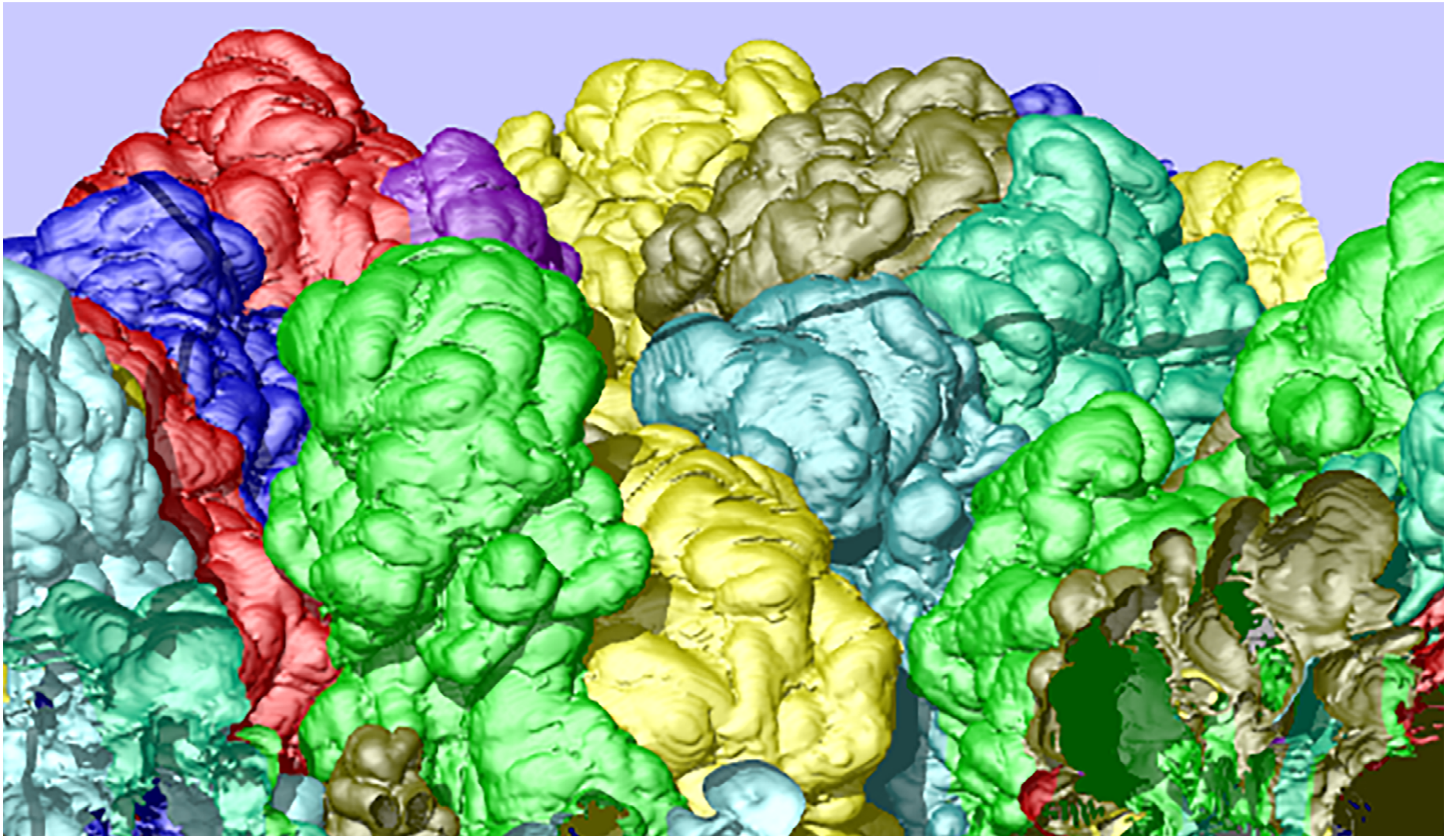


VISUALIZATION
for Scientific Discovery, Decision-Making, & Communication

U.S. DEPARTMENT OF ENERGY | Office of Science

doi.org/10.2172/1845708

TABLE OF CONTENTS

1. Executive Summary

Visualization—the use of visual elements to explore data, form hypotheses, or convey conclusions—is an integral part of the scientific process. Starting from an initial exploration of new data to illustrating outcomes for the general public, visualization is one of the most intuitive and powerful modes of communication. With the explosion of new data sources and types, unprecedented volumes of data, and new technologies, such as virtual reality (VR) and artificial intelligence (AI), visualization has become increasingly essential but also ever more challenging.

The Department of Energy's (DOE) Office of Advanced Scientific Computing Research (ASCR) sponsored a Basic Research Needs workshop in January 2022 to understand the major opportunities and grand challenges in visualization tools and technologies for scientific computing as well as for DOE-relevant applications and goals in general. The workshop identified five priority research directions (PRDs) for visualization to support scientific discovery, decision-making, and communication. The first three PRDs describe interconnected research themes addressing the need for new techniques to deal with complex data, uncertainty, and interpretability (PRD 1); the need for scalable and interoperable software stacks (PRD 2); and the challenges and opportunities inherent in new technologies, such as VR, cloud, or exascale computing (PRD 3). The remaining two PRDs describe foundational research themes that recognize the potential of visualizations to provide equitable access to information and to strengthen the scientific discourse (PRD 4); and the need to consider human factors when designing visualizations (PRD 5). Collectively, these PRDs form the pillars for a coherent, long-term research and development strategy in Visualization for Scientific Discovery, Decision-Making, and Communication in the context of the Office of Science's mission scope.

1.1 Priority Research Directions

The field of visualization presents a number of diverse challenges ranging from deeply technical questions about algorithmic innovations to graphical design, and from developing massively parallel frameworks to considering human cognition and trust. The workshop organized these heterogenous challenges into five PRDs conceptually organized in Figure 1.1. In the center are three broad technical thrusts within the area of visualization, roughly split into algorithms, software, and hardware and similar to ideas articulated in prior workshops. Flanking these are two cross-cutting themes that, for the first time, explicitly highlight the opportunities of visualization as a versatile communication tool for DOE and the need to deliberately consider how human perception and idiosyncrasies affect visualization results.

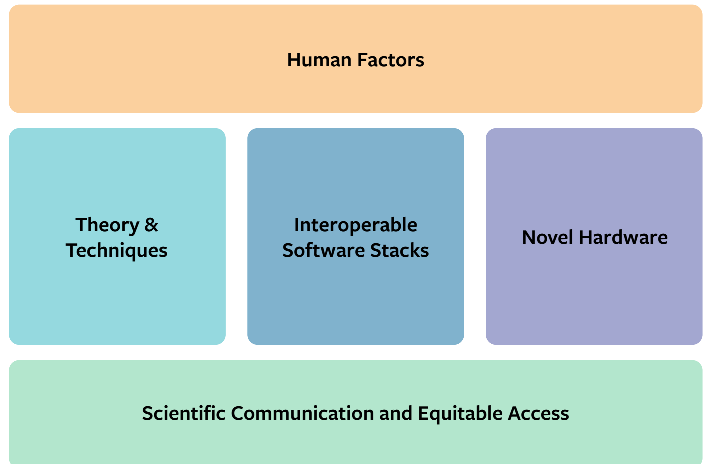


Figure 1.1. The workshop's five PRDs are conceptualized as three technical thrusts cross-cut with two themes. Together they form the pillars of a long-term strategy for ASCR's visualization research and development.

1.1.1 Advancing Theory and Techniques for Visualization to Support the Analysis and Understanding of Complex Scientific Data

New techniques and corresponding theory are needed to develop novel representations, algorithms, and systems to promote scientific understanding of the many different data types of interest to DOE. Many of the common scientific visualization techniques focus on 2D or 3D scalar fields. However, most of the data collected today has much more complex encoding, either because at each location multiple values (multivariate) or even multiple data types (multimodal) may exist, or because data is defined non-spatially (e.g., in high dimensions, in graphs). Furthermore, abstract information about the behavior of a complex system or the inherent uncertainty of a decision could be more accessible through intuitive visualizations.

