact of even tiny percentage-wise improvements in efficiency significant. Hydrocarbon-sourced energy is currently the dominant energy source and carrier. For example, natural gas is easily distributed for residential/commercial heating as well as industrial process heat and power generation. It is also relatively secure, with abundant domestic supplies, and has sufficient energy density that local storage to account for demand transients is tractable. Similarly, for the transportation sector, petroleum-based fuels such as diesel, gasoline, and kerosene are well suited because of the energy density. Such fuels currently supply 85% of the energy for the transportation sector with significant built infrastructure. The challenges of conducting this type of simulation are well known; previous workshops have looked at the requirements for industrially relevant computation in the context of internal engines (e.g., the PreCISE workshop in 2011 [5]) and many of the challenges raised there are still relevant.

In this landscape, several opportunities exist where combustion simulations at so far inaccessible scales will be able make outsized contributions to improving energy efficiency, reducing carbon intensity and atmospheric pollution emission, and boosting the economic growth that comes from enabling a technology shift already underway due to resource constraints. Both the automotive (IC engine) industry and gas turbine industry (both aero-propulsion and land-based power generation) are experiencing a shift in combustion technologies driven by the quest for higher efficiency, reducing emissions, and accommodating evolving fuel streams. As this change occurs, new technological challenges arise that require advances in fundamental understanding. For example, promising automotive engine designs in the laboratory with the potential to significantly improve efficiency involve near-homogeneous ignition by compression, resulting in a low-temperature combustion regime that is relatively poorly understood compared to the traditional regimes that have been the subject of a century of study. These technologies, especially reactivity controlled compression ignition (RCCI) [71], may be able to make use of the varying reactivity of new fuels resulting in an opportunity space for the uptake of new bio-derived and low-carbon fuels. On the gas turbine side, the drive for higher efficiency and lower emissions has manifested as a transition to lean, high-dilution combustion as well as the usage non-traditional fuels (e.g., hydrogen enriched fuels such syngas, typically as part CCS system). Traditional design concerns, such as determining the temperature distribution at the outlet of the combustion section, are more challenging to address in this environment because existing models for turbulence/chemistry interaction are based on more traditional combustion modes. New design concerns, such as stability, flashback, and efficiency at part load operation to offset supply side variability from renewable sources require new understanding that can be advanced through detailed simulation.

3.1.2 Mathematical models and numerical methods

Combustion problems are typically addressed with the full range of algorithms used for general purpose CFD, tailored to the particular concerns of turbulent reacting flows. The general problem is a multi-physics problem that is multi-scale in both time and space. While most of the models

employed can address a wide range of problems, they are typically chosen to exploit the particular features of the problem at hand for a more efficient solution. The specialization of codes for low-Mach vs. fully compressible simulation is a prime example of this.

Method-of-lines based finite-difference (e.g., S3D [17]), finite volume (e.g., LMC [20], RAPTOR [74]), and finite element (e.g., Nek5000 [3]) are all used in combustion research codes. Most implementations are either higher order, use mesh adaptivity, or both. High-order ($\geq$ 6th) structured approaches tend to be used more for fundamental research in canonical domains whereas the lower order (2nd, 4th) with adaptivity tend to address problems in non-trivial geometries, incorporated through both cut-cell and generalized curvilinear coordinates approaches. When geometry is treated with a cut-cell approach, adaptivity is useful to refine the grid to resolve boundary features. Both h- and r- refinement are used, although experience has shown that h-refinement is more suited to dynamic boundaries. For time discretization, the desire to preserve accuracy favors explicit algorithms for the flow field. Although industrial combustion simulation relies heavily on URANS turbulence models, petascale combustion simulations are typically research tools where either DNS or high-fidelity LES are appropriate and common. In DNS, no turbulence model is employed; in high-fidelity LES, the resolution requirement of exacting closures (e.g., [WP7]) can be met.

The multi-physics nature of combustion requires models for chemical reaction, diffusive transport, and, depending on the problem, non-ideal gas effects, spray (2 phase flow, droplet tracking), and radiative heat transfer. Chemical reactions appear in the form of a reaction network based on assembly of elementary reactions (from measurements, rate rules, QMD calculations) that provides a statistical description of chemical changes. Evaluating chemical rates is typically a dominant part of the computational cost: the Arrhenius model for the elementary reaction rates involves evaluating exponential functions of temperature, assembling the network typically involves sufficient computational state to put pressure on register and near cache, and the reaction set involves significant stiffness. The time-integration approach taken varies. For compressible DNS calculations, it is generally the acoustic timescale that limits the time step. In these situations explicit integration of the chemistry is used to reduce the cost per time step and facilitate high temporal order of accuracy. For low-Mach approaches, the relaxation of the flow constraint on the time step to the advective timescale justifies the higher setup cost per time step to be able to take larger chemical time steps. Implicit methods/backward differentiation (using, e.g., SUNDIALS/CVODE [42]) are useful to advance the ODEs resulting from considering only the rate of change due to chemical reaction; coupling of the physics through operator split formulations or alternate multi-rate strategies that preserve higher temporal order [10, 73] are necessary.

Petascale calculations generally employ reduced chemical mechanisms (e.g., [61]) for which direct solves of the linear system in the BDF formulation are efficient. With very large chemical mechanisms, adaptive preconditioning [62] can lead to a significant performance improvement when iterative methods are used for the implicit integration step.

The combustion community is beginning to experiment with UQ (e.g., the work of Braman et al. [12], Morrison [67], Khalil [54], and others); performing ensembles of predictive calculations, particularly in work of direct interest to industry, is an emerging need. Similarly, on-the-fly local analysis is an emerging technology to optimize models with multiple parameters and to accumulate statistics for events of interest to fundamental research activities.

3.1.3 State of the art in computational simulation

The scientific combustion community is experienced at using petascale resources and can fill the largest currently available machines with a single simulation; for research-oriented codes, running on 100k MPI ranks is routine. Science codes have kept up with Moore's law in a weak scaling

sense at 5-10% of maximum FLOPS, with memory bandwidth putting a limit on maximum local performance. The ability to use current generation machines is facilitated by the heavy local computation to evaluate reaction rates. Accessible problems reach up to 6–7 billion grid points with chemistry networks involving $\mathcal{O}(100)$ species. Reduced mechanisms are essential to address the costs associated with transporting a large number of species. A constraint on the complexity of the underlying mechanism that is more difficult to address is the underlying spatial stiffness; realistic liquid fuel surrogates or reasonable pure components introduce very fine spatial length scales that are a limiting factor (down to fractions of a micron), occupying a small fraction typically of a domain with relevant space and time horizons in internal combustion engines that are of order 10s of cm and 10s of ms. For this reason, adaptive grids that focus resolution to best resolve the flame are essential. Even so, DNS is currently restricted to canonical subsets of the problem. Example calculations are found in the work of Chen et al. [37, 94], while high-fidelity LES [74] (resolving down to c. 10x the DNS resolution requirement) is tractable for engine combustion using simplistic chemistry models and leadership computing. LES of realistic geometries is possible with much coarser resolutions, while the workhorse of the design cycle remains anchored in URANS to maintain acceptable turnaround [80]. In design-cycle calculations, single-cylinder simulations are generally coupled loosely to a 0D/1D acoustic model to capture multi-cylinder effects to create engine operating maps that can feed into fleet-level analysis with appropriate perturbations to capture transient operation [7, 36].

Heterogeneous computing (e.g., GPU-accelerated) has been addressed in several ways. OpenACC implementations of both entire codes and computationally intensive kernels have been developed to utilize GPU computing. For more careful control of scheduling, some groups have explored dynamic task-based programming models that enable use of both host and accelerator and analysis of data-flow graphs to maximize the benefit of asynchronous communication between host and device. Implementations for static grid codes exist, and there is research activity on using these programming models for dynamic grids. Commercial codes typically lag in scalability, with current codes routinely running up to $\mathcal{O}(1000)$ MPI ranks. In part, this is a result of the lower resolutions appropriate for lower-fidelity (e.g., RANS, mixing-length LES) simulations that move the problem into the strong scaling regime, but also due to a lower traditional emphasis on scalable algorithms and implementations. Commercial codes have made steps to use accelerators to offload computationally intensive chemistry updates.

3.1.4 Opportunities and challenges for exascale computing

There are multiple opportunities for exascale computing to impact combustion research at both the basic and applied levels. As with many areas of turbulence research, increasing the Reynolds number and range of scales that can be simulated will advance discovery science and increase the relevance of the conditions that can be treated with high-fidelity approaches from a turbulence closure standpoint. However, as high-fidelity combustion simulations move into ever more realistic turbulence regimes, the addition of new physics and new approaches using the data for model development are exciting drivers for discovery. As simulation moves into new flow regimes with increasing resolution the underlying models need to be reconsidered for appropriateness. For example, the turbulence-chemistry subgrid model for relatively coarse LES may not remain valid as the grid is refined to high-fidelity LES, and the physical models (e.g., mass action kinetics at elevated pressures) may need development.

Many of the additional physical processes beyond the flow relevant to combustion (local transport, reaction, non-ideal property evaluation) are particularly amenable to exascale computing because they are intrinsically local in nature. Others, such as soot, sprays, long wavelength acous-

tics, and radiation are non-local and may require a different solution paradigm than the flow. The resulting 'basket of algorithms' with different locality and computation/communication properties presents an opportunity to design a comprehensive solver that schedules the various facets of the communication to balance the available computational resources.

One avenue for utilizing exascale computing power in a transformative way is to incorporate multi-scale modeling inline. While combustion problems tend to have spatially varying resolution requirements that are readily addressed by AMR [9], there is also a tendency to have spatially varying requirements for model fidelity. In some regions, it may be advantageous to incorporate micro-scale models to compute transport coefficients on the fly, or higher complexity/fidelity reaction mechanisms. Such approaches would alter the local computational character in terms of FLOPS/memory bandwidth. Another manifestation of using a hierarchy of model fidelities would be to embed DNS in an LES calculation, or the reverse. Broadly, the gap between DNS and LES is closing and it may be feasible to couple them, potentially within an AMR framework where the refinement is in a physical model rather than only increased resolution. LES can provide useful boundary conditions for DNS, and DNS could be embedded in LES to treat phenomena such as soot formation, shocks, and combustion where existing models are uncertain. Similarly, space-time localized UQ in regions of interest could be used to guide development of the appropriate local model for chemistry and transport parameters to discover what is the appropriate fidelity for each model. The combination of increased resolution, Reynolds number, geometry, and comprehensive physics moves high-fidelity simulations ever closer to industrially relevant configurations. As this happens, there is an opportunity for data-driven model discovery of engineering closures and fundamental science advances. Autonomic closures [WP7], which yield a model-free subgrid representation for LES, are one example.

3.1.5 Barriers to progress

As with other applications, the level of effort needed to establish and maintain performance portability on emerging architectures is of concern. Performance portability implies a low-impact approach (insofar as expressing physics/algorithms is concerned) that manages load balance and minimizes communication at several layers. Between nodes, work needs to be assigned to handle potentially conflicting requirements to balance the computational requirements of different physics. For example, reaction is more expensive to evaluate when the stiffness in the mechanism is excited, Lagrangian and Eulerian components often have conflicting requirements due to non-uniform particle loading, and AMR codes have dynamic load balancing requirements. Within nodes, a potential mix of heterogeneous and homogeneous / many-core systems along with multi-level memory hierarchies require expressing parallelism in a way that maps naturally onto the architecture. Finally, at the finest level of granularity, one needs to be able to effectively use vector operations. Traditionally, combustion codes decompose spatially for all three levels; this may need to be mixed with some level of task based decomposition as well as optimizations for cache utilization such as tiling to make effective use of available hardware caches and map effectively onto the node.

Already, the combustion community is increasingly using *in situ* processing for analysis of the results to reduce the volume of data saved; as simulations become increasingly complex and costly, this will likely increase in importance. When doing analysis *in situ*, it is advantageous to build community consensus about what should be measured beforehand, while the community is accustomed to analysis of large datasets at will for individual aspects of interests. A forum to facilitate such discussion and planning beyond the PIs that will execute the runs would add significant value. This is also an opportunity space, where carefully crafted specific scientific questions vetted by the community provide a concrete measure of success for exascale simulations.

3.2 Atmospheric turbulence

3.2.1 Background and motivation

Atmospheric LES modeling is a fundamentally expensive computational task due to the small timesteps required within the LES compared to weather and climate models. As such, aspirations for LES use have often been limited by available computing resources. Ultimately, the time-to-solution is typically the bottleneck due to the large number of timesteps needed to do simulations over multiple days and longer time periods. Early LES modeling of the planetary boundary layer in the 1970's used domains on the order of 5 km with grid spacings around 150 m [21]. While useful, these early models were limited to the convective boundary layer and could not simulate the stable boundary layer due to Reynolds number limitations. As computing power and model sophistication have increased, modern atmospheric LES modeling is routinely used for domains on the order of 10s of km across with some researchers using domains the size of the Netherlands [78], albeit with simplified representations of clouds and radiation. Larger, faster computers have also enabled the use of higher Reynolds numbers and finer resolution, which has enabled simulation of some aspects of stable boundary layers. However, it is still impractical to simulate entire weather systems using LES over long time periods, which is what will be required for understanding some of the most vexing problems for cloud parameterization related to storm propagation, organization, and inter-cloud interactions.

Atmospheric turbulence plays a critical role for many aspects of both basic research and applied areas within DOE. For example, LES is used within the DOE ASR program to improve fundamental understanding of the planetary boundary layer, its interaction with and contribution to the formation of clouds, and tracer transport from the surface up into the free troposphere. LES is also critical for developing and evaluating parameterizations used in climate models of these same processes. These parameterizations are then used within the DOE ESM and RGCM programs for improving climate models, improving understanding of the larger earth system, and ultimately making better estimates of the impact of human activity on future climate. Additionally, the DOE ARM facility is in the process of developing an LES modeling workflow that will begin routinely simulating the atmosphere over the ARM Southern Great Plains megasite in Oklahoma using LES modeling combined with the detailed atmospheric measurements at the site [38, 39].

3.2.2 Mathematical models and numerical methods

Atmospheric LES models use a discretized representation of the atmosphere based on the Navier Stokes equations. A common assumption made in the model dynamical cores is the anelastic assumption, e.g., with the System for Atmospheric Modeling (SAM) and the Dutch Atmospheric LES (DALES) models, although the fully compressible system is also used in some LES models, such as the Weather Research and Forecasting (WRF) model. The anelastic approach is more efficient to solve because the pressure becomes a diagnostic variable; however, this introduces uncertainty and can cause difficulty when using a large domain where the pressure can deviate farther from the assumed basic state. Solving the fully compressible set of equations is more difficult due to the presence of sound waves. However, these can be handled by splitting the acoustic and gravity modes. This comes at the cost of the pressure becoming a prognostic variable.

