THE CFD VISION 2030 ROADMAP:
2020 STATUS, PROGRESS AND CHALLENGES

Andrew Cary, The Boeing Company
John Chawner, Pointwise, Inc.
Earl Duque, Intelligent Light
William Gropp, University of Illinois at Urbana-Champaign
Bil Kleb, NASA
Ray Kolonay, Air Force Research Lab
Eric Nielsen, NASA
Brian Smith, Lockheed Martin

Copyright © 2021 United States Government as represented by the Administrator of the National Aeronautics and Space Administration; Lockheed Martin Corporation; The Boeing Company; Intelligent Light; Pointwise, Inc., and The Board of Trustees of the University of Illinois. All Rights Reserved.
Abstract

The CFD Vision 2030 Study (referred to hereafter as Study) contracted by NASA was published in 2014 and has become an aspirational objective for focusing CFD development areas toward aerospace applications. This Roadmap identified the six main technical domains for aligning research as high performance computing (HPC), physical modeling, algorithms, geometry and grid generation, knowledge extraction, and multidisciplinary analysis and optimization (MDAO), with subtopics within each of these maturing from 2015 to 2030. These maturing technologies target the application of CFD to aerospace applications of increasing complexity through 2030. Development of these techniques is fundamental to the successful execution of a series of grand challenge problems. The grand challenge problems are a stepping stone to routine application of CFD to an array of design and development problems with significantly expanded scope and complexity.

Recognizing the ongoing utility of the Roadmap toward improving CFD technology, AIAA formed an Integration Committee around this vision. One of the charters of this committee is to maintain recognition of the Roadmap and highlight the ongoing aerospace CFD research needs. Shortly after a 2019 Aviation session reviewed the status of the Roadmap, a team was formed to provide a more detailed technical review of the Roadmap and identify progress with (and challenges to) following the outlined trajectory. This Report presents a summary of the developments through 2020 related to the technical elements in the Roadmap. While this Report cannot be exhaustive and surely some elements have been omitted, it includes inputs from a group of experts with extensive background in the diverse disciplines represented in the Study. In addition to the authors, other key contributors are acknowledged at the end of the document. Overall, significant progress has been made toward the Vision, but many areas are behind the schedule forecast in the original Study. Several elements and technologies that influence the Roadmap have emerged in the last five years, such as machine learning for turbulence modeling. Relevant technologies that were not included in the original version have been identified. Each of the six domains are reviewed in detail, both for 2020 accomplishments and for overall progress on the milestones identified in the Study.

Overview

This Report provides an assessment of the Technology Development Roadmap developed in the Study (hereafter referred to as the Roadmap) based on the state of CFD technology in 2020, six years after the release of the Study [1]. In addition to highlighting 2020’s accomplishments relative to the Roadmap, an overall review of the different tracks has been performed to assess progress to date.

As shown in Figure 1, the Roadmap charts the development of CFD technology as categorized into six domains.

1. High Performance Computing (HPC)
2. Physical Modeling
3. Algorithms
4. Geometry Modeling and Mesh Generation
5. Knowledge Extraction
6. MDAO
Within each technology domain, several specific elements are tracked along their own timeline. For example, within Algorithms there are elements for solver robustness and uncertainty quantification. These timelines identify key technology milestones and technology demonstrations.

Figure 1. Roadmap graphic from original Study [1].

The Roadmap is a compact and informative summary of future objectives and milestones, but cannot contain the details available within the text of the Study. This Report is intended to review progress toward the Vision since the release of the Study in 2014 and to provide additional details on some of the milestones. This assessment may result in adjusting milestones or adding new milestones based on unforeseen developments to ensure that the Roadmap accurately portrays progress and development needs and supports progress toward the Vision.

2020 marks the first occurrence of technology demonstrations on the Roadmap and hence provides a good opportunity for careful assessment of the status. The three technology demonstrations identified were: “Unsteady, complex geometry, separated flow at flight Reynolds number (e.g., high lift)” in the Physical Modeling Domain, “Grid convergence for a complete configuration” in the Algorithms Domain, and “On demand analysis/visualization of a 10B point unsteady CFD simulation” in the Knowledge Extraction Domain. Considerable progress has been made in each of these areas since the release of the Study and these demonstrations are nearly accomplished.
For the Physical Modeling Domain technology demonstration, there has been steady, incremental progress in the development of unsteady methods for high Reynolds number flight over complex configurations. However, progress is uneven, and significant development remains before these methods are routinely used. At the 3rd AIAA CFD High Lift Prediction Workshop in 2017 [2], only one entrant presented a hybrid RANS/LES method and the results from this method were positive, but not exceptional. Comparatively, the upcoming 4th AIAA CFD High Lift Prediction Workshop (HLPW4) appears to be having significant contributions from time-accurate simulations. While this demonstration is not yet met, NASA has a goal to demonstrate scale resolving methods with flight test certification accuracy for a transport aircraft by 2025. For the Algorithms Domain technology demonstration, grid convergence for simple configurations (such as ONERA M6 wing) have been demonstrated in different tools using different algorithms. Finite element discretizations [3-5] have used strong solvers and adaptive meshes to demonstrate progress toward grid convergence. These techniques are enabling the ability to identify concerns with multiple solutions for RANS problems [6] and what this may mean in terms of predictive capabilities. While progress is being made, there remain questions of numerical uncertainty between fixed grid results and different adaptation approaches for complete configurations. HLPW4 may provide a strong showcase for the objectives of both of these demonstrations through the use of multiple solvers and multiple meshes being applied to the same problem with strong international community involvement. Accomplishing the Knowledge Extraction Domain demonstration at SC19, NVIDIA used a FUN3D-generated data set (150TB, 6B elements/200,000 time steps) of a Mars landing simulation to demonstrate an on-demand visualization that attendees could interact with [7]. This demonstration used state of the art hardware (4 DGX2™ systems each with 16 V100™ GPUs and 16 SSDs for GPUDirect™ storage) to meet this objective.

2020 also saw progress on milestones in other Roadmap domains. A number of benchmark turbulence experiments have been performed to provide validation data to assess and refine turbulence modeling approaches for juncture flow, smooth body separation, and hypersonic canonical experiments for shock-shock, shock-boundary layer, and glancing shock interactions. Enhancements in solver convergence algorithms such as the implementation of the Hierarchical Adaptive Nonlinear Iteration Method developed at NASA Langley into FUN3D provided significant acceleration for complex solutions using CPU-based resources [8]. To support higher-order flow solvers, new approaches have been developed for mesh generation. Additionally, fixed mesh techniques have been improved to increase generation speed and robustness. The need for more flexible and robust geometry modeling is becoming better appreciated through the industry and tools to do this are emerging.

While good progress has been made through 2020 in many of the domains, most have not reached the predicted level of capability. While several of the 2020 technology demonstrations have seen progress, the effort required to accomplish these at present is far beyond what could be used routinely. Whether the milestone is associated with grid convergence of full configurations, automated robust solutions, massive-scale visualization, or UQ-informed optimization, these have not yet reached a level where they can be performed outside of concentrated efforts on leadership-class systems. With progress being demonstrated across multiple fronts, there are strong expectations for significant advances as we continue into the 2020s.
Technology Element Status Updates

This overview is based on the knowledge and data collected by the AIAA CFD Vision 2030 Integration Committee Roadmap subcommittee and contributors. It is intended to be a synthesis of main trends and not a comprehensive review of all published activity. Inevitably, certain endeavors have not been cited or may appear to have been overlooked. Product names and trademarks, where used herein, are for identification purposes only and should not be considered endorsements. This overview is primarily based on aerospace-related communities with the literature search primarily focused in 2019–2020. Each of the following sections are broken into subsections for “Recent Developments” focusing on 2020 and a longer term “Five-Year Perspective” providing additional details on the progress of milestones within each domain.
High Performance Computing

High Performance Computing (HPC) Progress

The High Performance Computing component of the Study is considered an enabling technology across all domains of the Roadmap. The HPC Domain is further organized into a primary element aimed at an evolutionary progression of more conventional hardware technologies, as well as a secondary element intended to monitor community progress toward potential game-changing use of revolutionary hardware technologies such as quantum and neuromorphic computing for CFD applications. The Roadmap specifies decision gates at the years 2020, 2025, and 2030 to perform periodic evaluations of such nascent hardware paradigms to determine if substantial effort should be refocused in such a direction.

Recent Developments

The primary HPC Domain shows three milestones across the 15-year span of the Roadmap. The first appears in the 2019 timeframe and calls for a demonstration of the implementation of algorithms for extreme parallelism in a NASA CFD application. At that time, the Department of Energy (DOE) Summit system located at Oak Ridge National Laboratory (ORNL) held the top ranking as the world’s most powerful HPC system, enabled predominantly by its six NVIDIA® Graphics Processing Units (GPUs) available on each compute node [9]. In Refs. [7] and [10], 2019 computational campaigns on Summit using the NASA FUN3D unstructured-grid solver are described. In these efforts, a comprehensive port to NVIDIA GPUs based on the CUDA™ programming model was used to demonstrate a performance advantage of 4.5x and 6.5x for the NVIDIA Tesla™ V100 GPU over dual-socket Intel® Xeon™ Gold 6148 (40 cores total) and IBM® POWER9® (44 cores total) CPUs, respectively. Excellent scaling to 1,024 nodes of Summit (6,144 GPUs) was observed, with absolute performance equivalent to approximately 1.2 million Intel Xeon Gold 6148 cores and a nominal per-node performance advantage of 36x for GPU-versus-CPU-based simulations on the Summit system. This capability was used throughout 2019 to perform parametric studies of long duration, high-resolution simulations of a supersonic retropropulsion concept for Entry, Descent, and Landing operations of a human-scale Mars lander, using spatial meshes of six billion elements and with each simulation producing several hundred terabytes of output data. A number of other recent efforts are also noted here based on their use of large-scale HPC resources.

In Ref. [11], a study of turbulence modeling strategies as applied to active flow control over a vertical tail/rudder assembly are presented. Simulations were performed using unstructured meshes consisting of approximately one billion elements running on 65,536 cores. Recent computational studies of canonical flows have also leveraged substantial leadership-class HPC resources. In Ref. [12], an asynchronous spectral approach is used to simulate isotropic turbulence on meshes containing as many as 6.3 trillion points using 3,072 nodes of Summit (18,432 GPUs). Refs. [13] and [14] present the computation and analysis of direct numerical simulations of stratified turbulence using 2.9 trillion grid points on 156,816 processor cores.
As demonstrated by the results described above, the NASA FUN3D solver has been successfully ported to NVIDIA GPU architectures and has shown effective scalability to 6144 GPU, with an equivalent computational throughput of roughly 1.2 million Intel Xeon cores. This activity was the result of several years of workforce development through strategic partnering, extensive software modifications, and the adoption of a steady progression of available hardware features. Minimizing data motion across complex memory hierarchies and identifying approaches to substantially increase node-level parallelism were critical. Applying these lessons learned to other NASA applications is ongoing, but the experience highlights some of the challenges the CFD community faces in transitioning to a new computing paradigm.

For the past two decades, application developers have enjoyed a relatively stable environment in terms of the fundamental hardware technology. Periodic hardware refreshes have generally brought additional processing cores, vectorization support, and improved memory performance, but have generally not demanded radical changes to large-scale software implementations. Compilers for common high-level languages such as Fortran, C, and C++ have delivered reasonable performance with minimal developer effort. Node-level parallelism has generally called for O(100) degrees of concurrency, which could be readily achieved through popular shared-memory programming models such as OpenMP [15,16] or POSIX™ Threads, or the ubiquitous message-passing model of MPI [16]. As the HPC community prepares for a new generation of exascale systems, the landscape is undergoing a fundamentally disruptive paradigm shift in the technologies driving today’s most powerful computing architectures. Looming manufacturing constraints and power requirements that grow as a strong function of clock speed have forced vendors to seek improved performance through vastly higher levels of concurrency, or parallelism, using processing elements that often operate at reduced clock speeds compared to prior architectures. Increasingly elaborate memory hierarchies abound and often include High Bandwidth Memory (HBM) ideal for the proliferation of memory-bound applications characterized by motifs with low arithmetic intensity. Trends suggest a steady increase in heterogeneity; that is, architectures over the next decade are likely to leverage highly diverse arrays of processing elements most amenable to very specialized tasks.

Such concerns were strongly noted in the Study, as these fundamental shifts in hardware direction have profound, far-reaching ramifications for application developers. The vast majority of today’s large-scale computational science applications will require a substantial rearchitecting to leverage the potential of emerging next-generation hardware systems. Applications that “stand pat” may in fact perform worse, if at all, on tomorrow’s systems. As an example, consider the GPUs present in many of the world’s current leadership-class HPC systems. In order to realize a significant fraction of peak performance for such processors, application developers must now expose several orders of magnitude more parallelism at the node level. Expressing this degree of concurrency in many of today’s workhorse algorithms can be a formidable challenge and forces developers to reformulate existing algorithms or even abandon them entirely.

The disruption in hardware technologies is in full swing among the upper echelon of leadership-class systems. The RIKEN Center for Computational Science in Japan recently debuted Fugaku, the world’s new top system, which delivers a 442-petaflop LINPACK rating [9] and is based on a new ARM processor from Fujitsu. Within the United States, the DOE will accept delivery of three state-of-the-art HPC systems over the next few years starting in late 2021 with the delivery of the Frontier system at Oak
Ridge National Laboratory. Frontier will use AMD CPUs and GPUs. Argonne National Laboratory will follow with the Aurora system, which will use Intel CPUs and new Intel GPUs. In 2023, the El Capitan system will arrive at Lawrence Livermore National Laboratory. Like Frontier, this system will also be based on AMD CPUs and GPUs and is expected to realize two exaflops in LINPACK performance. The European community hosts a number of systems incorporating similar CPU and GPU technologies, but also aims to field a new architecture of their own in the next few years through the European Processor Initiative [17].

Development teams are now faced with the daunting challenge of migrating, or implementing anew, large-scale applications to deliver performance portability across a broad swath of diverse, complex architectures, which generally call for more expressive programming models beyond the conventional approaches of Fortran, C, and C++. Today’s developer has a plethora of options available offering a variety of advantages and disadvantages, including high-level abstractions such as directive-based approaches and abstraction models, entirely new languages specific to a particular hardware architecture, and machine-specific intrinsics sometimes necessary to achieve peak performance. The mapping of application data and algorithms to hardware, as well as latency hiding associated with memory accesses, network traffic, and I/O subsystems is of paramount importance. Mixed-precision algorithms performed on specialized hardware driven by the rapid growth of the machine learning community can yield substantial performance benefits. The use of asynchronous task-based models leveraging knowledge of algorithmic graph dependencies and a run-time scheduler to dynamically allocate work to idle processing elements offers great promise but may also require a complete overhaul of the target application.