1.1.2 Introducing Interoperable and Adaptable Visualization to Support Diverse Scientific Workflows Across All Scales

New visualization approaches are required to leverage as much common infrastructure as possible, maintain data provenance and uncertainty information, and span the needs of multiple scientific communities while taking advantage of the unique resources and opportunities of the specific application. The need for informative visualizations is near universal, covering a plethora of scientific domains and systems ranging from embedded sensors to high performance computing (HPC) resources, rare and expensive data, and petabyte-sized data collections. Developing bespoke solutions for every unique combination is infeasible.

1.1.3 Harnessing Technology Innovations to Accelerate Science through Visualization

New techniques are needed to take advantage of novel displays, interfaces, and collaboration tools while exploiting state-of-the-art computing architectures. Rapidly evolving technology is creating tremendous opportunities for new visualization and collaboration paradigms with novel displays, VR, haptic interfaces, and more. However, realizing this potential to develop immersive yet intuitive environments or enable real-time, remote collaborations will require techniques and frameworks significantly beyond the current state of the art. Furthermore, visualization must effectively exploit diverse and disruptive computing modalities, including exascale platforms, cloud, sensors, and mobile chips, each of which may enable unprecedented visualization solutions if they can be integrated effectively into portable tools.

1.1.4 Improving Equity in Accessing and Engaging with Scientific Data and Processes

New approaches to visual communication are needed to significantly accelerate team science, integrate all stakeholders—from policy makers to the general public—and change the scientific discourse. Visualization is one of the most powerful ways to communicate complex ideas, concepts, and decisions across domains, educational backgrounds, or cultures. Distinct from the goal of providing new insights to subject matter experts, visual communication requires a new focus on accessibility, transparency, and trustworthiness to promote engagement and understanding.

1.1.5 Developing Intelligent Approaches for Adaptive, Context-Aware Visualization of Scientific Data and AI

Visualizations are designed to be consumed by humans. Understanding how individuals may interpret visual content differently depending on the application context and intended use is a key challenge for designing new tools and techniques. Recent advances in our understanding of perceptual learning and reasoning processes are crucial in developing general-purpose, intelligent visualization tools that are readily customizable to meet the evolving needs of the users (e.g., scientists, general public) and the task at hand (e.g., hypothesis generation, decision support, scientific information dissemination). Furthermore, as data and visualization become more embedded in scientific exploration and systemic decision-making processes, visualization tools must be properly validated to ensure broad and lasting impact. A growing need exists for human-centric design and evaluation methodologies, particularly those that properly account for minority viewpoints. Technical advances are needed across multiple fronts: (1) development of methods to understand perception and cognition to drive the development and adaptation of next-generation visualization tools and serve as surrogates for large-scale visualization evaluation studies; (2) methods to enable dynamic personalization of and collaboration in visualization tools to address the needs of individual users for scientific discovery and effective utilization; and (3) development of robust, scalable, and unbiased evaluation methods and metrics for visualization tools.

2. Introduction

At the heart of scientific progress lies our ability to communicate ideas and ultimately share insights that jointly advance our understanding of the universe. With as much as 50% of the human brain devoted to visual processing, visualization—the use of visual elements to explore data, form hypotheses, or convey conclusions—has always been an integral part of science. As technological advances have allowed us to produce ever more complex data, visualizations have evolved from rare and extravagant efforts that are now considered works of art [1] to a day-to-day necessity to many branches of science. The 1987 National Science Foundation Report on Visualization in Scientific Computing [2] is often considered the first official recognition of the crucial role visualization can play in scientific discovery. Subsequent years saw an explosion of new tools and techniques and other efforts highlighting both open challenges as well as the potential impact of visualizations [2][3]. Arguably the most well-known of these efforts is Johnson's 2004 list of the top 15 research challenges in scientific visualization [4].

DOE has been at the forefront of these research efforts with visualization being an integral part of both the NNSA Advanced Strategic Computing Initiative (ASCI) [5] as well as the Office of Science Scientific Discovery through Advanced Computing (SciDAC) and ASCR (Advanced Scientific Computing Research) programs. (See Figure 2.) Most notably, a chain of successful SciDAC engagements began with two visualization-focused SciDAC2 projects, the Visualization and Analytics Center for Enabling Technologies (VACET) [6][7] and the Ultra-Scale Visualization Institute [7], and continues to this day as part of the RAPIDS2 institute for Computer Science, Data, and Artificial Intelligence [8]. The SciDAC engagements alongside many smaller DOE-funded projects have produced a suite of visualization tools and techniques that are now indispensable in many applications. These programs were designed to, and still do, directly address the needs of various DOE missions and have been the seed for many successful collaborations between visualization and application experts.

For the field of visualization, an unintended consequence of this focus on direct and (relatively) immediate impact has been an increasing shift in research and even strategic planning towards more immediate practical concerns. From VACET in 2006 [6], the focus moved to “Data Management, Analysis, and Visualization” for exascale and experimental data [9][10] in the 2011 and 2016 DOE workshops to visualization becoming one of many concerns for “Exascale Requirements” [11], “In Situ Data Management” [9], and “Storage Systems” [12] in 2018. These efforts correctly identify the many practical challenges of visualization, especially at DOE scales. For example, many algorithms are bound by data movement costs; storing data is too expensive and thus visualization may have to occur in situ; and at the largest scale, automatic feature detection might be the only viable path to insight. Nevertheless, this shift in focus does not change the fact that many of the research challenges outlined by Johnson in 2004 [4] and even by Rosenblum in 1994 [3] remain open. In fact, in many cases the need for research-driven solutions is now significantly greater than it has ever been. DOE produces larger, more diverse volumes of data than ever before, and decisions on subjects ranging from national security to energy policy are increasingly based on this data. Especially when these decisions involve interdisciplinary teams, policymakers, or the public at large, “decisions based on data” usually translate to decisions based on visualizations of data. Yet we are not much closer to integrating uncertainty in these visualizations (Challenge #3), considering perceptual issues (Challenge #4), dealing with multifield visualization (Challenge #9), or addressing many of the other 20- to 30-year-old challenges.

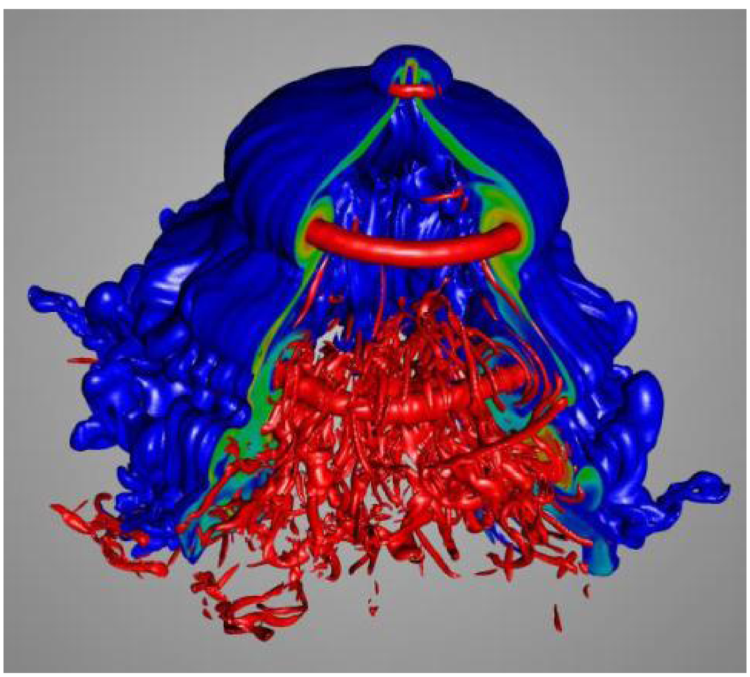


Figure 2 Initially developed under ASCI, the VisIt software tool provides DAV capabilities for scientific codes, such as this rendering of a complex surface—a dense spherical gas bubble subjected to a strong planar shockwave.