LES requires an appropriate subgrid-scale model (SGS) to handle turbulence smaller than the resolved grid. For atmospheric LES, a SGS that assumes a constant Smagorinsky coefficient is often used due to its cost efficiency. However, this assumption does not always work well. Dynamic approaches to adjusting the coefficient can increase accuracy but are computationally expensive and the appropriate methodology is a research topic. Regions of complex wall interactions, such as

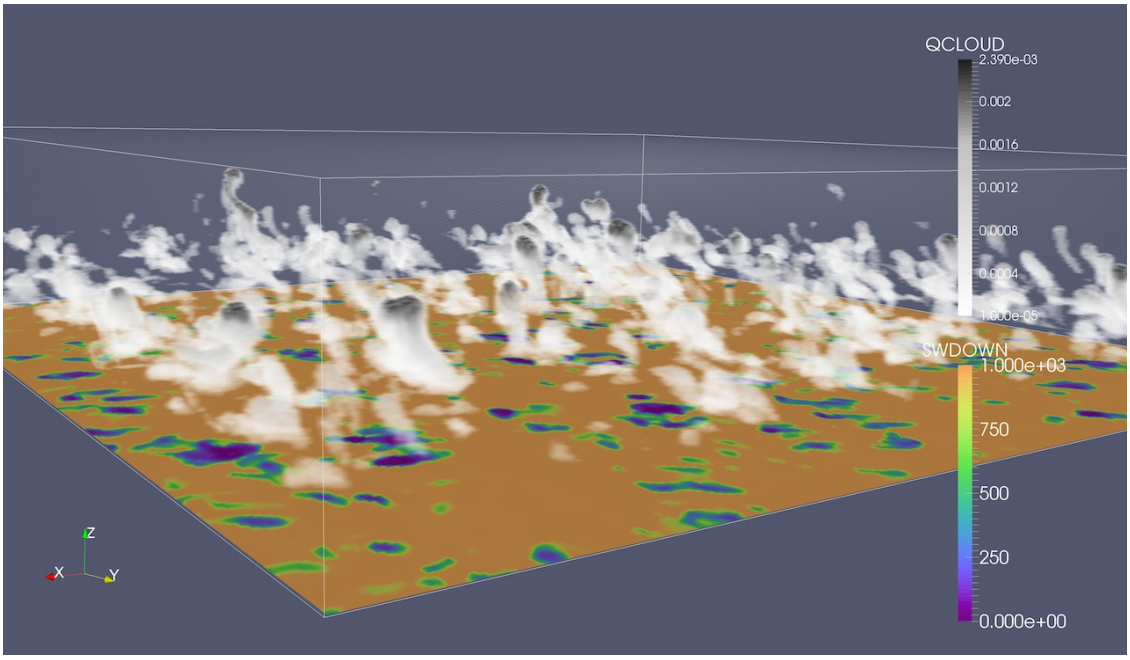


Figure 5: LES is commonly used to simulate clouds and the planetary boundary layer. Shown here are cloud water content (QCLOUD, kg kg^{-1}) and the resulting shadows that impact the sunlight reaching the ground (downwelling shortwave radiation, SWDOWN, W m^{-2}).

the earth's surface, make this more difficult [8, 64].

Coupled with the dynamical core of the atmospheric LES models are physics parameterizations to handle radiation, clouds, and surface interactions. These are critical components of the model, particularly when it is used for process studies, such as understanding cloud behavior. Aspects of the physics parameterizations are commonly the limiting factors when applying LES for improving GCMs, and are thus important for obtaining useful results. For example, LES-simulated clouds are highly sensitive to the parameterization of cloud microphysics, which handles the phase conversion of water between vapor, liquid, and ice. Detailed handling of microphysics can be done using spectral bin microphysics techniques, which explicitly predict size bins of cloud droplets. But, this has often been considered too expensive for LES, and researchers typically use bulk microphysics parameterizations. Both methods suffer from deficiencies in handling of cloud ice, which through interactions with dynamics results in too strong vertical updrafts within clouds [89]. This adds to biases in supersaturation calculations, which complicates using LES for multi-phase cloud simulations.

3.2.3 State of the art in computational simulation

The current state-of-the-art for atmospheric LES modeling lies in the pursuit of simulations over ever larger spatial scales in combination with increasing physical complexity. These two categories are often at odds in terms of available resources and, therefore, require a compromise based on particular research needs. In a weak scaling sense, the ability to use more processing power with larger spatial domains permits modelers to take advantage of new computers by using bigger domains, which bodes well for exascale. However, higher resolutions are not always possible because of their requirement to use shorter time steps for the model integration. This ultimately becomes the lim-

iting factor when balancing domain configurations against time-to-solution requirements. Added complexity also adds computational cost, which does not always scale well with added processors.

Global LES is the ultimate goal for atmospheric modelers so they can accurately calculate the full energy cascade from global to turbulent-eddy scales. The hope is that this will enable both a more detailed dynamical understanding of the Earth's atmosphere combined with a better understanding of the interacting processes within the atmosphere, such as clouds and radiation. The closest attempt to-date for a global LES simulation has been produced by Miyamoto et al. [65], who were able to achieve a grid spacing of 0.87 km for a global domain with the Nonhydrostatic Icosahedral Atmospheric Model (NICAM). While this resolution is not LES in the truest sense, this is the highest-resolution, realistic global simulation available. Their analysis of cloud characteristics reveals a distinct change in the nature of the simulated clouds when transitioning from 2-km grid spacing to finer grids, pointing to the value of these high-resolution simulations and motivating even higher resolutions. The computation cost to produce a one-day simulation using this grid is 36,800 PFlops, which they obtained from the 10 PFlop K computer in Japan. Extrapolation of their results point toward the ability to roughly halve the model grid spacing to 400 m on an EFlop machine versus the current K computer [85].

An alternative to a true global LES, as pursued by Miyamoto et al., is the use of a multi-scale modeling framework (MMF) approach that replaces the convective parameterization within a GCM with many high-resolution models [90]. In a typical MMF configuration, each GCM grid column contains its own cloud-resolving model with a grid spacing on the order of several kilometers. The embedded model can be viewed as a highly detailed cloud parameterization that is better able to capture the true physical behavior within the GCM column. MMF increases the overall computational cost of the GCM by two orders of magnitude. However, because all these embedded high-resolution models are independent, the model scales extremely well and can effectively use a high number of compute cores. Current development is ongoing for pushing this MMF technique to the next level by using LES within each grid column instead of cloud-resolving models, which essentially entails using much higher resolution for the embedded models [77]. The cost of this new Ultraparameterized Community Atmosphere Model (UP-CAM) is potentially 64,000 times a typical GCM if implemented naively by just increasing the resolution in the current off-the-shelf MMF. However, additional software engineering and judicious domain choice for the embedded LES domains brings the estimated cost down by orders of magnitude when UP-CAM is fully developed. The advantage of UP-CAM lies in the ability to more finely resolve temperature and moisture inversions, cloud-top boundaries, and the turbulent mixing where traditional parameterizations struggle to accurately reproduce reality. This model style will benefit greatly from exascale computing where the additional compute resources can be used to increase the size and resolution of the embedded LES models.

At the opposite extreme for pushing the limits of current computing capabilities for turbulent simulations of clouds is the use of spectral bin microphysics (SBM) to represent condensed cloud water. Unlike traditional bulk microphysics parameterizations that track approximately a half dozen cloud condensates, each with an assumed particle size distribution, SBM uses separate size bins to track each condensate type, such as cloud water or ice. The computational consequences are the need to advect many more tracers within the model and an increased load imbalance between cloud-free and cloudy grid cells. A typical SBM configuration uses 33 size bins per condensate type, which increases model run times by roughly an order of magnitude. Because of this, SBM at LES resolution has been rare in the past, but is beginning to be more common [33, 53]. Because typical atmospheric models parallelize across compute cores by decomposing the 3-D grid only in the two horizontal dimensions, the ability for models using SBM to effectively use very large computers is limited by the number of columns within the model. However, if methods can be

found to more efficiently treat tracer advection combined with increased parallelization within the SBM calculations, there is great potential for using exascale computers to simulate the turbulence within clouds to more accurately understand cloud processes.

3.2.4 Opportunities and challenges for exascale computing

With the increased computing capability of exascale computers, atmospheric turbulence modelers will have many new opportunities available to fundamentally address new aspects of current problems by extending simulations into realms currently considered impractical, at best, to impossible, at worst. Increased resolution will enable some improvements, but this incremental increase is not the whole story. The ability to simulate much larger areas will be the most important new capability for understanding boundary layer and cloud processes.

For fundamental research of clouds and the boundary layer, exascale will permit routine simulation of domains the size of the United States at LES resolutions, which will provide the statistics necessary to understand mesoscale organization of cloud systems. This is of particular importance because organized, propagating convection generates about 60% of the midsummer rain in the U.S. Midwest [14], yet both weather and climate models do not faithfully reproduce these systems, which limits the ability to confidently make statements regarding future changes in precipitation in the context of climate change. At the more basic level, the ability to use increased resolution will enable more detailed simulations of atmospheric mixing, which is needed to better understand cloud formation and dissipation. This is a fundamental process that is poorly constrained in climate model cloud parameterizations, and is one of the reasons that climate models exhibit spread in their ability to simulate shallow clouds, which in turn leads to a spread in their future climate predictions. The simple ability to use exascale to do large ensembles of LES will also permit extensive uncertainty quantification and characterization of the parameter space within LES models, which is currently impractical. This will enable model developers to better address weaknesses of atmospheric LES models, which for example, produce too strong updrafts and too much cloud ice [89]. Improving these weaknesses will permit more robust use of LES for increasing process understanding and development of climate model parameterizations.

While atmospheric modelers look forward to the increased resources that will come with exascale computing, they are not so naive as to expect to easily take advantage of the new potential. There are significant barriers to overcome. The most obvious is the fundamental fact that increased resolution requires a shorter timestep to numerically integrate the solution forward in time. This means more timesteps will be required to simulate the same period compared to using coarser resolution. As future computing clock speeds are expected to not increase, and more likely will decrease due to energy constraints, this will be a limiting factor for achievable model resolutions.

The anticipated limited growth in output bandwidth for exascale computing will also be a limiting factor that will require changing current modeling approaches and the culture of atmospheric scientists for the way they do their computational research. By its nature, atmospheric LES modeling is very concerned with time series of the model state, which is used to study the evolution of the atmosphere under various conditions. Currently, modelers meet this need by regularly outputting both metrics representing the atmosphere at each output time (or over a period of time such as since the previous output time) and the full 3-D state of selected variables. If this will no longer be possible, then other means will be necessary to record the model state over time to a sufficient level of detail. It has been suggested that *in situ* analysis can be used for this purpose. However, this does not address the fundamental issue that the evolving state is the important feature that is typically of interest. Possibly even more difficult than the technical issue of how to do appropriate *in situ* analysis during model integration is changing the culture of the atmospheric science

modeling community to accept this limitation. Many currently hope to still be able to archive as much data as possible and pay the extra price of limiting overall code scalability in the process. A strong emphasis on more efficient software libraries for saving data from exascale simulations, along with requisite hardware development, will be of critical importance for atmospheric LES codes to achieve exascale performance.

A second cultural issue that will need to be addressed within the atmospheric modeling community is reproducibility of results. It is presently expected that, given appropriate compiler options, a model can reproduce a simulation bit-for-bit on any given machine. If this cannot be done, the code is assumed to have an error, which is almost always the case, or the compiler or hardware have a bug, which also occasionally happens. Thus, reproducibility has become a critical check for code robustness. If exascale computers cannot achieve this level of reproducibility, the expectations of the community will need to be changed and new tools will need to be developed for testing code. This is particularly important given the nonlinearity of the atmosphere and its sensitivity to small perturbations. For example, if a cloud forms in one grid cell during one simulation but not for a subsequent simulation, the result can cascade into additional changes, which sometimes can change the overall model results. Non-reproducibility will be a particular problem for single deterministic simulations, as well as those whose intent is to look at detailed process-level questions where cause and effect are being evaluated. An increased use of ensembles will be required to address the sensitivity of simulations to the computing architecture, and methods will be needed to separate true signals from noise introduced by the computer.

Reformulation of the atmospheric LES codes will also be needed to make them amenable to exascale computers. Typical numerical techniques used to solve the anelastic set of equations used in many LES models employ FFTs when determining the pressure. This requires global communication to collect and distribute information throughout the domain, which may limit scalability of the code. The highly vectorized environment on these computers will also require reformulation of the codes to efficiently use many-core technologies, whether those be on GPUs, MICs, or other similar technologies that might become available. Parallelization currently is typically done through domain decomposition in the horizontal, which ultimately limits the number of concurrent tasks that can be used. Additional ways to parallelize the code will need to be identified and implemented in ways compatible with the anticipated hardware.

Another numerical issue that will become more prominent at the exascale is load imbalance. Current atmospheric LES models use static load balancing because it has been generally found that the increased communication cost of dynamic load balancing overwhelms any overall benefit. A feature somewhat unique to atmospheric modeling is clouds and radiation that occur within the simulation. Calculations related to these two processes dominate the cost of the model calculation, and to make matters more difficult, one anticipated opportunity allotted by exascale is the ability to use more complex and expensive physical packages. However, the cost of these calculations varies throughout the domain depending on the given meteorological state. For example, if a particular grid cell has low relative humidity, a cloud will not form. Alternatively, a grid cell with pre-exisiting cloud condensate, or a cell that forms cloud during a given timestep, incurs many more computations. This leads to an imbalance in computational cost throughout the domain, which current models typically do not attempt to balance. Development of more effective ways to equalize this load imbalance could improve the model efficiency and its ability to effectively use the computer. Alternatively, grid cells with fewer calculations could have their clock speeds throttled to save energy without greatly impacting the overall time-to-solution since these cells would otherwise end up waiting for the cloudy cells. Methodologies that can *a priori* anticipate the relative cost of grid cells versus the average cost will be of great use for this purpose. That would also allow the models to continue to use a static load balance methodology similar to current approaches.

Ultimately, new atmospheric LES models are needed for exascale computing. DOE is presently funding a major effort to develop a climate model to function on next-generation computing hardware through the Accelerated Climate Model for Energy (ACME) Project. A similar effort to develop the next-generation high-resolution, LES atmospheric model would also be highly valuable. The DALES model developed in the Netherlands is one example of an atmospheric LES model that can run on GPUs, but the U.S. does not have an equivalent model. The DALES model is also simplistic and is unsuitable for many types of process studies that require detailed cloud and radiation calculations. Assumptions made within its dynamical core also make it unsuitable for very large domains. Overcoming these limitations will require investment in a team to develop a new dynamical core and an appropriate physics package designed specifically for exascale-type computers.

3.3 Wind energy

3.3.1 Background and motivation

As described in the 2015 U.S. Department of Energy Wind Vision Report [87], a national objective is to have 20% of U.S. electricity being provided by wind power by 2020, with 30% being provided by 2030. Greater use of the nation's abundant wind resources for electric power generation, reaching 30% of U.S. electrical supply, will have profound societal and economic impact: strengthening U.S. energy security through greater diversity in its energy supply, providing cost-competitive electricity to key regions across the country, reducing greenhouse gas emissions, and reducing water used in thermo-electric power generation. While significant technological advances in blades, gearboxes, and individual turbines have enabled the wind industry to reach its current level of electrical supply, significant challenges still remain to allow increased electricity generated by wind. The infrastructure required to extract sufficient energy will be composed of many large wind farms, each composed of hundreds of multi-megawatt turbines. However, optimized performance of wind farms is elusive due to poorly understood turbulent flow within and around wind farms, including turbine-turbine wake interaction, complex terrain effects, and wind-plant-wind-plant interaction. Figure 6 shows the many complicating flow dynamics that can occur in a wind farm. Wide-scale deployment (high penetration) of wind plants without subsidies requires the reduction of plant-level energy losses, which are currently estimated to be 20% (and much higher in areas of complex terrain), and better quantification of the uncertainty in plant performance. Further, due to turbine-wake interactions, turbines within wind farms (compared to turbines in isolation) experience significantly higher failure rates and 33% higher extreme loads [88]. The layout, operation, and control of wind farms present significant opportunities to reduce the cost of energy. However, the realization of these opportunities is limited with today's simulation tools and our inadequate understanding of wind plant flow physics. High-fidelity predictive turbulent flow simulations provide an obvious path towards reducing the cost of energy produced from wind farms, by providing new understanding of wind plant flow and a foundation for new capabilities in computer-aided engineering.