These software development challenges are compounded by the observation that CFD development teams have historically been comprised of domain experts with formal training limited to fields such as fluid dynamics and numerical methods. There is little doubt that the success of future development efforts will strongly hinge on strategic partnerships with computational science experts immersed full-time within the traditional application development team.

Finally, an additional critical pacing item will be the availability of access to emerging and leadership-class architectures. Within the United States, the DOE is tasked with maintaining the nation’s standing at the forefront of the global HPC community and stewards a broad portfolio of leadership-class systems to this end. However, timely and sustained access to these systems can be a considerable challenge for other organizations. Lack of early access to new architectures severely curtails timely development and will result in applications and workflows perpetually lagging the state of the art. Proposal-based programs offer annual opportunities to gain access to production systems, but are generally very high-risk, require an immense amount of coordination to pursue, cannot accommodate sensitive data, and demand rigid project execution plans. Planning and executing long-term research and development programs with such inherent uncertainty in resource availability is untenable.

Nonetheless, if a group does decide to pursue such an allocation with DOE, proposal guidelines generally call for demonstrations showing the application is capable of scaling to at least 20% of the target system, a DOE mandate placed on the Leadership Computing Facilities at ANL and ORNL. Although guidelines permit the potential use of ensembles to meet this requirement, in general, this stipulation can present several obstacles.
First, this metric typically far exceeds the capabilities of an application not routinely used for leadership-class simulations. On the current Summit system at ORNL, this 20% metric corresponds to simulations using approximately 1,000 nodes (6,000 GPUs). To put this in perspective, the performance of a single node on Summit generally provides the computational throughput of over one thousand Intel Xeon cores. Therefore, proposal guidelines generally demand applications be prepared to run on the equivalent of one million Intel Xeon cores. Very few applications, particularly among those specialized engineering codes and workflows demanded by highly nonlinear flows over complex aerospace configurations, are capable of meeting such a requirement. The disparate nature of these simulation scales is often so large that it can be a challenge for development teams to secure sufficient midrange computing resources to begin addressing code scaling deficiencies at even moderate scales, say the equivalent of one hundred thousand Intel Xeon cores.

A more subtle complication arises in the specific application area of scale-resolving CFD, as is precisely called for in the Study. In a time-marching CFD simulation, the 20% metric may generally be addressed through spatial mesh size, p-refinement in the case of high-order methods, or perhaps more elaborate mixtures in the case of multispecies or multiphase flows. However, note that strong scaling can be particularly challenging on new generations of HPC systems where node-level performance can be substantially faster than conventional architectures, while network communication latencies remain similar. This implies that communication hiding — and therefore achieving reasonable strong scaling — can be considerably more difficult. Regardless of how the 20% metric is addressed, the physics of the problem, and perhaps numerical stability concerns, are then used to determine the size of the physical time step. However, one critical parameter selection remains, namely the temporal duration of the simulation. When performing scale-resolving simulations, one is often interested in time-averages over a sufficiently long duration to realize statistically-converged results. It is this combination of spatial and temporal mesh sizes, coupled with the need for statistically-meaningful run durations, that can be a considerable challenge to fit within recommended allocation requests. The situation is further exacerbated by problems that exhibit highly-disparate frequency content, a common trait for many large-scale aerospace applications.

Finally, note that with the imminent arrival of the Aurora, Frontier, and El Capitan exascale systems, the 20% metric will presumably raise the absolute performance requirement by another order of magnitude, such that candidate applications must demonstrate acceptable performance in equivalent environments of approximately 10 million Intel Xeon cores.

Although technical advances such as parallel-in-time methods [18] may eventually help address the challenge of mapping four-dimensional scale-resolving CFD simulations to large compute resources, it remains critical for the aerospace CFD community to establish dedicated access to computing resources managed according to its needs. It is incumbent on the scientists and engineers in the aerospace technical community to cast this advocacy in clear and convincing terms based on tangible mission benefits, such that senior leadership can readily appreciate the value proposition of establishing sustained investments in leadership-class computing facilities.

Looking Ahead to 2030
The third notional milestone in the Roadmap occurs in the year 2030 and calls for a 30 exaflop simulation of, for example, maneuvering flight including a comprehensive analysis of the propulsion system accounting for full combustion physics. While such a goal is still considered highly relevant, the
technical feasibility of this longer-term milestone is difficult to ascertain at this time and will be carefully revisited over the next 5-10 years. Although Japan’s Fugaku boasts an impressive 442-petaflop LINPACK rating, the more appropriate metric may be its HPCG (High Performance Conjugate Gradients) rating [9], widely considered to be more relevant in the context of today’s large-scale science and engineering applications. In this measure, Fugaku achieves 16 petaflops. Viewed in this light, the 2030 milestone calling for a 30-exaflop simulation lies some three orders of magnitude beyond the current state of the art. While this considerable gap may appear daunting at the present time, such a rate of growth is remarkably consistent with the historical trend of the past 25 years [9] and speaks to the meticulous deliberations and foresight of the authors of the Study.

**Revolutionary Systems**

The secondary track of the HPC Roadmap is intended to monitor developments in revolutionary hardware technologies, which may hold the potential to enable disruptive change across the fields of science and engineering. The field of quantum computing holds the promise of radically more efficient computation enabled by basic physical properties of quantum physics. Basic research has led to tremendous strides in this field in recent years and large efforts are actively funded by DOE, the Department of Defense, and the National Science Foundation, among others in the United States, as well as private industry and other nations around the world. NASA has long sponsored the Quantum Artificial Intelligence Laboratory, or QuAIL, project at the NASA Ames Research Center.

A recent overview article commissioned by NASA [19] surveys progress in hardware and software topics related to quantum computing, as well as demonstrations of the technology, which may lead to potential applications of quantum computing for aerosciences. Examples include applications to linear algebra, machine learning, differential equations, hybrid approaches combining quantum computing with classical HPC, and recent research simulating quantum computation of a Poisson solver as applied to CFD [20]. At least initially, however, applications will likely be limited to specific problems and the challenge lies in identifying which aerospace computational problems stand to benefit from quantum computing or can be reformulated to benefit from quantum computing. The authors conclude with an outline of a series of recommended steps to promote early impacts of quantum computing on aerosciences, and specifically state,

"...we need to generate more interaction between aerospace scientists and engineers and [quantum computing] researchers. We have reached a point in the development of [quantum computing] and other quantum technologies where interaction and collaboration beyond a specific narrow scientific field is going to be key to further developments. We need to develop strong engagement between people developing [quantum computing] software and the potential end users in aerospace and engineering (CFD, for example). This implies connections to industry and to agencies such as NASA directly, and also engagement between physicists and engineers to identify the most promising problems that can be addressed in the short- and medium-term..."

The field of neuromorphic computing seeks to realize highly energy-efficient processors inspired by the biological processes governed by neurons and synapses in the human brain. Such processors would appear a natural fit for the fields of artificial intelligence and machine learning, and may eventually find a niche role alongside more conventional hardware. Current implementations include the Intel Pohoiki...
Beach® processor based on its Loihi® architecture, as well as the IBM TrueNorth® processor. Under the ten-year European Human Brain Project [21], the University of Manchester has constructed the one-million core SpiNNaker system aimed at simulating the human brain [22]. A high-level overview of neuromorphic computing recently appeared in Ref. [23] and a comprehensive review of the field can be found in Ref. [24].
Physical Modeling

Physical Modeling Progress
The Physical Modeling Domain encompasses the key modeling technologies required to represent complex physical phenomena for air vehicles including turbulence, transition to turbulence, and complex chemical reaction phenomena related to combustion. The domain is subdivided into four elements: RANS, Hybrid RANS/LES, LES and Combustion. The first three elements related to computational simulation of turbulence are not completely distinct, as aspects of RANS modeling and LES modeling are contained in hybrid models. It is acknowledged that there is a long list of additional physical phenomena that are not included in the Roadmap but that are important for many applications. This list includes icing phenomena, two phase flows in turbulence, real gas effects and plasma phenomena, and modeling for high altitude rarefied gas applications.

Recent Developments
In the Physical Modeling Domain, the most notable highlights of 2020 are the development of numerous carefully defined test cases for testing and validation of both RANS models and scale resolving simulations of turbulent separated flows. These cases include well documented boundary conditions and flowfield mean and instantaneous velocity measurements. These experiments are critical to assessing and improving modeling and simulation to capture complex turbulent flows. There are two foci of these tests: (1) improved prediction of 2-D or 3-D flow separation off smooth bodies and (2) prediction of the flow in the recovery regions downstream of flow reattachment.

The NASA Juncture flow test case evaluates 3-D separation in the adverse pressure gradient region of a wing-fuselage juncture [25]. While the initial test campaign was completed in the fall of 2019, post processing and simulation of data continue into 2020 as well as a planned second test entry. Experimental programs in progress to investigate smooth body separation, reattachment and recovery include the VT bump experiment [26] and the Boeing speed bump [27]. Transonic shock induced separation evaluation of a bump on an axisymmetric cylinder, first tested in the 1970s, is being re-evaluated with carefully defined tunnel wall boundaries at a range of Reynolds numbers by Lynch et al. [28]. These experimental programs are a critical part of the efforts to improve the fidelity of physical models for CFD.

Five-year Perspective
The Physical Modeling Domain is subdivided into four elements: RANS, Hybrid RANS/LES, LES and Combustion. The following assessment largely follows the progress on the following five milestones in the Physical Modeling Domain: (1) Improved RST Models in CFD Codes (2016), (2) Highly accurate RST models for flow separation (2019), (3) Integrated transition prediction (2017), (4) Unsteady complex geometry separated flow at flight Reynolds Number (e.g., high lift) (2020) and (5) Chemical kinetics calculation speedup (2017).
**Improved RST models in CFD codes**

There have been continued efforts toward developing and implementing Reynolds stress transport (RST) turbulence models in CFD codes since 2014. There have been significant efforts at DLR [29] and NASA Langley [30] to implement and refine RST closures. Other notable efforts include implementations by China [31] [32], Superior Tech in Lisbon and Maritime Res. Inst. in the Netherlands [33]. Work from the UK was shown at ECCOMAS in 2016. Recent conference papers document additional RST development work in 2015 [34] and 2020 [35]. While these models are available for use in several codes, outside of DLR they have not penetrated extensively into aerospace industry application. In addition, several popular commercial flow solvers including ANSYS® Fluent® and Siemens’ STAR-CCM+® include RST closures. When applied to flows with swirl and/or curvature these models often give better results than one or two transport equation models, although the improvements are rarely dramatic. RST models tend to perform well in predicting the flow in low pressure turbines [36]. Use of RST models in other industries, including mechanical engineering and flow around buildings, is also widespread [37]. However, rarely do you see multiple codes or organizations using a consistent, “verified” RST model for production CFD applications.

Progress in development and application of RST models is likely to continue over the next decade. Improving predictions of separated flow is a focus of the Roadmap. While these models have shown significant benefit in flows with curvature and swirl as well as corner flows, they have not demonstrated significantly improved predictions of separated flows. The well-known underprediction of the Reynolds shear stress after separation also applies to RST models. In application to the High Lift workshops and some of the canonical test cases shown on the Turbulence Modeling Resource websites, RST model predictions have been comparable, but not superior to other RANS models. RST models did reasonably well in prediction of the Juncture flow experiment, comparable to some of the better RANS model predictions.

It is unclear if use of RST models will continue to expand. Their added computational cost and tendency to be less robust than one- and two-equation models make them less attractive for routine production CFD application. However, there are individuals and institutions who use them routinely and extensively. The Roadmap milestone date of 2016 seems largely irrelevant given that there have been incremental and continuing improvement and adoption of RST models over the past two decades. The inability of RST models to significantly improve predictions of most separated flows means that RST modeling is unlikely to negate the interest in scale resolving simulations for separated flow applications.

**Highly accurate RST models for flow separation**

The Roadmap makes 2019 a decision point for RANS turbulence models. Although incremental improvements based on known deficiencies have been made, the milestone implies that if highly accurate RST models for separated flow are not available by 2019, focus should move to hybrid RANS/LES methods. While it is reasonable to focus more effort on hybrid RANS/LES methods to improve predictions of separated flows, it is not recommended that further research into RANS methods for separated flows and other complex applications be eliminated or severely reduced. The Roadmap shows no milestones on the RANS timeline after 2019. However, RANS methods will continue to play an important role in many aircraft industry applications including conceptual design, optimization, and loads prediction. While RANS modeling, including RST models, are rightly designated to have a high technology readiness level (TRL) on the Roadmap, continued incremental improvements in RANS models are likely.
The development of machine learning (ML) methods to improve RANS modeling predictions was in its infancy in 2014, and this area was not mentioned in the Study. Over the past five years there has been a steady increase in work in this area by researchers around the world. Research is currently being funded by NASA and by the HiFi-Turb consortium in Europe [38]. The level of research in this area has two important implications. First, there continues to be a need for improved RANS methods given their high level of computational efficiency relative to scale resolving methods. If the computational efficiency of RANS over scale resolving methods were not an issue, there would be minimal impetus to develop improved approaches through machine learning. Second, it is possible that RANS models coupled with machine learning could improve predictions of separated flow, whether in a pure RANS solver or within a hybrid method.

The application of ML to turbulence modeling is currently low TRL and essentially unproven. However, as a new approach to a longstanding problem where only slow and incremental progress has been achieved through traditional methods of turbulence model development, continued research and regular assessment of progress in this area is warranted. The NASA “Symposium on Turbulence Modeling: Roadblocks, and the Potential for Machine Learning” planned for March 2022 will act as one of the needed assessments of ML methods. We recommend that ML be added to the RANS timeline with a five-year timeline toward a milestone, which could be a “practical demonstration of machine learning to simulation of a complex flow over an aircraft component”. However, the true need is for an improved model with potential for a wide variety of flows.