The purpose of the Visualization for Scientific Discovery, Decision-Making, & Communication workshop was to bring together experts from DOE, academia, and industry to discuss the current state of the art, update and re-formulate the list of research challenges in visualization, and ultimately suggest priority research direction for DOE to pursue. The workshop attracted 229 participants, which over the course of three days and 23 breakout sessions identified both gaps and opportunities for visualization (see Workshop Summary) in the context of DOE missions. The final outcome includes five high-level PRDs that encompass and extend previous lists of challenges. The first three PRDs describe interconnected research themes addressing the need for new techniques to deal with complex data, uncertainty, and interpretability (PRD 1); the need for scalable and interoperable software stacks (PRD 2); and the challenges and opportunities inherent in new technologies, such as VR, cloud, or exascale computing (PRD 3). The remaining two PRDs describe cross-cutting research themes that recognize the potential of visualizations to provide equitable access to information and to strengthen the scientific discourse (PRD 4); and the need to consider human factors when designing visualizations (PRD 5). Collectively, these PRDs form the pillars of a coherent, long-term research and development strategy in Visualization for Scientific Discovery, Decision-Making, and Communication in the context of the Office of Science’s mission scope.

3. Priority Research Directions

3.1 Advancing Theory and Techniques for Visualization to Support the Analysis and Understanding of Complex Scientific Data

New techniques and corresponding theory are needed to develop novel representations, algorithms, and systems to promote scientific understanding of the many different data types of interest to DOE. Many of the common scientific visualization techniques focus on 2D or 3D scalar fields. However, most of the data collected today has much more complex encoding, either because at each location multiple values (multivariate) or even multiple data types (multimodal) may exist, or because data is defined non-spatially (e.g., in high dimensions, in graphs). Techniques for preserving key features from novel data representation and reduced data formats, either from compression or feature extraction techniques, as well as for processing ensemble simulation data to encode variations, sensitivities, and uncertainty of the simulation outputs are still lacking. Furthermore, abstract information about the behavior of a complex system or the inherent uncertainty of a decision could be more accessible through intuitive visualizations. Meanwhile, conversion algorithms between all of these representations are necessary but challenging, especially when facilitating them all with corresponding algorithms.

3.1.1 Data Representations

While visualizations are valuable in virtually all branches of science, they rarely are the primary focus. Accordingly, whichever data should be visualized is typically optimized for simulation code that produced it, the experiment that collected it, or, more generally, the most expensive processing step necessary to curate it. As a result, visualization techniques must be able to ingest a huge variety of data representations, converting them into either standard formats or directly extracting the salient information, and doing so without inflating the already large input data beyond the available computational resources. This section discusses the associated algorithmic and methodological challenges, while Section 3.2 provides a related discussion on the need for portable and reusable software.

Key challenges

Visualization of large-scale scientific data is facing three key challenges related to efficient representations of data (e.g., simplification, resampling, topological decomposition, feature extraction, compressions). The first challenge is to develop effective representations for complex data from emerging scientific applications such as point clouds, high order meshes with novel geometric primitives, and solutions associated with multi-dimensional probability density functions. (See Figure 3.1.1.) Developing efficient methods that allow conversion of data between representations is also important. The second challenge concerns the conservation of features of importance to domain scientists. Depending on the simulation or experiment and the analysis performed, users have different requirements. Often users consider the preservation of local properties (e.g., point-wise field values and their derivatives) and global properties (e.g., statistics, spectral profiles) as critical. Another important requirement concerns the preservation of structures (e.g., topological or geometric descriptors). So, from a visualization perspective, one needs to know how the loss of accuracy in the representation from the original dataset will impact the ability of visualization algorithms to preserve the quantities of interest for the viewer. The final challenge is to keep the data in a compact form to the extent possible during visualization. Reconstructing the data in its entirety before visualization is often not practical or possible because of memory and bandwidth constraints.

State of the art

To convert data that have different representations, Voronoi and Delaunay tessellations [13] are often used to convert point data to a mesh. Sampling methods [13][14] are used to extract individual data points from a collection, and density estimation is often used to convert point data (e.g., from an

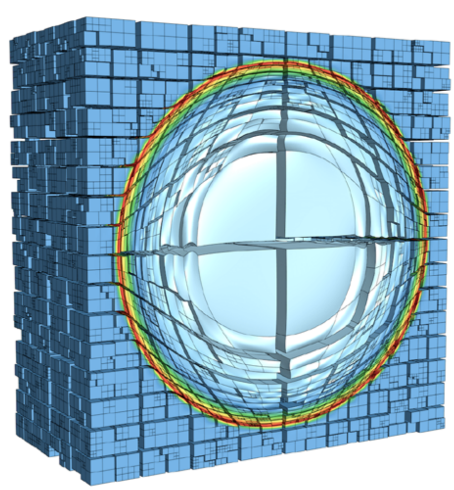


Figure 3.1.1
Propagation of a spherical shockwave through a random non-conforming mesh in the MFEM-based BLAST shock hydrodynamics code. MFEM is a discretization library used for building a variety of simulations in physics codes.

N-body simulation) to a preset discretization. Recently we have seen both functional and neural representations appear in the visualization literature. Representative work in implicit neural networks includes SIREN [15], ACORN [16], NeRF [17], and Mip-NeRF [18], and a comprehensive survey of this topic has been published [19]. Regarding more compact data formats, recent research [20][21] has demonstrated the possibility of preserving topological structures (e.g., critical points, persistence pairs). Most advanced data representations via compressors such as MGARD [22], SZ [23][24], and ZFP [24] support the preservation of point-wise values with absolute and relative error controls. MGARD [22], SZ [23][24], and TTHRESH [25] also provide control over aggregate (statistical) error measures such as the peak signal-to-noise ratio or Sobolev norms. Furthermore, other more domain-based data representations have been proposed to reduce meshes while preserving some of their geometric (e.g., volume, surface area) or topological (e.g., genus, scalar field topology) properties; see the recent survey by Li et al. [26]. Adaptive multilinear meshes (AMM) decompose a field into piecewise multilinear cells of varying size, allowing traditional visualization algorithms (e.g., isocontouring, direct volume rendering) to operate on the reduced representation [27].

Research directions

From users' requirements on efficient data representation and the current state of the art, we derive three important research directions. The first direction concerns the design and development of new data representations that complement existing feature preservation mechanisms with preservation of topological, geometric, and other feature descriptors, the related error quantification and analysis tools; and understanding the impact of lossy data representation on visualization algorithms. For example, users should be able to express tolerances and, when appropriate, error bounds as well as assess errors on topological features (e.g., critical points, contour trees, Morse complexes), curves (e.g., ridge lines, vortex core lines, streamlines), surfaces (e.g., isosurfaces, stream surfaces, fiber surfaces), and volumes (e.g., superlevel sets) in addition to point-wise error bounds. The second research direction concerns the design of new principles and methods and development of new software for novel data representations. This research direction relates to programming interface and high performance implementations of reduction techniques as well as to novel lossy reduced data representations that can be directly visualized without un-reducing the data first. Finally, the third research direction concerns understanding how different types of inaccuracies introduced by different representations of data will impact image generation and, thus, how users will perceive the images. This research direction will involve user studies and the design of new perception-based models that can correlate error loss with perception.

3.1.2 Multivariate and Multimodal Data

With few notable exceptions the majority of existing scientific visualization algorithms focus on analyzing one property at a time, which can be a single scalar field, a vector field, or, more rarely, a higher order tensor. However, both simulations and experiments typically generate a plethora of information at each point in space (or time), each one of which might itself be a complex entity (e.g., an image, a spectrum). Finding visual encodings of such multivariate (many properties) and multimodal (of different types) data remains difficult.

Key challenges

Almost all scientific applications generate multiple variables. Those variables often carry complex relationships, and jointly they represent important scientific features. When analyzing multivariate data, scientists face several key challenges. One is determining an accurate and succinct summary of the relationships among the variables, in particular when the correlations between variables are nonlinear and the volume of data is large. Also challenging is the nontrivial task of comparing and contrasting the features generated from different variables (e.g., isocontours) or their topological summary (e.g., contour trees). (See Figure 3.1.2a.)