We describe here the opportunities and challenges for exascale computing and the prediction of wind farm turbulent flows. The material in this section was drawn from the workshop's working session 2, "Impacts of turbulent flow simulations at the exascale for DOE applied programs (wind)," and the five workshop white papers that had a wind-energy component: [WP3, WP5, WP8, WP10, WP15]. It is worth noting the Atmosphere to electrons (A2e) initiative [1, WP15], which was started in 2015 by the U.S. Department of Energy under the Wind and Water Program Technology Office (WWPTO). At the heart of A2e is the goal to reduce the cost of energy produced by large wind plants that comprise tens to hundreds of wind turbines by using high-fidelity computational

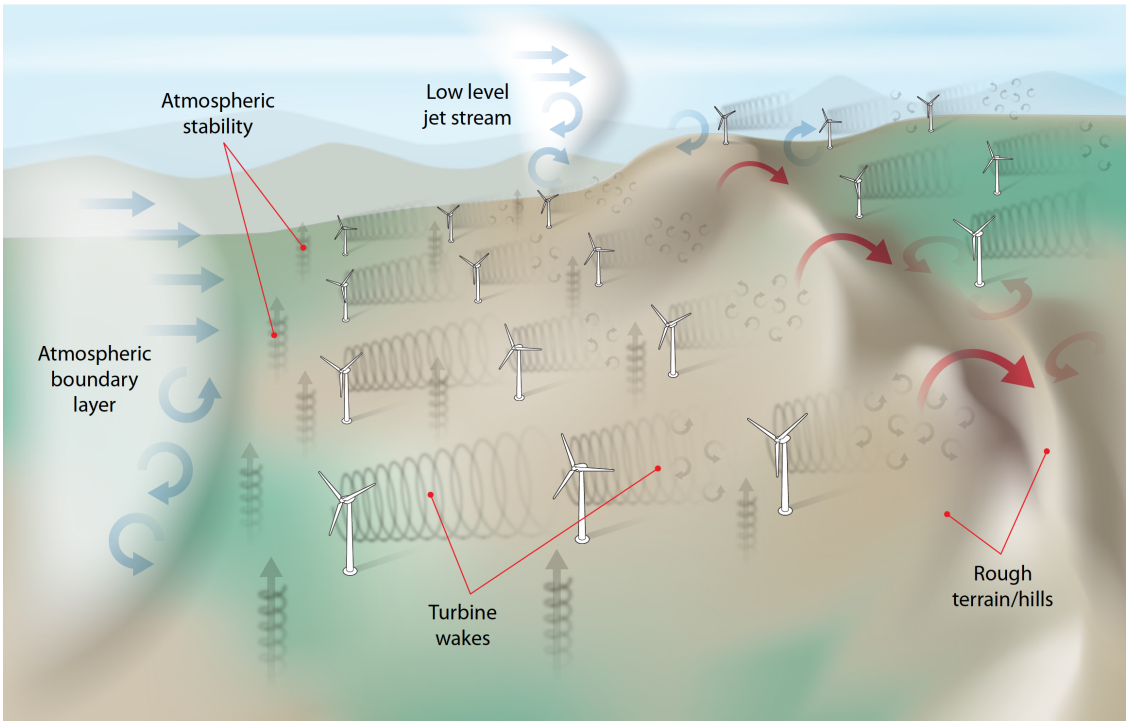


Figure 6: Schematic illustrating the sources that complicate the flow in wind farms, including atmospheric boundary layer dynamics, turbine-wake and wake-wake interaction, and complex terrain. Our inability to predict with confidence wind farm flows presents a barrier to reducing the cost of wind energy. (Illustration: Al Hicks, NREL)

modeling in conjunction with a validation-focused experimental campaign. Over half of the 21 wind-working-session participants also attended at least one of the two 2015 A2e strategic planning meetings on high fidelity modeling. Ideas and points in the working-session discussion align with those in the A2e strategic planning meetings, and supporting material was drawn from the A2e meetings' report [41].

3.3.2 Mathematical models and numerical methods

Wind plant flows are highly turbulent and they span a tremendous range of spatial and temporal scales. Looking at a whole wind plant, which can cover a hundred square kilometers, there is a daunting range of spatial scales, going from the blade boundary layer at $O(10^{-5})$ m up to regional weather at $O(10^{5})$ m [WP5]. We focus here on the "micro-scale domain," which is taken as the "box" encompassing the wind farm. The "meso-scale domain" is taken as the larger domain required to capture atmospheric turbulence and regional weather flows; meso-scale flow simulations (i.e., numerical weather prediction) are described in detail in §3.2. Of course, physically relevant simulations in the micro-scale domain depend on its interaction with, and forcing from, the larger meso-scale domain. This is known as the meso-scale micro-scale coupling (MMC) problem (see, e.g., [70]), which is described further below.

Wind farm flows have relatively small Mach numbers. Regardless of turbine size, the maximum Mach number is found at blade tips and is no larger than about 0.3, which is an upper bound due primarily to aeroacoustics constraints. As such, it is accepted that, for wind farm simulations, flows will be well described by solutions to either the compressible (e.g., [79]) or incompressible/low-Mach

(e.g., [13]) Navier-Stokes equations [41]. For wind-farm-scale simulations, models should include atmospheric boundary layer (ABL) effects including buoyancy and the planetary-rotation Coriolis force. The stability of the ABL has a strong influence on the dynamics of the ABL turbulence, which then affects the wind turbine wake evolution. For example, in a stably stratified flow, there is a significant vertical shear across the turbine rotor, which is largely absent in an unstable ABL.

Wind energy simulations are only feasible with turbulence being modeled to some level – DNS of wind energy flows is wholly impractical. At the lower-fidelity end of the turbulence-modeling spectrum, Reynolds-Averaged-Navier-Stokes (RANS) approaches are incapable of predicting the inherently unsteady and multi-scale wind farm flows. At the higher-fidelity end of the spectrum is wall-resolved large-eddy simulation (WRLES). In simulations where the blade surface is resolved (as opposed to an actuator-line treatment described below), the small scales of the boundary layer at the blade surface make WRLES resolution requirements also impractical. The boundary-layer region necessitates either a local RANS treatment (in the context of a larger RANS/LES or Detached Eddy Simulation (DES) approach) or wall-modeled LES (see, e.g., [WP3]). Addressing the transition zone between RANS and LES is an active research area, as is wall modeling, especially in complex terrain.

As described above, meso-scale micro-scale coupling deals with the interaction between the micro-scale wind-farm "box" and the larger meso-scale. Many in the wind energy community consider one-way coupling, where a meso-scale model drives a micro-scale model, as being of prime importance [41] in the near term; we maintain that view here. While important questions exist around the influence of large wind farms on regional weather and the interaction between large wind farms, those questions will likely be addressed with numerical-weather prediction codes where wind farms are parameterized (see, e.g., [34]). That said, performing a simulation where the micro-scale-domain CFD model is two-way coupled to a meso-scale-domain code will be a tremendous challenge, with unresolved questions around model coupling. Two-way coupling will be necessary for the most physically accurate micro-scale simulations, and will create the foundation for creating new meso-scale wind farm parameterization. Focusing on one-way coupling, meso-scale forcing for a "finite" micro-scale wind farm simulation is typically handled in one of three ways:

1. A synthetic turbulences is superimposed on an appropriate mean flow.

2. A "precursor" simulation that is the same as the micro-scale wind farm simulation, but without turbines and with periodic boundary conditions in the horizontal directions, is run to statistically steady state (see, e.g., [18, 93] and Figure 7). Those flow and temperature data are extracted as an initial condition and flow boundary conditions for the wind-farm simulation. While this approach is well suited for studying wind farm flows in canonical ABL conditions, it is not clear how to extend the method to unsteady weather events or complex terrain.

3. A numerical-weather-prediction (NWP) simulation is performed (or field measurements are gathered) and flow and temperature data are extracted as boundary and initial conditions for the the micro-scale simulation (see, e.g., [79, 93]). This approach faces several challenges. The NWP grid will typically be much coarser than the micro-scale simulation grid and will have a much different turbulence model. Further, finding an effective method for "spinning up" the turbulence in the micro-scale domain is an active research topic (see, e.g., [70, 72, 92]).

Alternatively, one may simulate an "infinite" windfarm as in Meneveau et al. [13], where the fluid box has periodic boundary conditions in the horizontal directions.

Individual turbines within a wind farm simulation have been modeled with varying success at several levels of fidelity. Accurate simulation requires a model of the turbine itself, including, e.g.,

Figure 7: (a) Aerial view of the Lillgrund wind farm, and (b) simulation results showing velocity magnitude. Results were calculated with the SOWFA tool. [18]

structural dynamics, pitching/yaw, and control system dynamics. There are several whole-turbine models, including the open source whole-turbine model FAST [51, 52]. The most recent version of FAST, version 8, includes a high-fidelity finite-element model appropriate for modeling large elastic deformations of modern, flexible blades [91]. We focus here on the aerodynamics of the blade. The following are three common approaches, listed by increasing level of fidelity:

1. Actuator Disc. In the actuator disc approach, i.e., the drag disc approach, a total thrust force acts on fluid based on the turbine thrust coefficient and the induction at the rotor. This is a simplified approach appropriate for examining large-scale effects, but it is incapable of capturing blade-scale tip-vortex dynamics. An example of wind farm simulations with actuator discs can be found in [13].

2. Actuator Line. The actuator-line approach, proposed by Sørensen and Shen [81], represents the wind turbine blade as a line force in the fluid domain. Based on local fluid velocities, fluid forces (drag and lift coefficients) are calculated from two-dimensional (2-D) airfoil data (typically a look-up table). The forces are typically corrected for rotational effects and tip loss. Examples of wind farm simulations with actuator lines can be found in [18, 60, 93]. The actuator-line approach is appealing for its balance between simplicity, efficiency, and fidelity. Wind farm simulation can be performed on stationary structured grids, and blade-surface boundary layers do not need be resolved. Further, the actuator line can be coupled to deflections predicted by a structural model. However, the reliance on 2-D airfoil data and the representation of a turbine blade as a line force are clearly significant approximations.

3. Body-Conforming Mesh. In this approach, the discretized fluid domain conforms to all or some of the turbine, i.e., blades, hub, nacelle, and tower. While this is arguably a straightforward approach, the associated mesh-resolution requirements can be extreme in order to capture well the blade boundary layer. Further, the meshes must accommodate motion due to rotation of the rotor (hub and blades), yaw rotation of the nacelle-rotor, pitching of individual blades, and large elastic deflections of blades and the tower. These large motions must be accounted for, and several approaches exist, including sliding meshes (e.g., [44]), overset meshes (e.g., [79]), and immersed-boundary methods.

3.3.3 State of the art in computational simulation

With the daunting span of relevant scales and physics in wind energy, high-fidelity computational simulation of a single wind turbine (beyond a single blade) and wind farms has only become possible

in the last decade. Single-turbine simulations include those with actuator-line blade representation and LES turbulence modeling [76, 86], and blade-resolved simulations with RANS [95], variational multiscale (VMS) [44], and DES [58, 59] turbulence models. The highest fidelity wind plant simulations to date have been accomplished with LES and actuator-line turbine representation [18, 60, 93]. However, wind plant simulations where the blades are resolved are becoming possible, as in the simualations of Sitaraman et al. [79], which employed DES turbulence modeling. Table 2 lists the codes being exercised for wind energy computing that were discussed at the workshop. While two of the codes solve the fully compressible Navier-Stokes equations, most solve the incompressible/low-Mach NS equations. A wide range of spatial discretization methods are employed in both high-order and low-order configurations.

Table 2: Summary of the codes used for wind energy that were discussed at the workshop. Nomenclature: CNS - compressible Navier-Stokes, INS - incompressible/low-Mach Navier-Stokes, FD - finite difference, FE - finite element, FV - finite volume, CVFEM - control-volume finite element, DES - detached-eddy simulation, RANS - Reynolds Averaged Navier Stokes, LES - large eddy simulation, ILES - implicit LES, PISO - pressure implicit splitting operation,

Code	Flow Model	Spatial Disc.	Temporal Disc.	Turbulence Model	License
Helios See, e.g., [79]	CNS	Overset FD	Implicit	DES	Closed
Nalu	INS	FV & CVFEM	Implicit	DES/LES	Open-Source[1]
Nek5000 See, e.g., [76]	INS	spectral FE	Split op.	Spec. filtering LES	Open-Source[2]
OpenFOAM See, e.g., [18]	INS	FV	Split-Op. PISO	RANS,LES	Open-Source[3]
Phasta	INS CNS	FE	Implicit	LES	Open-Source[4]
Rex See, e.g., [59]	INS	Struct. overset FD/FV	Split-Op. PISO	RANS,DES	Closed
SHARP See, e.g., [96]	CNS	Spectral difference	Explicit, Implicit	ILES	Closed
VWiS See, e.g., [93]	INS	FD	Split-op.	LES	Closed

[1]github.com/spdomin/Nalu, [2]nek5000.mcs.anl.gov, [3]www.openfoam.org, [4]github.com/PHASTA

3.3.4 Opportunities and challenges for exascale computing

Looking forward, a wind plant simulation capability that is truly predictive will have the following features [41]:

- Blade structural dynamics model that includes complicated composite structure and large, nonlinear deflections that can address, e.g., bend-twist coupling
- Blade/nacelle/tower conforming fluid meshes that deform with large deflections
- Overset/sliding fluid mesh capabilities that accommodate the rotor rotation, and nacelle yaw

- Fluid meshes and models that accommodate complex terrain
- Hybrid LES/RANS turbulence modeling, where LES captures the dynamics of wakes and RANS captures sufficiently the boundary layer at the blade surface
- Coupling of mean flow and subgrid turbulence from the meso-scale via numerical weather prediction or experimental measurements.

Simulations will require extensive storage (models with $O(10^9)$ to $O(10^{11})$ grid points will be common), *in situ* data-analysis capabilities, and tools for uncertainty quantification. A simulation for a single wind turbine that encompasses the above features and is sufficiently resolved in space and time is a petascale-class computing problem [WP5]. As such, scaling up a single-turbine model to a wind plant that comprises hundreds of turbines is a weak scaling of the single-turbine model, which makes wind plant simulation very well suited for exascale computing.

To effectively utilize extreme-scale systems, the end-to-end modeling and simulation system will have to be reworked. The scale of future architectures and challenges associated with I/O, fault-tolerance, etc., will make the current sequential workflow paradigm of create-run-output-analyze impossible. There is a need for *in situ* analysis, in particular, and an integrated approach starting with problem setup (meshing) through analytics. Known challenges include mesh generation at scale, in situ visualization, and uncertainty quantification algorithms that efficiently exploit future machine architectures. In-situ data analytics with the explicit goals of enabling scientific discovery and improving plant-scale design and controls will need to be incorporated into the end-to-end workflow.

Workshop participants voiced the need for exascale-class hero calculations as one necessary component in understanding and predicting wind plant flows. However, participants also noted that the other necessary components are uncertainty quantification (UQ) and sensitivity analysis (SA). The vision for UQ/SA in the exascale regime is of tightly coupled ensembles of petascale class simulations. Integrated calculations of uncertainty and sensitivities at the exascale will enable credible prediction of full wind plant performance. Other opportunities include improved reduced-order/lower-fidelity models founded on exascale simulations and ensembles of highly coordinated petascale simulations that intelligently explore the parameter space and improve plant design and operation

3.4 Fusion

3.4.1 Background and motivation

World energy consumption doubled over the last 30 years, currently exceeding 150,000 terawatt-hours per year. The International Energy Agency projects that the energy demand will further increase by 37% by 2040. More than 80% of energy consumed in the world is provided by fossil fuels: oil, natural gas, and coal. Even with the development of low-carbon alternatives such as solar, wind, and nuclear energy, this fraction has seen very little decline. The reliance on non-renewable fuels is not sustainable in the long run. Besides economic benefits, curbing the fraction of fossil fuels is essential for energy security, for reduction of air pollution, and for mitigation of climate change. The 2015 UN Conference on climate change sets the goal of reducing greenhouse gas emissions in order to limit global warming to below 2C. With the current level of fossil fuel consumption and carbon dioxide emission, this goal cannot be met.

