

Outcomes from the DOE Workshop on Turbulent Flow Simulation at the Exascale

Michael A. Sprague*

National Renewable Energy Laboratory, Golden, CO, 80401

Stanislav Boldyrev†

University of Wisconsin–Madison, Madison, WI, 53706

Choong-Seock Chang‡

Princeton University, Princeton, NJ, 08543

Paul Fischer§

University of Illinois at Urbana–Champaign, Urbana, IL, 61801

Ray Grout¶

National Renewable Energy Laboratory, Golden, CO, 80401

William I. Gustafson Jr. ||

Pacific Northwest National Laboratory, Richland, WA, 99352

Jeffrey A. F. Hittinger**

Lawrence Livermore National Laboratory Livermore, CA, 94551

Elia Merzari ††

Argonne National Laboratory, Argonne, IL, 60439

and Robert Moser ‡‡

University of Texas at Austin, Austin, TX, 78712

This paper summarizes the outcomes from the *Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop*, which was held 4-5 August 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.

*Senior Scientist, Computational Science Center, 15013 Denver West Parkway, michael.a.sprague@nrel.gov, AIAA Member

†Professor, Physics Department, 1150 University Avenue

‡Managing Principle Research Physicist, Princeton Plasma Physics Laboratory, P.O. Box 451

§Professor, Department of Computer Science, 201 North Goodwin Avenue

¶Senior Scientist, Computational Science Center, 15013 Denver West Parkway

||Senior Scientist, Atmospheric Sciences and Global Change Division, P.O. Box 999, K9-30

**Group Leader, Scientific Computing Group, Center for Applied Scientific Computing, PO Box 808, L-561

††Principal Nuclear Engineer, Mathematics and Computer Science, 9700 S. Cass Avenue

‡‡Professor, Institute for Computational Engineering and Sciences and Department of Mechanical Engineering, 201 East 24th St.

I. Introduction

THE U.S. Department of Energy (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. 1). The anticipated hardware changes in next-generation platforms (NGPs), will, in many cases, require changes in the algorithms embodied in today's HPC simulation software. The HPC 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. To better understand the opportunities and challenges in computational fluid dynamics (CFD) and turbulent flow simulation that will come with exascale computing, the 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*.² Building upon previous ASCR workshops on exascale computing, participants were tasked with defining a research agenda and path forward that enable 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 summarize in this paper the most important outcomes of the turbulent flow workshop.

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,³⁻¹⁰ the Advanced Scientific Computing Advisory Committee (ASCAC) report¹¹ on "The Opportunities and Challenges of Exascale Computing," and the 2014 Workshop on Exascale Applied Mathematics.¹² The importance of exascale computing is also called out by the Executive Order establishing the National Strategic Computing Initiative¹³ (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:¹³

- 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.
- 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.

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,¹¹

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 1000 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;¹² 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.¹⁴ The goal has therefore been set to achieve exaFLOPS performance with a power limit of 20 MW. 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 determinations 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. Thus, 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 sheer 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 talk¹⁵).

	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

The *Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop* was held on 4–5 August 2015 at the Autograph Mayflower Hotel, Washington, DC. About sixty people attended from national labs, academia, and industry. The workshop was composed of four plenary talks^{15–18} and four break-out sessions that were structured to address four 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 (DNS), engineering simulations in turbulent flows are performed using Large Eddy Simulations (LES) and Reynolds Averaged Navier-Stokes (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 petascale codes. We propose examining this question for turbulent fluid mechanics.

II. Challenges for CFD at Exascale

Several of the well-known issues associated with computational modeling of systems governed by PDE's on anticipated exascale hardware have special consequences when considering turbulence flow simulations. 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 Ref. 19 required 2TB, and an analogous exascale calculation might have a 64TB state. A months-long simulation would generally need to be divided up and performed in smaller pieces (say 12 to 24 hours apiece). 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 checkpointed 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.

III. Opportunities with Exascale Computing

In defining the opportunities that will arise with exascale computing, the workshop focused on scientific discovery and DOE applied programs. The applied programs included (i) combustion, (ii) fusion energy, (iii) nuclear energy, (iv) atmospheric flows, and (v) wind energy. The most important meeting outcomes are described in the following sections.

III.A. Opportunities for Scientific Discovery

III.A.1. Background & History

Since its inception in the early 70's,^{20,21} 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 the identification of intense vortex tubes with diameters that scale with the Kolmogorov scale as the dominant small-scale vortical structure;²² the discovery of an autonomous dynamical mechanism that sustains near-wall turbulence;²³ and the characterization of the scaling and non-uniqueness of turbulent wakes.²⁴ In addition, turbulence simulations have been used to validate and improve turbulence models²⁵ 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 Ref. 26.

The utility of turbulence simulation in fundamental turbulence research is due to two unique characteristics.²⁶ 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 imposed 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 through 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 (N-S) equations is such a reliable description of the dynamics of most flows of simple fluids. The N-S modeling assumptions of Newtonian viscous stress (compressible or incompressible N-S), Fourier heat conduction, and ideal gas thermodynamics (compressible N-S) 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.²⁷ 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 of technological importance will not be accessible to DNS even with exascale resources or for the foreseeable future (see Sections III.E, III.F 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 (UQ).

The simulation of more complicated flow phenomena, such as turbulent combustion, plasma turbulence, and multi-phase turbulence, also involve additional modeling assumptions that 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 Sections III.B–III.F 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.

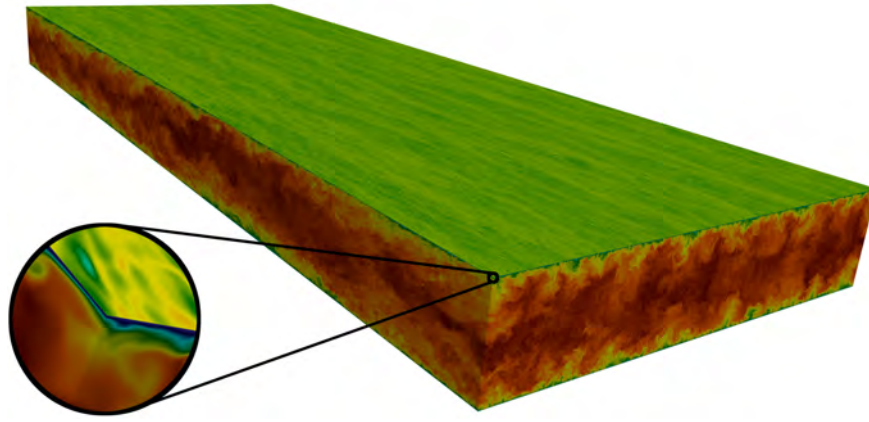


Figure 1. Visualization of the turbulent flow in a planar channel at friction Reynolds number 5200 from Ref. 19. Shown is the computation domain, with streamwise velocity visualized on the sides and the streamwise wall shear stress on the top. Apparent is the large difference in the length scales of the turbulent fluctuations at and away from the walls.

III.A.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 flow in which it is embedded and on the other physical phenomena (e.g., chemical reactions) with which 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.

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 in which such an increase takes the flow into a new regime. An example of this in wall-bounded shear flows is the simulation by Lee and Moser,¹⁹ which, with the increased computational capability offered by the Mira system at the Advanced Leadership Computing Facility (ALCF) at Argonne National Laboratory, 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 1 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.