The other challenge is that datasets are often extremely large, which increases the difficulty of performing interactive queries even for a single variable, let alone a combination of variables. When the number of variables is large, the combination of variables to consider when analyzing a given scientific phenomenon is often unknown. Queries by trial and effort often lead to high computational costs with no guarantee that salient features can be identified. Furthermore, effective visual encoding that can simultaneously display information from multiple variables is lacking. Also lacking are methods that can quantify the uncertainty in the multivariate correlation and track its temporal variation. Multimodal datasets are heterogeneous and defined over different data types (e.g., video, audio, text, and simulation output), domains (e.g., space, time, and spectra), and sources (e.g., simulations, observations, annotations) that in combination describe specific scientific phenomena. Numerous DOE Office of Science applications—cosmology, energy science, climate, nuclear physics, and more—use multimodal data

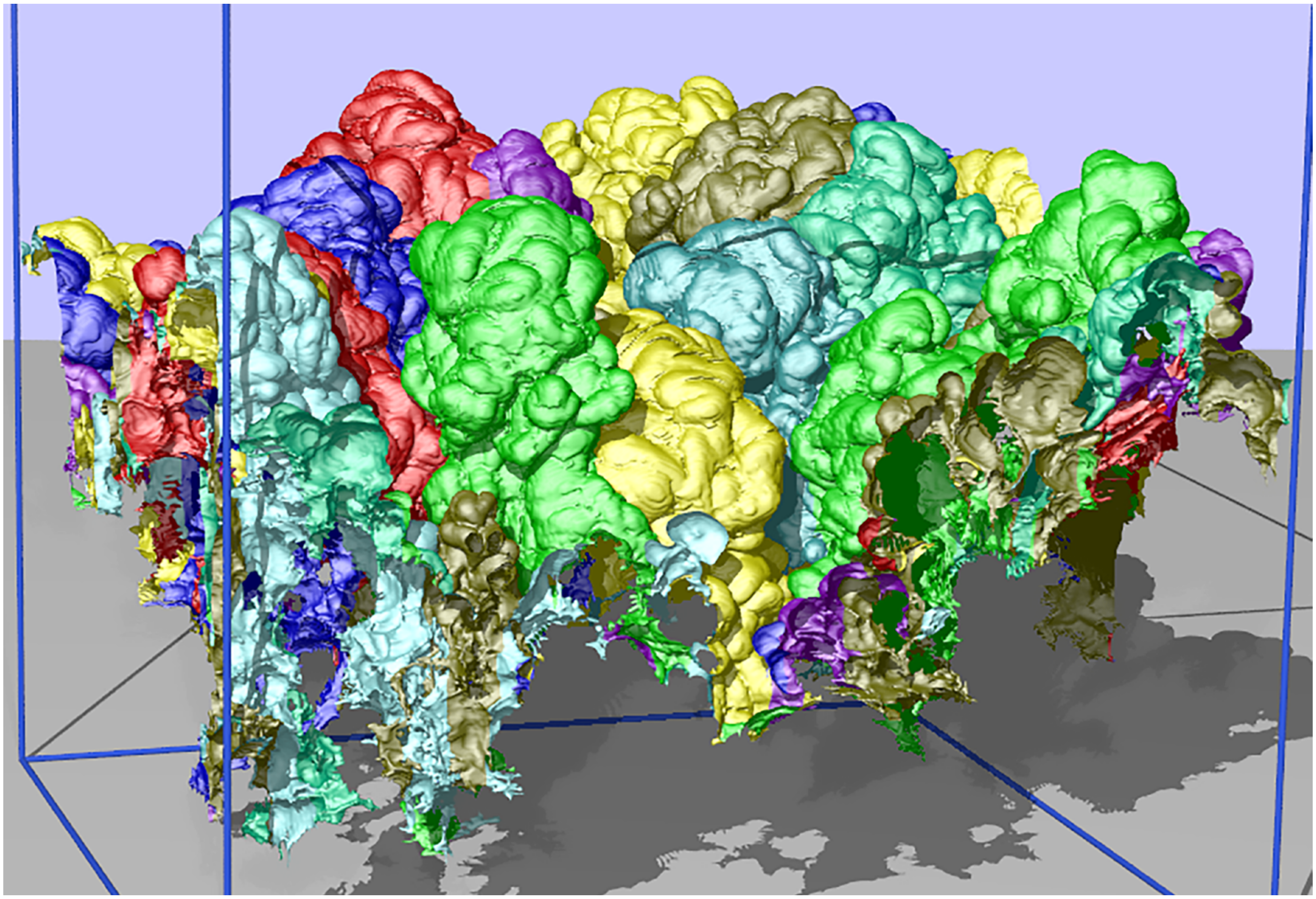
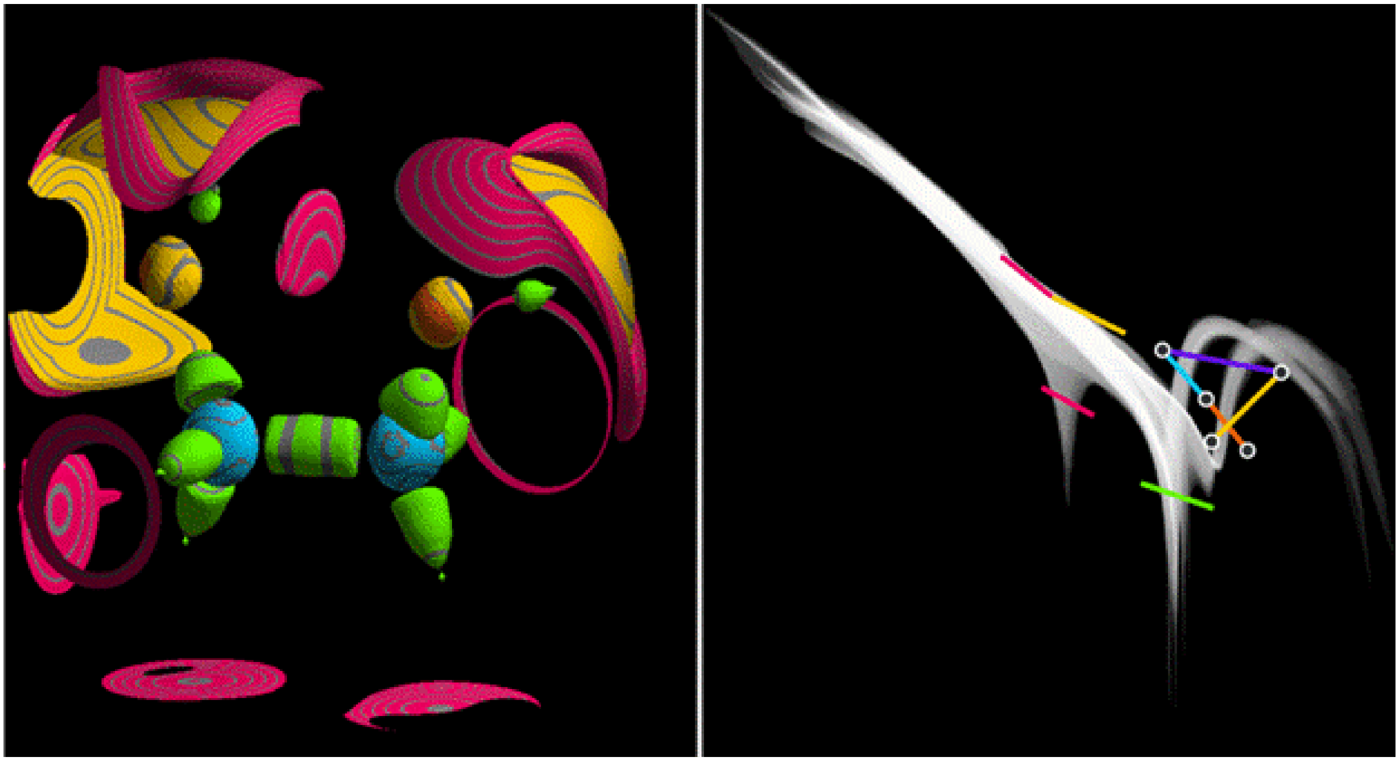


Figure 3.1.2a Topological features extracted from large-scale combustion simulations indicating connected regions of a burning flame front in a high-turbulence cross-flow.

to investigate hypotheses. Visualizing (e.g., comparatively) multimodal data remains a grand challenge for our community, and further research is needed to satisfy the evolving needs of scientific applications. In principle, combining more than one data modality should increase understanding, but the incompatibility of multimodal data often has the opposite effect, complicating processing, obscuring underlying patterns, complicating validation and reproducibility, and ultimately hindering communication and decision-making.