Fusion is a process that powers stars like our sun. In star interiors, hydrogen plasma is held and squeezed together by the gravitational force until it reaches very high density and temperatures of hundreds of million degrees. Hydrogen nuclei then ignite the fusion reaction that turns hydrogen

into helium, and releases enormous amount of energy in the process. This source of energy, tamed in a laboratory, can provide most of the needed energy supply for at least millions of years. Fusion is complementary to other alternative energy sources and can even be a dominant energy, with a number of attractive advantages. The hydrogen isotopes deuterium and tritium used in fusion experiments are easily available. Deuterium is abundant in sea water, and tritium can be bred from lithium, whose currently available amount is sufficient for thousands of years of operation. Fusion energy stations will not have carbon dioxide emission, they will be relatively compact in size compared to solar or wind plants, and they will provide low risk of radioactive contamination. Fusion reaction requires precisely tuned conditions, so any malfunction of the plasma confining device will immediately shut off the fusion process. Fusion plants will thus be intrinsically safe, with no risks of blowing up or melting down similar to those existing in fission nuclear plants. They are not vulnerable to terrorist attacks or natural disasters like the Fukushima nuclear accident of 2011.

The basic principles of fusion require confining hot and dense plasma for long enough time. One scheme is the inertial confinement fusion (ICF), where the deuterium-tritium plasma is compressed in a small volume by many converging powerful laser beams. The National Ignition Facility (NIF) in the Lawrence Livermore National Laboratory uses 192 laser beams in order to compress and ignite plasma inside a mm-sized pellet. As the central portion of the plasma is heated to ignition, the thermonuclear fusion burn will then propagate outward. Due to inertia, particles will stay close together before they fly apart, for the time sufficient to maintain the fusion reaction.

The practical implementation of this procedure reveals deviations from the ideal picture. As the pellet gets compressed by the laser beams, it implodes with acceleration. The implosion needs to reach the velocities of 300-400 km/s, and compress the pellet to a low percentage of its original diameter. As the fuel gets accelerated and compressed, an instability develops at the boundary between the fuel and the ablated material. This instability is analogous to the well-known Rayleigh-Taylor instability that develops due to gravitational force when a heavy fluid is placed atop a light one. This hydrodynamic instability tries to destroy the shell of the implosion, leading to development of fluctuations that have detrimental effects. In addition, as the laser beams propagate through a gas, laser-plasma interaction leads to laser-plasma instabilities that can scatter off the laser light, thus causing loss of power and degradation of the implosion symmetry. Laser-plasma interactions also lead to generation of very energetic electrons that can heat the fuel too early, before sufficient force is built up to compress it. As the fuel temperature and pressure increase the compression becomes less efficient. In order to tame the unwanted instabilities, many methods have been employed, including variations of the laser pulse profile and intensity.

Another scheme for plasma fusion is to confine particles by a strong magnetic field, a method principally different from the inertial confinement. The motion of charged particles across the magnetic field lines is inhibited, so magnetic field lines forming toroid-like surfaces can confine a hot deuterium-tritium plasma for sufficiently long time for the fusion reaction to occur. The major magnetic fusion energy effort is concentrated around the $20B-scale international project called the International Thermonuclear Fusion Reactor (ITER), currently under construction in southern France. Its fusion chamber will have the vacuum vessel that is 10 times as large in volume as any previous tokamak. Seven countries including the United States participate in this project. The goal of ITER is to demonstrate, for the first time, scientific and technical feasibility of a sustained fusion reactor, where about 500 MW of fusion thermal power will be produced for 50 MW consumed (Q=10), for about 400 seconds. First deuterium-tritium burn is planned for 2030, with the power scaled up to 500 MW by 2034. The actual power plant, DEMO, will be constructed after the ITER construction as the demonstration fusion reactor. It may achieve the thermal power output of 25,000 MW, with the similar size and principle of operation as ITER.

In a magnetic fusion reactor the plasma is kept very hot (well over 100 million degrees) in the central part of the device, where the fusion reaction is initiated, and it is significantly colder at the outer parts that are close to the surface of the toroidal chamber. The radial gradients of the temperature and the density of the plasma provide the source of free energy for instabilities and turbulence. The plasma turbulence turns out to be the major cause of the confinement degradation.

3.4.2 Mathematical models and numerical methods

A variety of mathematical models are employed in addressing the effects of instabilities and turbulence in fusion plasmas. In the case of inertial confinement fusion, numerical simulations that have predictive capability of implosion are crucial for guiding experiments. The large-scale flows and the large-scale Rayleigh-Taylor instability can be modeled with the aid of compressible magnetohydrodynamic codes, which describe the evolution of the conducting plasma fluid and the evolution of the magnetic field that interacts with the currents flowing in the fluid. At small scales, however, the laser-plasma interactions are most precisely described by the fully kinetic simulations, which model the evolution of the distribution functions of the plasma particles simultaneously with solving the Maxwell equations for the electric and magnetic fields (e.g., particle-in-cell codes). Numerical simulations utilizing the models intermediate between the one-fluid magnetohydrodynamics and the fully kinetic particle description are also employed. These models include continuum MHD two-fluid description, where the electron component and the ion component of the plasma are described as independent fluids, and the hybrid simulations where the electron component is modeled as a continuous fluid, while the ions are described kinetically. Opportunities that may be realized by exascale computing include the more realistic 3-D simulations of whole plasma volumes and broad ranges of spatial and temporal scales. They should explore the nature of unsteady developing non-symmetric 3-D turbulent flow, capture intermittency effects, and quantify uncertainty.

One of the codes used for this purpose is the VPIC code that solves the relativistic Vlasov-Maxwell system of plasma kinetic equations [e.g., 11]. This code has been used for various problems of laser-plasma interactions, as well as for the studies of magnetic turbulence in space and astrophysical plasmas. The code can treat various boundary conditions. The code has been optimized for petascale architectures, and can scale up to 720k cores.

In the case of magnetic confinement, plasma turbulence in a fusion device is an extremely complicated phenomenon. It couples the multi-scale dynamics among microturbulence at gyroradius scale, plasma profile evolution at device-size scale, zonal flow dynamics at intermediate size scale, plasma instabilities at device or intermediate size scale, and atomic physics at the edge of the plasma chamber. The time scales of these phenomena span from microseconds to seconds, with significant overlap among them. A steady-state fusion plasma condition is determined from a nonlinear self-organization among these complicated phenomena. The most significant, and mysterious, nonlinear self-organization process is a spontaneous turbulence bifurcation in the edge when the plasma is given enough heating power. In this process the self-generated plasma flow appears to suppress the plasma turbulence. When this bifurcation happens, the transport in the edge decreases to a negligible level, a steep plasma pedestal forms (called "low to high" [L-H] confinement-mode transition), and the core plasma gets into a fusion burn condition. ITERs $Q = 10$ goal relies upon this H-mode of operation.

Macroscopic plasma dynamics at the scales larger than the ion gyroradius may be described using the equations of magnetohydrodynamics (MHD), which neglect the discrete nature of the plasma medium and treat the ions and the electrons as fluids. MHD-like activities play an essential role in magnetically confined plasmas, and they can affect the performance of the fusion devices. One of the codes used to simulate macroscopic plasma dynamics in TOKAMAK relevant

configurations is the NIMROD code [82], which has been used in studies of laboratory magnetic-fusion concepts. The code solves an initial problem of the extended MHD in both two and three dimensions. The extended MHD equations can include the two-fluid effects (Hall and diamagnetic terms), they can incorporate the finite Larmor radius effects such as gyroviscosity, and generalized Ohm's law. In the 3-D case, the only restriction to the geometry is the requirement to have periodic boundary conditions in at least one direction. In the non-periodic dimensions the high-order spectral finite element discretization is used, while in the third periodic dimension the pseudo-spectral method is used. A semi-implicit operator is used in the time-advance algorithm. At each step the code solves large ill-conditioned sparse matrices, which present challenges for the scalability of the code, and it presents a bottleneck to exascale. Parallel scaling up to 10K cores has been achieved.

At scales comparable to the ion gyroscale, the fluid-like description is not adequate, and the kinetic codes must be used. One of such codes is the GENE code [47] that uses the gyrokinetic description, which is a reduced kinetic description where the fast space-time gyro motion is analytically eliminated. As a result the six-dimensional phase space of the particle distribution function is reduced to a five-dimensional space. This code provides an efficient description of the phenomena that respect the gyrokinetic ordering, such as the processes whose frequency is smaller than the ion gyrofrequency, and the spatial scales are strongly anisotropic with respect to the background magnetic field. Reduction of the dimensionality allows for computational efficiency of several orders of magnitude compared to the full kinetic codes. The scalability to pre-exascale and large memory requirements present a bottleneck. The code can be scaled up to 9K nodes on Titan without using GPUs, with some performance degradation.

As a representative kinetic code we consider here in more detail the first-principles-based code that simulates the whole volume tokamak fusion plasma, the XGC1 gyrokinetic particle-in-cell code [15, 16]. The XGC1 simulation domain extends from the material wall boundary to the magnetic axis across the magnetic separatrix, with the neutral particle recycling at wall, and it covers a broad range of relevant scales, including the H-mode layer. XGC1 uses the fully nonlinear Fokker-Planck collision operator for modeling of the non-thermal equilibrium plasma physics in the edge region of a magnetic fusion reactor [40, 56]. Unlike most of the existing fusion gyrokinetic codes, XGC1 does not assume the scale separation between the background plasma profile and turbulent fluctuations, that is, it does not assume that the perturbation of the distribution function and the resulting turbulence are driven by a fixed fluid background gradient (the so-called perturbed delta-f method). Rather, XGC1 calculates the full kinetic plasma all together (the total-f method). Since XGC1 solves for the total kinetic distribution function instead of a small perturbed part of the distribution function, it requires more computing power than the conventional delta-f methods. The more powerful the computers become, the more physics XGC1 can contain. Even an exascale computing may not be enough to give the full answers, but only the important answers. Figure 8 shows the nonlinear blobby turbulence fluctuation in the edge of a tokamak plasma in a multiscale simulation with gyrokinetic ions, driftkinetic electrons, and neutral particles.

3.4.3 State of the art in computational simulation

The XGC1 gyrokinetic particle-in-cell code designed for simulations of magnetically confined plasma has been used to study various regimes of plasma turbulence and flows in multiple tokamak geometries including ITER. XGC1 scales very well to the full-scale heterogeneous Titan (27 PF theoretical peak, with 299,008 CPU cores and 18,688 GPUs) with workload sharing between GPUs and CPUs [see Fig. 9], and to the near full-scale homogeneous Mira (10 PF theoretical peak with 786,432 CPU cores; at the time of test, 589,824 cores were available). In order to cover the complicated edge geometry, XGC1 uses unstructured triangular grid that approximately follows the magnetic

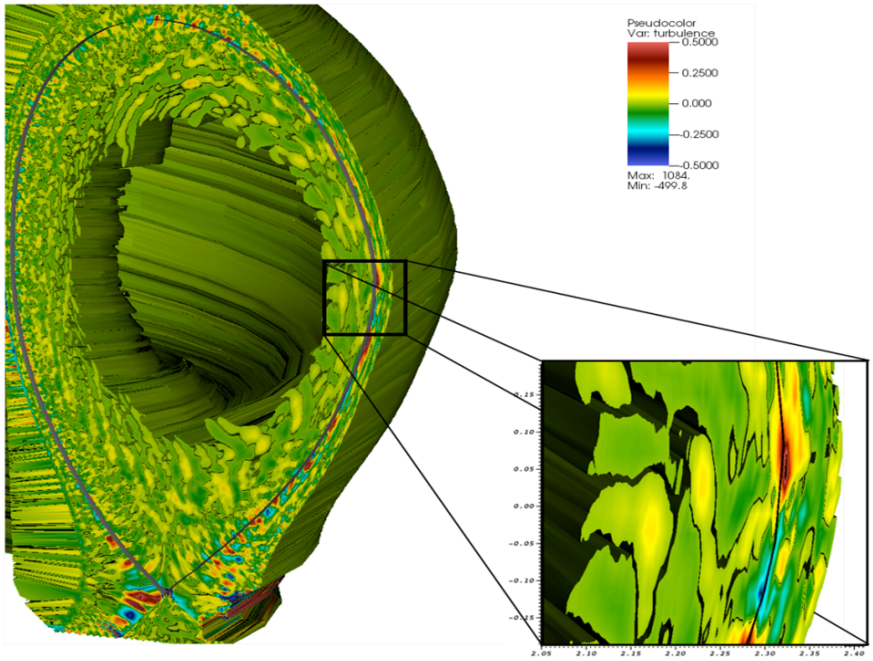


Figure 8: Nonlinear coherent structures called blobs from XGC1 in the edge of a tokamak plasma, regulated by self-generated plasma flows. In the inserted enlarged figure at the right bottom, the flow shearing effect of the turbulent blobs can be seen. The black line in the edge shows the magnetic separatrix surface. From OLCF Featured Highlight in February 2014. (image courtesy of C.S. Chang)

field lines. For production runs, XGC1 uses over 300 billion particles on 88% of the full Titan capability for the simulation of ITER in realistic whole-volume geometry. Unlike other codes, the solver scalability is not a bottleneck in XGC1 since the PETSc solver takes only $\sim 3\%$ of the computing time. The only two bottlenecks seen so far are from the GPU-CPU communication overhead and from the large amount of physics data it generates. The physics data size on the present peta flop LCFs is peta bytes and one check-point file size is about 100TB. However, it is expected that with the HBM, with the in-memory *in situ* data analysis technique, and NVRAMs, these bottlenecks will be eased. XGC1 is in two pre-exascale programs: CAAR at OLCF and NESAP at NERSC.

3.4.4 Opportunities and challenges for exascale computing

On a future exascale machine, XGC1 could include most of the important whole-device fusion physics for predictive simulation of ITER and fusion reactor performance. The opportunities that may be realized by exascale computing include simulation of nonlinear multi-scale self-organized turbulence in burning plasma, and covering the whole fusion chamber, from magnetic axis to material wall. The whole device modeling must also include the edge transport bifurcation physics since ITER must include it to achieve the goal. The modeling should also include the fusion reaction and the generation of alpha particles and neutrons that will convert their energy to heat and turbine. For the most complete first-principles modeling of a magnetic fusion reactor without approximations, it is expected that a 10 exascale computer is needed. Those studies are important to gain high fidelity understanding and prediction of the burning plasma behavior, as it can yield more economical and accelerated achievement of the ITER goal (Q=10), and help realize commercial reactors sooner.

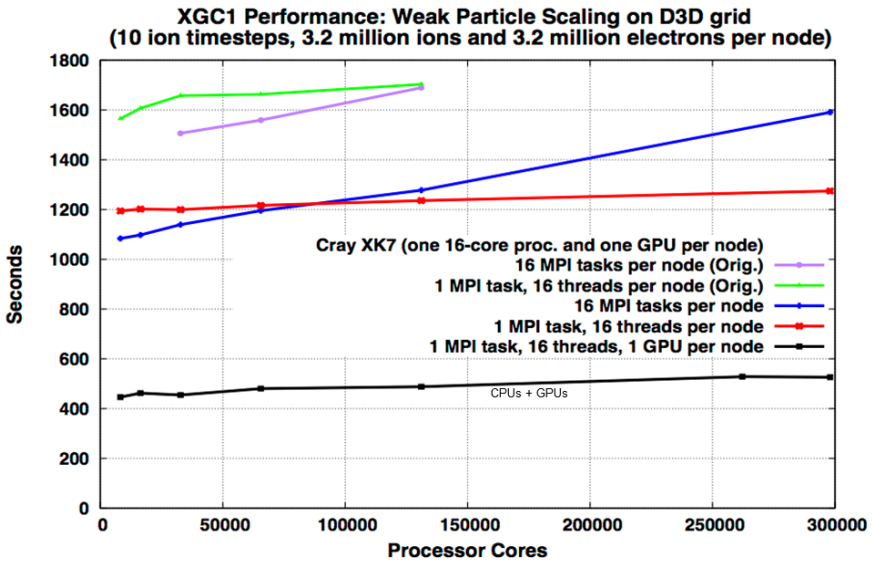


Figure 9: Near perfect weak scalability of XGC1 on Heterogeneous Titan to the maximal core (299,008) and GPU (16,384) counts, which allows the multiscale turbulence simulation of ITER. (image courtesy C.S. Chang)

3.5 Nuclear energy

3.5.1 Background and motivation

Nuclear power promises to be a plentiful, carbon-free energy resource essential to any energy mix. All traditional and advanced reactor concepts rely on fluid flow to cool and transport heat from the reactor core to energy conversion systems. These flows are typically complex, highly turbulent, and involve separation and a multitude of scales. Numerical simulation has been an intrinsic part of nuclear engineering research since its inception. In recent years a transition is occurring toward predictive, first-principle-based tools such as computational fluid dynamics.