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 are often 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:

- 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 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. Such model improvements are needed because important combustion applications (e.g., internal combustion and gas turbine engines) are at high pressure with complex fuels and in complex flow configurations. Furthermore, 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 III.B) 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 Ref. 28) 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 than 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 III.C and Ref. 29). 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 Ref. 29).

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 currently 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 Ref. 30 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.

Complex turbulent flows: While 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.

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.

III.B. Applied Research: Combustion

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 of 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 existing infrastructure.

In this landscape, several opportunities exist where combustion simulations at (so far) inaccessible scales will be able to make outsized contributions to improving energy efficiency, reducing carbon intensity and atmospheric pollution emission, and to 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 improve efficiency significantly 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),³¹ may be able to make use of 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 non-traditional fuels (e.g., hydrogen enriched fuels such syngas, typically as part Carbon Capture and Storage systems). Traditional design concerns, such as determining the temperature distribution at the outlet of the combustion section that ultimately determines overall efficiency, 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 intermittency from renewable sources, require new understanding that can be advanced through detailed simulation.

Combustion, a multiphysics and multi-scale extension to the broader turbulence simulation problem, is tackled with a wide variety of methods. Method-of-lines based finite-difference (e.g., S3D³²), finite volume (e.g., LMC,³³ RAPTOR³⁴), and spectral-element (e.g., Nek5000) methods are all used. Most implementations are either higher order, use mesh adaptivity, or both. High-order (≥ 6 th) structured approaches tend to be used more for fundamental research in canonical domains, whereas the lower order (2nd, 4th) methods with adaptivity tend to address problems in non-trivial geometries, incorporated through both cut-cell and generalized curvilinear coordinates approaches. Both h- and p-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 unsteady-RANS (URANS) turbulence models, petascale combustion simulations are typically research tools where either DNS or high-fidelity LES are appropriate. In DNS, no turbulence model is

employed; in high-fidelity LES, the resolution requirement of exacting closures (e.g., Ref. 35) can be met.

Incorporating the multi-physics nature of combustion requires models for chemical reaction, diffusive transport, and, depending on the problem, non-ideal gas effects, spray (two-phase flow, droplet tracking), and radiative heat transfer. Chemical reactions are incorporated by evaluating a reaction network based on an 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 calculations, the time-step is generally limited by the acoustic CFL condition or nearly so, and explicit methods are common. For low-Mach approaches, the relaxation of the flow constraint on the time-step to the advective CFL condition justifies the higher cost per time-step to be able to tackle the chemistry with implicit-methods/backward-differentiation (using, e.g., SUNDIALS/CVODE³⁶). Coupling of the physics through operator-split formulations or alternate multi-rate strategies that preserve higher temporal order^{37,38} are necessary. The combustion community is beginning to experiment with UQ (e.g., the work of Braman et al.,³⁹ Morrison,⁴⁰ Khalil⁴¹ 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.

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 being 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 $O(100)$ 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 reasonably pure components introduce very fine spatial lengthscales that are a limiting factor (down to fractions of a micron), typically occupying a small fraction of a domain in internal combustion engines with relevant space and time horizons of order 10's of cm and 10's 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.^{42,43}), while high-fidelity LES³⁴ (resolving down to c. $10\times$ 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.⁴⁴ 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 be fed into fleet-level analysis with appropriate perturbations to capture transient operation.^{45,46}

There are multiple opportunities for impact of exascale computing on 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 acoustics, 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,¹⁶ 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 the physical model rather than only in 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 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, an opportunity exists for data-driven model discovery for engineering closures such as autonomic closure with model-free determination of subgrid models as well as fundamental science advances. Autonomic closures³⁵ are attractive because the necessary form of assumptions for model validity are typically weaker than for phenomenological models with the tradeoff of a potentially more stringent resolution requirement.

As with other applications, the level of effort needed to establish and maintain performance portability on emerging architectures is of concern. Strategies for including optimizations such as tiling for efficient cache utilization with minimal disruption to existing codes could reduce the effort. As more complex physics are added, dynamic load balancing is needed to redistribute the computation as different physics components exhibit different computational intensities. 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. As simulations become increasingly complex and costly, establishing community consensus about objectives and formulation of a given simulation prior to execution is valuable to be able to perform data reductions *in situ* and to reduce the volume of data that needs to be saved for eventual analysis.

III.C. Applied Research: Fusion

More than 80% of energy consumed in the world is provided by fossil fuels. Even with the development of low-carbon alternatives such as solar, wind, and nuclear energy, this fraction has seen a 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. Fusion is a process that powers stars like our sun. This source of energy, tamed in a laboratory, can provide an energy supply for at least millions of years. Fusion can be a dominant energy, with a number of attractive advantages. The hydrogen isotopes used in fusion experiments, deuterium and tritium, 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 power plants will not have carbon-dioxide emissions, will be relatively compact in size compared to solar or wind plants, and will provide little 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. 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 a sufficiently long enough time. One scheme is 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) at the Lawrence Livermore National Laboratory uses 192 laser beams in order to compress and ignite a 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 for a time sufficient to maintain the fusion reaction before they fly apart. Another, and arguably the most promising, scheme for plasma fusion is to confine particles by a strong magnetic field: a method principally different from inertial confinement. The motion of charged particles across the magnetic-field lines is inhibited to the lowest order, so magnetic-field lines forming toroid-like surfaces can confine a hot deuterium-tritium plasma for sufficient time for the

fusion reaction to occur. The major magnetic fusion energy effort is concentrated around the international project ITER (International Thermonuclear Experimental Reactor), currently under construction in southern France. The goal of ITER is to demonstrate, for the first time, the 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. The actual power plant, DEMO (DEMONstration Power Station), 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 gradient of the plasma pressure provides the source of free energy for instabilities and turbulence. The turbulence and its interaction with the plasma flow turns out to be the major source of confinement degradation.

A variety of mathematical models is employed in addressing the effects of turbulence in fusion plasmas. In the case of magnetic confinement, plasma turbulence couples the multi-scale dynamics among micro-turbulence at gyro-radius scale, plasma profile evolution at device-size scale, zonal flow dynamics at intermediate size scale, and atomic physics at the edge of the plasma chamber. A steady-state fusion plasma condition is determined from a nonlinear self-organization among these complicated multiscale 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 suppresses the edge 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 reaches an efficient fusion burn condition. ITER’s $Q = 10$ goal relies upon this H-mode of operation.

To include turbulence scales comparable to the gyroradius scale, the fluid-like description is not adequate, and kinetic codes must be used. An example is the GENE code⁴⁷ that uses the gyrokinetic description, which is a reduced kinetic description where the fast 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. Gyrokinetic codes provide an efficient description of the physics phenomena whose frequency is smaller than the ion gyrofrequency. Reduction of the dimensionality to 5-dimension allows for computational efficiency of a few orders of magnitude compared to the full 6-dimensional kinetic codes.