Integrated transition prediction (RANS, Hybrid, LES)

The milestone for development of integrated transition prediction has not been met for most critical applications. This is recognized as an area of major importance to expanding the fidelity of CFD for many applications. Prior to 2014, a transport model-based method for prediction of Tollmien-Schlichting 2-D transition was available through the two-transport equation Langtry-Menter model. Over the past six years there has been active research resulting in the development of additional transport-based transition prediction models. Some of these models include additional transport equations meant to improve the accuracy of predictions. There has also been some work to extend transport-based predictions to crossflow instabilities [39]. These models are significantly less robust numerically than RANS turbulence models and extend computational simulation times significantly. Outside of the Langtry-Menter model, there has been little convergence among groups or CFD codes around a single model or approach. There is an active and continuing AIAA working group supporting workshops to assess the accuracy of transition models that can be integrated with RANS CFD flow solvers. The level of activity in this area is indicative of the importance of this subject and the immaturity of the available models.

Current methods based on the Langtry-Menter or other similar models are potentially appropriate for full scale, flight Reynolds number applications to transport and other large manned vehicles where Tollmien-Schlichting disturbances dominate transition. For these applications, with large chord Reynolds numbers, transition typically occurs soon after the streamwise pressure gradient becomes adverse. Current transport equation-based transition prediction models tend to capture these characteristics well. For many other applications, current prediction models are insufficient. Use of these methods for crossflow dominated subsonic, transonic and supersonic flow over swept wings is limited. The extended zones of laminar flow in natural laminar flow vehicles present significant challenges due to weak pressure gradients. The presence of wing-body junctures, pylons, antennae, air data probes and other
protuberances are difficult to include in an automated analysis. Prediction of swept wing laminar flow designs are another challenge. There is limited practical modeling capability for hybrid laminar flow systems. Extended regions of lower Reynolds number boundary layers in small and moderate scale UAVs require high fidelity prediction methods. There are not established, automated methods to account for curvature and wake effects in low pressure turbines. More accurate prediction of transition with significantly less empiricism for this range of flows requires stability-based methods. These methods typically entail difficult to automate processes. Base flow solutions with much greater grid resolution and accuracy than RANS simulations for engineering purposes are required for input to a stability analysis. Transition fronts then must be mapped back onto the geometry for a combined laminar, transitional and turbulent flow simulation.

Hypersonic flows represent a particularly significant challenge. In these applications there are multiple, potentially interacting modes of instability. In addition, hypersonic transition is highly sensitive to environmental disturbances including aerosol particles and freestream turbulence. There has been significant progress in understanding the receptivity processes and in measuring environmental disturbances over the past ten years. Other important factors include real gas effects and ablative processes. There has been significant progress in direct simulation of transition flows, see for example Thome et al. [40]. While this represents a significant advance, it doesn’t really qualify as “integrated transition prediction” in a modeling sense.

There are other practical factors to integrated transition prediction that, while there is some level of understanding of the phenomena, have largely not been included in prediction methods. These include the effects of insects and other contamination, surface erosion, flap gaps and seams, manufacturing tolerances and unintended surface waviness due to flight loads or other factors. These are items that could be potentially part of a detailed transition prediction modeling Roadmap. A Roadmap of this type might be needed as a supplement to the Roadmap, when a reader wants to drill down into this area.

Prediction of transition in wall-modeled LES presents some unique challenges. Methods have been proposed by Bodart and Larsson [41] and by Park and Moin [42]. However, while these show some promise, they are far from mature and are not currently likely to give high fidelity predictions of transitional flow. Published work in this area largely predates the Study.

The 2017 date for this milestone seems far too early. There is a risk that it implies that this area has reached a level of maturity that does not require significant effort or funding. In fact, this is an area where incremental progress has been made in only some of the critical problem areas over the past six years. Failure to make significant progress will limit the ability of CFD-based predictions to meet the ambitious objectives of the Study for many classes of vehicles and flight regimes.

Unsteady complex geometry separated flow at flight Reynolds numbers
The success of hybrid RANS-LES methods for complex geometry separated flows is uneven. Well before the Study, hybrid methods had been successfully applied to high angle of attack tactical fighter applications. This application is simpler than many others because locations of separation onset are largely dictated by sharp leading edges in the thin, swept wings. For other applications, particularly off-design flight regimes such as high lift and lift break conditions, highly accurate simulations are elusive. Hybrid methods are currently most effective where the point of flow separation is fixed by a sharp edge or a shock. This creates a “niche” application, in which VLES and similar methods also have had
preliminary successes. Simulation of flight regimes with separation from a smooth surface without a priori knowledge of the separation location remains a challenge.

The 3\textsuperscript{rd} AIAA CFD High Lift Prediction Workshop \cite{2} documents the state of the art as of 2017 for complex high-lift applications, “Consistently accurate computations near maximum lift conditions remain collectively elusive.” Only one contributor to the workshop applied a hybrid RANS/LES method. Those results did not show superior performance relative to RANS predictions. Since this is the result of a single contributor, it cannot be concluded that hybrid methods do not show great promise or might not be more effective in the near- to midterm future. However, it can be concluded that hybrid methods are not proven to be highly accurate as of today for many separated flow applications.

Applications with extended regions of laminar flow present a challenge for hybrid applications. In particular, flows where separation occurs in a transitional flow regime are extremely challenging for hybrid methods given the difficulties in simultaneously modeling transitional flow and predicting separation. Applications where this can be a problem include hypersonic applications, reentry vehicles, and laminar flow designs.

Unsteady methods such as Lattice-Boltzmann and wall modeled LES (WMLES) show promise for these flows, but additional computational cost is a barrier, and their accuracy in predicting smooth-body separation with current grid counts is even more debatable than that of hybrid methods. However, some codes using these types of models may benefit from GPU implementations, possibly at the cost an extensive rewrite. NASA has committed to “develop and demonstrate computationally efficient, eddy-resolving modeling tools that predict maximum lift coefficient ($C_{L,\text{max}}$) for transport aircraft with the same accuracy as certification flight tests” (by 2025). These could be either hybrid RANS/LES or WMLES. Given the current state of the art, this is an ambitious goal.

The placement of this milestone in 2020 is possibly appropriate if the intention is the demonstration of a low TRL capability for a limited set of applications. However, given the NASA 2025 commitment, it may make sense to add another milestone representing higher TRL, more widely applicable separated flow applications.

Chemical kinetics calculation speedup

The complexity of the chemistry models (chemical kinetics schemes) to be included is determined by the quantities of interest in the combustion simulation. Though the following statements are meant to be general they may be more applicable to gas turbine or automotive combustion. For several time-averaged combustor performance parameters, such as overall heat release and spatial temperature distribution, it is sufficient to include very simple (1- to 4-step) chemical kinetics models, or pre-tabulated chemistry models such as used in flamelet combustion models where the flamelets themselves are computed a priori with detailed chemical kinetic models. These can be done in the RANS or LES frameworks, with LES simulations being more typical for combustor designs that include large recirculation zones and/or strong jet-in-cross flow phenomena. These computations are not very compute-intensive and are performed routinely and chemical kinetics calculation speedup is not a roadblock to performing these simulations.

Other combustor phenomena do require more detailed chemistry approaches and chemical kinetics calculation speedup is required. Examples include predictions of emissions (oxides of Nitrogen [NOx], carbon monoxide [CO], soot), ignition and extinction events, and assessing impacts of unconventional
fuels with significant difference in composition relative to conventional fuels. While NOx, and to some extent CO, can be reasonably well predicted with an extension of simple chemistry models (and extensions to the detailed chemistry in flamelet models), predictions of soot entails more complexities, not only in chemical kinetic models, but also in the models that describe the physics of soot production and oxidation. Soot emission simulations can also be done in the framework of flamelet combustion models with precomputed chemistry. Soot concentration predictions using simple chemistry models computed in situ or using flamelet combustion models are also routinely performed, though the underlying soot models still need significant improvement and predicted soot concentrations often differ from measurements by one or more orders of magnitude, as well as challenges in even producing correct soot emission trends.

For simulations of inherently unsteady processes such as flashback, lean blow-out, and ignition, LES is invariably needed and sufficient details including many minor species and radicals need to be included in the chemistry model. It may be important to include in situ detailed chemistry computations in the simulations. Further, if we need to differentiate based on fuel composition, for example while evaluating the performance of conventional or alternate fuels, it is all the more important to include the detailed kinetics that describe the different fuels.

There are several approaches to include detailed chemistry that consists of thousands of reactions and hundreds of species. One approach is to use “laminar-chemistry”, which neglects the effects of turbulence and uses very efficient ordinary differential equation (ODE) solvers. Stochastic approaches, such as stochastic field methods, transported probability density function methods or linear eddy model, can be used to include turbulence-chemistry interactions. There are also adaptive approaches that employ different chemistry models and/or different turbulence-chemistry models in different regions of the spatial domain. Methods such as artificial neural networks (ANN), in situ adaptive tabulation (ISAT) and similar methods can be used to store and access chemistry computations in-line during the simulations. Additionally, approaches to reduce the size of the chemistry to hundreds or tens of reactions and tens or single-digits of species (skeletal and reduced mechanisms, respectively) can be employed. These reductions can be done a priori or done on the fly using techniques like intrinsic manifolds or rate-controlled constrained equilibrium (RCCE). These methods are currently available and are being used. However, newer approaches and ways of solving them more efficiently in parallel environments continue to be devised.

There are two other outstanding issues in practical combustion CFD that are being addressed but need more attention. First, modeling of soot and other nonvolatile particulate matters is becoming more important due to increased environmental concerns. Second, modeling of atomization at relevant operating conditions is key to accurate descriptions of the spray droplet distributions and other properties are key to obtaining accurate predictive simulations of emissions and unsteady phenomenon such as ignition and blowout. With increased interest in scramjet and ramjet combustors these methods need to be extended to combustion in the high-speed compressible flow regime. Of course with the increase in the complexity of the simulations that these advances will entail, we need to devise efficient computational techniques that can leverage massively parallel exascale computing architectures.
Algorithm Progress
The Algorithm Domain of the Roadmap includes timelines for both numerical algorithms convergence/robustness and uncertainty quantification (UQ). The overall objectives identified in this domain are associated with estimating the uncertainty in CFD simulations, both through assessment of the sensitivities in the results as well as reducing numerical errors associated with grid resolution and iterative convergence. Technology milestones targeted to have been reached by 2020 include development of technology for automated robust solvers, characterization of uncertainties in aerospace, and reliable error estimates in CFD codes. These were to lead to a 2020 technology demonstration of grid convergence for a complete configuration.

Recent Developments
Improvements to solver robustness have been demonstrated through the use of the Hierarchical Adaptive Nonlinear Iteration Method (HANIM) implemented in both USM3D and, this year, FUN3D. This enhanced solver algorithm has led to significant improvement in flow solver convergence on a number of NASA programs. For example, in the Transonic Truss Braced Wing (TTBW) project [43], USM3D HANIM was able to provide accurate solutions in 15-20 minutes using 520 Ivy Bridge cores at the NASA Advanced Supercomputing (NAS) facility’s Pleiades supercomputer. Other CFD codes may require 30X the runtime for attaining solutions on grids with similar degrees of freedom and using similar computational resources. USM3D HANIM was also used by NASA’s Aerosciences Evaluation and Test Capabilities program for Unitary Plan Wind Tunnel simulations that encountered challenging flows with the Mach number range from 0.01 in the settling chamber through 4.6 in the test section. For such flows, USM3D HANIM required 10 minutes or less to compute solutions on 1040 NAS Ivy Bridge cores, while other codes used in the project needed hours or days to compute similar solutions. This approach has also demonstrated more than an order of magnitude improvement in convergence time for FUN3D over the legacy iterative solver on several benchmark flows, including the juncture-flow experiment configuration [8]. FUN3D HANIM was also used by the Global Center for Medical Innovation task force for N95 mask design in the global response to COVID-19 pandemic.

Progress in using UQ with CFD simulations has not proceeded as rapidly as envisioned in the Study, but advancements are being made in both development of appropriate methods for performing UQ analysis within the context of CFD simulations and in applying these methods to CFD analysis. Most CFD applications focus on nonintrusive techniques and many use polynomial chaos to develop surrogate models for propagating uncertainty. These methods are being applied to two to three uncertain variables for engineering assessment [44,45], but applications looking at turbulence model or geometric sensitivities have pushed toward order 10. An important aspect of CFD uncertainty is determination of discretization error uncertainty; most often it is assumed to be smaller than other uncertainties and is therefore neglected. This is not clearly valid for all aerospace applications as illustrated by the AIAA CFD Drag Prediction Workshop series and the AIAA CFD High Lift Prediction Workshop series. Virginia Tech and NASA are performing benchmark experiments involving smooth walls and turbulent separation with care to meet exacting requirements for CFD validation [46]. This information will be useful to assess
Algorithm

model form error uncertainty. Because of the cost of individual simulations and the number of simulations required for uncertainty assessment, an increasing focus is being brought to multifidelity techniques using CFD [47,48]. General awareness of UQ techniques beyond Monte Carlo is increasing across the community and multiple organizations are collaborating to provide a consistent terminology and recognition of different viable approaches [49].

Five-year Perspective

The improvements identified in the Algorithms Domain require significant advancement in several areas of CFD, beginning with enhancing solver robustness leading to improved convergence levels in terms of both iterative convergence and grid convergence. Improvements in grid convergence have been further aided by advances in solution-based mesh adaptation. However, the typically highly-skewed meshes resulting from mesh adaptation increase the challenges for robust solver convergence, leading to further demands for robust and efficient algorithms. The increasing demand for scale-resolving simulations is also creating demand for robust low-dissipation numerical schemes.

Flow solver algorithms have seen advances in not only individual elements of the solution, but also with efficient combinations. Linear solvers for unstructured grids are increasingly extending beyond point-relaxation techniques in production codes while pseudotransient continuation with backtracking line search has become widespread, leading to increased robustness. Newton-Krylov [50,51] solvers have also been demonstrated to lead to effective convergence in monolithic/tightly coupled multidisciplinary problems [52]. However, these techniques often require reaching basically infinite CFL numbers in the pseudotime marching to steady-state, which makes them susceptible to instabilities in the underlying equations. This presently limits their ability to be applied to some physical models where traditional solver techniques can work. In some cases, such as the SA-negative model, the physical model is modified to enable the use of Newton-Krylov methods.