3.5.2 Mathematical models and numerical methods

Due to the massive scale separation (Table 3) involved in nuclear reactors, modeling has been limited to simplified one-dimensional analysis in previous decades. Computational fluid dynamics (CFD) is increasingly used to simulate nuclear reactor flows. Most CFD analysis, especially within industry, relies on the Reynolds averaged Navier-Stokes (RANS) approach and traditional two-equation turbulence models. Higher-fidelity approaches to the simulation of turbulence such as wall-resolved large eddy simulation (LES) and direct numerical simulation (DNS) remain to large supercomputing platforms.

3.5.3 State of the art in computational simulation

Petascale architectures are enabling the simulation of physical systems of increasing size and complexity. Current supercomputers have been used to simulate nuclear systems with grids that reach tens of billions of points, enabling the simulation of entire rod bundles (Fig. 1) with wall-resolved

Table 3: Simulation data of four representative cases [63]. U is the bulk velocity; D is a characteristic spatial scale.

Simulation	Points / $(Pin \times 10D)$	Re	Shortest Time Scale	Longest Time Scale
37-pin bare rod bundle	216 million	66,000	$2 \times 10^{-4}\ D/U$	$10^{2}\ D/U$
MASLWR core (59 pins)	1 million	5000	$3 \times 10^{-3}\ D/U$	$10^{6}\ D/U$
19-pin SFR bundle	6 million	15,000	$10^{-3}\ D/U$	$10^{2}\ D/U$
2×2 LWR bundle	16 million	15,000	$10^{-3}\ D/U$	$10^{2}\ D/U$

LES. Codes that have achieved this grid sizes include the spectral-element code Nek5000, PHASTA, and CODE_SATURNE. Codes that have been used for petascale simulations of nuclear systems include HYDRA-TH, Open-FOAM, and DREKAR.

The above-mentioned codes have been used to simulate, at ever increasing Reynolds numbers, part of the fuel assemblies that constitute the reactor core or other nuclear components (e.g., T-junctions, portions of the upper plenum) by using DNS or LES (Figure 10 and Figure 11), as well as large portions of nuclear reactor vessels by using RANS or URANS. In both cases the trend has been toward simulating larger and larger systems rather than making problems run faster. Yet in the case of DNS/LES in particular, there are fundamental limitations to the margin for accelerating current turbulent transients. In addition to the formidable range of temporal scales that needs to be resolved [63] for examples of scale separation, a fundamental computational science issue exists. Because of power constraints, high-performance architectures are being designed to support extreme concurrency. Unfortunately, little can be done to reduce inter-node latency, which sets the node-level granularity of simulations and, ultimately, the rate at which work can be done. Extreme concurrency, however, provides an avenue to solve larger problems rather than to solve today's problems faster (assuming, as in the present case, that we are already running at the strong-scale limit). Unfortunately, this situation has consequences also at the exascale. The presence of these two constraints means that when running cases that are orders of magnitude larger (at higher Reynolds numbers or for larger domain size) for longer integration times, the time-scale separation will increase. Consequently, accelerating transients will increasingly become an imperative on larger architectures [63], not only because solving today's problems faster will be of increasing importance if industry is going to leverage HPC, but also because without accelerating transients the rate of return of increasing computational power will diminish.

3.5.4 Opportunities and challenges for exascale computing

Current petascale codes share several characteristics on extreme-scale architectures:

- Meshing is a bottleneck in the workflow at large scales.
- Algorithms are latency bound at the strong-scaling limit.
- The on-node performance is memory-bandwidth bound.
- Transients in turbulence-resolved simulations are characterized by massive time-scale separation due to the large system size and inherent transient scales (Figure 3).

These characteristics will likely remain and intensify at exascale, posing significant challenges. In fact, given the sheer scale of nuclear systems and the limited reach of current high-fidelity

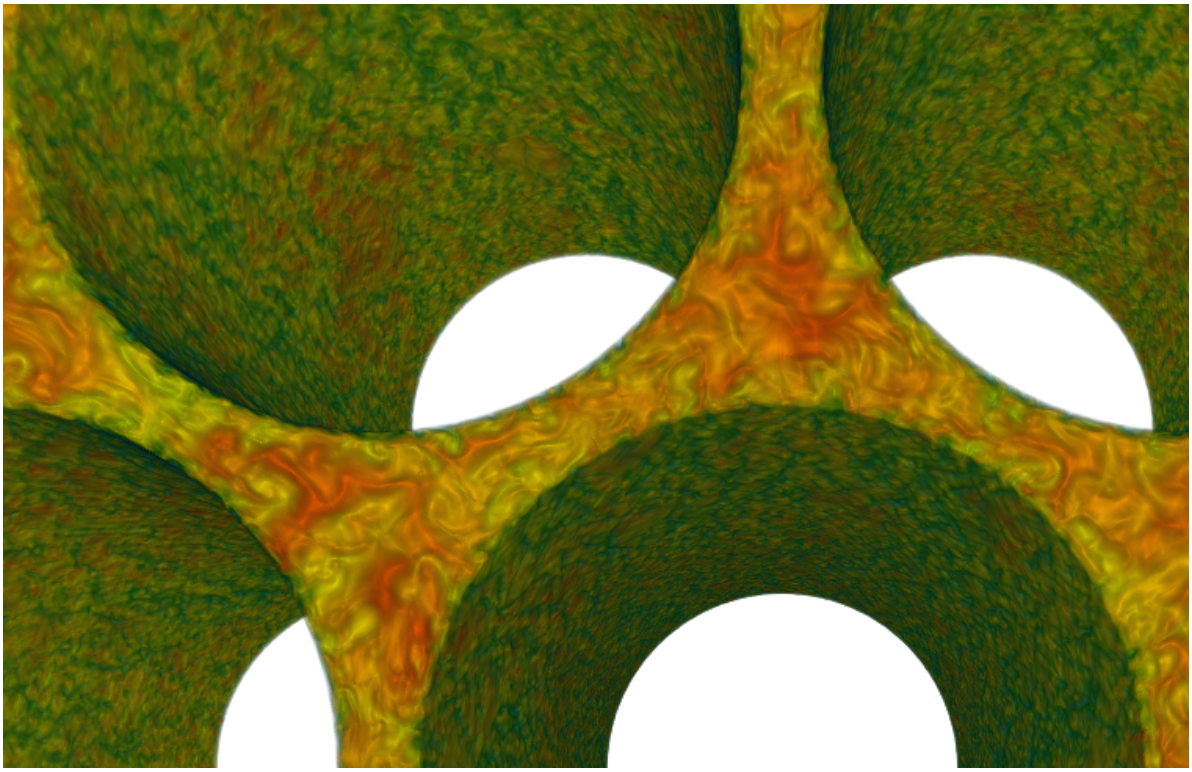


Figure 10: Flow in 37-pin rod bundle typical of sodium fast reactors, volume rendering of the velocity magnitude; Wall-resolved Large Eddy Simulation, 8 billion grid points. (image courtesy of E. Merzari)

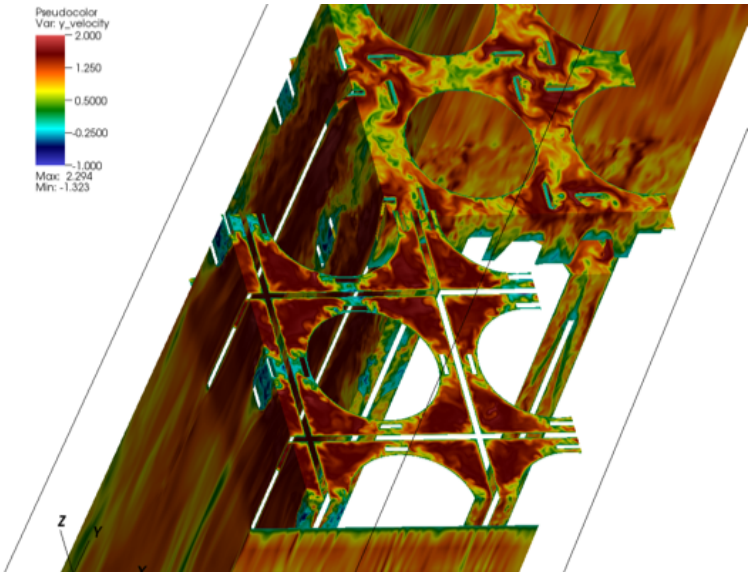


Figure 11: Contour plot of the velocity magnitude (normalized by bulk) for the flow through a fuel-bundle array with spacer grid. (image courtesy of E. Merzari)

simulations, weak scaling likely will still play an important role in the near future, although probably at a lower rate of return.

Exascale applications: Foreseeable future applications include the following:

1. Full-core fluid calculations aimed at better predicting the steady-state performance. These will likely be conducted with hybrid RANS or LES. LES may be used to simulate a portion of a core, while the rest will be handled by RANS. Results of these calculations may provide power spectral densities for structural calculations aimed at predicting flow-induced vibration (Figure 12).

2. DNS or LES for smaller sizes to benchmark RANS/hybrids and potential multiscale applications involving LES informing RANS/design-level models dynamically, perhaps employing machine learning.

3. Multiphase simulations of fuel assemblies, increasing from the current single sub-channel simulations.

4. System-level coupling between petascale models (e.g., two petascale models of assemblies coupled by a system-level algebraic constraint). This may involve integration with system (1D) models.

5. Conjugate heat transfer and coupling with neutronics (related to cases 1 and 4) and other multiphysics aspects (e.g., fuel performance). This may involve short transients and may be achieved for small cores at startup.

6. Uncertainty quantification of current petascale problems, which would benefit from nearly perfect weak scaling.

7. Shape optimization of nuclear components (related to case 6).

These directions will provide enhanced methods to study a subset of advanced reactor designs and to aid the design of new reactors. No obvious theoretical limit to achieving these goals exists. Some algorithmic work will be necessary in order to achieve good scaling in cases 4, 5, and 6. Porting and optimization of codes to new architectures will be necessary. All these applications will benefit from natural and smooth transition from petascale to exascale because the level of resources can dictate the amount of resolution employed. For instance, in case 1 all assemblies will need to be simulated with RANS in order to model a full core on the Argonne Blue Gene/Q Mira supercomputer, while on Aurora already a few assemblies can be simulated by using LES (see Figure 12).

Additional directions: Besides weak scaling, some important additional applications merit special attention:

1. Nuclear transients at full-core level to improve the accuracy of safety predictions and reduce margins. At present, safety analysis is performed only with massive simplifications (1D, porous media).

2. Inverse problems in safety analysis.

3. Acceleration of current petascale-level simulations.

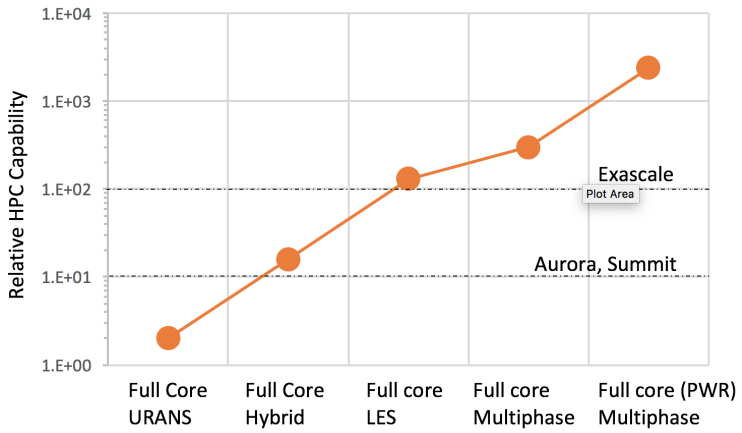


Figure 12: Progress of achievable simulations for a small modular reactor core normalized by current simulation size. (image courtesy E. Merzari)

All these applications will benefit from algorithmic advances in the area of reduced-order modeling to accelerate transients (although RANS may already be used today). Multiple ensembles, or ensemble averaging, should also be considered as a way to accelerate the collection of turbulent statistics as a means to achieve case 4. While these applications remain more challenging and less certain, they provide a higher potential to affect the nuclear industry and bridge the gap between supercomputing and engineering practice.

Conclusions: In summary, the advent of exascale promises to broaden the range of applications of high-fidelity fluid dynamics in nuclear engineering, which have traditionally been limited by resource restrictions. Exascale computing will, however, pose significant workflow and algorithmic challenges (i.e., acceleration of transients) to achieve this potential.

4 Opportunities and challenges in adaptation of existing codes and algorithms for turbulent flow simulations at the exascale

Strategies to make productive use of current codes have been considered extensively both in the workshop discussed here and in other venues (e.g., [22]). In general, the topic has been given a great deal of attention by the computational science community. There are actually two related questions. The first is how current simulation capabilities and workflows will be maintained with anticipated new hardware environments. This is an issue because it is anticipated that hardware evolution will affect all high performance computing, not just exascale. The second is how the existing software base can be used and adapted to anticipated exascale hardware to address simulation problems that are not currently feasible (bigger, more complex simulations).

Until recently, advances in node performance have been accomplished through a continuous increase in the processor clock speed, with minimal changes to the architecture. It has thus generally been straightforward to port existing simulation capabilities to new hardware, with resulting improved performance. However, with clock-rate increases stalled, node improvements are now attained by increased parallelism, with the number of processor cores and hardware threads per node set to increase by orders of magnitude as we move to exascale. Further, power considerations are leading to decreases in processor clock rates, while core counts are increasing, memory

hierarchies are getting more complex, and available memory-bandwidth per core is decreasing. Without measures to ensure *core-level* strong-scaling, the HPC node of the future will require a significant commensurate increase in node-level granularity (i.e., more work per node) to realize high performance. The advent of shared-resource parallelism will require adapting codes in a way that maps well to the anticipated hierarchical hardware architecture, with some mechanism for expressing memory locality. Furthermore, limited global resources, such as network bandwidth and file-system capacity, may require closer coordination with other jobs executing concurrently when the system is shared between multiple jobs. Thus, the anticipated evolution of hardware architecture toward exascale will require that current turbulence simulation codes be re-tooled even to maintain current turbulence simulation capabilities.

To harness the power of exascale hardware to perform larger turbulence simulations than is possible on current platforms would appear to be easier. One could simply use the increase in available hardware parallelism to increase the size of the problem while keeping the number of simulation degrees of freedom per process the same as in current simulations, as in weak scaling. However, two things work against this. First, the number of time steps required for a simulation will increase with the number of degrees of freedom, with number of time steps usually growing like number of degrees of freedom to a power between a third and one, depending on the problem. This means that for constant time to solution, the number of degrees of freedom in the solution must grow more slowly than the number of available cores or hardware execution threads. Assuming that the computational complexity per time step is linear in the degrees of freedom, which is generally the best case in turbulence simulations, the number of degrees of freedom per execution thread will scale like the number of threads to a power between $-1/4$ and $-1/2$. Second, as discussed above, reduced clock rates, reduced available memory per core, and reduced available memory bandwidth per core on exascale nodes means that the number of degrees of freedom per thread must decrease even faster as the number of available threads increases. Thus, one is again left having to retool current codes to expose greater parallelism in the simulation algorithms.