As a representative example of an extreme-scale computing, we consider here in more detail the XGC1 multiscale gyrokinetic particle-in-cell code.^{48,49} XGC1’s simulation domain uniquely extends from the material wall boundary to the magnetic axis across the magnetic separatrix, with the neutral particle recycling at the wall, and it covers a broad range of relevant physics, including the H-mode physics. 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.^{50,51} Since XGC1 solves for the total kinetic distribution function instead of a small perturbed part of the distribution function that is used in the scale-separated simulations, it requires more computing power. The more powerful the computers become, the more physics XGC1 can contain. Even an exascale computer may not be enough to give full answers, but it will provide important answers. Figure 2 shows the nonlinear blobby turbulence fluctuation in the edge of a tokamak plasma in a multiscale simulation with gyrokinetic ions, drift-kinetic electrons, and neutral particles.

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 Figure 3), and to the near full-scale homogeneous Mira (10 PF theoretical peak with 786,432 CPU cores). In order to cover the complicated edge geometry, XGC1 uses an unstructured triangular mesh that approximately follows the magnetic field lines. For production runs, XGC1 uses over 300 billion particles on 90% 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 solver takes only a few percent 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 petaFLOPS LCFs is peta-bytes, and one check-point file size is over 20TB. However, it is expected that with high bandwidth memory, in-memory *in situ* data analysis techniques, and non-volatile RAM, these bottlenecks are expected to be eased.

On a future exascale machine, a fusion gyrokinetic code could include most of the important whole-device

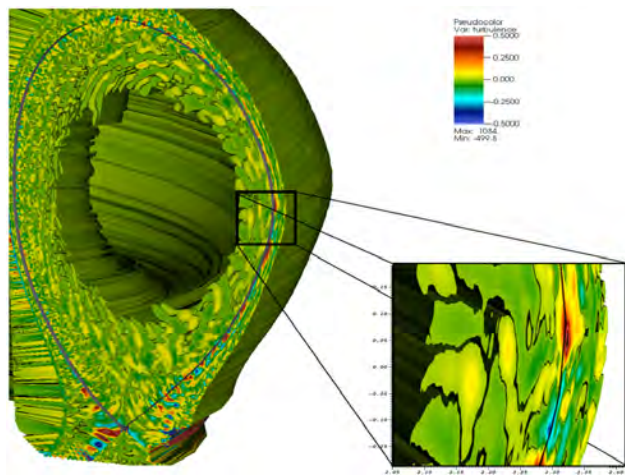


Figure 2. Nonlinear coherent structures called blobs from XGC1 in the edge of a tokamak plasma, regulated by self-generated plasma flows. In the inserted enlarged figure at the right bottom, the flow shearing effect of the turbulent blobs can be seen. The black line in the edge shows the magnetic separatrix surface. (From OLCF Featured Highlight in February 2014 by CS Chang.)

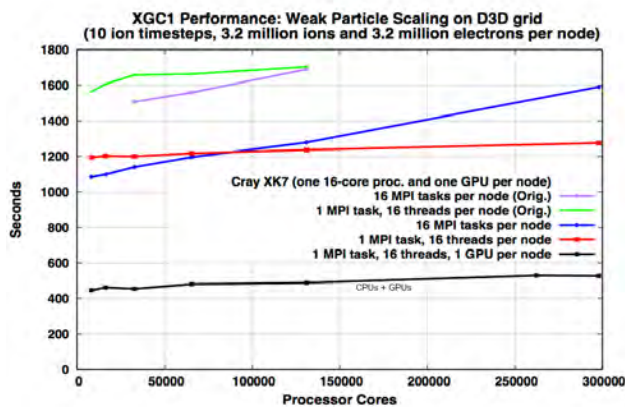


Figure 3. Near perfect weak scalability of XGC1 on heterogeneous Titan to the maximal core (299,008) and GPU (16,384) counts, which allows the multiscale turbulence simulation of ITER. The black bottom graphs are from using both CPUs and GPUs with workload sharing. Strong scaling is similarly good.

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, covering the whole fusion chamber, from the magnetic axis to material wall. The whole-device modeling should also include the plasma heating and current drive, plasma-material interactions, fusion reactions, and the generation of alpha particles and neutrons that will convert their energy to heat. For a complete first-principles modeling of a magnetic fusion reactor without approximations, it is expected that 10 exascale computers are needed. Studies at exascale and post-exascale computers will be 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 the commercial reactor goal.

III.D. Applied Research: Nuclear Energy

Computational fluid dynamics is increasingly used to simulate nuclear reactor flows. Most CFD analysis, especially in industry, relies on the RANS approach and traditional two-equation turbulence models. Higher-fidelity approaches for the simulation of turbulence, such as wall-resolved LES and DNS, can be done only on large supercomputing platforms. In fact, since the Reynolds number dictates the local resolution in wall-bounded flows, large machines are currently necessary to simulate engineering systems with turbulence-resolving techniques.

Nonetheless, 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. 4) with wall-resolved LES. Codes that have achieved these grid sizes include the spectral-element code Nek5000,⁵² PHASTA, and CODE.SATURNE. Codes that have been used for petascale simulations of nuclear systems include HYDRATH, OpenFOAM, and DREKAR.⁵³ These codes share several characteristics on extreme-scale architectures:

1. Meshing is a bottleneck in the workflow at large scales.
2. Algorithms are latency bound at the strong-scale limit.
3. The on-node performance is memory-bandwidth bound.

Each of these performance challenges is expected to be amplified on future architectures. Issues of scaling related to multiphysics coupling are largely unexplored.

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, 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 need to be resolved (see Ref. 54 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 internode 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,⁵⁴ 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.

Given the sheer scale of nuclear systems and the limited reach of current high-fidelity simulations, however weak scaling likely will still play an important role in the near future, although probably at a lower rate of return. 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.

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 5). 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.

All these 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.

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.

III.E. Applied Research: Atmospheric Turbulence

Atmospheric LES models play a critical role in the hierarchy of models. For example, they are important within the development process of physics parameterizations for climate models, and they are important for driving micro-scale-domain models for wind energy. Therefore, the increased computing that will come with exascale is greatly anticipated for increasing both the fundamental understanding of turbulence and clouds in the atmosphere, as well as contributing to improved climate simulations. Toward this end, DOE is funding several efforts to increase the use of LES for atmospheric research. One example is the current project by the DOE Atmospheric Radiation Measurement (ARM) Climate Research Facility to develop 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.⁵⁵ This project is using LES to bridge the gap between observations and large-scale models. Another example is the use of LES to replace traditional physics parameterizations within a climate model. Initial formulations of the multiscale modeling framework (MMF) use cloud resolving models with grid spacings on the order of 4 km to replace the turbulence, cloud, and radiation parameterizations within each climate

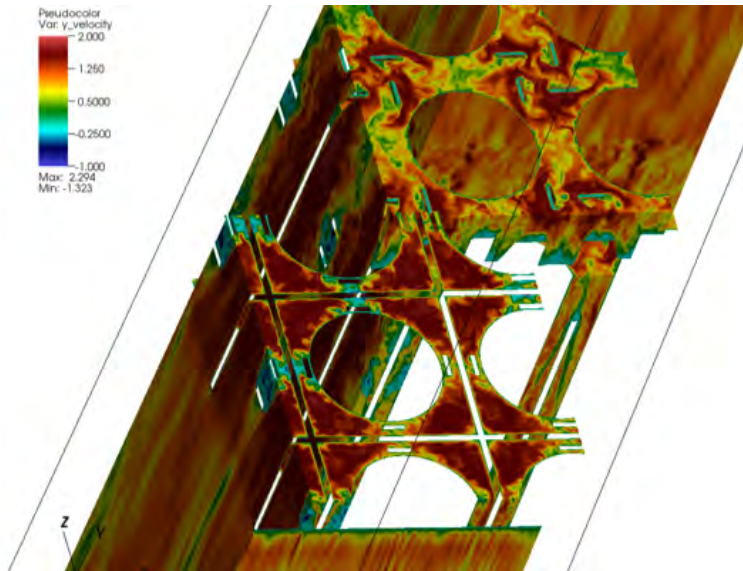


Figure 4. Contour plot of the velocity magnitude (normalized by bulk) for the flow through a fuel-bundle array with spacer grid.