State of the art

For multivariate data, researchers have developed statistical methods that calculate the information distance between variables, thus identifying variables that are more important to visualize. These methods contribute to identifying salient information among the variables and, in turn, reducing scientists' cognitive load for data analysis and visualization (DAV). For features that can be identified by the joint distributions among variables, methods also exist to facilitate efficient search and queries [28]. To extract and exploit the correlations in the multivariate datasets and to enable effective sampling and reconstruction, Copular or Gaussian mixtures based methods were proposed for data sets that can be modeled by the underlying parametric models [29]. Local relationships can be extracted by identifying sets of biclusters [30]. Finally, a method also exists that can display multiple variables simultaneously based on blending or visual encoding [31]. The topology community has developed many works for multivariate data comparison (e.g., largest contour comparison [32], quantizing the variations using joint contour nets [33], local and global comparisons [34][35], and generating fiber surfaces [36][37]). (See Figure 3.1.2b.) For multimodal data, to date no uniform way exists to fuse or transform data of different modalities; different data models are treated individually, and various methods convert from one modality to another. As mentioned in the previous sections, methods [9] to convert point data to meshes and density estimation are often used to convert point data (e.g., from an N-body simulation) to a preset discretization.



Research directions

To enable flexible exploration of multivariate data, future research directions include automatic identification of the most salient variables and their relations This recommendation requires detailed analysis of the information content and distance between variables both in space and time. To facilitate effective visualization of multivariate data, new visual mapping and encoding schemes across different modalities (scalars, vectors, and tensors), as well as novel interaction techniques that allow probing, selection, and projection, will be needed. A strong need also exists for novel data representation and management schemes that allow efficient out-of-core data retrieval and query. Finally, the ability to allow scientists to define, extract, track, and visualize geometric and topological features across scales will be crucial for scientific discovery.

Future research directions of multimodal visualization will focus on the integration of information from various sources to enable decision-making. As data of different modalities are often defined in different domains, successful information fusion from multimodal data will need to involve: (1) information aggregation and amplification; (2) information comparisons to identify unique and salient data features from each source; (3) information fusion with a guarantee of minimum information losses. To facilitate the necessary comparison or fusion, an important research problem becomes how to convert the multimodal data into

Figure 3.1.2b Interactive extraction of fiber surfaces from bivariate fields [36]. Right: Continuous scatterplots with colored selection of the fiber surface control polygon; Left: resulting colored fiber surfaces.

features that are invariant to the sources of data. In addition, effective visual substrates that can display multimodal data in a unified space to enable feature identification and tracking features will be proven very important. Finally, novel data sampling, matching, and dimensionality techniques for multimodal datasets to allow efficient queries will be crucial.

3.1.3 High-Dimensional Data

High-dimensional data arise naturally from the scientific analysis of complex systems (from simulations to observations) as well as machine learning (ML; from feature representations to loss landscapes). A variety of approaches have been introduced for high-dimensional data visualization including dimensionality reduction, regression analysis, direct visual encoding, and interactive exploration. However, new visualization techniques are needed to address the increasing size and complexity of these high-dimensional data to obtain actionable insights.

Key challenges

Learning high-dimensional data representations emerges as a central task in various scientific analysis and ML applications. The "curse of dimensionality" and the "empty space phenomenon" affect not only the efficiency of learning, but also the effectiveness of visualization. Certain high-dimensional information is invariably lost during data transformation, while human intuition no longer easily applies to visual mapping. In addition, often little (or no) ground truth exists, making validation and evaluation difficult. While high-dimensional surrogate models such as neural networks and Gaussian processes models can be trained to map input parameters to observed outputs, sample selection is still constrained by the high intrinsic dimension and is oftentimes not amenable to dimensionality reduction. Furthermore, learning and visually communicating data representations with uncertainty remain a challenge. Finally, verification and validation are essential in science, and they rely on comparisons between different data and models. The key challenges include understanding the fundamental differences between high-dimensional simulations and experiments, as well as aggregating such differences for visualization.

State of the art

A number of surveys cover different aspects of high-dimensional data visualization, including parallel coordinates [38][39], quality measures [38]–[40], clutter reduction [41], visual data mining [42][43], and interactive techniques [44]. For high-dimensional scientific data, Burger and Hauser discussed techniques from various disciplines in visualizing multidimensional scalar, vector, and tensor datasets. Kehrer and Hauser explored representation, analysis, and interaction of multidimensional data with spatial and temporal characteristics [45]. The most recent survey of Liu et al. [46] classified recent advances based on different stages of an enriched visualization pipeline, including data transformation, visual mapping, view transformation, and user interaction. For data transformation, techniques include dimension reduction, subspace clustering capturing various aspects of data, regression analysis, and topological data analysis for gaining insights of data. For the task of visual mapping, methods are developed to convert the original data or analyzed result into visual structures. According to the structure patterns, axis-based methods, glyphs, pixel-oriented methods, hierarchy-based approaches, animations and evaluations are typical categories. Then view transformation includes various approaches to finally render the visual content to the screen space (e.g., illustrative rendering for a specific visual style, continuous visual representation to deal with visual clutter and computational cost, color blending considering perception of order and structure, and image space metrics for quality measures). Finally, user interaction includes many efforts focusing on computation-centric, interactive exploration, and model manipulation.

Research directions

Three potential research directions relate to high-dimensional data. First, visualizing high-dimensional data representations aims to preserve structural information while maintaining interpretable visual encodings. However, these two objectives are often at odds with each other. For instance, non-linear dimensionality reduction can capture good intrinsic structures, but the dimensions of the embeddings are often hard to interpret. Providing meanings to dimensions, such as discovering semantic or concept directions, remains under-explored. A potential solution is creating a summary from the data to serve as an abstraction for visualization—i.e., mapping individual samples onto abstract summaries, which are obtained by leveraging analysis tools from graph theory, topological data analysis, subspace learning, representation learning, and more. Data derived from real-world applications often include various anomalies and outliers, and effectively identifying and interpreting them is an integral part of high-dimensional data visualization. Furthermore, the visualization should be designed in a way that retains and reflects the application information for advancing domain understanding. Then, with the wide adaptation of neural network models in scientific applications, some of the pressing questions to be addressed include how to model and assign semantics in the space of high-dimensional learned representations, and where and how visualization and human interaction can help with their interpretations [47].

Second, high-dimensional surrogate modeling captures a mapping between a model input and its output. Utilizing visualization to explore model behavior and evaluate model performance (e.g., for model steering, debugging, and adaptive refinement) are crucial for building better surrogates. In particular, understanding the structure of such a mapping (e.g., via the loss landscape) is important for constructing intuitive visual encodings, such as its local minima, local maxima, and basins of attraction before, during, and after training. The next step is performing visual analytics of model ensembles [48] to better understand the underlying physical phenomena. Finally, scientific insight generation requires comparing and contrasting data of interests—in particular, between simulations and experiments, between different surrogate models, and between configurations that lead to failures and successes. To provide a meaningful comparison of high-dimensional structures, the first step is to develop novel statistical, geometric, and topological metrics, which facilitate local and global comparative analysis. Such comparisons are also crucial for understanding data distribution shift and domain adaptation [49], particularly for evaluating and generalizing ML models. Furthermore, visualizing the evolution of samples in a high-dimensional space (e.g., the phase space of a physical system or the nonlinear transformation between neural network layers) is a relatively unexplored area that could inspire new solutions and applications.

3.1.4 Ensembles

Despite tremendous advances in computational sciences, a single simulation is rarely considered to be conclusive because of the many, often uncertain, assumptions that are required. Furthermore, codes deemed to be predictive are often used in design optimization or scenario exploration. Such considerations invariably lead to computational ensembles rather than individual simulations, just as experimental science has long thrived on repeats and complex designs of experiments. With thousands or millions of ensemble members, visualizing any one example becomes less effective, and instead, new techniques are required to explore ensembles as a whole.

Key challenges

Ensemble simulations are used in various DOE science applications (e.g., cosmology, Earth systems, fluid dynamics) to understand simulation model sensitivities/variations, quantify uncertainties, and optimize model designs. Each run in an ensemble of simulation executions can cost millions of core hours and generate terabytes to petabytes of diverse data in a very high-dimensional space, making simulation of all possible configurations of the input parameters prohibitively expensive. Furthermore, attempting to analyze exhaustively the data generated from large-scale ensembles is equally expensive. Effective, efficient analysis of ensemble data with tractable computational and input/output (I/O) costs remains a grand challenge in the DAV community due to the following difficulties: (1) the capability to explore simulation parameter spaces is limited; (2) the ability to transform ensemble simulation data into compact representations is missing; and (3) scalable algorithms for in situ analysis and visualization of ensemble data are generally lacking.