A significant challenge to exposing more parallelism in many turbulence simulations is the global linear solver that is required for implicit and incompressible/low-Mach algorithms. Because of the long tails of the associated Green's functions (which are a reflection of the disparate time scales in the physics), optimal linear solvers for these problems require hierarchical preconditioners such as multigrid or multi-level Schwarz methods. Moreover, finding the best approximation in the iteration space requires vector reductions. Both of these requirements, which drive optimal (i.e., fast) algorithms, imply global communication. The latter can be reduced through communication-reducing strategies and/or Chebyshev iteration, which has the same asymptotic complexity as conjugate gradient iteration (but is not optimal). The issue of multilevel preconditioners and in particular the global coarse-grid solve, presents a more significant challenge. The overhead of the global communication could potentially be mitigated by increased hardware support for collectives and parallel prefix operations. Such support has existed for vector-reductions on IBM's BG series for over a decade. All-reduce times on these machines are essentially P-independent, even for $P > 10^5$, so extreme-concurrency is already realized on this architecture. In addition to hardware support, availability of hardware all-reduce relies critically on *convex partitions* of the interconnect, which is a feature that will in any case be essential for good strong-scaling of applications.

By far the most pressing issue in migrating to new architectures is expressing concurrency and memory locality at the intra-node level to the same degree as has been realized through MPI at the inter-node level. Methods to access the multi-level memory hierarchies need to be expressed in a programming model that is amenable to implementation in current codes. Experience at the petascale suggests that existing PDE codes can effectively make use of several of the 'MPI+X' programming models. For example, both the high order spectral element code nek5000 and the BoxLib

based suite of codes use MPI between nodes and get good on-node scalability using OpenMP or OpenACC within the nodes. Both of these represent methods where there is a natural decomposition between intra- and inter-node decompositions. In the case of spectral element algorithms, it is natural to spatially decompose the elements (MPI) and perform per-element computation within node (OpenMP/OpenACC). In the case of block-structured AMR, the blocks map naturally onto nodes with intra-block computations threaded across cores. This also provides a natural framework for dealing with intra-node memory bandwidth non-uniformity (NUMA) issues; the blocks provide subdomains where halo regions can be explicitly managed to ameliorate memory-access issues in a way that is transparent to the legacy code governing the physics within the block.

Finally, the anticipated evolution in node hardware continues the trend of recent years in which floating-point operations are increasingly less scarce, while bandwidth for data movement becomes more so. This changes the objective in the design and evaluation of computational algorithms. For example, this generally makes higher order discretizations more favorable and may make more complex linear solver pre-conditioners more attractive. Therefore, when porting existing turbulence simulation codes to the anticipated exascale hardware, there may be opportunities to improve performance by adopting new algorithms and numerical representations to achieve the same function.

5 Opportunities and challenges of new simulation approaches for turbulent flow at the exascale

With a 100 to 1,000 fold increasing in computing capability, there is a choice to be made in each field regarding how to use the new capability. On one hand, exascale computing opens the door to evolutionary advances to existing calculations by enabling a 'larger' version of some calculation. As discussed in the previous section, this requires either adapting existing algorithms to exascale hardware or formulating new algorithms to achieve the same functionality. On the other hand, increases in available computing could be used to solve approximately the same scale physical problem that is tractable today, in the context of a broader problem that cannot currently be addressed, resulting in transformative advancements.

Some examples of such reformulation are to perform *system level simulations*, to undertake *optimization* of a device or process, or to solve *inverse problems* involving parameter estimation and uncertainty quantification to combine simulations with measured data.

System level simulations can take several forms. First, multiple petascale simulations can be combined to a single larger system. For example, simulations of individual wind-turbines could be integrated to simulate a wind-farm with hundreds of turbines, or simulations of individual gas turbine combustor sectors could be integrated to simulate a full annular combustion chamber. By integrating the unit simulations into a larger aggregate, it becomes possible to simulate phenomena that arise from interactions of the individual units that would not otherwise be captured. Another form of system level simulation results from tightly coupling hierarchical models that are usually loosely coupled. For example connecting in-cylinder combustion processes to vehicle fleet-level performance and emissions is usually done sequentially. If cylinder calculations for a set of predetermined operating points are fed into a low-dimensional model of the gas exchange processes, then turbocharging and drivetrain performance can be handed off to a fleet-based drive-cycle model to obtain an estimate of vehicle performance. Tightly coupling all of these models would enable a significant improvement in overall fidelity together with improved prediction of the in-cylinder processes under realistic conditions. Such an integrated simulation capability could be used, for example, to connect fuel composition to emissions from either the existing or a new vehicle fleet that could then be used to provide a scientific basis for policy decisions.

Another way to deploy exascale resources is to solve an optimization problem based on petascale turbulence simulations. In this case, a 100-fold increase in computing capability allows 100s of simulations to be performed as part of the optimization algorithm. Several issues arise here. First, gradient-based optimization algorithms generally do model evaluations one at a time; that is, once a gradient is approximated, a step is taken and a new gradient is approximated. In this context, the exascale resource needs to be applied efficiently to the petascale problem for each model evaluation, which leads to the strong scaling issues discussed in section 4. Second is the issue of adjoint methods, which use additional computational capability to improve gradient evaluation over numerical derivative approximations. The adjoint methods are of limited utility for chaotic systems, such as turbulence. An ability to apply adjoint methods productively to the time or ensemble averaged solutions of chaotic systems would be advantageous here. Another useful optimization approach uses surrogates constructed based on an ensemble of independent samples of simulation outputs. A similar collection of samples is also used in Jacobian-free optimization methods. Computing such an ensemble would map easily onto exascale hardware as a concurrent set of petascale simulations. There might also be opportunities for some calculations to be reused across multiple simultaneous simulations.

Inverse problems, in which one infers inputs to a simulation model from observations of the system that the model is intended to represent, are another challenging extension of petascale turbulence simulations. Deterministic inverse problems can be posed as optimization problems, so all the issues discussed in the above paragraph apply. However, the observations generally include uncertainties, and if the observation is a turbulence statistical quantity, the estimate of that quantity from the simulation will also include sampling uncertainty. To account for these uncertainties, the inverse problem is naturally cast as a Bayesian inference, and Markov-chain Monte Carlo algorithms are commonly used. These are sampling algorithms, but the samples are derived from a Markov chain which is computed sequentially. So the exascale issues are similar to those for gradient-based optimization, with the exception that one can compute multiple independent chains.

6 Closing remarks

This report describes the outcomes of a workshop that brought together scientists and engineers working on the numerical modeling and simulation of turbulence, both as a pure research topic and in the context of several application areas, to consider the implications of exascale-class computing. The participants discussed impacts on their research areas and the necessary advances and modifications to existing codes required to effectively use upcoming computer architectures based on expected trends.

The prevailing view is that the biggest impacts of exascale computing will be realized by making "large" and "complex" turbulence problems accessible, with either larger domains or a richer treatment of physical phenomena addressed than what is currently possible. This may require significant reformulation of existing codes, implementation of new physics, and development of a more nuanced problem formation, but has the potential to produce significant advances. This is distinct from moving to higher Reynolds numbers, which equates to doing a currently tractable problem "better"; the modest increase in Reynolds number possible is only expected to be transformative for the limited number of situations where the increase will expose transitions to new flow characteristics. Examples of the types of simulations that may become tractable include:

- Simulation of nonlinear multi-scale self-organized turbulence in burning plasma, covering a whole fusion chamber

- Simulation of combustion in high pressure internal combustion engines and gas turbine sectors at pressure with real geometries and sufficient physical complexity to address fuel and operational flexibility

- Nuclear reactor simulation to generate experimental-quality simulations of steady-state and transient reactor behavior with extreme-fidelity resolution to inform and validate engineering-scale simulation tools; this will enable dramatic advances in nuclear technology that will ultimately increase safety, reduce margins, and lead to an increase in economic competitiveness

- Simulation of boundary-layer turbulence and clouds over large areas, such as continental scales, to better understand the impact of cloud formation on the surrounding environment, and how this feeds back to subsequent cloud formation; this will ultimately improve our ability to parameterize the clouds for climate models

- Simulation of an entire wind farm under realistic atmospheric flow conditions and in complex terrain where turbine geometry is well resolved, thereby exposing pathways to optimization and reduced cost of energy.

Although limited in number, there are still examples where a modest increase in Reynolds number may lead to significant new insights because new regimes with different behavior can be explored. A particular example discussed in this report is wall bounded shear flows.

To pursue opportunities typified by the problems above and to maintain the existing research capability on next-generation computer hardware, it is essential for the scientists developing the algorithms and computational tools to adapt to the changes in the computer hardware. Without algorithmic and implementation changes, today's petascale class turbulence simulations will take longer to run on anticipated exascale hardware. This suggests it is important that scientists and engineers engage in that transition and understand the implications for turbulent-flow simulations. Power-usage restrictions are expected to dramatically change the makeup of computing architectures and lead to decreases in processor clock rates, while core counts are increasing, memory hierarchies are getting more complex, and available memory bandwidth per processor core is decreasing. These changes will affect HPC on the largest leadership systems down to small clusters. Mapping turbulence-simulation algorithms and software onto future exascale hardware will require exposing parallelism at the node level, as well as between nodes, and making good use of multi-level memory hierarchies.

A final observation that came out of the workshop is that there is an opportunity to make a transformative change to the algorithmic and modeling approach to turbulent flows. This arises in two particular aspects. First, uncertainty quantification (UQ) and design-space exploration frequently involve multiple simulations with perturbed parameters or initial/boundary conditions. This results in a "closely coupled ensemble" of simulations. Algorithmic strategies to perform these runs in a coordinated way—for example by sharing core parts of the solution algorithm that are redundant across the simulations in the ensemble, or by evaluating multiple realizations in a single vector operation—have significant value. Although it may be possible to perform such an ensemble of runs on a collection of less capable computers, tightly coupling them on an exascale machine has the promise of faster time to solution and lower total power consumption. Second, hybrid turbulence modeling could enable several of the above problems. Having a simulation utilize, adaptively, the best combination of refinement in space, time, or model fidelity—either hybrid DNS/LES in a fundamental research setting, or hybrid LES/RANS in a more applied scenario—may be an enabling pathway to the long time-horizon / complex-domain simulations that are

necessary to overlap computation with experimentation. Such overlap is essential for validation and provides design decision support of relevance.

The ability to understand and, ultimately, control the interaction between turbulence and the applications considered here remains a challenge. With the advent of exascale computing, researchers in this field will have a valuable tool to provide increased fidelity and richness to the simulations that are possible. While realizing that taking advantage of exascale computing will require significant effort, the future is bright. New ways of thinking about identified problems and a community rallied around the development necessary to use exascale computing most effectively provides fertile ground for breakthrough and discovery.

References

[1] A2e: Atmosphere to electrons. https:a2e.pnnl.gov. [Online; accessed 29 March 2016].

[2] Exascale Initiative. http://www.exascaleinitiative.org/pathforward, April 2016. [Online; accessed 23 May 2016].

[3] Nek5000, a spectral element code for CFD. https://nek5000.mcs.anl.gov/. [Online; accessed 23 August 2016].

[4] Executive order – creating a national strategic computing initiative. https://www.whitehouse.gov/the-press-office/2015/07/29/executive-order-creating-national-strategic-computing-initiative, July 2015. [Online; accessed 1 April 2016].

[5] A workshop to identify reserach needs and impacts in predictive simulation for internal combustion engines PreSICE. Report sponsored by the Office of Basic Energy Sciences, Office of Science and the Vehicle Technologies Program, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, March 3, 2011.

[6] Turbulent flow simulation at the exascale: Opportunities and challenges workshop. https://www.orau.gov/turbulentflow2015, August 2015. workshop website [Online; accessed 23 May 2016].

[7] Argonne National Laboratory. Autonomie. http://www.autonomie.net [Online; accessed 24 May 2016].

[8] S. Basu, J.-F. Vinuesa, and A. Swift. Dynamic LES modeling of a diurnal cycle. *Journal of Applied Meteorology and Climatology*, 47:1156–1174, 2008. doi:10.1175/2007JAMC1677.1.

[9] J. Bell. Block-structured AMR for exascale applications. https://www.orau.gov/turbulentflow2015/plenary-talks/Bell_John.pdf, August 2015. plenary talk at Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop [Online; accessed 23 May 2016].

[10] A. Bourlioux, A. T. Layton, and M. L. Minion. Higher-order multi-implicit spectral deferred correction methods for problems of reacting flow. *Journal of Computational Physics*, 189: 351–376, 2003.

[11] K. J. Bowers, B. J. Albright, L. Yin, B. Bergen, and T. J. T. Kwan. Ultrahigh performance three-dimensional electromagnetic relativistic kinetic plasma simulation. *Physics of Plasmas*, 15(5):055703, May 2008. doi:10.1063/1.2840133.

[12] K. Braman, T. Oliver, and V. Raman. Bayesian analysis of syngas chemistry models. *Combustion Theory and Modelling*, 17(5):858–887, 2013.

[13] M. Calaf, C. Meneveau, and J. Meyers. Large-eddy simulation study of fully developed wind-turbine array boundary layers. *Physics of Fluids*, 22:015110, 2010. doi:dx.doi.org/10.1063/1.3291077.

[14] R. E. Carbone and J. D. Tuttle. Rainfall occurrence in the US warm season: The diurnal cycle. *Journal of Climate*, 21:4132–4146, 2008. doi:10.1175/2008jcli2275.1.

[15] C. S. Chang, S. Ku, P. Diamond, M. Adams, R. Barreto, Y. Chen, J. Cummings, E. D'Azevedo, G. Dif-Pradalier, S. Ethier, L. Greengard, T. S. Hahm, F. Hinton, D. Keyes, S. Klasky, Z. Lin, J. Lofstead, G. Park, S. Parker, N. Podhorszki, K. Schwan, A. Shoshani, D. Silver, M. Wolf, P. Worley, H. Weitzner, E. Yoon, and D. Zorin. Whole-volume integrated gyrokinetic simulation of plasma turbulence in realistic diverted-tokamak geometry. *Journal of Physics Conference Series*, 180(1):012057, July 2009. doi:10.1088/1742-6596/180/1/012057.

[16] C. S. Chang, S. Ku, P. H. Diamond, Z. Lin, S. Parker, T. S. Hahm, and N. Samatova. Compressed ion temperature gradient turbulence in diverted tokamak edge. *Physics of Plasmas*, 16(5):056108, May 2009. doi:10.1063/1.3099329.

[17] J. H. Chen, A. Choudhary, B. deSupinski, M. DeVries, E. R. Hawkes, S. Klasky, W.K.Liao, K. L. Ma, J. Mellor-Crummey, N. Podhorszki, R. Sankaran, S. Shende, and C. S. Yoo. Terascale direct numerical simulations of turbulent combustion using S3D. *Computational Science and Discovery*, 2(015001), 2009.

[18] M. J. Churchfield, S. Lee, P. J. Moriarty, L. Martínez, S. Leonardi, G. Vijayakumar, and J. G. Brasseur. A large-eddy simulation of wind-plant aerodynamics. In *50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, Nashville, TN, Jan. 9–12, 2012*, number AIAA Paper 2012-537. AIAA, Washington D.C., 2012. URL http://www.nrel.gov/docs/fy12osti/53554.pdf.

[19] M. J. Churchfield, J. K. Lundquist, E. Quon, S. Lee, and A. Clifton. Large-eddy simulations of wind turbine wakes subject to different atmospheric stabilities. Poster presented at the American Geophysical Union Fall Meeting, December 2014.

[20] M. S. Day and J. B. Bell. Numerical simulation of laminar reacting flows with complex chemistry. *Combustion Theory and Modelling*, 4:535–556, 2000.

[21] J. W. Deardorff. Three-dimensional numerical study of the height and mean structure of a heated planetary boundary layer. *Boundary-Layer Meteorology*, 7:81–106, 1974. doi:10.1007/bf00224974.