In the past five years, advanced solution methods for RANS equations on unstructured grids have been implemented and assessed in USM3D, which is a prominent cell-centered finite-volume RANS solver and a part of the NASA Tetrahedral Unstructured Software System [50,51,53]. An overriding attribute of the USM3D flow solver has been its speed and robustness in providing solutions for a broad class of aerospace vehicles. In 2016, a new strong nonlinear solver named Hierarchical Adaptive Nonlinear Iteration Method (HANIM) [54,55] was implemented in the mixed-element USM3D to further improve robustness and efficiency of RANS solutions. HANIM is a highly scalable iterative solver that uses an adaptive pseudo-time approach to accelerate iterative convergence. HANIM extends the simple preconditioner of USM3D by providing two additional hierarchies around the preconditioner. The hierarchies are a matrix-free Newton-Krylov linear solver for the exact linearization of discrete RANS equations and nonlinear control of the solution update. The HANIM methodology has been assessed on several turbulent flow benchmark cases [54,55,56] involving multiple grid families and linear [57,58] and nonlinear [59] turbulence models. HANIM has been able to achieve machine-zero residuals on all grids, including a grid of 1.5 billion cells [60], while the baseline solver could not attain target low levels of residuals on many grids. HANIM speedup factors up to 72 have been demonstrated relative to the USM3D baseline solver.

In 2019, the USM3D HANIM methodology was extended to unsteady flows and high-speed steady-state flow applications involving complex geometries, strong shocks, and highly irregular grids. These applications spurred additional HANIM enhancements improving robustness and maturing capabilities
Algorithm

(flux limiters, a variety of boundary conditions, flux functions, etc.). USM3D HANIM was successfully used in several NASA projects where it provided robust and dramatically accelerated solutions for NASA configurations. In 2020, this methodology was transferred to another flagship NASA CFD code, FUN3D [61-63], featuring a node-centered finite-volume formulation. FUN3D HANIM was applied to challenging unsteady and low-speed steady turbulent flows. Initial demonstrations of FUN3D HANIM for rotary-wing applications indicate significant improvements in efficiency and robustness of rotorcraft-flow solutions. FUN3D HANIM was successfully used by the NASA Low Boom Flight Demonstrator project for analysis of cooling flow supply. In this application, HANIM made the difference between converged solutions and no-solution for very complex flows, grids, and geometries.

In addition to enhancements for traditional finite volume CFD discretizations, enhancements have been made with other schemes. There are multiple codes successfully using $p=1$ Streamline Upwind Petrov/Galerkin (SUPG) discretizations for RANS applications [6,64,66]. These schemes are demonstrating higher accuracy for the same degrees of freedom as finite volume schemes and use robust Newton-Krylov solvers to achieve reliable convergence. Other methods such as Galerkin/Least squares and discontinuous Galerkin formulations are also showing progress. The net solution cost for a given level of accuracy is still debatable [52]. High-order workshops have demonstrated third, fourth, and fifth order ($p=2,3,4$) solutions on complex industrial problems, laying the foundation for production applications. The results of 2D multielement high-lift computations demonstrate the computational savings potential from using higher-order stabilized continuous Galerkin discretizations in combination with the MOESS output-based mesh adaptation method. Higher-order methods, in particular $p=2$ and $p=3$ continuous Galerkin variational multiscale with discontinuous subscale discretizations, provide accurate outputs with an order of magnitude less computational time than $p=1$ methods [66]. These approaches lead to high accuracy on coarse grids, but require curved-elements—a new grid generation challenge. Additionally, the solution cost remains an issue for these approaches and improvements in solution techniques will be necessary to make them competitive for steady RANS problems.

Furthermore, Lattice Boltzmann codes have been demonstrated to provide industry-level solutions on multiple aerospace applications including aeroacoustics and general unsteady/separated flow applications [67]. The Eddy code [68,69] is a space-time spectral-element Discontinuous-Galerkin solver for complex separated flows. This software is intended to enable researchers to collaborate through a common framework, which enables three-dimensional simulations on practical problems.

There has also been an increase in the availability of adjoint solvers within commercial codes with applications for error estimation and multidisciplinary problems. Time-dependent adjoint solutions are also seeing more widespread development and use within the research community. These additional solvers enable a number of different applications that were not feasible prior.

Various researchers [52,66] have shown that high-order methods provide a benefit in terms of accuracy versus cost for a range of problems including scale-resolving simulations. For these high-order schemes, nonlinear stability can become a limiting factor and has led to an increased focus on nonlinearly stable discretizations. One class of these appropriate for scale-resolving methods are those that preserve total kinetic energy (TKE) and are entropy stable. These algorithms have also been shown to benefit low-order methods because of their low dissipation. For the high-order algorithms, the emergence of tensor product discontinuous Galerkin (DG) high-order methods with TKE-preserving or $S$-stable properties provide promising candidates. Additional enhancements have been developed for implicit temporal
discretizations to support scale-resolving methods including space-time DG and associated preconditioners for the solution of implicit time steps.

In addition to the methods, which allow extension to higher orders, progress has been made on schemes similar to traditional methods that are less computationally expensive than methods extensible to higher orders but have lower dissipation and become formally third order on tetrahedral meshes [70,71]. Additionally, several unstructured grid MUSCL schemes have become popular based on their reduced dissipation compared to typical approaches and minimal additional cost [72].

Highly automated CFD solutions using mesh adaptation have been demonstrated [73], but these solutions tend to use meshes that are very irregular and highly skewed, necessitating a robust discretization and solver. Hyperbolic methods for the Navier-Stokes equations have demonstrated advantages for this type of mesh in terms of both accuracy and iterative convergence times [74-80].

Uncertainty quantification has continued to see slow penetration into CFD problems over the last five years. The JANNAF guide [81,82] from 2015 remains a key reference. In 2016, Barth [83] received the Fluid Dynamic Best Paper award for his overview of application of uncertainty estimates to CFD computations. Although there has been an increase in the number of publications each year, the community is still limited. With the cost of each CFD simulation and the number of uncertain quantities, the use of Monte Carlo techniques has been regarded as infeasible for most applications. This has led to applications of UQ in CFD relying on various types of surrogate modeling, including radial basis functions, Gaussian Processes, and nonintrusive polynomial chaos (NIPC) [84]. The AIAA Community of Standards released a preview [85] of their updated standard that provides a general framework for describing CFD uncertainty. This framework provides indications of identifying, characterizing, propagating, and analyzing uncertainties. Important aspects of this characterization are separating random (aleatory) uncertainty from lack of knowledge (epistemic) uncertainty. Some of the key epistemic uncertainties in CFD include the discretization error estimation and model form uncertainty.

An area that has seen particular focus is assessing the uncertainty of turbulence model closure coefficients [86-89]. These examinations provide an efficient starting point to systematically examine the consequences of different choices for model coefficients to assess one aspect of model form uncertainty. One of the first examinations of this type of problems for full configurations was performed by Schaefer [90]. A similar application of this approach was applied to the High-Lift Common Research Model (CRM) [91]. An alternate approach for estimating turbulence uncertainty is based on the Lumley triangle and has demonstrated appropriate trends for capturing differences between experiment and simulation [92].

Another key aspect of trying to leverage CFD data for UQ is the use of multifidelity techniques where different levels of CFD approximation or other data can provide cumulative insight into uncertainties by leveraging their individual strengths. Examples of this type of approach are given by West [48] and Wendorff [93]. The DARPA-sponsored Scalable Environment for Quantification of Uncertainty and Optimization in Industrial Applications (SEQUOIA) [94-96] considered several representative applications using multifidelity techniques for both UQ and robust design considerations.

As this field continues to grow, benefit will be achieved by pursuing the milestones outlined in the Roadmap for characterizing uncertainties that are likely to appear in aerospace applications, developing
reliable error estimates for CFD results, and expanding propagation methods to be particularly appropriate for simulations with relatively high individual costs.
The Geometry Modeling and Mesh Generation Domain of the Roadmap (formerly Geometry and Grid Generation) contains two elements for Fixed and Adaptive Meshing with milestones for achieving tighter coupling with CAD software and the addition to CFD code of production-level adaptive mesh refinement from the period before 2020. Future milestones include large-scale parallel meshing in the 2021 timeframe and automated, adaptive meshing in 2030. Given that the Study places geometry modeling and mesh generation at the center of its call for autonomous and reliable CFD simulations, progress in these technology areas is crucial for achieving the Vision.

**Recent Developments**

**Mesh Generation**

High-order, curved meshing has been the focus of much fixed mesh research in the past year. Roca [97] continued development of a quality-preserving, linear-to-curved mesh morphing algorithm. Karman [98] evolved a technique for optimization-based smoothing to allow for element-by-element variation in polynomial degree thereby opening the door for hp-adaptation. In a novel approach, Marcon [99] used cross fields to generate curved, quadrilateral meshes.

The general trend in fixed meshing research is the resolution of specific robustness or quality issues via new techniques, support for specific cell types and topologies, or meshing strategies. As an example of the latter, Ito [100] implemented a remeshing strategy that was shown to facilitate rapid geometry model changes. Steinbrenner [101] extended a boundary layer resolving hybrid meshing technique with the addition of isotropic, size-field-based hexahedral mesh away from the boundaries. Hu et al. [102] extended a novel and robust method for generating a tetrahedral mesh within any “triangle soup” of surface facets.

With respect to meshing software’s exploitation of HPC resources, commercial CFD and meshing codes are commonly running cases with cell counts in low billions [103-105]. Tens of billions of cells are reported as technically possible, but few customers are reported as using meshes that large. Most HPC meshing is being done on local servers and experimentation is ongoing in the cloud as well [106]. National labs have access to larger computers than commercial customers, but the reported meshing is still in the low tens of billions. The national labs seem to be I/O bound with their problems. They have spent quite a bit of research compressing their CFD data so that it can fit on existing compute clusters and the workstations they use to post-process their results [107-108]. The U.S. Dept. of Defense is directly supporting the HPCMP CREATE program with meshing software said to be generating "10s of billions" of cells at the upper end [109-110].

The automation of initial and adaptive mesh generation has improved for complex boundary representation (B-Rep) features, surface curvature, and underlying surface parametrization [111]. The verification of adaptive unstructured mesh methods has continued for analytic functions [112], ONERA M6 wing [113], and the High-Lift Common Research Model (HL-CRM) airfoil [114]. These cases show more convincing mesh convergence of engineering outputs than is typically seen (e.g., at prediction workshops) with fixed meshes. Improvements to an a priori anisotropic output-based metric [115] and...
Mesh Optimization via Error Sampling and Synthesis (MOESS) [116-117] were published. Application papers are starting to emerge [73] including comparisons to fixed mesh approaches showing mesh adaptation reducing human interaction time as compared to a fixed-mesh approach [118].

Solvers that are not traditionally used for external aerodynamics such as CONVERGE [119] and Cart3D [120] continue to evolve a highly-capable Cartesian cut-cell approach to adaptive mesh refinement. Dual mesh (unstructured near-body and Cartesian off-body) approaches are implementing mesh adaptation techniques for the near-body mesh [121].

Geometry Modeling
Geometry modeling is proposed to be introduced as a new element for the Roadmap in 2020. Thanks in part to prior products associated with the Study [122], the needs for simultaneous access to multiple forms of geometric representation are becoming better appreciated throughout the aerospace engineering community. Communication of these needs continues [123] and applications in commercial off-the-shelf (COTS) software are emerging with increasing frequency. An example of the latter is the incorporation of topology optimization workflows into the latest release of PTC® Creo® Mechanical CAD (MCAD) software [124]. In Ref. [122], it was noted that “we find ourselves in a situation where we can no longer design objects that we can make using the MCAD modelling systems that form the backbone of most industrial MBE enterprises.” With the advent of Creo v7.0, this is, arguably, no longer the case, at least for structural design. However, considerable work remains to be done before such workflows are suitable for use with CFD or the sensitive aerodynamic design activities envisaged in the Study.

While COTS MCAD platforms fail to provide access to the underlying geometry representation in the manner required by advanced CFD applications, interest in furthering the capabilities of bespoke geometry modelling systems continues. For instance, improved capabilities for generating and repairing B-Rep models have been incorporated into CREATE™ Capstone [125]. Also, development of the Geode™ geometry kernel [126] and the associated MeshLink™ mesh-geometry associativity [127] provide a virtual topology interface that makes B-Rep models more suitable for meshing. In addition, researchers at the University of Bristol continue to develop improved ways of defining and parameterizing Outer Mold Lines (OML) for aerodynamic design. Publications in the last year [128] have focused on new methods for defining orthogonal modes and constraints for preserving continuity in both two and three dimensions. Most of the published examples apply to airfoil test cases, although their initial application to wing OML (in isolation) have also been reported.

Five-Year Perspective
In reviewing the progress made in geometry modelling and mesh generation, the intent is to provide a brief synthesis of what the authors consider to be the main trends rather than a catalogue of all published activity in the area. The text in this section has been adapted from a recent survey [122].

Geometry Modeling
In recognition of the need for improved awareness, preparation of an AIAA Guide explaining the underlying concepts of geometry modelling as they apply to computational simulation has commenced. A technical paper with preliminary information has been published [129].

The 1st AIAA Geometry and Mesh Generation Workshop (GMGW-1) [130] provided a compelling illustration of geometry model interoperability issues. A model of the NASA High Lift Common Research Model (HL-CRM) was developed for the workshop and disseminated to participants in a range of widely
used file formats [131]. All participants reported that some level of modest repair was required and the portions of the model that were problematic varied from mesh generator to mesh generator [132]. However, when details of a geometry model developed to support the manufacture of a wind tunnel model of the HL-CRM were presented to GMGW-2 participants 18 months later [133], the reaction was one of general concern that the workload involved in preparing it for meshing would likely make the model unsuitable for use in a workshop setting. GMGW-1 also showed that interoperability problems can be exacerbated by mesh adaptation, since the maximum permissible modelling tolerances to be used when trimming B-Rep surfaces (which are influenced by the minimum surface mesh spacing) are not generally known a priori. This, together with several other geometry-related problems highlighted by mesh adaptation (errors in the assessments of local tangency or curvature potentially becoming attractors for error estimates, resulting in pockets of excessive mesh refinement, for instance) have been reviewed in Ref. [134].

A growing awareness of the potential geometric ambiguities inherent in all B-Rep models has led many researchers who are pursuing the Study’s goals to either build their own B-Rep geometry modelling kernels (and provide bespoke control of the attendant consequences) (e.g., Refs. [135] and [136]) or to seek alternative geometry modelling techniques which yield geometrically watertight models (e.g., Refs. [137] and [138]). Others, particularly those involved in multidisciplinary studies, have preferred to exploit some of the other benefits of commercial MCAD modelling systems (rich, feature-based parametric modelling capability, compatibility with contemporary industrial infrastructure, etc., such as Ref. [139]). However, even in these circumstances, the models used are usually much simpler than those described in Ref. [133] and are often subject to various artificial constraints.