State of the art

The DAV community has been extensively developing ensemble data visualization techniques, and a comprehensive survey of state-of-the-art developments can be found in [48], which categorizes ensemble visualization techniques into location- and feature-based approaches. Location-based approaches compare ensemble members at fixed spatiotemporal locations, while feature-based methods first identify features and then compare features across ensemble members. For example, spaghetti plots are a well-known method that overlays curve features such as contours to visualize differences between ensemble runs; such visual representations may be further abstracted with contour boxplots [50]. Beyond the visual representation of ensembles, an essential task in ensemble visualization is to help understand the impact of simulation parameters to the simulation results. For example, Sedlmair et al. [51] provided a conceptual framework that contains a data flow model, a set of navigation strategies, and a characterization of analysis tasks for visual parameter space exploration. More recently, researchers have investigated AI models for parameter space exploration. For example, InSituNet [52] is a deep image synthesis model for predicting visualization results with given simulation parameters, which help users preview the impact of parameter changes. In addition, distribution-based methods are used to model ensemble data. Liu et al. [53] used a Gaussian mixture model per sample to represent scalar field ensembles for visualization. Bayesian model averaging is also used to establish a statistical aggregation of ensembles for ensembles [54]. More recently, He et al. [55] modeled samples in ensemble data as range-likelihood values, which further enables clustering and visual exploration with likelihood volumes.

Research directions

A primary broad-scale research direction in ensemble DAV is to develop generalized theoretical bases for ensemble

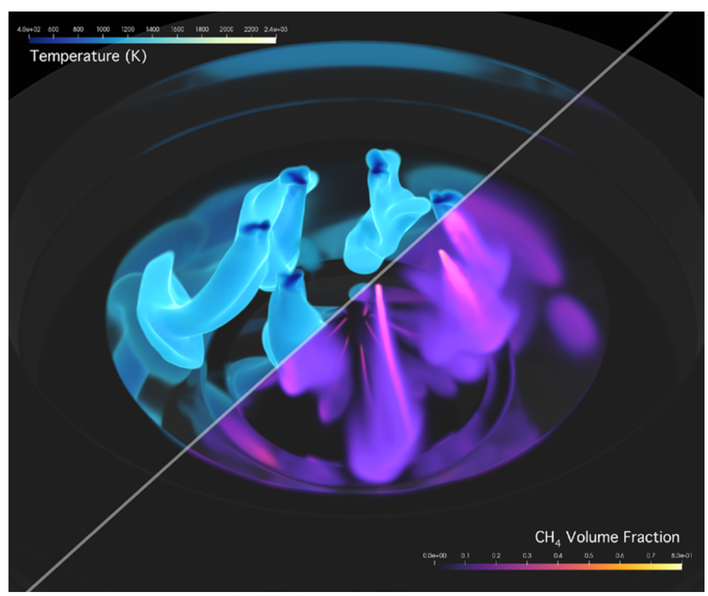


Figure 3.1.5 Predictive simulations of the complex incylinder processes of internal combustion engines that are being used to improve performance and decrease pollutant emissions.

representation, analysis, and visualization as the existing works are mainly domain-specific and often application-driven. Another important research direction is to develop ensemble analysis frameworks that have the capability to perform uncertainty quantification and visualization at every step of the analysis pipeline. Close collaborative research in ensemble analysis will further bridge the gap between cognitive science, uncertainty visualization, and application sciences. Apart from these broad-scale general research directions, precise task-specific needs for ensemble data visualization can be threefold. The first is simulation parameter space exploration. In contrast to traditional simulation surrogate research, a need exists to develop AI and statistical surrogates to predict the visualization of ensemble simulation output with any given configuration of input parameters. Visualization surrogates could skip generating simulation data and directly produce visualizations, allowing scientists to obtain a quick preview of their simulations to determine more salient parameters. Research is also needed to enable sensitivity analysis of the visualization surrogate models. The second direction is efficient feature representation. Advanced statistical and AI models could represent the spatiotemporal distributions of ensemble data in situ to reduce them for down-the-line interactive post hoc analysis, feature exploration, tracking, and visualization. The third direction is ensemble visualization infrastructure. For example, reinforcement learning (RL) can improve load balancing of data- and task-parallel visualization algorithms, such as particle tracing, density estimation, and feature tracking. Such algorithms can minimize resource utilization in the in situ analysis of ensemble simulations. Finally, sharing and making community-wide open ensemble data repositories will strongly benefit ensemble data visualization research.

3.1.5 Uncertainty Quantification

Visualization is a powerful tool not only because it can provide an intuitive understanding of information, but also because humans have evolved to believe what they see. However, scientific data is often inherently uncertain; thus, a single image at best represents one of many possible interpretations. This implied certainty can lead to misinterpretation and opens the door to deliberate bias and misleading conclusions. In general, understanding uncertainty is one of the most important scientific challenges of our time, and visualization techniques should be developed to help address it.

Key challenges

Scientific simulations and data are increasingly complex due to their scale, dimensionality, and modality. Associated uncertainty—inherent in all data and computational processes—makes analysis and decision-making increasingly challenging. Effectively conveying uncertainty to domain scientists is, therefore, an important challenge to enable trusted decision-making. Several aspects of decision-making under uncertainty require further research: (1) Although uncertainty visualization has been practiced in multiple scientific disciplines, a theoretical foundation of uncertainty visualization is in the early stages. The lack of theory and generalization prevents researchers from integrating uncertainty into standard large-scale visualization software, such as ParaView and VisIt. (2) Uncertainty exists at each stage in the visualization pipeline, from inaccuracies in measured and simulated data to uncertainties introduced by visualization algorithms: filtering, mapping, and rendering. Understanding the impact of uncertainty on individual algorithms and how uncertainty propagates through visualization algorithms is a significant research challenge and important for science-based decision making. (3) The evaluation of uncertainty visualizations is another significant challenge. How to visually represent uncertainty without impacting users' cognitive load is not fully understood. We still are not clear how humans perceive uncertainty, and the bridge between cognitive and visual analysis needs further research. (4) Uncertainty visualizations are often computationally expensive, thus limiting human capabilities to interactively explore uncertainties and make decisions.

State of the art

Research in uncertainty visualization dates back to the late 20th century, with a few studies emphasizing the importance of understanding errors in computational and visualization pipelines [56][57]. Since then, researchers have made advances in uncertainty visualization of scalar-, vector-, and tensor-field visualizations and their topology, which are summarized in a few literature reviews [58][59]. Uncertainty has also been analyzed in specific use-cases or disciplines, such as biomedical imaging [60]–[62], climatology [63], and oceanology [64]. A few researchers have evaluated perception of uncertainties through studies on human subjects [65][66]. Overviews of the current state of the art in uncertainty visualization can be found in recent survey papers [48]–[67].

Research directions

A large spectrum of complex scientific workflows and use cases would need to be addressed by researchers to establish the theory behind probabilistic and evaluation models of uncertainty. Access to scientific workflows, use cases via public repositories, and integration of uncertainty techniques into visualization tools such as ParaView and VisIt would foster growth in uncertainty visualization research and application. Quantification and tracking of uncertainties through the visualization pipeline requires developing new probabilistic models for each visualization stage corresponding to filtering, mapping, and rendering in addition to understanding the interaction among probabilistic models. The evaluation of visual representations of quantified uncertainty can be facilitated by collecting human feedback via user studies on the aspects of uncertainty, including perception, interactivity, and decision-making quality. Such evaluation of a wide range of use cases designed through a collaboration among visualization, cognitive science, and applied science researchers would drive research in novel visual mappings for effective representation of uncertainty. Lastly, developing AI or surrogate models combined with HPC could help automate and speed up uncertainty quantification and decision-making processes, thus enabling efficient analysis in the case of large-scale simulation data.

3.1.6 Interpretability of Complex Systems

An often overlooked visualization use case is not focusing on any individual outcome or piece of data, but rather trying to understand the behavior of a complex system as a whole. This might be a trained ML model, the interaction of various control algorithms, or the behavior of a quantum system. The goal is to provide insight into how inputs of the system affect the outputs, despite the inner workings being too involved or even fundamentally unknowable.