[22] DOE ASCAC Subcommittee Report. ASCAC Subcommittee Report: The Opportunities and Challenges of Exascale Computing, Fall 2010. URL http://science.energy.gov/~/media/ascr/ascac/pdf/reports/Exascale_subcommittee_report.pdf.

[23] DOE Workshop Report. Scientific Grand Challenges: Discovery in Basic Energy Sciences: The Role of Computing at the Extreme Scale, August 2009. URL http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Bes_exascale_report.pdf.

[24] DOE Workshop Report. Scientific Grand Challenges: Challenges for Understanding the Quantum Universe and the Role of Computing at the Extreme Scale, December 2008. URL http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Hep_report.pdf.

[25] DOE Workshop Report. Scientific Grand Challenges: Opportunities in Biology at the Extreme Scale of Computing, August 2009. URL http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Biology_report.pdf.

[26] DOE Workshop Report. Scientific Grand Challenges: Challenges in Climate Change Science and the Role of Computing at the Extreme Scale, November 2008. URL http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Climate_report.pdf.

[27] DOE Workshop Report. Scientific Grand Challenges: Fusion Energy Sciences and the Role of Computing at the Extreme Scale, March 2009. URL http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Fusion_report.pdf.

[28] DOE Workshop Report. Scientific Grand Challenges for National Security: The Role of Computing at the Extreme Scale, October 2009. URL http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Nnsa_grand_challenges_report.pdf.

[29] DOE Workshop Report. Science Based Nuclear Energy Systems Enabled by Advanced Modeling and Simulation at the Extreme Scale, May 2009. URL http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Sc_nework_shop__report.pdf.

[30] DOE Workshop Report. Scientific Grand Challenges: Forefront Questions in Nuclear Science and the Role of Computing at the Extreme Scale, January 2009. URL http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Np_report.pdf.

[31] J. Dongarra, J. Hittinger, J. Bell, L. Chacon, R. Falgout, M. Heroux, P. Hovland, E. Ng, C. Webster, and S. Wild. Applied mathematics research for exascale computing. Report, U.S. Department of Energy, ASCR, March 2014. URL http://science.energy.gov/~/media/ascr/pdf/research/am/docs/EMWGreport.pdf.

[32] P. A. Durbin and B. A. Pettersson-Reif. *Statistical Theory and Modeling of Turbulent Flows*. Wiley, 2 edition, 2011.

[33] J. W. Fan, M. Ovtchinnikov, J. M. Comstock, S. A. McFarlane, and A. Khain. Ice formation in Arctic mixed-phase clouds: Insights from a 3-D cloud-resolving model with size-resolved aerosol and cloud microphysics. *Journal of Geophysical Research-Atmospheres*, 114:D04205, 2009. doi:10.1029/2008jd010782.

[34] A. C. Fitch, J. B. Olson, J. K. Lundquist, J. K. Dudhia, A. K. Gupta, J. Michalakes, and I. Barstad. Local and mesoscale impacts of wind farms as parameterized in a mesoscale NWP model. *Monthly Weather Review*, 40:3017–3038, 2012. URL http://dx.doi.org/10.1175/MWR-D-11-00352.1.

[35] D. G. Fox and D. K. Lilly. Numerical simulation of turbulent flows. *Reviews of Geophysics and Space Physics*, 10(1):51–72, 1972. doi:10.1029/RG010i001p00051.

[36] Z. Gao, S. Curran, J. Parks, D. Smith, R. Wagner, M. R. M, C. Daw, K. Edwards, and J. Thomas. Drive cycle simulation of high efficiency combustions on fuel economy and exhaust properties in light-duty vehicles. *Applied Energy*, 157:762–776, 2015.

[37] R. Grout, A. Gruber, H. Kolla, P.-T. Bremer, J. Bennett, A. Gyulassy, and J. Chen. A direct numerical simulation study of turbulence and flame structure in a round jet in cross-flow. *Journal of Fluid Mechanics*, 706:351–383, 2012. ISSN 15407489. doi:10.1017/jfm.2012.257.

[38] W. I. Gustafson, A. M. Vogelmann, X. Cheng, S. Endo, B. Krishna, Z. Li, T. Toto, and H. Xiao. The LES ARM Symbiotic Simulation and Observation (LASSO) Alpha 1 Release. *U. S. Department of Energy Atmospheric Radiation Measurement Climate Research Facility*, 2016. doi:10.5439/1256454. URL http://www.arm.gov/science/themes/lasso/releases.

[39] W. I. Gustafson Jr. and A. M. Vogelmann. LES ARM Symbiotic Simulation and Observation (LASSO) Implementation Strategy. Technical Report DOE/SC-ARM-15-039, U. S. Department of Energy Atmospheric Radiation Measurement Climate Research Facility, September 2015. URL http://www.arm.gov/publications/programdocs/doe-sc-arm-15-039.pdf. 33 pages.

[40] R. Hager, E. Yoon, S. Ku, E. D'Azevedo, P. Worley, and C. Chang. A fully non-linear multi-species Fokker-Planck-Landau collision operator for simulation of fusion plasma. *Journal of Computational Physics*, 315:644–660, 2016.

[41] S. W. Hammond, M. A. Sprague, D. Womble, and M. Barone. A2e High Fidelity Modeling: Strategic Planning Meetings. Technical Report NREL/TP-2C00-64697, National Renewable Energy Laboratory, 2015.

[42] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward. SUNDIALS: Suite of Nonlinear and Differential/Algebraic Equation Solvers. *ACM Transactions on Mathematical Software*, 31(3):363–396, 2005.

[43] J. Hittinger. Challenges on the path to exascale. https://www.orau.gov/turbulentflow2015/plenary-talks/Hittinger_Jeffrey.pdf, August 2015. plenary talk at Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop [Online; accessed 23 May 2016].

[44] M.-C. Hsu and Y. Bazilevs. Fluid-structure interaction modeling of wind turbines: simulating the full machine. *Computational Mechanics*, 50:821–833, 2012. doi:10.1007/s00466-012-0772-0.

[45] T. Ishihara, T. Gotoh, and Y. Kaneda. Study of high-Reynolds number isotropic turbulence by direct numerical simulation. *Annual Review of Fluid Mechanics*, 41:165–180, 2009. doi:10.1146/annurev.fluid.010908.165203.

[46] A. M. Jaffe. The millennium grand challenge in mathematics. *Notices of the AMS*, 53:652–660, 2006.

[47] F. Jenko, W. Dorland, M. Kotschenreuther, and B. N. Rogers. Electron temperature gradient driven turbulence. *Physics of Plasmas*, 7:1904–1910, May 2000. doi:10.1063/1.874014.

[48] J. Jimenez. Turbulent flows over rough walls. *Annual Review of Fluid Mechanics*, 36:173–196, 2004. ISSN 0066-4189. doi:10.1146/annurev.fluid.36.050802.122103.

[49] J. Jiménez and A. Pinelli. The autonomous cycle of near-wall turbulence. *Journal of Fluid Mechanics*, 389:335–359, 6 1999. ISSN 1469-7645. doi:10.1017/S0022112099005066. URL http://journals.cambridge.org/article_S0022112099005066.

[50] J. Jimenez and A. A. Wray. On the characteristics of vortex filiments in isotropic turbulence. *Journal of Fluid Mechanics*, 373:255–285, 1998.

[51] J. Jonkman and B. Jonkman. Fast v8. https://nwtc.nrel.gov/FAST8, October 2015. [Online; accessed 29 March 2016].

[52] J. M. Jonkman. The new modularization framework for the FAST wind turbine CAE tool. In *Proceedings of the 51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, Grapevine, Texas, January 2013.

[53] A. Khain, T. V. Prabha, N. Benmoshe, G. Pandithurai, and M. Ovchinnikov. The mechanism of first raindrops formation in deep convective clouds. *Journal of Geophysical Research-Atmospheres*, 118:9123–9140, 2013. doi:10.1002/jgrd.50641.

[54] M. Khalil, G. Lacaze, J. C. Oefelein, and H. N. Najm. Uncertainty quantification in LES of a turbulent bluff-body stabilized flame. *Proceedings of the Combustion Institute*, 35(2): 1147–1156, 2015.

[55] P. Kogge, K. Bergman, S. Borkar, D. Campbell, W. Carlson, W. Dally, M. Denneau, P. Franzon, W. Harrod, K. Hill, J. Hiller, S. Karp, S. Keckler, D. Klein, R. Lucas, M. Richards, A. Scarpelli, S. Scott, A. Snavely, T. Sterling, R. S. Williams, and K. Yelick. Exascale computing study: Technology challenges in achieving exascale systems, 2008. URL `http://www.cse.nd.edu/Reports/2008/TR-2008-13.pdf`.

[56] S. Ku, R. Hager, C. Chang, J. Kwon, and S. Parker. A new hybrid-Lagrangian numerical scheme for gyrokinetic simulation of tokamak edge plasma. *Journal of Computational Physics*, 315:467–475, 2016.

[57] M. Lee and R. D. Moser. Direct numerical simulation of turbulent channel flow up to $Re_\tau \approx 5200$. *Journal of Fluid Mechanics*, 774:395–415, 2015. ISSN 0022-1120. doi:10.1017/jfm.2015.268.

[58] Y. Li, K.-J. Paik, T. Xing, and P. M. Carrica. Dynamic overset CFD simulations of wind turbine aerodynamics. *Renewable Energy*, 37:285–298, 2012. doi:10.1016/j.renene.2011.06.029.

[59] Y. Li, A. M. Castro, T. Sinokrot, W. Prescott, and P. M. Carrica. Coupled multi-body dynamics and CFD for wind turbine simulation including explicit wind turbulence. *Renewable Energy*, 76:338–361, 2015. doi:10.1016/j.renene.2014.11.014.

[60] H. Lu and F. Porté-Agel. Large-eddy simulation of a very large wind farm in a stable atmospheric boundary layer. *Physics of Fluids*, 23, 2011. doi:10.1063/1.3589857.

[61] T. Lu and C. Law. Toward accommodating realistic fuel chemistry in large-scale computation. *Prog. Energy Combust. Sci.*, 35:192–215, 2009.

[62] M. J. McNenly, R. A. Whitesides, and D. L. Flowers. Adaptive preconditioning strategies for integrating large kinetic mechanisms. In *8th US National Combustion Meeting*, 2013.

[63] E. Merzari, A. Obabko, P. Fischer, N. Halford, J. Walker, Y. Yu, and A. Siegel. Large Scale Simulation of Nuclear Reactors: Issues and Perspectives. In *Proceedings of the 16th International Topical Meeting on Nuclear Reactor Thermal-Hydraulics*, Chicago, Illinois, August 2015. American Nuclear Society.

[64] J. Mirocha, B. Kosovic, and G. Kirkil. Resolved turbulence characteristics in large-eddy simulations nested within mesoscale simulations using the Weather Research and Forecasting model. *Monthly Weather Review*, 142:806–831, 2014. doi:10.1175/mwr-d-13-00064.1.

[65] Y. Miyamoto, Y. Kajikawa, R. Yoshida, T. Yamaura, H. Yashiro, and H. Tomita. Deep moist atmospheric convection in a subkilometer global simulation. *Geophysical Research Letters*, 40: 4922–4926, 2013. doi:10.1002/grl.50944.

[66] P. Moin and K. Mahesh. Direct numerical simulation: A tool in turbulence research. *Annual Review of Fluid Mechanics*, 30:539–78, 1998.

[67] R. Morrison, R. Moser, and T. Oliver. Representing model inadequacy in combustion mechanisms of laminar flames. *Bulletin of the American Physical Society*, 60, 2015.

[68] R. D. Moser. On the validity of the continuum approximation in high reynolds number turbulence. *Physics of Fluids*, 18(7):078105, 2006. doi:http://dx.doi.org/10.1063/1.2222398. URL http://scitation.aip.org/content/aip/journal/pof2/18/7/10.1063/1.2222398.

[69] R. D. Moser, M. M. Rogers, and D. W. Ewing. Self-similarity of time-evolving plane wakes. *Journal of Fluid Mechanics*, 367:255–289, 7 1998. ISSN 1469-7645. doi:10.1017/S0022112098001426. URL http://journals.cambridge.org/article_S0022112098001426.

[70] D. Muñoz-Esparza, B. Kosović, J. Mirocha, and J. van Beeck. Bridging the transition from mesoscale to microscale turbulence in numerical weather prediction models. *Boundary-Layer Meteorology*, 153:409–440, 2015.

[71] M. P. Musculus, S. L. Kokjohn, and R. D. Reitz. Evaluating temperature and fuel stratification for heat-release rate control in a reactivity-controlled compression-ignition engine using optical diagnostics and chemical kinetics modeling. Technical report, Sandia National Laboratories (SNL-CA), Livermore, CA (United States), 2014.

[72] H. Nakayama and T. T. adn H. Nagai. Large-eddy simulation of urban boundary-layer flows by generating turbulent inflows from mesoscale meteorological simulations. *Atmospheric Science Letters*, 13:180–186, 2012. doi:10.1002/asl.377.

[73] A. Nonaka et al. A deferred correction coupling strategy for low Mach number flow with complex chemistry. *Combustion Theory and Modeling*, pages 1–36, 2012. doi:10.1080/13647830.2012.701019.

[74] J. Oefelein. Large eddy simulation of turbulent combustion processes in propulsion and power systems. *Progress in Aerospace Sciences*, 42(1):2–37, 2006.

[75] S. A. Orszag and G. S. Patterson. Numerical simulation of three-dimensional homogeneous isotropic turbulence. *Physical Review Letters*, 28:76–79, 1972. doi:10.1103/PhysRevLett.28.76. URL http://link.aps.org/doi/10.1103/PhysRevLett.28.76.

[76] Y. Peet, P. Fischer, G. Conzelmann, and V. Kotamarhi. Actuator line aerodynamics model with spectral elements. In *Proceedings of the 51st AIAA Aerospace Sciences Meeting*, Grapevine, TX, January 2013.

[77] M. S. Pritchard, H. Parishani, M. Wyant, C. Bretherton, M. Khairoutdinov, and B. Singh. Ultraparameterization: A new strategy for including explicit low cloud physics in global climate models at the GPU-accelerated petascale, January 2016. URL https://ams.confex.com/ams/96Annual/webprogram/Paper286965.html. 28th Conference on Climate Variability and Change, American Meteorological Society Annual Meeting, New Orleans.

[78] J. Schalkwijk, H. J. J. Jonker, A. P. Siebesma, and E. Van Meijgaard. Weather forecasting using GPU-based large-eddy simulations. *Bulletin of the American Meteorological Society*, 96: 715–724, 2015. doi:10.1175/bams-d-14-00114.1.

[79] J. Sitaraman, D. Mavriplis, and E. P. Duque. Wind farm simulations using a full rotor model for wind turbines. In *Proceedings of the AIAA SciTech 2014 Meeting*, National Harbor, MD, 2014.

[80] S. Som and D. E. Longman. Numerical study comparing the combustion and emission characteristics of biodiesel to petrodiesel. *Energy and Fuels*, 25(4):1373–1386, 2011.

[81] J. N. Sørensen and W. Z. Shen. Numerical modeling of wind turbine wakes. *Journal of Fluids Engineering*, 124:393–399, 2002. doi:10.1115/1.1471361.

[82] C. Sovinec, A. Glasser, T. Gianakon, D. Barnes, R. Nebel, S. Kruger, S. Plimpton, A. Tarditi, M. Chu, and the NIMROD Team. Nonlinear magnetohydrodynamics with high-order finite elements. *Journal of Computational Physics*, 195:355, 2004.

[83] P. R. Spalart. Detached-eddy simulation. *Annual Review of Fluid Mechanics*, 41:181–202, 2009. doi:10.1146/annurev.fluid.010908.165130.

[84] R. Stevens et al. A decadal DOE plan for providing exascale applications and technologies for DOE mission needs. http://science.energy.gov/~/media/ascr/ascac/pdf/meetings/mar10/Awhite.pdf, March 2010. [Online; accessed 23 August 2016].