One of the problems associated with many geometry kernels is that they were not originally designed to operate in HPC or distributed environments. There are two aspects to this limitation. First, most support only sequential execution for build and querying. Second, many are implemented in software that is in some way remote from the rest of the CFD workflow (either obscured by licensing constraints or operating on disparate hardware or operating systems). Various mitigations against these limitations have been postulated, such as the use of mesh and geometry databases [134]. Some researchers developing bespoke geometry kernels have considered the needs of HPC operation from the outset. For instance, Ref. [140] describes a scalable client-server implementation of a B-Rep modelling system. Several organizations have sought to exploit the potential offered by alternative geometry modelling techniques – those based on spatial occupancy in particular – as part of the development of end-to-end parallelized simulation workflows – see e.g., Ref. [137].

At the time of the Study’s publication, the AIAA Aerodynamic Design Optimization Discussion Group provided a forum for assessing the efficiency and effectiveness of the methods being used to manipulate and control Outer Mold Lines (OML) at the time. For the reasons outlined in Ref. [129], most approaches affected shape change indirectly (i.e., via the mesh rather than the geometry), using spatial deformation techniques like free-form deformation, or relatively simple feature-based parameterization schemes. It quickly became apparent, however, that the most successful techniques contained inherent mechanisms for adaptive, localized refinement (e.g., Ref. [141]). As a consequence of their geometric watertightness, techniques using subdivision were particularly amenable to this and yielded markedly better results than their competitors [142]. (Although, because of the equivalence of subdivision curves and B-Splines in two dimensions, similar results have subsequently been obtained for airfoils using classical knot insertion techniques [143].)
More recently, the focus of research interest has moved away from modelling techniques in which OMLs are parameterized a priori to those where the desired geometry emerges as a result of other, physics-based drivers. The published examples all rely on spatial-occupancy-based geometry modelling techniques – some explicit [144], others implicit [137]. Here, there has been some transference of the lessons originally learned in structural modelling (via topology optimization). Compelling demonstrations are beginning to appear in a wide range of practical applications, not confined to the aerospace sector (or, for that matter, engineering). It is difficult to conceive how some of the resulting shapes could be derived using classical B-Rep modelling techniques.

Some of the issues to be addressed have been described in Ref. [138], where researchers describe the problems they encountered when trying to use subdivision surfaces in an industrial MCAD modelling system. Particular attention is drawn to the limited access provided to the underlying parameterization and the difficulties in providing feature-like control. Since subdivision surfaces are a generalization of NURBS [145], there would appear to be no mathematical reason why this level of support cannot be provided.

The task of transferring the results of aerodynamic (or multidisciplinary) analyses to the primary MCAD model will continue to pose challenges. Several developments have been made on this front over the last 5 years (e.g., Refs. [137] and [146]).

Many of the geometric reasoning developments originated in the Computational Structural Mechanics (CSM) community, driven partly by the need to accommodate disparate levels of dimensional reduction in adjacent components of analysis models. For instance, medial axis based spatial decomposition and midsurfacing algorithms have been developed to (a) distinguish between the various components in complex turbomachinery MCAD assemblies and (b) assign the appropriate forms of geometric abstraction required for structural simulations [147]. More recently, these have been augmented by feature recognition schemes that facilitate OML generation [148]. An important aspect of these developments is that these new models are built without altering the master model from which they were derived.

Another factor that has guided these developments has been a desire to provide simultaneous access to more than one representation of geometry and to allow information to be transferred between them. To these ends, the concepts of Virtual Topology and Equivalence have been developed [149]. Here, the mappings (or virtual topologies) between different geometry models are defined in a way that allows modifications to one representation to be reflected directly in others (equivalence). An alternative approach, favored by those who have developed bespoke geometry modelling kernels (e.g., Refs. [135] and [136]), is to address these requirements during construction of the master geometry. In each case, the intention is to allow various instances of the model to be generated from the same feature-based parameterization. In this way, there is not a single resultant model, but several models may be generated, one for each type of simulation, at the appropriate level of fidelity. Matching and equivalence between models generated from the same Master Geometry is accomplished by attribution on the topological entities.

**Mesh Generation**

While not exhaustive, the authors’ review of published mesh generation research finds that it is dominated (as measured by number of publications) by work that addresses the issue of suitability. This work involves a range of topics that is too broad to cover completely herein but includes achieving
validity (e.g., untangling crossed meshes), improving quality as defined by a variety of cell metrics, or achieving certain levels of refinement. The techniques used to do so cover an equally broad range from partial differential equation methods, octree, point insertion criteria for Delaunay unstructured meshes, smoothing and swapping, and techniques targeted at specific subregions of the OML such as ridge lines and cylinders.

Mesh adaptation is a special, or perhaps the ultimate, expression of suitability in that it accounts for the flow physics, the geometry model, flow solver algorithms, and usually an HPC computing environment. Paradoxically, adaptive meshing techniques have been discussed, researched, and published since virtually the day after mesh generation was invented. The perceived benefits of an adapted mesh are generally agreed upon. Yet, during these intervening decades one finds relatively little adoption of adaption among practitioners.

An excellent overview of adaptation as it pertains to the Study is provided in Ref. [134] and need not be repeated here. Since publication of that document, an open group of researchers formed the Unstructured Grid Adaptation Working Group (UGAWG), published the results of their first benchmarks for adaptivity [150], and established an online presence [151]. The working group is notable because its goals are derived directly from the Study and because its membership is cross-organizational, international, and includes representatives from industry, academia, and government.

Reference [152] documents the latest developments from members of the UGAWG. Its titular focus is use of geometry models for surface adaptation but the work documented therein covers all aspects of the process. Highlights from this research include the implementation of persistent mesh-to-B-Rep associativity, the use of high-order grids as a surrogate geometry model when necessary, and implementation of two adaptation tool chains that begin with a robustly generated yet relatively simple surface mesh and end with adapted surface and volume meshes and converged CFD solutions for several aerospace benchmark problems.

A broader look at mesh adaptation finds work that covers the gamut of techniques including surface and volume meshes; r- (redistribution), h- (refinement), and p- (polynomial) adaptation; and adaptation to solutions and moving boundaries. Within the realm of h-refinement methods, a technique such as nonconformal (i.e., hanging node/edge) refinement exemplified by Ref. [153] offers a flexible form of adaptation for flow solvers that permit such mesh topologies. While deemed a complicating factor by the flow solver, hanging edges provide a great deal of flexibility in the mesh generator. Adaptation techniques that use r-refinement combine the benefits of adaption with the economy of a fixed set of points and connectivity. As discussed in Ref. [154] in which a spring analogy approach to equidistribution is employed, the flow solver’s algorithms have to be accounted for when selecting a metric field to which to adapt.

Importantly, the degree of anisotropy typically seen in a CFD mesh poses challenges that exceed much of the published research on adaptation methods, even those that specifically claim to be designed for anisotropy. A number of researchers focus specifically on achieving CFD levels of anisotropy during adaption including Refs. [155] and [156].

As adaptation technology has matured, a companion trend is evident in the implementation of systems and frameworks for the entire adaptation process. In addition to Ref. [111] cited above, other examples of adaptation frameworks include Refs. [157] and [158].
Geometry Modeling and Mesh Generation

As a final example of progress in mesh adaptation that also addresses another of the Study's elements – multidisciplinary simulations, a fluid-structure interaction problem was simulated with adaptation in Ref. [159].

Given that adapted meshes have been a component of most of the recent AIAA CFD Workshops (e.g., Ref. [160]) and other applications, the trend toward wider adoption of adaptation is evident. However, adaptation still is not de rigueur across aerospace CFD as desired by the Study.

Improving meshing's robustness (i.e., improving the likelihood of obtaining a suitable mesh on the first attempt) is often the province of research into Emerging Methods.

One area of emerging mesh generation research that appears to be nearly as energized as mesh adaptation (measured by publication count) is the generation of high-order (H-O) meshes for use in H-O CFD solvers. H-O meshing's emergence is a direct response to the emergence of H-O CFD solvers such as COFFE [161] and PyFR [162] that are being pursued because of their potential to produce better results more economically than traditional CFD methods. These and other flow solver implementations have matured to the level where meshes on realistic geometries are needed as evidenced by the two meshing challenge cases (the HL-CRM and NASA Rotor 67) at the most recent H-O workshop [163].

The vast majority of H-O meshing techniques start from a linear mesh in a four-step process: generate a linear mesh, insert boundary and interior nodes on each cell to elevate its polynomial degree above linear, curve the OML-adjacent cells to conform to the geometry model, and blend that curvature onto the mesh's interior to remove inverted cells. The method by which cell curvature is blended from the boundary onto the mesh's interior is perhaps the most challenging step of the process due to the fact that the highly anisotropic mesh cells used to resolve the boundary layer are likely to exhibit crossing as the mesh is curved. That curving can be performed using linear-elasticity methods [164], weighted condition number smoothing [165], pointwise optimization of a distortion metric [166], Winslow smoothing [167], and other methods. Like adaptation, some H-O frameworks are being developed such as Ref. [168] in which a unique NURBS surface fitting and subdivision technique ensures that the OML is modeled accurately.

The enthusiasm for H-O methods, both meshing and solver, is encouraging. Widespread adoption significantly lags adaptation, which is not surprising as several issues with this emerging technology remain to be solved. Open H-O issues include determining what polynomial degree of the mesh is sufficient (given that some H-O solvers further elevate the order), deciding whether mesh curving is only necessary for OML matching (versus curving to flowfield features), identification of the appropriate metrics to guide the meshing process, and a feel for exactly how coarse the original linear mesh needs to or can be.

Mesh generation methods with integrated domain decomposition and partitioning such as GridPro [169] have been available for many years. Medial Object Technology (MOT) has developed rapidly during the past five years as demonstrated by Refs. [170] and [171] in which MOT is used to partition the domain around the NASA CRM aircraft for rapid generation of a hybrid mesh. MOT is based on use of an object’s medial axes (the points in the fluid domain that have more than one closest point on the boundary) and medial radius (the distance from the medial axes to the boundary) to partition the fluid domain into meshable subregions. If that sounds familiar, it should. It is the crux of multiblock, structured grid generation for which MOT offers the tantalizing possibility of automating the creation of structured grid
topology as demonstrated in 2D in Ref. [172]. Reference [173] includes an analysis indicating that MOT and related technologies such as frame/cross fields and paving/plastering may be seen as different views of the same fundamental issue. Given that it is generally agreed that hexahedra are preferred over the other canonical element types for solution accuracy and given that structured grid flow solvers generally run faster than unstructured mesh flow solvers, one is tempted to imagine a resurgence in the use of structured grids in applied CFD. Of course, MOT’s benefits are limited to structured grids. Maturation of this technology will involve mitigation of singularities [174], quality optimization [175], resolution of OML features, and more.

Less emerging and more emerged is the use of strand grids for near-wall meshing as exemplified by Ref. [176]. Relative to advancing layer methods in which each extruded grid line off the surface mesh can be unique in its shape, length, and distribution of grid points, each strand is virtually identical. This vast simplification pays dividends in terms of computational efficiency. Reference [176] describes the fully automated application of strand grids to complex aerospace configurations such as the JAXA Standard Model aircraft and a UH-60A helicopter.

The Study calls for use of HPC systems in order to a) generate a large number of moderately-sized meshes as quickly as possible and b) generate a few massively-sized meshes when needed. It is a positive sign that published research on meshing suitability and robustness often implicitly includes a parallel implementation. The discussion that follows includes only research with an explicit goal of parallelization. With respect to adoption of this research, the authors note that nothing is less portable between software packages than a parallel implementation that raises a number of issues with respect to how meshing as a whole will move toward broad adoption of HPC platforms.

The published research on parallel implementations is virtually all for unstructured meshing techniques, especially Delaunay. There is a further bifurcation between parallelized techniques that start from the boundaries and generate the entire mesh versus those that perform operations on an existing mesh. The latter can demonstrate extreme speedups as in Ref. [177] in which a given set of points are tetrahedralized at a rate approaching one billion cells per minute. An example of the former is documented in Ref. [156] in which one billion tetrahedra are generated in 20 minutes on 120 compute cores.

Generally, the approach taken to parallelization is hybrid: coarse grain at the distributed memory level using MPI and finer grain at the compute node level with OpenMP. There is a large amount of published research on the use of GPUs with CFD; little of it pertains to mesh generation. What has been published about GPUs for meshing [178-180] demonstrates good performance improvements over serial code while at the same time citing the challenges of moving data to and from the GPU.

At the coarse level of parallelization two issues dominate the implementations: dynamically partitioning the mesh across processors and handling the mesh points on the partition boundaries. While these issues are no different than those for the CFD solver, in Ref. [181], one finds a concise summary of the challenges for unstructured meshing, the competing needs of data locality and algorithm effectiveness.

It was also observed during the authoring of this paper that the current use of HPC resources almost exclusively targets the Vision’s first call: generate a large number of moderately sized meshes as quickly as possible. One exception was a test case from GMGW-2 [182] that involved exascale (i.e., 31 billion cell) meshing of the NASA HL-CRM, a mesh size that will be considered “medium resolution” by 2030.
according to the Study. This case was specifically designed as a stress test of contemporary meshing software’s ability to handle that level of data. The four participants who tackled this exascale case all came up short on cell count with meshes ranging from 1.7 to 7.9 billion cells. Volume mesh computational time ranged from 1 to 28 hours, a range indicative of the types of computing platforms used from a basic serial workstation to a distributed system with over 10,000 cores. This case identified weaknesses in the exascale CFD toolchain including exceeding the computer’s available RAM (all 4 participants), large integer support for counting all the components of an exascale mesh (1 participant), and an inability to export the resulting mesh to file (2 participants).
Knowledge Extraction Progress

The Knowledge Domain of the Roadmap includes timelines for Integrated Databases and Visualization. The overall objectives identified in this domain are associated with the development of integrated database systems that manage and process large numbers of unsteady simulation data and test data to enable Uncertainty Quantification workflows and the ability to perform on demand analysis and visualization of ever larger datasets. The technology milestone targeted to have been reached by 2020 is the ability to analyze and visualize a 10-Billion point unsteady CFD simulation.