Key challenges

Complex systems including ensemble simulations, black box models (AI/ML models in particular), and quantum systems present computing and data capabilities that are core to ASCR's mission. However, for these capabilities to maximize their impact in high-consequence science and policy application scenarios, these technologies must address fundamental gaps in interpretability, characterization, and control of their outputs. Visualization can play an important role in addressing these gaps, and research advancing the theory of visualization for AI, quantum, ensemble, and other complex systems is required.

For classical complex systems (e.g., black box, ensemble output) and quantum systems, model interpretability within DOE-relevant scientific domains presents three unique challenges. First, the data—whether experimentally acquired or produced through numerical simulation—takes on a fundamentally different form. For instance, an AI model that aims to predict a quantity of interest (e.g., yield), given a sampling of a parameter space, presents fundamentally new challenges for visual exploration, where the method for visually conveying model prediction/explanation pairings can greatly impact an end user's exploration of the parameter space. Moreover, AI models that function as physics-based surrogates tend to produce large-scale field data as a computationally cheap alternative to running a full numerical simulation. Thus, all of the traditional visualization challenges faced with analyzing data of high spatial and temporal resolution, and multivariate in nature, are exacerbated by model-based replacements. More specifically, these models might be wrong, and how or where they make mistakes is unknown. Even when presented with an interpretation of the model's predictions, we are still left with challenges of how to integrate these interpretations with a visualization of the field itself. On the other hand, noisy physics experiments such as state-of-the-art quantum computing produce large, probabilistic, high-dimensional data during the era of noisy-intermediate-scale quantum [68]. Interpreting and analyzing the massive amount of noisy data is important in characterizing and controlling them effectively, optimizing their performance, verifying their computational output, and operating them in a stable manner. Meanwhile, underlying this stochastic, discrete-valued, high-dimensional data is a continuous-in-time, non-Markovian dynamical system.

The second key challenge stems from humans. Numerous stakeholders with various roles influence model interpretability. First, model builders with AI expertise require interpretability as support for model development (e.g., diagnosing why a model was wrong). Second, consumers of AI models wish to gain trust that a model performs well and for the right reasons. Third, laypersons (i.e., citizen scientists) can contribute to the construction of a model (e.g., through data annotation/gathering/curation) without requiring any background in AI. Though much research has focused on solutions in these individual categories, little attention has been paid to facilitating collaboration between stakeholders. The human-centered design of visual interfaces thus plays an essential role in establishing a medium for people of different backgrounds to express their expertise, abilities, and goals in collaboratively interpreting AI models. More details are discussed in the collaborative visualization in Section 3.5.2.

The third key challenge is ensuring that interpretability instills human trust in models. Any interpretation technique is subjective, making assumptions about the behavior of a model. Such assumptions could be about how a model makes predictions, the knowledge captured by a model, and the nature of an interpretation (e.g., whether features in the input are correlated with, or cause, output predictions). A lack of transparency about how interpretations are formed (e.g., treating interpretation as a black box) can inhibit trust. Methods for interacting with interpretations thus become critical for users to express their own working knowledge on model behavior and, in turn, reveal how this relates to a given model interpretation.

State of the art

Despite the successes of recent advances in AI and quantum computing, the black box nature or the intrinsic complexity make these systems largely uninterpretable or controllable. To increase transparency, one commonly used form of visual explanation for AI models involves saliency maps [69][70], which indicate in the form of a heatmap which input features are important to an AI model in generating its output results. Supporting these explanations, several open-source packages can help with the computation of saliency maps, such as Captum [71] and XAITK-saliency [72]. Others extract high-level concepts that are connected to the model's decision but more intuitive to understand by non-experts [73][74]. Additionally, efforts to build inherently interpretable algorithms include explainability as intrinsic to the model design, rather than as an afterthought and in a post-hoc manner [75][76]. Another branch of research stems from the visual analytics domain that typically builds an interactive visualization system for a specific network type, application, or quantum computer type in order to present and connect the system-related details and learned features [77][78]. While the capability to interact with neural network details is more advanced, for the quantum information system, in-house data visualization tools are mainly used to benchmark the computers while having limited exploratory functionality [79]–[82]. Typically, tools present static 2D representation and statistical summary plots of high-dimensional data to convey benchmark quantum program performance, information flow through quantum algorithms [83][84], circuit comparison [85], qubit states [86], and results of quantum computations [87].

Research directions

Several promising research directions are associated with interpretability and eXplainable AI (XAI), though they often apply to any black box system. These research directions can be grouped into the following thrusts: (1) domain-informed XAI, (2) collaborative XAI, and (3) actionable XAI. Given the domain-specific knowledge required to understand the data (e.g., training data, results) of a scientific experiment or simulation, new methods for domain-informed XAI may prove useful by providing domain-specific languages (e.g., Tempura, Polyjuice) or concepts for users to interactively specify and modify interpretations. To reduce the annotation burden required to incorporate these user-defined concepts, the development of new semi-supervised or weakly supervised interpretation methods will also be required. For more complex systems like quantum systems, continuous innovation on scalable visualization and exploration of quantum-related data forms and device performance assessment will accelerate the community's ability to extract insights that inform future technology designs. For example, methods to effectively analyze, visualize, and glean insight from distributions or functions of such distributions underlie many assessment and algorithmic tasks in quantum computing, including variational optimization, adaptive execution, and verification of correctness. As AI models and their explanations must interface with a wide range of users, collaborative forms of XAI are critical. These forms will allow for the catering of interpretations to specific groups of stakeholders with the appropriate level of abstraction, as well as for bridging interpretations between domains to facilitate communication and collaboration between users. Finally, actionable XAI addresses the question of what to do with a model interpretation. Generating model explanations should not be the end goal of interpretability, but should instead drive additional learning and discovery towards producing actionable scientific insights. For example, counterfactual explanations can suggest the potential actions to overturn the model decision. To this end, interactive and human-in-the-loop methods for manipulating model explanations in causal manners to test different hypotheses in the context of prior knowledge will be critical. PRD 5 includes more discussion for human related factors and interactive visualization.

3.2 Interoperable and Adaptable Visualization to Support Diverse Scientific Workflows across All Scales

The accessibility of visualization tools and relevant data varies across a wide spectrum of users. Users include researchers and domain scientists whose expertise ranges from casual to advanced, including the ability to contribute code as a developer. Visualization scientists also range in expertise from casual to advanced in their ability to use a broad set of tools. Both domain and visualization scientists might also integrate visualization tools into specific use cases.

Significant barriers exist to applying these tools effectively in a rapidly changing computing ecosystem. Distributed computing ecosystems are being developed that couple experimental, observational, and computational facilities. These systems are operated using complex workflows to control and orchestrate the processing and movement of data, and visualization is a critical component of these workflows for understanding the scientific questions being explored. Visualization tools must be easily deployed onto diverse workflows across multiple domains running on varying hardware. Users of all experience levels need to be able to easily configure, connect, and deploy visualization tools in different ways according to their particular needs. Solutions to these challenges include increased usability and accessibility of visualization software, support for a wider array of data types, and efficient use of modern technologies.

3.2.1 Data Interoperability

Data management, organization, and access are essential components for any medium that deploys visualization systems. An effective visualization process, in part, relies on having data that aligns effectively with the visual representation's requirements. For instance, a scientist performing a scientific experiment goes through many iterations of data acquisition, preprocessing, analysis, and insight across multiple data modalities and across different institutions (sometimes combining observations and simulations). Components in a workflow need to understand how to interpret the data they ingest as well as to ensure an understanding of the results they produce. Advances in compute and sensor technology, processing techniques (e.g., autonomous experiments), and incorporation of new modes of analysis (e.g., digital twins, ML) has resulted in a proliferation of data types, format, and tools. Unfortunately, this landscape has led to a proliferation of bespoke solutions that, while solving specific problems, are difficult to use outside their native contexts. One fundamental challenge in combining or generalizing these tools is that any connection of the bespoke data formats and descriptions requires significant effort and insight. To address this challenge, standardization has become not only necessary but can also be the difference between success and failure of a DOE project or effort. The description of data throughout the entire data lifecycle needs to be expressive, clear, and concise. Failure to do so impacts all layers of workflow orchestration—from resource allocation to meeting metrics for successful execution.