[85] H. Tomita. Global convection permitting/resolving model: Towards the global LES, 2013. URL http://www.ecmwf.int/sites/default/files/elibrary/2013/14767-global-convection-permittingresolving-model-towards-global-les.pdf. ECMWF Annual Seminar, Reading UK.

[86] N. Troldborg, J. N. Sørensen, and R. Mikkelsen. Actuator line simulation of wake of wind turbine operating in turbulent inflow. *Journal of Physics: Conference Series*, 75:012063, 2007. doi:10.1088/1742-6596/75/1/012063.

[87] *Wind Vision: A New Era for Wind Power in the United States*. U.S. Department of Energy, DOE/GO-102015-4557, April 2015. URL http://www.nrel.gov/docs/fy12osti/53554.pdf.

[88] *Complex Flow Workshop Report*. U.S. Department of Energy, Energy Efficiency and Renewable Energy, Wind Program, January 2012. URL http://energy.gov/sites/prod/files/2013/12/f5/complex_flow_workshop_report.pdf.

[89] A. Varble, E. J. Zipser, A. M. Fridlind, P. Zhu, A. S. Ackerman, J.-P. Chaboureau, S. Collis, J. Fan, A. Hill, and B. Shipway. Evaluation of cloud-resolving and limited area model intercomparison simulations using TWP-ICE observations: 1. Deep convective updraft properties. *Journal of Geophysical Research: Atmospheres*, 119:13891–13918, 2014. doi:10.1002/2013JD021371.

[90] M. Wang, S. Ghan, R. Easter, M. Ovchinnikov, X. Liu, E. Kassianov, Y. Qian, W. I. Gustafson, V. E. Larson, D. P. Schanen, M. Khairoutdinov, and H. Morrison. The multi-scale aerosol-climate model PNNL-MMF: model description and evaluation. *Geoscientific Model Development*, 4:137–168, 2011. doi:10.5194/gmd-4-137-2011.

[91] Q. Wang, M. A. Sprague, J. Jonkman, and N. Johnson. Beamdyn: A high-fidelity wind turbine blade solver in the FAST modular framework. In *Proceedings of SciTech 2015*, Kissimmee, Florida, 2015. Also published as NREL tech. report NREL/CP-5000-63165.

[92] Z.-T. Xie. Modelling street-scale flow and dispersion in realistic windstowards coupling with mesoscale meteorological models. *Boundary-Layer Meteorology*, 141:52–75, 2011. doi:10.1007/s10546-011-9629-x.

[93] X. Yang, F. Sotiropoulos, R. J. Conzemius, J. N. Wachtler, and M. B. Strong. Large-eddy simulation of turbulent flow past wind turbines/farms: the virtual wind simulator (VWiS). *Wind Energy*, 2014. doi:10.1002/we.1802.

[94] C. S. Yoo, R. Sankaran, and J. H. Chen. Three-dimensional direct numerical simulation of a turbulent lifted hydrogen jet flame in heated coflow: flame stabilization and structure. *Journal of Fluid Mechanics*, 640:453, Dec. 2009. ISSN 0022-1120. doi:10.1017/S0022112009991388. URL http://www.journals.cambridge.org/abstract_S0022112009991388.

[95] F. Zahle, N. N. Sørensen, and J. Johansen. Wind turbine rotor-tower interaction using an incompressible overset grid method. *Wind Energy*, 12:594619, 2009. doi:10.1002/we.327.

[96] B. Zhang and C. Liang. A simple, efficient, high-order accurate sliding-mesh interface approach to the spectral difference method on coupled rotating and stationary domains. *Journal of Computational Physics*, 295:147–160, 2015. doi:10.1016/j.jcp.2015.04.006.

White Papers

Electronic versions of these position papers are available for download from the Turbulent Flow Simulation at the Exascale workshop website: https://www.orau.gov/turbulentflow2015/whitepapers.htm.

[WP1] S. Balachandar, T.L. Jackson, and P. Fischer, Exascale Simulations of Canonical Turbulent Disperse Multiphase Flows

[WP2] I.A. Bolotonov, Exascale Applications for Nuclear Reactor Thermal-Hydraulics: Fully Resolved Bubbly Flow in Realistic Reactor Core Fuel Assembly

[WP3] A.M. Castro and P.M. Carrica, Impact of Exascale Computers on the Simulation of Complex Turbulent Flows of Engineering Interest

[WP4] C.S. Chang, XGC1

[WP5] M.J. Churchfield and M.A. Sprague, Exascale Computing and Wind Energy

[WP6] J.B. Freund, D.J. Bodony, and W.D. Gropp, Multi-Physics Turbulence Simulation: A Holistic Approach to Performance

[WP7] P.E. Hamlington, R.N. King, and W.J.A. Dahm, Autonomic Subgrid-Scale Closure for Large Eddy Simulations

[WP8] R. Jacob, Exascale Simulations of the Urban Atmospheric Boundary Layer for Understanding Urban Climate Change

[WP9] A.G. Kritsuk, H.C. Yee, and B. Sjogreen, Understanding Compressible Turbulence with Time Varying Random Forcing by Low Dissipative High Order Schemes

[WP10] C. Liang, Massively Parallel High-Order Spectral Difference Method for Large Eddy Simulation of Vertical Axis Wind Turbine and Oscillating Wing Wind Turbine

[WP11] E. Merzari, P. Fischer, and A. Siegal, Current Trends in the Large-Scale Simulation of Nuclear Reactor Flows

[WP12] R. Sankaran and J.C. Oefelein, Development of Predictive Models through Coupled Application of Simulations and Uncertainty Quantification on Exascale Systems

[WP13] T.M. Smith, Full System Scale Simulations of Reactor Cores Enabled by ExaScale Computing

[WP14] D. Sondak, J.N. Shadid, L.M. Smith, F. Waleffe, T.M. Smith, and A.A. Oberai, Turbulence Explorations at the Exascale: Turbulence Structures, Models and Scalable Solution Methods

[WP15] M. Sprague, S. Hammond, D. Womble, and M. Barone, The Atmosphere to Electrons (A2e) Initiative, Wind Plant Simulations, and Exascale Computing

[WP16] J.A. Templeton, Scale-Based Parallelization for Turbulent Flow Simulations Enabled by Extreme-Scale Computing

[WP17] Q. Wang, Sensitivity of Turbulence: Can Exascale Solve It?

[WP18] J. Wu, *In Situ* Workflow Management for Extreme-Scale Turbulent Flow Simulation

[WP19] H.C. Yee, Quantification of Numerical Uncertainty in DNS Simulations of Compressible Turbulence with Strong Shocks – A Dynamics of Numerics Approach

Plenary Talks

Electronic versions of these position papers are available for download from the Turbulent Flow Simulation at the Exascaleworkshop website: https://www.orau.gov/turbulentflow2015/plenarytalks.htm.

[PT1] *Challenges on the path to exascale*

Jeffrey A.F. Hittinger
Center for Applied Scientific Computing, Lawrence Livermore National Laboratory
http://people.llnl.gov/hittinger1

Abstract: The move to exascale computing is expected to be disruptive due to significant changes in computer architectures. Computational scientists will need to address new challenges in extreme concurrency, limited memory, data locality, resilience, and overall system and software complexity. Advances in applied mathematics will be necessary to realize the full potential of these supercomputers, but will these advances be incremental changes to existing methods or will exascale computing require a substantial rethinking of how we compute? Will the transition to exascale be evolutionary or revolutionary? Reflecting on the findings of the DOE Advanced Scientific Computing Research Program Exascale Mathematics Working Group, Dr. Hittinger will provide his perspective on the path to exascale and the opportunities for new applied mathematics research that will enable exascale computing.

[PT2] *Block-structured AMR for exascale applications*

John Bell
Center for Computational Sciences and Engineering, Lawrence Berkeley National Lab
https://ccse.lbl.gov/people/jbb

Abstract: Block-structured adaptive mesh refinement (AMR) has been successfully used in a number of multiphysics petascale applications. The ability of AMR to localize computational effort where needed, reducing both memory requirements and computational work, makes it a key technology for exascale applications. In this talk, I will discuss the basic concepts of AMR and how discretizations of simple partial differential equations on adaptively refined grids are constructed. I will then discuss how to combine the basic discretization methodology into algorithms for more complex multiphysics applications based on decomposition of the problem into component processes. Finally, I will discuss some of the implementation issues associated with AMR and describe some of the research issues needed to make AMR an effective tool for exascale computing.

[PT3] *Path to high-order unstructured-grid exascale CFD*

Paul Fischer
Mathematics and Computer Science, Argonne National Laboratory, and Mechanical Science and Engineering, University of Illinois at Urbana-Champaign
http://www.mcs.anl.gov/person/paul-fischer

Abstract: Petascale computing platforms currently feature million-way parallelism and it is anticipated that exascale computers with billion-way concurrency will be deployed in the early 2020s. In this talk, we explore the potential and difficulties of PDE-based

simulation at these scales with a focus on turbulent fluid flow and heat transfer in a variety of applications including nuclear energy, combustion, oceanography, vascular flows, and astrophysics. Using data from leading-edge platforms over the past 25 years, we analyze the scalability of state-of-the-art solvers to predict parallel performance on exascale architectures. With the end of frequency scaling, the principal avenue for increased performance is through greater concurrency, which favors solution of larger problems rather than faster solution of today's problems. We analyze these trends in order to shed light on the expected scope of next generation simulations and to provide insight to design requirements for future algorithms, codes, and architectures.

[PT4] *Simulation of turbulent flow on emerging HPC – An aerospace perspective*
Jeffrey Slotnick
Boeing Research & Technology, The Boeing Company

Abstract: As the use of high-fidelity, physics-based computational fluid dynamics (CFD) numerical simulation continues to expand into the conceptual and preliminary design phases of aerospace systems, which, in many cases, involve highly complicated geometry with complex flow physics, the need to accurately and reliably predict turbulent, viscous flow fields with significant flow separation is paramount. Further, efficient use of these CFD tools and processes on emerging exascale computing architectures presents some significant challenges. This talk will focus on the current state of turbulent flow CFD simulation and key technical issues, and will draw heavily from the NASA Vision 2030 CFD study report, as well as from recent discussions on this topic among the aerospace CFD community. Recommended research thrusts will be described, along with opportunities for collaboration between DOE ASCR and the aerospace engineering discipline.

Workshop Participants

Name	*Organization*
Shashi Aithal	Argonne National Laboratory
Hussein Aluie	John Hopkins University
Ramesh Balakrishnan	Argonne National Laboratory
Jon Baltzer	Los Alamos National Laboratory
Yuri Bazilevs	University of California, San Diego
John Bell	Lawrence Berkeley National Laboratory
Myra Blaylock	Sandia National Laboratories
Daniel Bodony	University of Illinois
Igor Bolotnov	North Carolina State University
Alejandro Castro	The University of Iowa
Choong-Seock Chang	Princeton Plasma Physics Laboratory
Jacqueline Chen	Sandia National Laboratories
Matt Churchfield	National Renewable Energy Laboratory
Phillip Colella	Lawrence Berkeley National Laboratory
Werner Dahm	Arizona State University
Nicholas Featherstone	University of Colorado
Paul Fischer	Argonne National Laboratory
Travis Fisher	Sandia National Laboratories
Brian Friesen	Lawrence Berkeley National Laboratory
Jonathan Freund	University of Illinois at Urbana-Champaign
Timothy Germann	Los Alamos National Laboratory
Ray Grout	National Renewable Energy Laboratory
William I. Gustafson Jr.	Pacific Northwest National Laboratory
Scott Harper	Office of Naval Research
Willam Harrod	U.S. Department of Energy
Jeffrey Hittinger	Lawrence Livermore National Laboratory
Robert Jacob	Argonne National Laboratory
Kenneth Jansen	University of Colorado, Boulder
Ki-Han Kim	Office of Naval Research
Seung Hyun Kim	The Ohio State University
Douglas Kothe	Oak Ridge National Laboratory
Carolyn Lauzon	U.S. Department of Energy
Chunlei Liang	The George Washington University
Julia Ling	Sandia National Laboratories
Mujeeb Malik	National Aeronautics and Space Administration
Michael Martin	U.S. Department of Energy/AAAS
Georgios Matheou	California Institute of Technology
Sally McFarlane	U.S. Department of Energy
Elia Merzari	Argonne National Laboratory
Robert Moser	The University of Texas at Austin
Joseph Oefelein	Sandia National Laboratories
Katherine Prestridge	Los Alamos National Laboratory
Vadim Roytershteyn	Space Science Institute
John Sabino	The Boeing Company
Ramanan Sankaran	Oak Ridge National Laboratory

Name	*Organization*
John Shadid	Sandia National Laboratories
William Shaw	Pacific Northwest National Laboratory
Jeffrey Slotnick	The Boeing Company
Thomas Smith	Sandia National Laboratories
David Sondak	University of Wisconsin-Madison
Fotis Sotiropoulos	University of Minnesota
Michael Sprague	National Renewable Energy Laboratory
Madhava Syamlal	National Energy Technology Laboratory
Jeremy Templeton	Sandia National Laboratories
Dominic von Terzi	GE Global Research
Qiqi Wang	Massachusetts Institute of Technology
David Womble	Sandia National Laboratories
John Wu	Lawrence Berkeley National Laboratory
Helen Yee	National Aeronautics and Space Administration

Workshop Agenda

Tuesday, 4 August 2015

Time	Speaker	Topic
7:15 AM - 8:00 AM		Check in and continental breakfast
8:00 AM - 8:15 AM	Michael Sprague	Welcome, meeting overview, charge to participants
8:15 AM - 8:45 AM	William Harrod & Michael Martin	Comments from DOE ASCR
8:45 AM - 9:30 AM	**Plenary Talk #1 & Discussion**	
	Jeff Hittinger	Challenges on the path to exascale
9:30 AM - 9:45 AM	Break	
9:45 AM - 11:45 AM	**Working Session #1**: Exascale computing for scientific problems of DOE interest and our fundamental understanding of turbulence	
		Breakouts 1-4
11:45 AM - 1:00 PM	Lunch (provided)	
1:00 PM - 1:15 PM	Breakout Leads	Briefing from Working Session #1
1:15 PM - 2:00 PM	**Plenary Talk #2 & Discussion**	
	John Bell	Path to block structured-grid AMR exascale CFD
2:00 PM - 2:45 PM	**Plenary Talk #3 & Discussion**	
	Paul Fischer	Path to high-order unstructured-grid exascale CFD
2:45 PM - 3:00 PM	Break	
3:00 PM - 4:45 PM	**Working Session #2**: Impacts of turbulent flow simulations at the exascale for DOE applied programs (wind/atmosphere, nuclear, fusion, combustion)	
		Breakouts 1-4
4:45 PM - 5:00 PM		Briefing from Working Session #2 and Day 1 Closing Remarks
5:00 PM	Adjourn	

Wednesday, 5 August 2015

Time	Speaker	Topic
7:15 AM - 8:00 AM		Continental breakfast
8:00 AM - 8:15 AM	Program Chair	Recap Day 1 and Goals for Day 2
8:15 AM - 9:00 AM	**Plenary Talk #4 & Discussion**	
	Jeffrey Slotnick	Simulation of Turbulent Flow on Emerging HPC An Aerospace Perspective
9:00 AM - 9:15 AM		Break
9:15 AM - 11:30 AM	**Working Session #3**: Challenges in adaptation of existing codes and algorithms for turbulent flow simulations at the exascale	
		Breakouts 1-4
11:30 AM - 11:45 AM	Breakout Leads	Briefing from Working Session #3
11:45 AM - 1:00 PM	Lunch (provided)	
1:00 PM - 2:45 PM	**Working Session #4**: Opportunities and challenges of new simulation approaches for turbulent flow at the exascale	
		Breakouts 1-4
2:45 PM - 3:00 PM		Briefing from Working Session #4
		Next Steps
		Closing Remarks