Recent Developments

In 2019, NASA and NVIDIA® demonstrated progress toward the 10 billion element visualization achievement as highlighted in Figure 2. At the SC19 conference for HPC, they demonstrated an interactive visualization of an entire time series from NASA’s Mars Lander FUN3D simulation. Users were able to interactively animate a 6B node simulation with 150 TB of time series data in real time. These large unsteady simulations were performed on the Oak Ridge National Lab Summit computer. This capability required the use of four dedicated NVIDIA DGX-2™ systems each with 16 NVIDIA Tesla™ V100™s and coupled with 16 SSDs to hold the data using NVIDIA’s GPUDirect™ to move the data directly across the network to the GPU memory.

Figure 2. Six Billion Element Interactive Visualization of Mars Lander [183]. Image courtesy of NVIDIA, used with permission.
Knowledge Extraction

Five-year Perspective

Over the past six years since the release of the Study, significant advances have been made in CFD knowledge extraction (KE). The Study recognized that Integrated Databases and Visualization technologies would be key to achieving the 2030 goals. Under Integrated Databases, by 2025, a Technology Demonstration for the creation of real-time multifidelity databases with 1000 unsteady CFD Simulations plus test data with complete UQ of all data sources would be achieved. Under Visualization, a technology demonstration by 2020 specifies an on-demand analysis/visualization of a 10-billion-point, unsteady CFD simulation.

There are several efforts in place that lead to the 2025 Database demonstration. Graham Pullan at Cambridge University has been developing dbslice, which is a web-based database approach for CFD Post-Processing [184]. Similarly, Intelligent Light has been developing Spectre-UQ™, which is a web-based environment for uncertainty quantification [185] and ANSYS® has been developing Nexus, an automated report-generation data analytics capability through its ANSYS EnSight™ product [186]. Sandia Analysis Workbench (SAW) [187] is another capability that allows users to share data between simulation tools from model building, grid generation, solutions, post-processing and UQ. All four of these tools use web-server-based data processing and analysis. Typically, a relational database system (e.g., Oracle®, PostGRES, SQLite) is integrated to store and serve meta data, simulation and experimental data needed at different parts of an analysis, UQ or design workflow.

There currently exist publicly available data repositories within the CFD community. However, there is a large gap in making that data easily accessible across the community and there remains a pressing need for a repository for hypersonic simulation verification and validation (V&V) data [188]. For example, in the hypersonic simulation community, there are current efforts using DNS simulations and experiments as “truth” to drive the development of improved Wall Resolved Large Eddy Simulation (WR-LES), Wall Modeled Large Eddy Simulation (WM-LES), Unsteady Reynolds-averaged Navier-Stokes (URANS) or RANS for use in hypersonic vehicle simulations and design. However, the existing databases are distributed and do not use a consistent/coordinated database schema, which makes it difficult to easily share and use the data [189-192].

Knowledge extraction through visualization has made good progress toward “on-demand analysis/visualization” of a 10 billion cell, unsteady dataset. To reduce the I/O overhead of writing large volumes of data to disk and reading that back into a post-processing tool to visualize, there has been considerable development of in situ/in transit methods. With in situ processing, the flow solver is instrumented with a data processing/visualization library, such as VisIt-Libsim [193] or Paraview-Catalyst [194], which shares the memory space of the flow solver. This capability often requires that the simulation code pauses while the data are processed. With in-transit processing, the simulation data are transferred to a separate set of compute nodes, which processes the data and allows the simulation to resume after the data have been transferred. VisIt-Libsim, Paraview-Catalyst and other scientific data processing libraries such as ALPINE [195] have unique visualization and analysis capabilities that a user may want to use in their specific simulation. However, developers may not want to be locked into only one library. Sensei is an in situ/in transit infrastructure that once a solver is instrumented with it, the solver can use the functionality of any in situ/in transit method [196].

Visualization can either be via data extracts or direct to an image. With data extracts, the visualization graphic objects are created in situ, in-transit or post-processing and then written to disk. The user can
Knowledge Extraction

then load the extract into a traditional post-processor for further analysis or render an image as illustrated in Figure 3. Since the extract is just a subset of the volume data (i.e., a coordinate cut plane, or line extract), it can be many orders of magnitude smaller than the volume data, reducing the amount of data stored and reducing the processing time during user interaction. An example of an extract workflow was presented by Kirby et al. [197] in 2018. Using the WAKE3D code instrumented with Libsim, they created extracts that were subsequently visualized and manipulated in FieldView. In this simulation, they extracted knowledge about the wake breakdown of the vortical wake structure in a wind turbine simulation and using Proper Orthogonal Decomposition (POD) Methods. They were able to correlate wake breakdown features to POD modes.

Figure 3: Lillgrund wind farm wake structures composed of 48 Siemens SWT-2.3-93 wind turbines visualized in FieldView with XDB workflow. Velocity magnitude and Q-criterion iso-surfaces consisting of 48 blade-resolved turbines, 1.2 Billion Degrees of Freedom, 22,464 compute cores on the Cheyenne NWSC supercomputer and 12 rotor revolutions requiring approximately 90 hours of wall-clock time [309]. Used with permission (Intelligent Light).
Data extracts can also be images. While the simulation executes, an image file can be produced. However, a user will not be able to explore the data (zoom in/out, pan, rotate, animate) if they save only a single image or movie. To enable interactive visualization of large datasets, the Cinema Technology from Los Alamos National Laboratory makes use of a large collection of image extracts from many different viewpoints collected from the solver via in situ/in transit pipelines [198]. The images can then be viewed via the Cinema viewer, a web app or even through a Jupyter notebook. VisIt-Libsim and Paraview-Catalyst are both enabled to output Cinema extract databases.

Figure 4: Cinema viewer showing a wind farm dataset produced by the University of Wyoming and using the WAKE3D code with 508 Million Degrees of Freedom on 5,328 cores [197] and exported to Cinema using VisIt (visualization from Whitlock). Used with permission (Intelligent Light).

Over the last 5 years the visualization technologies described above plus others demonstrated progress toward “on-demand analysis/visualization of 10 Billion cells unsteady simulations” culminating in the achievement of Visualization Technology Achievement demonstrated by NASA and NVIDIA.
In 2017, Jansen et al. [199] demonstrated the use of Sensei with Paraview to visualize and analyze a 5 billion element DDES simulation of separated flow over a vertical tail with active jets as shown in Figure 5.

Figure 5: Visualization of 5 billion element DDES simulation using Phasta with Sensei and Paraview, Frame image taken from https://vimeo.com/221008870 [199]. Used with permission.

In 2015, Tecplot® began to explore the requirements for visualizing large scale datasets. Using their commercial product Tecplot on desktop sized computer systems, they visualized datasets from a canonical 1 trillion cell dataset [200].
MDAO Progress

CFD capabilities in 2030 will play a significant role in routine, multidisciplinary analysis (MDA) and optimization (MDAO) that will be typical of engineering analysis and design. Most of the aerospace engineering problems of interest will be of a multidisciplinary nature and CFD will have to interface seamlessly with other high-fidelity analyses such as, but not limited to, acoustics, structures, heat transfer, reacting flow, radiation, dynamics and control, and even ablation and catalytic reactions in thermal protection systems. Not only will fast, accurate and robust coupled analysis be required but also sensitivities of these analyses with the same aforementioned attributes will be necessary to perform fully automated design of complex systems.

Recent Development

The current progress will focus primarily on the 2020 milestone, incorporation of UQ for MDAO. Design of next-generation aerospace systems is pushing toward higher efficiency while ensuring safety, which presents significant challenges due to the presence of uncertainties and increasingly expensive-to-evaluate simulation models. Clark [202] gives a comprehensive review of recent work done in this area and classifies optimization under uncertainty techniques into three primary areas; reliability-based design optimization (RBDO), robust optimization, and distribution matching techniques. RBDO formulation creates designs that have a low probability of failure due to probabilistic variables, while a robustness formulation creates an objective or constraint that is ultimately less sensitive (i.e., containing small variation) to changes in the design. Finally, distribution matching techniques incorporate distance measurements between a design’s cumulative or probability density function to a desired distribution. All of these techniques become computationally challenging as the complexity of the modeled physics increases. Recently a large research effort has been funded in the area of UQ by DARPA to address many of the challenges associated with incorporating uncertainty in design optimization. From 2015—2018, DARPA funded EQUIPS (Enabling Quantification of Uncertainty in Physical Systems) [203] acknowledging the importance in characterizing model, parametric, environment, and measure uncertainty when designing complex physical systems, devices and processes. As part of EQUIPS, a Scalable Environment for Quantification of Uncertainty and Optimization in Industrial Applications (SEQUOIA) [204] was established as a collaboration between Sandia National Labs, Colorado School of Mines, Stanford University, and University of Michigan. The basis of the SEQUOIA project was to develop a strategy for UQ and Design Under Uncertainty (DUU) that advances the scale and complexity of problems [205]. Generally, the results of the SEQUOIA project have pushed to create a tool set to explore large design problems under uncertainty through the uses of dimension reduction, surrogate models, and multifidelity techniques. The three thrusts of the program were: 1) scalable algorithms, 2) model-form uncertainty, 3) design and decision making under uncertainty. Within thrust one most effort was on the development of dimension reduction techniques [206-209], active subspace identification [210-215], reduced-order modeling [216], surrogate models [96,217-220] and Monte Carlo acceleration methods [96,221-222] to allow for problems with a large number of design variables and random parameters. Thrust two identified model-form discrepancies through perturbation methods [223-224], physics error estimates [95,225-228], and multifidelity methods for forward UQ [225,227]. Thrust three, integrates thrusts one and two developing mathematical foundations for forward DUU integrating a handful of the
previous techniques [229-234]. Additional work was conducted to improve experimental design practices using adaptive inverse algorithms [210,235-236] and model calibration [211].

Additional progress outside of SEQUOIA has been made in the area of multifidelity methods that enhance information with cheaper lower-fidelity information sources are the key enabler to make multidisciplinary design optimization under uncertainty computationally feasible. Researchers from the Massachusetts Institute of Technology and the University of Texas at Austin have advanced the state of the art in multifidelity MDAO under uncertainty through new methods that leverage Monte Carlo variance reduction techniques and machine learning along with effective reuse of information from past optimization iterations. They introduced multifidelity Efficient Global Reliability Analysis (mfEGRA) [237], an efficient active learning reliability analysis method using multiple information sources. They presented a new reliability-based design optimization (RBDO) method that reuses existing information from past optimization iterations in multifidelity active learning through mfEGRA at AIAA SciTech in January 2019 [238]. In 2020, they demonstrated a method for computationally efficient, reliable design of a single element model rocket combustor by reusing existing information for Monte Carlo variance reduction via IRIS-RBDO (Information Reuse for Importance Sampling in RBDO) [239]. These new methods enable computationally efficient implementations of robust optimization and RBDO for multidisciplinary systems to ensure safe, efficient designs.

Five-year Perspective

Coupling CFD with Other Physics

Solving multiphysics problems can be achieved primarily in two ways. One approach is to solve a large system of equations for all unknowns as a single system or alternatively, in a partitioned approach, one has separate solvers for the each of the participating disciplines. For this discussion, since much of the recent work has been on developing the partitioned approach, the partitioned approach will be the focus. In the partitioned approach, boundary conditions for multiple single physics problems at the coupling surfaces have to be defined. The respective interface values are passed from one solver to the other. This approach requires mapping methods for physical variables between (in general) nonmatching solver grids at coupling surfaces. Important features desired of such mappings are accuracy, consistency, and conservation of energy and momentum. There have been numerous techniques developed over the years to transfer information from one computational domain to another. To date, no standard way of doing this transfer of information between the domains has been developed and most likely never will be since new techniques continually emerge. The goal should be to develop a standard application programming interface (API) that describes what information is required in order to transfer the information between the domains. Other information that should be made available through the API should include how to obtain sensitivities, accuracy, consistency, and conservation of energy and momentum of a specific technique. A comprehensive study by Smith, Hodges and Cesnik [240] evaluated the infinite-plate spline, finite-plate spline, thin-plate spline, biharmonic-multiquadrics, nonuniform B-splines and the inverse isoparametric method for transferring loads and displacements between CFD and computational structural dynamics (CSD) solvers. Although many of these early techniques were successful, they were not developed with MDO in mind (for example, supplying the necessary sensitivity information required to perform MDO with coupled analysis). Recently Kiviaho and Kennedy [241] classified load and displacement transfer techniques into two categories; projection-based methods and interpolation-based methods. Projection-based methods associate each aerodynamic surface node with a unique structural element, typically through a closest-
point projection, and then use the local element displacement data to evaluate the motion of the surface node. This technique requires information not only about the structural mesh but also the elemental shape functions information that may not be readily available. The interpolation-based techniques construct a global representation of the displacements at the structural nodes with only the structural coordinates. Kiviaho and Kennedy [241] developed a new technique that achieves localization similar to projection-based methods for computational efficiency with nonintrusiveness similar to interpolation-based methods. This technique, called the method of matching-based extrapolation of loads and displacements, or MELD, also addresses the development of the necessary sensitivity information required to perform MDO with coupled analysis. This work is currently being extended to include thermal analysis quantities in a three-way coupled aerothermoelastic analysis within the AFRL Sponsored Sensitivity Analysis for Multidisciplinary Systems (SAMS II) project. Another on-going effort in this area of state data transformation is the Mphys generalized coupling library being developed at the NASA Glenn Research Center. This work could also contribute to the development of standard interfaces in this area.

The above discussion focused on the transformation of information between disciplines. We now shift focus to specific functionality that is desired by a CFD application in order for it to participate in the MDAO environment. Once again no formal standard has been developed by the community but progress has been made on developing standard interfaces. Mader et al. [242] put forth a standard way of developing a CFD application that is suitable to be incorporated into an MDAO environment along with an open source reference implementation, ADflow. The work focuses on development of a highly efficient computational method that can be used repeatedly in an automated fashion. Their approach proposes developing the solver as a compiled library with direct memory access via a well-defined application programming interface (API). One of their key observations is that file I/O should be avoided at all cost and all information transfer take place through memory. Mader also points out that not only does the solver have to be efficient but that the necessary sensitivities of the flow solver needed for performing multidisciplinary gradient-based optimization needs to be computed and easily accessed. The API definition developed states that at a minimum must supply the following functions with the ability to:

- Manipulate the surface of the CFD geometry
- Specify the flow conditions
- Solve for the flow state variables
- Evaluate the functions of interest
- Recover the solution from a failure state
- Evaluate the solver derivatives

Mader goes on to define an advanced API that has the additional functionality to:

- Get and set the solver state variables
- Evaluate the solver residuals
- Get and set the adjoint variables
- Get and set coupling variables
- Evaluate matrix-vector products with the state Jacobian and its transpose
- Solve the adjoint equation with arbitrary right-hand sides
This work has been extended by Yildirim et al. in Ref. [243] and incorporated into openMDAO [244].