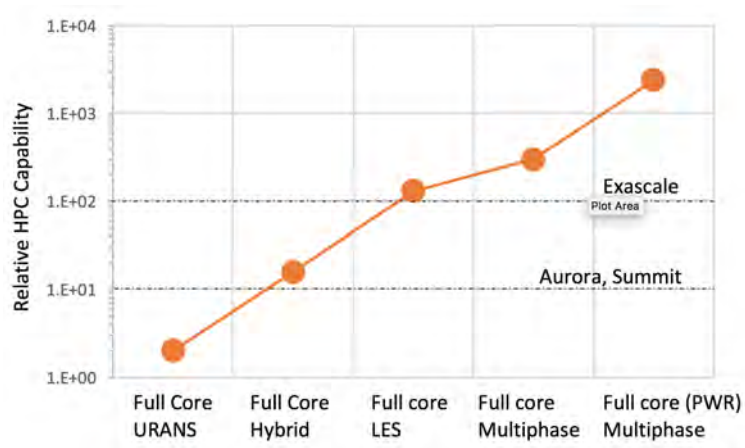


Figure 5. Progress of achievable simulations for a small modular reactor core normalized by current simulation size.

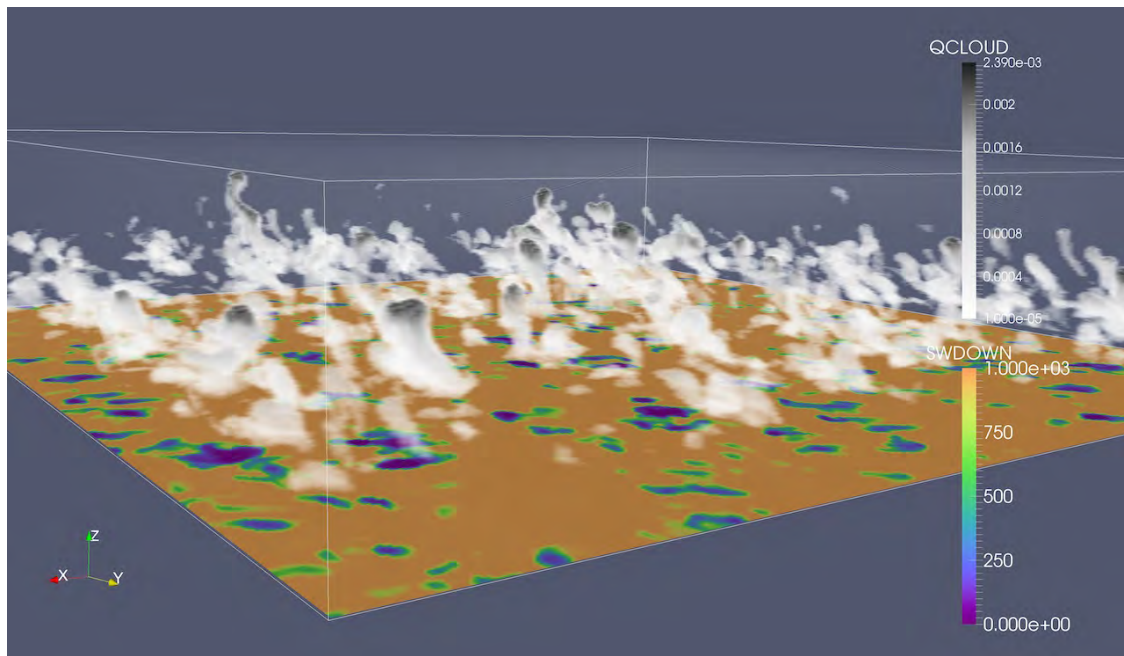


Figure 6. LES is commonly used to simulate clouds and the planetary boundary layer. Shown here are cloud water content (Q_{CLOUD} , kg kg^{-3}) and the resulting shadows that impact the sunlight reaching the ground (downwelling shortwave radiation, $SWDOWN$, W m^{-2}).

model grid column.⁵⁶ Recent funding from the DOE Earth System Modeling Program is being used to investigate the possibility of increasing the resolution of the embedded models to that of an LES as part of the Ultraparameterized Community Atmosphere Model (UP-CAM). The advantage of UP-CAM lies in the ability to resolve more finely 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 and where the problem will scale readily with increased compute cores.

Readying the atmospheric LES models for exascale will not be easy, and the model developers are not so naive as to expect the process to go smoothly. There is a constant tension in modeling between increasing domain size, decreasing grid spacing, and increasing model complexity to take advantage of increasing computing capabilities. The difficulty is that these do not all scale the same way. Increasing domain size typically can scale well in a weak scaling sense, and thus can readily use larger computers. However, decreasing grid spacing requires concurrent decreases in model timesteps. This increases overall integration times, which typically cannot easily take advantage of new computing capabilities except for increased clock speeds, which are expected to be stagnant, or even become slower, at the exascale. Increased physics complexity lies between these two extremes. There is the potential to design the models to use increased concurrency to calculate more complex physics, which will become a balance of how to accomplish this within the constraints of the exascale architecture(s). Given that the limitation of many atmospheric LES results is due to uncertainties in the physics, such as ice microphysics, better physical fidelity is one area where exascale could be very helpful.

One of the largest challenges for growing LES to the exascale will be finding enough ways to increase concurrency to use the available cores. For a given domain size, the limited amount of concurrent tasks will become the ultimate bottleneck. A related technical issue that may become more prevalent at exascale when distributing the work across more cores is load balancing. Cloud and radiation calculations can be a significant portion of the computational expense, and the cost of each within a given grid column tend to be correlated. So, columns with clouds tend to take longer to calculate than those without clouds. The impact of this on load balance will need to be considered carefully as models are redesigned to increase overall concurrency. Currently, static load distributions are used in atmospheric LES models, but other methods may be needed going forward.

There are also two cultural issues that will need to be addressed going forward if the exascale computers

are to be fully embraced for atmospheric LES. The first is the anticipated limited growth in output bandwidth. This will require changing current modeling approaches and the way atmospheric scientists do their computational research. By its nature, atmospheric modeling is concerned with time series of the model state, which are 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. Thus, this has become a critical check for code robustness. If exascale computers cannot achieve this level of reproducibility, even if only for a debug test “at scale,” 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.