The proposed standard in Refs. [242] and [243] works well if developing a new CFD application but does not address the use of existing applications where one may or may not have access to source code or an API to make the necessary modifications or access the necessary data. In these cases, depending on how the solver stores its data, the integrator is forced to use file I/O and file parsing to communicate with the flow solver. This is extremely inefficient, error prone and lacks robustness.

The commercial software vendors have begun to incorporate multiphysics simulations into their products but once again there is no agreed upon standard or standard interfaces. Commercial products such as Dassault Systèmes’ SIMULIA® [245], ANSYS [246] and COMSOL® [247] have a chained analysis capability and a coupled analysis capability for multiphysics simulations between fluids, structures, acoustics and heat transfer. Some do supply single discipline sensitivities but none compute fully coupled sensitivities for steady and transient responses that can be used in gradient-based optimization.

**MDA/MDAO Coupling Frameworks**

The previous milestone, “Define standards for coupling CFD to other disciplines”, focused on the development of standard interfaces and the transfer of information between CFD application and other disciplines. Here the discussion on computational frameworks that are necessary to execute MDA and MDAO using multifidelity CFD in a loosely coupled fashion will be addressed. The below discussion assumes a partitioned approach described above to solving the MDA. The partitioned approach or also referred to as the “best in class approach” (BCA) typically uses a scripting language such as python or Perl or Java to “glue” together several independent applications/tools that provide the “best” functionality for a given discipline to define a process and solve a selected problem. Before discussing the current state of frameworks, a discussion on what are the requirements for such a framework will be examined [248-250]. Framework states and requirements will be discussed in the context of the full vehicle MDAO milestone. When performing full vehicle MDAO, vehicle requirements associated with strength, stiffness, buckling, cruise performance, maneuver performance, static aeroelastic stability, dynamic aeroelastic stability, controllability, propulsion performance, thermal management, and more must be assessed for a given configuration. This is done by carrying out a loads survey, based on the mission profile determined in the concept of operations, to identify the critical set of flight conditions that drive the design of the specified configuration. This results in 100–200 critical conditions that must be considered while performing the design space exploration. These critical conditions are associated with different ground and air maneuvers throughout the mission and have a wide range of Mach number, altitude, dynamic pressure and control surface settings. In order to perform the design refinement at the critical set of flight conditions, higher fidelity, coupled, and/or chained analyses are required. This produces constraints for each of the disciplines associated with each of the critical flight conditions resulting in potentially 1x10^4—1x10^5 number of constraints. The number of design parameters can be quite large for a system level, full vehicle MDAO when all disciplines are considered. Examples of type and number of design parameters include,

- aerodynamic design parameters (1x10^2) (wingspan, chords, thickness to chord ratios),
- structural sizing parameters (1x10^3) (skin thicknesses, spar thicknesses, spar cap cross-sectional areas and moments of inertia),
- control effector parameters (1x10^2) (the number, size, and location of control effectors),
- system level propulsion parameters (1x10^2),
- system thermal management ($1 \times 10^2$) parameters

This brings the total number of design parameters on the order of $10^3 - 10^4$. Hence a large scale system level, full vehicle MDAO problem consists of potentially 10 multifidelity disciplines, 10 couplings, $10^4$ design variables, and $10^5$ constraints. Finally, since such a problem creates a demanding computational problem along with the fact that different organization may be responsible for different disciplines, the framework has to support a distributed heterogeneous computing environment.

Frameworks that support the BCA can be found in commercial offerings such as Phoenix Integration’s ModelCenter® and Analysis Server® [251], ESTECO’s modeFRONTIER and VOLTA [252], VR&D Visual Doc [253] and Dassault Systèmes’ SIMULIA Isight Simulation Engine [254]. There are also some U.S. government or open source efforts such as OpenMDAO [244], Dakota [255], GEMS [256], and MSTC Engineering [257] that can support BCA MDAO. Fundamentally, the BCA approach “wraps” each application with a scripting language and defines input and output that a given application requires or generates. A major benefit of this approach is that it gives the end user access to the “best” technology available in a given domain and thus supports the “plug and play” paradigm to a certain extent. This approach is much more palatable to engineering domain experts since they can include their “best in class” application for a specific problem being solved. In theory, this approach appears quite attractive but in practice many problems do arise. Each application has different data structures and formats. This can lead to difficult and inefficient data transfer between applications. Scripts for large-scale problems tend to become unruly/problem specific, and hence fragile, difficult to maintain, and not reusable. Also, more importantly, there is no well-defined manner to trap errors that occur during the process. This may not be an issue if combining only a few applications but when the number begins to approach 10s or 100s the ability to determine when, where, and why a failure occurs becomes the critical element in the success of the BCA.

Many of the commercial offerings such as ModelCenter, Isight and to an extent modeFRONTIER were originally developed for low-fidelity analyses and small design optimization problems (fewer than 100 design variables). Each application was assumed to have relatively simple inputs defined in an ASCII file along with outputs that again were in ASCII files of small size. The transfer of information from one application to the other is done using a wrapper that parses ASCII input and output files. This works well for simple applications but when integrating higher-fidelity applications such as CFD, CSM or CAD systems this approach becomes quite inefficient. Another issue with some of the commercial offerings is the ability to accept user supplied sensitivities when performing gradient-based optimization and computing coupled sensitivities. One would have to rely on finite difference sensitivities that are inefficient and prone to step size errors. This again results in a limit in the number of design variables that can participate in the optimization (less than 100) when considering high-fidelity analyses. Often the approach to bring higher fidelity information into these environments is done using surrogate models such as response surfaces. Unfortunately, surrogate models suffer from the curse of dimensionality, once again limiting the number of design variables to less than 100. Many of the vendors have begun to address some of these issues by creating more sophisticated wrappers to integrate using APIs (if available) and accepting user defined sensitivities. Further, they are developing the capability to perform distributed MDA and MDAO on heterogeneous computing environments across organizational and corporate boundaries. With that said, there is still a considerable amount of functionality that needs to be incorporated into the commercial products before they can fully support large scale multidisciplinary, multifidelity, distributed design exploration under uncertainty (for example, the ability
to compute coupled system sensitivities, ability to support multifidelity, support optimization under uncertainty and to allow tight efficient integration between high-fidelity applications such as CFD and CSM).

The open source frameworks have focused on some of the issues identified with the commercial offerings. OpenMDAO excels in solving design problems involving coupled numerical models of high-fidelity, complex engineering systems. It also has a framework for the computation of the derivatives of these coupled models for use with gradient-based optimization algorithms. This enables the solution of high-fidelity, large-scale \(10^3—10^4\) design variables, gradient-based design optimization. Dakota supports BCA integration with user supplied sensitivities but its primary strengths are in UQ, optimization algorithms, and surrogate modeling. Dakota’s UQ methods primarily focus on the forward propagation of uncertainty where probabilistic or interval information on parametric inputs are mapped through the computational model to assess statistics or intervals on outputs. Further, Dakota supports optimization under uncertainty (OUU) and reliability based design optimization (RBDO). GEMS focus is on developing an automatic programming of MDO processes along with distributed and multilevel MDO formulations (or MDO architectures). Finally, MSTC-Engineering is primarily a research code exploring efficient ways of performing multifidelity large scale geographically distributed MDAO.

Other computational frameworks, which are not general purpose MDAO frameworks but specific to CFD analysis, sensitivity analysis and optimization are: FUN2FEM [258], SU2 [259], and CREATE-AV KESTREL [260]. FUN2FEM is a coupled, high-fidelity, steady aeroelastic framework for analysis and design optimization that uses the adjoint method to efficiently solve a gradient-based optimization problem. SU2 is similar to FUN2FEM, it is a computational analysis and design environment enabling the solution of high-fidelity multiphysics analysis and optimization problems. It too computes coupled adjoint sensitivities for efficient gradient-based optimization. CREATE-AV KESTREL is an integrating product in the HPCMP CREATE™ program that allows crossover between simulation of aerodynamics, thermochemistry, dynamic stability and control, structures, propulsion, and store separation. It provides a robust and accurate multidisciplinary simulation capability targeting fixed-wing aircraft. Kestrel does not support any sensitivity analysis or design optimization capability.

**Full System MDAO**

MDAO simulation of an entire aircraft has an incredibly wide scope. When evaluating this milestone, the following will be considered.

- The number of disciplines participating in the MDAO
- The number of couplings or chainings between the disciplines
- The fidelity of each discipline
- The number of design variables
- The number of constraints
- The use of surrogate models to represent the objective functions or constraints

Terms used in the preceding list are defined as follows.

- **Coupling** is defined as the need to simultaneously solve disciplines such as aerodynamics and structures to perform an aeroelastic analysis. Coupling implies a bidirectional dependency of system state information.
Chaining is defined as the need to sequentially chain analyses together to obtain the necessary result. For example, performing a prestressed structural analysis prior to executing an eigenvalue analysis. This is a one-way dependency.

Fidelity refers to the level of physics included in a specific domain. A number from 0-5 is assigned to represent fidelity for a given engineering discipline, where 0 is the lowest fidelity and 5 the highest fidelity. For aerodynamics, the fidelity levels are:
- 0: traditional empirical representation
- 1: linear potential/panel methods
- 2: Euler equations
- 3: RANS
- 4: LES
- 5: DNS.

The selection of the fidelity, coupling, and chaining is critical. The appropriate levels of fidelity, coupling, and chaining are those that are required to capture the phenomena that are essential in designing the specified configuration. This implies that the appropriate fidelity, coupling, and chaining are dependent on the configuration and the flight conditions.

At this point in time, there appears to be two primary vectors that are being followed in the aerospace community. The first is focused on incorporating more disciplines into the vehicle level design process and the second is to bring higher fidelity, coupled analysis into the process. Both are aimed at capturing physics that are being missed during early design. In time, it will be necessary that these two vectors merge to solve the system level full vehicle MDAO problem.

First, the discussion will concentrate on the vector that is focused on bringing higher fidelity and coupled analysis into the design environment. This vector is heavily dependent on two technology areas: robust parametric associative geometric modeling of the inner and outer mold line of the vehicle, and analytic sensitivities. Other parts of this document have addressed the OML parametric geometry requirements but have not discussed the additional need to parametrize the IML and carry out robust automated discretization of this domain as well. An effort to address the specific needs to create parametric and associative geometry for MDO is the Engineering Sketch Pad (ESP) [261]. ESP has the ability to develop multiple, multifidelity linked models for the OML and the IML that can be used by multiple engineering disciplines throughout the design process. ESP also computes analytic geometric sensitivities that are required for shape and topology optimization.

Single discipline aerodynamic shape optimization at the RANS level of fidelity has become a standard practice in the aerospace industry. Many in-house, academic and government RANS [61,259] solvers are available that have analytic discrete aerodynamic adjoints that enable gradient-based, vehicle-level shape optimization for steady conditions. This problem is classified as a single discipline, high fidelity, large design space ($10^3$—$10^4$ design variables) with few constraints ($10^1$). Mangano and Martins [262] used a RANS-based CFD tool coupled with an adjoint solver to efficiently capture flow characteristics and its sensitivities at supersonic and subsonic speed, a robust geometry and mesh perturbation routine and an efficient gradient-based optimizer. They performed multipoint optimization to minimize the drag over an ideal supersonic aircraft flight. Munu, et al. [263] presents a framework for supersonic aerodynamic shape optimization, which uses algorithmic differentiation (AD) to obtain design variable sensitivities. He coupled the SU2 solver to a sonic boom propagation code based on the Thomas
algorithm in order to predict the sonic boom signature at the ground. Rallabhandi et al. [264] also presents an approach to design of the supersonic aircraft OML by optimizing a weighted, loudness-based objective of the sonic boom’s signature predicted on the ground. The optimization process uses the sensitivity information obtained by coupling the discrete adjoint formulations for the augmented Burgers’ equation and CFD equations.

There are many high-fidelity, two-discipline efforts found in the literature with the steady fluid-structural problem receiving the most attention. All of these high-fidelity optimizations depend on the ability to compute analytic sensitivities of the coupled problem. In most cases, since the number of design variables exceeds the number of responses, an adjoint approach is used. Some aero-structural examples can be found in Refs. [265-267], aero-propulsive configurations [268-269], and for rotorcraft aero-structural problems [270-271]. For the rotorcraft papers, efforts still tend to focus on blade-only problems, whereas the cited fixed-wing papers are more vehicle-centric. Once again with the availability of analytic gradients these problems use high fidelity analysis, have large design spaces ($10^3-10^4$ design variables) with few constraints ($10^1$). The capability of high-fidelity static/steady aero-structural gradient based optimization for full vehicle configurations with multiple flight conditions has been well established in the literature over the past five years and is now in the process of transitioning to industry. The team of AFRL, Northrop Grumman, and Stanford is executing the quantification of the utility of aerospace derivatives (QUAD) [272] program to accelerate this transition.

Currently, transient single discipline and coupled problems are still a topic of research. This is due to the challenge of developing efficient and accurate coupled sensitivities. Full-vehicle, transient, two-discipline, coupled, high-fidelity (level 2—3) MDAO is several years away from being demonstrated. A full discussion on the development of sensitivities can be found in the new milestone “Sensitivity Analysis”.

We will now shift the discussion to concentrate on the vector that is focused on bringing more disciplines into the design environment. The majority of this work has discipline fidelity in the 0-1 range and typically uses surrogate models with non-gradient based optimization techniques to search the design space. This approach suffers from the curse of dimensionality so the number of design parameters is usually restricted to a few dozen.

Efforts that focused on including subsystem models into the MDAO environment can be found in works by Chakraborty and Mavris [273-274]. They concentrated on the integration of aircraft and engine sizing with explicit sizing and analysis of the subsystems, using models and methods that are suitable for early design. They developed a methodology to automatically define the subsystem architecture in terms of the connectivity required among the various architecture components in order to satisfy redundancy requirements. This is performed by an architecting algorithm using identified subsystem architecture heuristics.