Overall, exascale offers a great opportunity to the atmospheric research community. Large regions using LES scales could potentially be simulated, which would enable increased understanding of clouds and turbulence along with how these interact with the surrounding environment. To reach this potential, a concerted effort will be required to redesign the LES models to be able to scale to the exascale machines, as few can even use today’s petascale machines to their full potential.

III.F. Applied Research: Wind Energy

As described in the 2015 U.S. Department of Energy Wind Vision Report,⁵⁷ a national objective is to have 20% of U.S. electricity being provided by wind power by 2020, with 30% being provided by 2030. 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 7 shows the many complicating flow dynamics that can occur in a wind farm. 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 limited understanding of wind plant flow physics. High-fidelity predictive turbulent flow simulations provide the most obvious path towards reducing the cost of energy produced from wind farms, by providing new understanding wind plant flow and a foundation for new capabilities in computer-aided engineering.

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.⁵⁸ 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. Of course, a physically relevant simulation in the micro-scale domain depends on its interaction with, and forcing from, the larger meso-scale domain. This is known as the meso-scale micro-scale

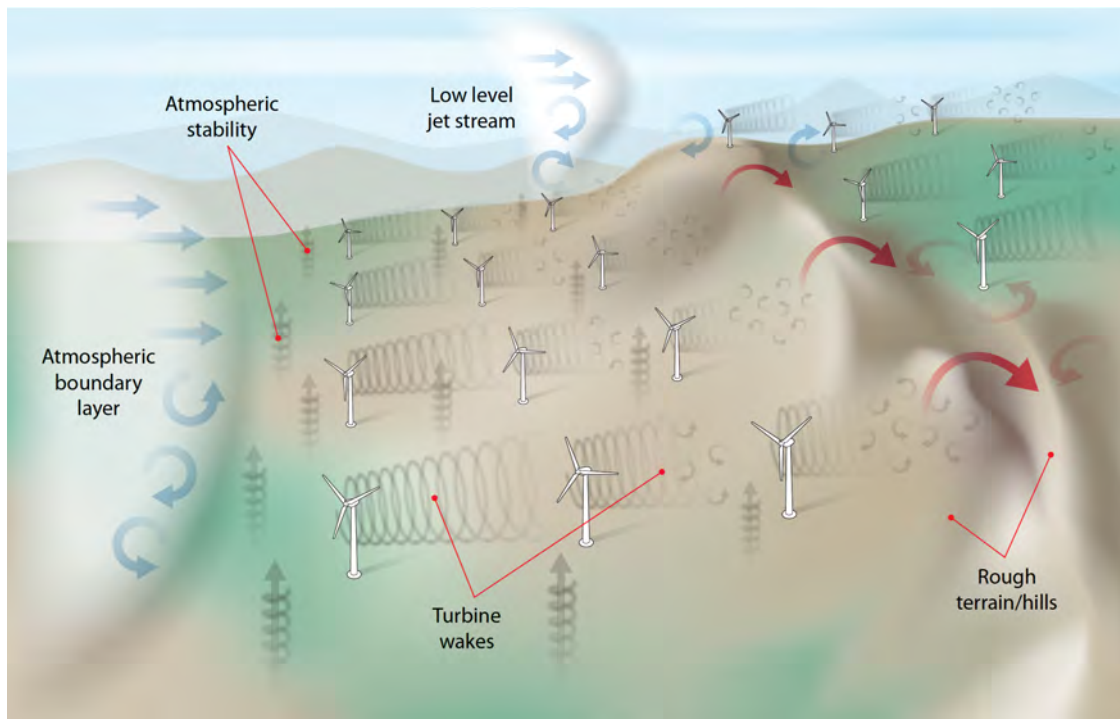


Figure 7. Schematic illustrating the sources that complicate the flow in wind farms, including atmospheric boundary layer dynamics, turbine-wake and wake-wake interaction, and complex terrain. Our inability to predict with confidence wind farm flows presents a barrier to reducing the cost of wind energy. (image source: A. Hicks, NREL)

coupling (MMC) problem (see, e.g., Ref. 59).

For wind farm simulations, flows will be well described by solutions to either the compressible (e.g., Ref. 60) or incompressible/low-Mach (e.g., Ref. 61) Navier-Stokes equations.⁶² However, 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. 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 under unstable atmospheric conditions.

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). However, in simulations where the blade surface is resolved (as opposed to an actuator-line treatment described below), the small scales in the blade boundary layer make WRLES resolution requirements impractical. The boundary-layer region necessitates either a local RANS treatment (in the context of a larger hybrid-RANS/LES or Detached Eddy Simulation (DES) approach) or wall-modeled LES (see, e.g., Ref. 63).

As described above, meso-scale micro-scale coupling deals with the interaction between the micro-scale wind-farm “box” and the larger, meso-scale. 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 turbulence is superimposed on an appropriate mean flow (see, e.g., Ref. 64).
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., Refs. 65,66 and Figure 8). 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 micro-scale simulation (see, e.g., Refs. 60,66). 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., Refs. 59,67,68)

Alternatively, one may simulate an “infinite” wind farm as in Meneveau et al.,⁶¹ where the fluid box has periodic boundary conditions in the horizontal directions.

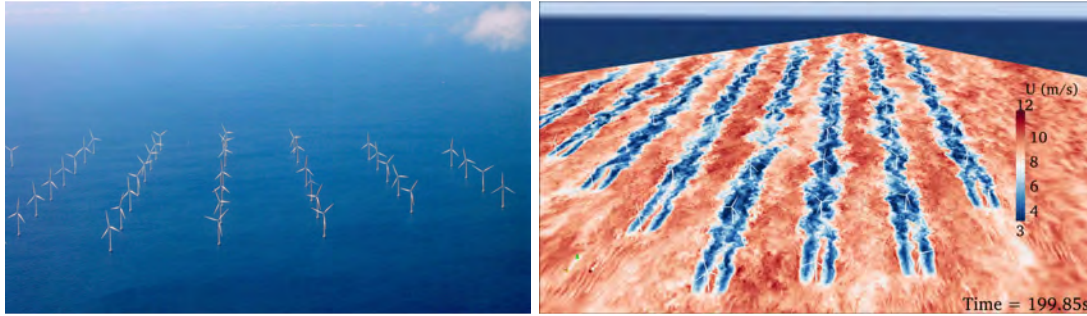


Figure 8. (a) Aerial view of the Lillgrund wind farm, and (b) simulation results showing velocity magnitude. Results were calculated with the SOWFA tool.⁶⁵ (Image provided by M. Churchfield)

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., structural dynamics, pitching/yaw, control system dynamics, etc. There are several whole-turbine models, including the open-source whole-turbine model FAST.^{69,70} 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.⁷¹

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^{72,73} and blade-resolved simulations with RANS,⁷⁴ variational multiscale (VMS),⁷⁵ and DES^{76,77} turbulence models. The highest fidelity wind plant simulations to date have been accomplished with LES and actuator-line turbine representation.^{65,66,78} However, wind plant simulations where the blades are resolved are becoming possible, as in the simulations of Sitaraman et al.,⁶⁰ which employed DES turbulence modeling.

Looking forward, a wind plant simulation capability that is truly predictive will have the following features:⁶²

- 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, and
- 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.⁵⁸ As such, scaling up a single-turbine model to a wind plant composed of hundreds of turbines is a weak scaling of the single-turbine model, which makes wind plant simulation very well suited for exascale computing.

IV. Preparing Today’s CFD Codes for Exascale Computing

Strategies to make productive use of current codes have been considered extensively both in the workshop discussed here and in other venues (e.g. Ref. 11). 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 the 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 straight-forward 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 into intra- and inter-node components. In the case of spectral element algorithms, it is natural to decompose the elements spatially (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 block structure 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.

V. Closing Remarks

The realization of exascale simulations of turbulence 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—high pressure internal combustion engines and gas turbine sectors at pressure—with enough chemical fidelity to address the fuel and operational flexibility concerns that are hampering long-term efforts to drive emissions to near-zero levels while simultaneously increasing efficiency;
- 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; and
- 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.

The prevailing view is that the biggest impacts of exascale computing will be in making “large” problems accessible, where “large” may refer to, e.g., large domains or many physical processes. However, we can only expect modest gains in moving to higher Reynolds numbers. In using future-generation platforms to explore higher Reynolds number flows, focus should be on those flows where a modest increase in Reynolds number will expose transitions to new flow characteristics.

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, 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. All of this suggests that the evolution of hardware architecture that is advancing toward exascale will require that current turbulence simulation-codes be re-tooled just to maintain current turbulence simulation capabilities.

Acknowledgments

The authors wish to thank Michael Martin, the DOE ASCR point of contact, for his assistance in preparing this paper and organizing the workshop. 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. 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 US DOE ASCR and FES through the SciDAC program and by OLCF for the computing resources. Fischer's contribution was supported by US 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.

References

- ¹<http://www.exascaleinitiative.org/pathforward>, April 2016, [Online; accessed 23 May 2016].
- ²"Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop," <https://www.ornl.gov/turbulentflow2015>, August 2015, workshop website [Online; accessed 23 May 2016].
- ³"Scientific Grand Challenges: Challenges in Climate Change Science and the Role of Computing at the Extreme Scale," November 2008, DOE Workshop Report, http://science.energy.gov/~media/ascr/pdf/program-documents/docs/Climate_report.pdf.
- ⁴"Scientific Grand Challenges: Challenges for Understanding the Quantum Universe and the Role of Computing at the Extreme Scale," December 2008, DOE Workshop Report, http://science.energy.gov/~media/ascr/pdf/program-documents/docs/Hep_report.pdf.
- ⁵"Scientific Grand Challenges: Forefront Questions in Nuclear Science and the Role of Computing at the Extreme Scale," January 2009, DOE Workshop Report, http://science.energy.gov/~media/ascr/pdf/program-documents/docs/Np_report.pdf.
- ⁶"Science Based Nuclear Energy Systems Enabled by Advanced Modeling and Simulation at the Extreme Scale," May 2009, DOE Workshop Report, http://science.energy.gov/~media/ascr/pdf/program-documents/docs/Sc_network_shop_report.pdf.
- ⁷"Scientific Grand Challenges: Fusion Energy Sciences and the Role of Computing at the Extreme Scale," March 2009, DOE Workshop Report, http://science.energy.gov/~media/ascr/pdf/program-documents/docs/Fusion_report.pdf.
- ⁸"Scientific Grand Challenges: Opportunities in Biology at the Extreme Scale of Computing," August 2009, DOE Workshop Report, http://science.energy.gov/~media/ascr/pdf/program-documents/docs/Biology_report.pdf.
- ⁹"Scientific Grand Challenges for National Security: The Role of Computing at the Extreme Scale," October 2009, DOE Workshop Report, http://science.energy.gov/~media/ascr/pdf/program-documents/docs/Nnsa_grand_challenges_report.pdf.
- ¹⁰"Scientific Grand Challenges: Discovery in Basic Energy Sciences: The Role of Computing at the Extreme Scale," August 2009, DOE Workshop Report, http://science.energy.gov/~media/ascr/pdf/program-documents/docs/Bes_exascale_report.pdf.
- ¹¹"ASCAC Subcommittee Report: The Opportunities and Challenges of Exascale Computing," Fall 2010, DOE ASCAC Subcommittee Report, http://science.energy.gov/~media/ascr/pdf/program-documents/docs/Exascale_subcommittee_report.pdf.
- ¹²Dongarra, J., Hittinger, J., Bell, J., Chacon, L., Falgout, R., Heroux, M., Hovland, P., Ng, E., Webster, C., and Wild, S., "Applied mathematics research for exascale computing," Report, U.S. Department of Energy, ASCR, March 2014.
- ¹³"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].
- ¹⁴Kogge, P., Bergman, K., Borkar, S., Campbell, D., Carlson, W., Dally, W., Denneau, M., Franzon, P., Harrod, W., Hill, K., Hiller, J., Karp, S., Keckler, S., Klein, D., Lucas, R., Richards, M., Scarpelli, A., Scott, S., Snavely, A., Sterling, T., Williams, R. S., and Yelick, K., "Exascale Computing Study: Technology challenges in achieving exascale systems," 2008.
- ¹⁵Hittinger, J., "Challenges on the Path to Exascale," https://www.ornl.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].
- ¹⁶Bell, J., "Block-structured AMR for Exascale Applications," https://www.ornl.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].
- ¹⁷Fischer, P., "Path to high-order unstructured-grid exascale CFD," https://www.ornl.gov/turbulentflow2015/plenary-talks/Fischer_Paul.pdf, August 2015, plenary talk at Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop [Online; accessed 23 May 2016].
- ¹⁸Slotnick, J., "Simulation of Turbulent Flow on Emerging HPC - An Aerospace Perspective," <https://www.ornl.gov/>

turbulentflow2015/plenary-talks/Slotnick_Jeffrey.pdf, August 2015, plenary talk at Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop [Online; accessed 23 May 2016].

¹⁹Lee, M. and Moser, R. D., “Direct numerical simulation of turbulent channel flow up to Re-tau approximate to 5200,” *Journal of Fluid Mechanics*, Vol. 774, 2015, pp. 395–415.

²⁰Fox, D. G. and Lilly, D. K., “Numerical Simulation of Turbulent Flows,” *Reviews of Geophysics and Space Physics*, Vol. 10, No. 1, 1972, pp. 51–72.

²¹Orszag, S. A. and Patterson, G. S., “Numerical Simulation of Three-Dimensional Homogeneous Isotropic Turbulence,” *Physical Review Letters*, Vol. 28, 1972, pp. 76–79.

²²Jimenez, J. and Wray, A. A., “On the characteristics of vortex filaments in isotropic turbulence,” *Journal of Fluid Mechanics*, Vol. 373, 1998, pp. 255–285.

²³Jiménez, J. and Pinelli, A., “The autonomous cycle of near-wall turbulence,” *Journal of Fluid Mechanics*, Vol. 389, 6 1999, pp. 335–359.

²⁴Moser, R. D., Rogers, M. M., and Ewing, D. W., “Self-similarity of time-evolving plane wakes,” *Journal of Fluid Mechanics*, Vol. 367, 7 1998, pp. 255–289.

²⁵Durbin, P. A. and Petterson-Reif, B. A., *Statistical Theory and Modeling of Turbulent Flows*, Wiley, 2nd ed., 2011.