Another example of including subsystem models earlier into full vehicle MDAO is the Optimized Integrated Multidisciplinary Systems (OPTIMUS) program [275-276]. OPTIMUS expanded the disciplines that participated in the early design stages. OPTIMUS integrated subsystems models including: full numerical propulsion system simulation (NPSS) parametric advanced propulsion system model, various detailed thermal (and fuel/thermal) management system (TMS & FTMS) models, electrical power system (EPS) models and actuation system models. In addition to the subsystem models linear static and dynamic aeroelastic (level 1) responses were included using surrogates. Finally, control power
requirements and availability analyses were integrated into a modeling framework to assess control system capability and ensure flight control feasibility. OPTIMUS demonstrated the capability on a full vehicle fighter configuration. OPTIMUS used surrogate models to represent the responses in the optimization hence the number of design degrees of freedom are limited to a few dozen.

One of the latest efforts funded by AFRL and executed by Lockheed Martin is the Expanded MDO for Effectiveness-Based Design Technologies (EXPEDITE) [277-278] program. EXPEDITE is aimed at improvement in industry’s early conceptual MDAO capabilities in several key areas: state based modeling, effectiveness-based design, path dependency, transient operation of systems and subsystems, uncertainty quantification, using high performance computing, cost, and the assessment of manufacturing issues. An important part of the program goals is advancing the state of the art for effectiveness-based design (EBD) [279]. EBD uses mission scenario measures of effectiveness (MoEs) and shifts away from the traditional paradigm of performance-based design, which uses aircraft performance metrics (weight, range, drag, etc.) as the design objective. EBD on the other hand uses mission effectiveness objectives such as cost per available seat mile for commercial air travel. A feasible implementation of EBD was created and tested across multiple mission scenarios. A vehicle designed for optimal, performance-based metrics differs significantly from the best effectiveness-based vehicle. Although EBD under EXPEDITE addresses military vehicle design, a similar approach for commercial aircraft has been demonstrated by Hwang et al. [280] where they maximized airline profit.

Currently, the standard practice in industry to carry out MDAO is to execute a Design of Experiment study and construct surrogate models to represent the various responses needed. The surrogate models typically run in seconds and are incorporated into one of the computational environments described above to carry out either a gradient or non-gradient optimization. This approach allows industry to include more disciplines earlier into the design process but limits the number of free design parameters to tens due to the curse of dimensionality associated with sampling based surrogates. Industry has begun to incorporate higher-order analyses (level 2-3) into the design process by using HPC resources to execute the design of experiments and construct surrogates of the high-fidelity responses. Once again, the number of free design parameters is limited and using sampling based surrogates to develop global surrogates of highly nonlinear systems is very difficult.

In order to achieve full vehicle MDAO, it is critical that these two primary vectors — incorporating more disciplines and bringing higher fidelity coupled analysis into the process — must merge. In order for that to occur, continued advances must be made in the area of modeling and sensitivity analysis. Efforts need to be initiated to develop analytic sensitivities for many of the subsystems being modeled such as generators, actuators, electrical systems and thermal management systems along with the continued development of high-fidelity physics sensitivities for coupled transient analysis.

**Sensitivity Analysis for MDAO**

While the Study does discuss the importance of sensitivities for MDAO and UQ, it did not create a milestone for sensitivities. It could be reasonably stated that high-fidelity, CFD-based MDAO and UQ will only be possible with analytic derivatives, as every other non-gradient-based technique is too expensive if the number of free parameters exceeds a few dozen. Hence it is recommended that a milestone for sensitivity analysis be added to the Roadmap. In high-fidelity tool development, computations of analytical sensitivities need to be considered up-front during the development of the tool architecture, as opposed to after the tool is completed. It is extremely difficult to “bolt-on” an analytic sensitivity
capability after a tool has been developed. The standard way of verifying sensitivities is via complex steps, and code complexification is again something that needs to be considered at a preliminary phase of code architecting. Although the focus in recent years has been on developing the sensitivities of coupled high-fidelity physics, efforts have to be initiated to compute analytic sensitivities of geometry and subsystem models as well. Subsystem models for propulsion, electrical systems, thermal management systems etc. need to supply analytic sensitivities. This is also true for geometric solid modeling CAD applications. They too need to supply analytic sensitivities of geometric primitives with respect to model parameters. A recent effort to supply analytic sensitivities for propulsion can be find in PyCycle [281-282], which supplies analytical derivatives to the chemical equilibrium and engineering calculations. For CAD, an open source effort, ESP/CAPS [283-285] addresses the development of analytic sensitivities for solid geometry. ESP/CAPS supports the generation of geometric models for aerospace design and analysis. Three of the key features of ESP/CAPS are its ability to develop multiple, linked models for use throughout the design process, to use user-defined features that are central to aerospace vehicles (such as airfoils, flaps, and spoilers), and to compute analytic geometric sensitivities for optimization and uncertainty quantification.

To compute analytic sensitivities of coupled physics systems, there are two approaches that have emerged: continuous and discrete. The primary difference between these two approaches is that under the continuous approach the original differential equations are first differentiated and then discretized while the discrete approach discretizes the differential equations and then differentiates them. The continuous approach eliminates the need to compute mesh sensitivities (for shape optimization) but requires the differentiation of the boundary conditions. The discrete approach does not require this but does require the mesh sensitivities. Further distinction when computing sensitivities is the difference between the direct and adjoint approach. Either direct or adjoint can be used with continuous or discrete formulations. The difference between direct and adjoint is primarily the order of operations when computing the derivatives where direct is more computationally efficient if the number of responses exceeds the number of design variables and adjoint is more efficient if the number of design variables is greater than the number of responses.

Hwang and Martins [286] give a unified view of the discrete approach for both direct and adjoint computations with a focus on the adjoint. Their work presents a new equation, the unifying chain rule, from which all the discrete methods can be derived. The computation of derivatives is described as a two-step process: the evaluation of the partial derivatives and the computation of the total derivatives, which are dependent on the partial derivatives. This theoretical development is followed by the implementation of this approach in the modular analysis and unified derivatives (MAUD) architecture [287]. MAUD formulates the multidisciplinary model as a nonlinear system of equations, which leads to a linear equation that unifies all methods for computing derivatives. This enables flow-based frameworks that use the MAUD architecture to provide a common interface for the chain rule, adjoint method, coupled adjoint method, and hybrid methods. MAUD has been implemented in the open source framework OpenMDAO. The discrete adjoint has been demonstrated on many high-fidelity steady coupled problems and is well established in the literature. The following references described previously demonstrate this capability for aerostructural [265-267], for aeropropulsive configurations [268-269], and for rotorcraft aerostructural problems [270-271].
The continuous approach for coupled problems is not nearly as mature as the discrete approach but good progress has been made recently by Canfield and Sandler [288] on coupled aeroelastic 2-D shape sensitivities. Other work on 2-D continuum sensitivities can be found in Refs. [289-292].

As mentioned previously, the transient or unsteady sensitivity analysis is also immature at this time. Currently, transient single discipline and coupled problems are still a topic of research. This is due to the challenge of developing efficient and accurate coupled sensitivities. Mani and Mavriplis [293] used the unsteady, discrete, adjoint equations to construct objective function gradients for use in geometry optimization in large-scale, 3-D, unsteady flow problems. They demonstrated the technique on a helicopter blade. Zhou [294] presents an unsteady aerodynamic and aeroacoustic optimization framework in which algorithmic differentiation (AD) is applied to the open-source multiphysics solver SU2 to obtain design sensitivities. Zhou demonstrated the technique on a 2-D airfoil. Prasan and Choi [295] use a time spectral method to compute conditions for both flutter onset and limit cycle oscillations with their design sensitivities in a computationally efficient way. One of the advantages of the time-spectral-based formulation is the computational efficiency obtained by eliminating transient flow solutions to reach a periodic steady state, through the solution approximation of a discrete Fourier series. Once again this was demonstrated on a 2-D airfoil. Other efforts that address transient aeroelastic responses can be found in Refs. [296-298]. Again, only 2-D airfoil cases demonstrated. Some 2-D topology optimization in the area of transient fluid and heat transfer using coupled adjoints was done by Makhija and Beran [299-300]. Makhija and Beran [301] also developed a general purpose framework called Spiral that generalizes coupled transient parameter sensitivity analysis equations, provides functionality for coupling multidisciplinary simulations, and facilitates extensible variable relationships. The framework was demonstrated on 2-D fluid-structure interaction problems. Finally, under the SAMS project [302], the static aeroelastic and dynamic aeroelastic coupled adjoint sensitivities have been implemented and verified on some simple 2-D and 3-D cases. With the current state of coupled transient sensitivities, vehicle transient two discipline coupled high-fidelity (level 2—3) MDAO is several years away from being demonstrated.

The above discussion concerns high-fidelity coupled sensitivities for two disciplines. There is also a need for three- and four-way coupling. Problems associated with aerothermoelastics and aerothermoacoustoelastics are examples of three and four way couplings. Recently, Kamali et al. [303] developed discrete coupled adjoint sensitivities for static and transient aerothermoelastic problems. They demonstrated the capability on a 2-D panel.

Other work in the area of three-way coupling can be found in the Sensitivity Analysis for Multidisciplinary Systems II (SAMS II) program. This is an effort between AFRL, Georgia Tech, and NASA aimed at extending the SAMS effort to include heat transfer in the coupled problem to supply sensitivities for the aerothermoelastic problem. This effort is in early development and has yet to demonstrate any capability.

The approaches discussed up to this point have been the continuous or discrete direct or adjoint techniques. These approaches work well for fidelities up to RANS (from a fluids perspective) but break down when applied to longtime averages of chaotic systems. This breakdown is a serious limitation because many aerospace applications involve physical phenomena that exhibit chaotic dynamics: most notably high-resolution large-eddy and direct numerical simulations of turbulent aerodynamic flows.
MDAO

[level 4 and 5]. Development in this area is very early on but two notable efforts can be found by Bhatia and D. Makhija [304] and Blonigan et al. [305].

Unless efficient and accurate sensitivities are available for all responses (objectives and constraints) participating in the optimization, it will not be possible to perform full gradient-based design, which is currently the only realistic way to handle the number of design variables ($10^3—10^5$) that will be present in full vehicle MDO problems. With that in mind, the additional MDAO milestones are proposed. By 2025, analytic sensitivities for all relevant subsystems will need to be developed, analytic geometric sensitivities for full vehicle configurations for the inner mold-line and outer mold-line will need to be defined, resulting in full vehicle 3-D steady and transient two, three, and four way coupled analytic sensitivities. By 2030, full vehicle 3-D analytic sensitivities for chaotic systems should be demonstrated.
Summary

This paper has reviewed key developments in each of the domains identified in the original Study to highlight progress that has been made that is aligned with the Vision Roadmap. Significant progress has been made in many of the areas through 2020, but there are several milestones that are not yet at the desired level. 2020 represented the first set of technology demonstrations identified on the Roadmap and all have been successful to some level, but the individual technologies are still at the demonstration level of maturity and require leadership-class computational capabilities. Mesh adaptation and large grids remain pacing items for both grid convergence and separated flow. Multiple varieties of scale-resolving simulations, including low dissipation algorithms, have made large progress in capturing separated flow, but conclusive comparisons have not yet been made, partially due to the complexity associated with detailed nonintrusive experimental measurement in these conditions. As these efforts continue, developing approaches for extracting meaningful data from these simulations will become critical. Associated with each of these challenges is the requirement to effectively integrate multidisciplinary analysis as aerospace system analysis includes more than fluid dynamic considerations. The Roadmap highlights all of these trends and, while progress has not been completely predicted, appropriate trends are highlighted and it has helped to provide direction for technical advancement since its release.

To preserve the Roadmap as an aspirational guide for CFD development needs, a revision is planned that will use the information gathered in this paper to provide updates to technology readiness level assessments of the identified milestones and make required adjustments to the path toward the vision. Furthermore, annual updates on the progress of the technologies identified in the Roadmap will continue to be made; please refer to www.cfd2030.com for more details or to find information on how to contribute to these updates.
Contributors

The authors want to acknowledge individuals who made significant contributions to the different sections. Your assistance in this endeavor is very much appreciated.

High Performance Computing: Kenneth Jansen (University of Colorado), PK Yeung (Georgia Institute of Technology), Stephen de Bruyn Kops (University of Massachusetts)

Physical Modeling: Chris Rumsey (Computational AeroSciences Branch, NASA Langley Research Center), Philippe Spalart (The Boeing Company, retired), George Huang (Wright-State University), M.S. Anand (Rolls-Royce), Jeffrey Moder (NASA Glenn Research Center)

Algorithms: Dimitri Mavriplis (University of Wyoming), Boris Diskin (NIA), Dmitry Kamenetskiy (The Boeing company), Mory Mani (The Boeing Company, retired), Hiro Nishikawa (NIA)

Geometry Modeling and Mesh Generation: Michael A. Park (Computational AeroSciences Branch, NASA Langley Research Center), Nigel Taylor (MBDA UK Ltd), David McLaurin (Siemens Digital Industries Software), Nick Wyman (Pointwise)

Knowledge Extraction: Steve M. Legensky (Intelligent Light),

MDAO: Anirban Chaudhuri (University of Texas at Austin), Daniel Clark (AFRL), Clifton Davies (Lockheed Martin), Edwin Forster (AFRL), John Gallman (Northrop Grumman), Ramana Grandhi (Air Force Institute of Technology), Justin Gray (NASA), Jim Guglielmo (The Boeing Company), Joaquim Martins (University of Michigan), Scott Sellers (The Boeing Company), Nagendra Somanath (Pratt & Whitney), Bret Stanford (NASA), Marc Stelmack (Lockheed Martin), Karen Wilcox (University of Texas at Austin)

Acknowledgements

The section on MDAO has been cleared for public release as AFLR-2021-0179 – CLEARED on 25 January 2021.

Works Cited


Y. Hu and e. al, "Fast Tetrahedral Meshing in the Wild," *ACM Trans Graph*, vol. 39, no. 4, p. Article 117.


[163] [Online]. how5.cenaero.be.


[189] [Online]. https://turbmodels.larc.nasa.gov/other_dns.html.


[191] [Online]. https://ctr.stanford.edu/research-data.


[307] [Online]. https://www.paraview.org/in-situ/.