²⁶Moin, P. and Mahesh, K., “Direct Numerical Simulation: A Tool in Turbulence Research,” *Annual Review of Fluid Mechanics*, Vol. 30, 1998, pp. 539–78.

²⁷Moser, R. D., “On the validity of the continuum approximation in high Reynolds number turbulence,” *Physics of Fluids*, Vol. 18, No. 7, 2006.

²⁸Balachandar, S., Jackson, T., and Fischer, P., “Exascale Simulations of Canonical Turbulent Disperse Multiphase Flows,” White Paper submitted to the Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop, August 2015, papers available at <https://www.ora.gov/turbulentflow2015/whitepapers.htm>.

²⁹Chang, C., “XGC1,” White Paper submitted to the Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop, August 2015, papers available at <https://www.ora.gov/turbulentflow2015/whitepapers.htm>.

³⁰Jimenez, J., “Turbulent flows over rough walls,” *Annual Review of Fluid Mechanics*, Vol. 36, 2004, pp. 173–196.

³¹Musculus, M. P., Kokjohn, S. L., and Reitz, R. D., “Evaluating Temperature and Fuel Stratification for Heat-Release Rate Control in a Reactivity-Controlled Compression-Ignition Engine using Optical Diagnostics and Chemical Kinetics Modeling,” Tech. rep., Sandia National Laboratories (SNL-CA), Livermore, CA (United States), 2014.

³²Chen, J. H., Choudhary, A., deSupinski, B., DeVries, M., Hawkes, E. R., Klasky, S., W.K.Liao, Ma, K. L., Mellor-Crummey, J., Podhorski, N., Sankaran, R., Shende, S., and Yoo, C. S., “Terascale direct numerical simulations of turbulent combustion using S3D,” *Computational Science and Discovery*, Vol. 2, No. 015001, 2009.

³³Day, M. S. and Bell, J. B., “Numerical simulation of laminar reacting flows with complex chemistry,” *Combustion Theory and Modelling*, Vol. 4, 2000, pp. 535–556.

³⁴Oefelein, J., “Large eddy simulation of turbulent combustion processes in propulsion and power systems,” *Progress in Aerospace Sciences*, Vol. 42, No. 1, 2006, pp. 2–37.

³⁵Hamlington, P., King, R., and Dahm, W., “Autonomic Subgrid-Scale Closure for Large Eddy Simulations,” White Paper submitted to the Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop, August 2015, papers available at <https://www.ora.gov/turbulentflow2015/whitepapers.htm>.

³⁶Hindmarsh, A. C., Brown, P. N., Grant, K. E., Lee, S. L., Serban, R., Shumaker, D. E., and Woodward, C. S., “SUNDIALS: Suite of Nonlinear and Differential/Algebraic Equation Solvers,” *ACM Transactions on Mathematical Software*, Vol. 31, No. 3, 2005, pp. 363–396.

³⁷Bourlioux, A., Layton, A. T., and Minion, M. L., “Higher-order multi-implicit spectral deferred correction methods for problems of reacting flow,” *Journal of Computational Physics*, Vol. 189, 2003, pp. 351–376.

³⁸Nonaka, A. et al., “A deferred correction coupling strategy for low Mach number flow with complex chemistry,” *Combustion Theory and Modeling*, 2012, pp. 1–36.

³⁹Braman, K., Oliver, T., and Raman, V., “Bayesian analysis of syngas chemistry models,” *Combustion Theory and Modelling*, Vol. 17, No. 5, 2013, pp. 858–887.

⁴⁰Morrison, R., Moser, R., and Oliver, T., “Representing model inadequacy in combustion mechanisms of laminar flames,” *Bulletin of the American Physical Society*, Vol. 60, 2015.

⁴¹Khalil, M., Lacaze, G., Oefelein, J. C., and Najm, H. N., “Uncertainty quantification in LES of a turbulent bluff-body stabilized flame,” *Proceedings of the Combustion Institute*, Vol. 35, No. 2, 2015, pp. 1147–1156.

⁴²Yoo, C. S., Sankaran, R., and Chen, J. H., “Three-dimensional direct numerical simulation of a turbulent lifted hydrogen jet flame in heated coflow: flame stabilization and structure,” *Journal of Fluid Mechanics*, Vol. 640, Dec. 2009, pp. 453.

⁴³Grout, R., Gruber, A., Kolla, H., Bremer, P.-T., Bennett, J., Gyulassy, A., and Chen, J., “A direct numerical simulation study of turbulence and flame structure in a round jet in cross-flow,” *Journal of Fluid Mechanics*, Vol. 706, 2012, pp. 351–383.

⁴⁴Som, S. and Longman, D. E., “Numerical study comparing the combustion and emission characteristics of biodiesel to petrodiesel,” *Energy and Fuels*, Vol. 25, No. 4, 2011, pp. 1373–1386.

⁴⁵Gao, Z., Curran, S., Parks, J., Smith, D., Wagner, R., M, M. R., Daw, C., Edwards, K., and Thomas, J., “Drive cycle simulation of high efficiency combustions on fuel economy and exhaust properties in light-duty vehicles,” *Applied Energy*, Vol. 157, 2015, pp. 762–776.

⁴⁶Argonne National Laboratory, “Autonomie,” <http://www.autonomie.net> [online; accessed 24 May 2016].

⁴⁷Jenko, F., Dorland, W., Kotschenreuther, M., and Rogers, B. N., “Electron temperature gradient driven turbulence,” *Physics of Plasmas*, Vol. 7, May 2000, pp. 1904–1910.

⁴⁸Chang, C. S., Ku, S., Diamond, P. H., Lin, Z., Parker, S., Hahm, T. S., and Samatova, N., “Compressed ion temperature gradient turbulence in diverted tokamak edge,” *Physics of Plasmas*, Vol. 16, No. 5, May 2009, pp. 056108.

- ⁴⁹Chang, C. S., Ku, S., Diamond, P., Adams, M., Barreto, R., Chen, Y., Cummings, J., D’Azevedo, E., Dif-Pradalier, G., Ethier, S., Greengard, L., Hahn, T. S., Hinton, F., Keyes, D., Klasky, S., Lin, Z., Lofstead, J., Park, G., Parker, S., Podhorszki, N., Schwan, K., Shoshani, A., Silver, D., Wolf, M., Worley, P., Weitzner, H., Yoon, E., and Zorin, D., “Whole-volume integrated gyrokinetic simulation of plasma turbulence in realistic diverted-tokamak geometry,” *Journal of Physics Conference Series*, Vol. 180, No. 1, July 2009, pp. 012057.
- ⁵⁰Ku, S., Hager, R., Chang, C., Kwon, J., and Parker, S., “A new hybrid-Lagrangian numerical scheme for gyrokinetic simulation of tokamak edge plasma,” *Journal of Computational Physics*, Vol. 315, 2016, pp. 467–475.
- ⁵¹Hager, R., Yoon, E., Ku, S., D’Azevedo, E., Worley, P., and Chang, C., “A fully non-linear multi-species Fokker-Planck-Landau collision operator for simulation of fusion plasma,” *Journal of Computational Physics*, Vol. 315, 2016, pp. 644–660.
- ⁵²Fischer, P., Loettes, J., Pointer, D., and Siegel, A., “Petascale algorithms for reactor hydrodynamics,” *Journal of Physics: Conference Series*, Vol. 125, No. 1, 2008.
- ⁵³Smith, T., “Full System Scale Simulations of Reactor Cores Enabled by ExaScale Computing,” White Paper submitted to the Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop, August 2015, papers available at <https://www.ornl.gov/turbulentflow2015/whitepapers.htm>.
- ⁵⁴Merzari, E., Obabko, A., Fischer, P., Halford, N., Walker, J., Yu, Y., and Siegel, A., “Large Scale Simulation of Nuclear Reactors: Issues and Perspectives,” *Proceedings of the 16th International Topical Meeting on Nuclear Reactor Thermal-Hydraulics*, American Nuclear Society, Chicago, Illinois, August 2015.
- ⁵⁵Gustafson Jr., W. I. and Vogelmann, A. M., “LES ARM Symbiotic Simulation and Observation (LASSO) Implementation Strategy,” Tech. Rep. DOE/SC-ARM-15-039, DOE ARM Climate Research Facility, September 2015, 33 pages.
- ⁵⁶Wang, M., Ghan, S., Easter, R., Ovchinnikov, M., Liu, X., Kassianov, E., Qian, Y., Gustafson, W. I., Larson, V. E., Schanen, D. P., Khairoutdinov, M., and Morrison, H., “The multi-scale aerosol-climate model PNNL-MMF: model description and evaluation,” *Geoscientific Model Development*, Vol. 4, 2011, pp. 137–168.
- ⁵⁷U.S. Department of Energy, DOE/GO-102015-4557, *Wind Vision: A New Era for Wind Power in the United States*, April 2015.
- ⁵⁸Churchfield, M. and Sprague, M., “Exascale Computing and Wind Energy,” White Paper submitted to the Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop, August 2015, papers available at <https://www.ornl.gov/turbulentflow2015/whitepapers.htm>.
- ⁵⁹Muñoz-Esparza, D., Kosović, B., Mirocha, J., and van Beeck, J., “Bridging the transition from mesoscale to microscale turbulence in numerical weather prediction models,” *Boundary-Layer Meteorology*, Vol. 153, 2015, pp. 409–440.
- ⁶⁰Sitaraman, J., Mavriplis, D., and Duque, E. P., “Wind farm simulations using a full rotor model for wind turbines,” *Proceedings of the AIAA SciTech 2014 Meeting*, National Harbor, MD, 2014.
- ⁶¹Calaf, M., Meneveau, C., and Meyers, J., “Large-eddy simulation study of fully developed wind-turbine array boundary layers,” *Physics of Fluids*, Vol. 22, 2010, pp. 015110.
- ⁶²Hammond, S. W., Sprague, M. A., Womble, D., and Barone, M., “A2e High Fidelity Modeling: Strategic Planning Meetings,” Tech. Rep. NREL/TP-2C00-64697, National Renewable Energy Laboratory, 2015.
- ⁶³Castro, A. and Carrica, P., “Impact of Exascale Computers on the Simulation of Complex Turbulent Flows of Engineering Interest,” White Paper submitted to the Turbulent Flow Simulation at the Exascale: Opportunities and Challenges Workshop, August 2015, papers available at <https://www.ornl.gov/turbulentflow2015/whitepapers.htm>.
- ⁶⁴Troldborg, N., Larsen, G. C., Madsen, H. A., Hansen, K. S., Sørensen, J. N., and Mikkelsen, R., “Numerical simulations of wake interaction between two wind turbines at various inflow conditions,” *Wind Energy*, Vol. 14, 2010, pp. 859–876.
- ⁶⁵Churchfield, M. J., Lee, S., Moriarty, P. J., Martínez, L., Leonardi, S., Vijayakumar, G., and Brousseau, J. G., “A large-eddy simulation of wind-plant aerodynamics,” *50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, Nashville, TN, Jan. 9–12, 2012*, No. AIAA Paper 2012-537, AIAA, Washington D.C., 2012.
- ⁶⁶Yang, X., Sotiropoulos, F., Conzemius, R. J., Wachtler, J. N., and Strong, M. B., “Large-eddy simulation of turbulent flow past wind turbines/farms: the Virtual Wind Simulator (VWiS),” *Wind Energy*, 2014.
- ⁶⁷Xie, Z.-T., “Modelling street-scale flow and dispersion in realistic windstowards coupling with mesoscale meteorological models,” *Boundary-Layer Meteorology*, Vol. 141, 2011, pp. 52–75.
- ⁶⁸Nakayama, H. and adn H. Nagai, T. T., “Large-eddy simulation of urban boundary-layer flows by generating turbulent inflows from mesoscale meteorological simulations,” *Atmospheric Science Letters*, Vol. 13, 2012, pp. 180–186.
- ⁶⁹Jonkman, J. and Jonkman, B., “FAST v8,” <https://nwtc.nrel.gov/FAST8>, October 2015, [Online; accessed 29 March 2016].
- ⁷⁰Jonkman, J. M., “The new modularization framework for the FAST wind turbine CAE tool,” *Proceedings of the 51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, Grapevine, Texas, January 2013.
- ⁷¹Wang, Q., Sprague, M. A., Jonkman, J., and Johnson, N., “BeamDyn: A high-fidelity wind turbine blade solver in the FAST modular framework,” *Proceedings of SciTech 2015*, Kissimmee, Florida, 2015, also published as NREL tech. report NREL/CP-5000-63165.
- ⁷²Troldborg, N., Sørensen, J. N., and Mikkelsen, R., “Actuator line simulation of wake of wind turbine operating in turbulent inflow,” *Journal of Physics: Conference Series*, Vol. 75, 2007, pp. 012063.
- ⁷³Peet, Y., Fischer, P., Conzelmann, G., and Kotamarhi, V., “Actuator line aerodynamics model with spectral elements,” *Proceedings of the 51st AIAA Aerospace Sciences Meeting*, Grapevine, TX, January 2013.
- ⁷⁴Zahle, F., Sørensen, N. N., and Johansen, J., “Wind turbine rotor-tower interaction using an incompressible overset grid method,” *Wind Energy*, Vol. 12, 2009, pp. 594619.
- ⁷⁵Hsu, M.-C. and Bazilevs, Y., “Fluid-structure interaction modeling of wind turbines: simulating the full machine,” *Computational Mechanics*, Vol. 50, 2012, pp. 821–833.
- ⁷⁶Li, Y., Paik, K.-J., Xing, T., and Carrica, P. M., “Dynamic overset CFD simulations of wind turbine aerodynamics,” *Renewable Energy*, Vol. 37, 2012, pp. 285–298.

⁷⁷Li, Y., Castro, A. M., Sinokrot, T., Prescott, W., and Carrica, P. M., “Coupled multi-body dynamics and CFD for wind turbine simulation including explicit wind turbulence,” *Renewable Energy*, Vol. 76, 2015, pp. 338–361.

⁷⁸Lu, H. and Porté-Agel, F., “Large-eddy simulation of a very large wind farm in a stable atmospheric boundary layer,” *Physics of Fluids*, Vol. 23, 2011.