In addition, the complexity, scope, and power of scientific workflows will grow alongside increasing scope and size of scientific instruments. The challenges of these workflows are twofold. First, the workflows themselves are inherently complex and difficult to understand, and visualization is a natural way of representing the intention of a scientific workflow as a whole. The visual medium provides a window into a complex process and clearly informs the scientist about what information is known at every step, where the information exists, what form the data exists as, and most importantly, how effectively the scientific workflow is functioning across all scales of execution. Second, as experiment, observation, and simulations are coupled, the wide variety of data types and formats are a significant complication to performing analysis and visualization.

Key challenges

State-of-the-art methods are ill equipped to provide insight, keep up with the complexity, or operate at the scales and modes of workflow systems, which are becoming more common for managing the increasing complexity of scientific campaigns. When multiple facilities are coupled across a campaign (e.g., experimental and HPC), the complexity of data processing and visualization increases. Data types, models, and representations can vary significantly in such scenarios. While a simulation of a scientific process may be based on a finite element mesh, an experiment or observation of this same scientific process may produce a large quantity of 1D and 2D images, signals, or spectra. These data are often captured in different formats and may exist in completely different spaces (e.g., frequency versus spatial domains). The data from AI and machine learning are different yet again, and often stored in a number of ways from the columnar and tabular to array-based and hierarchical. Data preparation, training, and model serialization of ML pipelines have seen an explosive growth in ways of describing and storing content, and efforts are under way to consolidate to more open standards. Additionally, autonomous experiments and digital twins produce higher order elements and often incorporate or leverage multimodal data. Additionally, coupling simulation with experiment requires precise tuning and control to ensure that all aspects of the pipeline effectively transform data between the necessary visual and analytical mediums. Further, data values may contain uncertainty or be incomplete for many reasons, including sensor tolerances, faults, acquisition noise, simulation input parameters, solver type, grid type, and resolution.

These outputs may be final products, or they may be further processed by other tasks in the workflow. Visualization tools and methods must be able to rapidly adapt to the real time constraints of the data which includes quality of the data and quantity and respond with results that match the experimental conditions. Without expressive descriptions of the data, generalized connections among components in a workflow is difficult. A workflow that couples multiple sources of data, for example coupled simulations, and coupling between simulation and/or experiments and/or observations complicates the ability to visualize these combined results. Further, variations among scientific domains increase the challenge of providing interoperable, broadly useful solutions, which results in stove-piped solutions and missed opportunities for community building.

State of the art

As described by Mackinlay [87][88], "The expressiveness criteria determine whether a graphical language can express the desired information, while the effectiveness criteria determine whether a graphical language exploits the capabilities of the output medium and the human visual system." These two criteria aid in determining which visual representation best illuminates the data. Thus, capturing the varied forms of data enables visualization systems to best capture the search space.

Traditional grid-based data types (e.g., structured and unstructured) have been well defined in the visualization community (see VTK, the Visualization Toolkit [89]). The VTK project has spearheaded the development of standards to describe, share, and visualize in a common way. The project's visualization systems often take a plugin-based approach to incorporating future formats and types. Although they enable functionality at a tool level, these new data types and formats can lead to fragmentation because they are not core lexicons. Some of these new types of data may require new representation and standards. OpenPMD [90] provides a schema for describing mesh- and particle-based data. The eXtensible Data Model and Format (XDMF) [91] is a schema using XML for the standardized exchange of scientific data written in HDF5 between HPC codes and analysis tools. Fides [91][92] is a general-purpose schema for streaming and file-based data that supports the VTK-m [93] data model and uses ADIOS middleware [94] for access to streaming data. The Conduit [95] library provides a data model for scientific data using Blueprint [96] and is used for I/O, serialization, and code coupling. ONNX [97] provides an open format to represent ML models across a range of AI frameworks, runtimes, and compilers.

CF2 [98] provides metadata extensions in netCDF files for climate data, whereas Nexus is a common data exchange and archival format for neutron, X-ray, and muon experiments. Vega and Vega-Lite [99] provide a visualization grammar for interactive graphics. In current and next-generation pipelines, data increasingly comes in nontraditional forms. Systems that support autonomous, streaming, ML, and other data-driven approaches easily overwhelm the visual medium through complexity and volume. Novel techniques to sift through these enormous search spaces will be key.

For streaming and ML data, efforts such as ViSUS [100] and ParaView [101] data filters enable users to visually investigate ML data and model building. These platforms allow for creation and evaluation of ML pipelines.

Research directions

The quest for a single set of standards for scientific data has proved daunting. Even standards among similar types of applications are difficult. Research to aid in bridging these gaps include the following.

Broader support for data types. Effective representation for all variants of possible data types is likely not possible. However, broader support is needed with visualization tools to increase usability. Abstractions and motifs for classes of types can help broaden supported use cases, and should include descriptions of multivariate, multimodal, and high-dimensional data. Visualization of missing or incomplete data is another research direction to explore, as it will help answer questions about how comprehensively or precisely data-driven approaches cover the science space. These solutions can also play a key role when decisions are made without complete information, and having the ability to express how well a simulation or ML model fitting performs could be transformative. Data generated by simulations, experiments, observations, or AI-driven techniques will contain uncertainty that should be represented and used to guide decision-making processes.

Transformations and conversions. Multiple standards will likely be required within complex scientific workflows. Methods for the transformation or conversion from one data model to another provide a means of interoperability among visualizations. Methods for zero-copy transformations and conversions are ideal when possible. Conversions may require changes to the underlying data (e.g., resampling an unstructured grid onto a uniform grid for a visualization task that only supports uniform grids). In such cases, controls are needed to specify the acceptable error, data size, and other parameters. Methods for graceful failure are needed when a conversion between data types is not possible.

Support for higher order mesh and fields. Upcoming exascale systems and the increased use of accelerators such as GPUs have provided opportunities to rethink the solvers used by simulations. High-order numerical methods are ideally suited to take advantage of this changing computing landscape because they expose fine-grain parallelism and maximize the ratio of floating-point operations to energy-intensive data movement. High-order finite elements, in particular, have become a win–win proposition with respect to both simulation accuracy and HPC efficiency, and a growing number of large-scale simulation codes at DOE and in industry have now shifted to high-order finite element discretizations [102]–[105]. Accurate visualization of both the geometry and field data is critical in many DOE applications.

ML-based models. As AI and ML become more integrated into the visualization process, they must be represented and visualized by themselves as well as in conjunction with associated simulation, experimental, and observational data. Associations and relationships among the different types of data are important, and methods for bridging these gaps will require new research.

3.2.2 Software Interoperability

Current visualization software and approaches are strained by today's scientific community and ongoing changes to modern hardware and software infrastructure. The move towards computing ecosystems that couple HPC resources and experiments will require complex, distributed workflows and place significant new stresses on visualization software. Easily configuring and applying a visualization tool across a variety of scientific domains, use cases, and scales will be critical for success. Further, rapid technological developments in computing hardware, displays, and networking provide significant opportunities for innovation and improvements to scientific inquiry while posing a disruptive challenge. Recent trends in computer architectures have delivered a range of chips, both general purpose and specialized. Advances in networking provide increased connectivity among scientific instruments and scientists. Improvements in display technology and the development of additional modes of interaction provide unique opportunities for scientists to explore and collaborate.

Key challenges

Complexity in scientific campaigns is accelerating due to rapid technological advances, the distributed nature of computing ecosystems, and the large variety of data types involved. Visualization is a critical tool for helping scientists gain insight from this complexity. However, visualization scientists cannot create and maintain bespoke solutions for every scientific team. A shift towards a domain scientist–centric paradigm is called for. Specifically, instead of depending on visualization researchers, domain science users want to quickly adapt and innovate visualizations to solve particular problems. Examples include the coupling of experiments and simulations in novel ways or multidisciplinary investigations of a broad range of data. As a result, these users' workflows are becoming more diverse, project specific, and oftentimes even task specific.

The situation would be more efficient and sustainable if visualization research could pivot from a focus on full-featured visualization applications to serving as enablers of such scientist-centric customer workflows. A pressing research challenge is the absence of science users being able to easily configure and apply visualization tools across multiple scientific domains, use cases, and scales. Scientists want to mix and match various tools and analysis approaches. A common example is using data science environments such as Jupyter for simulation, analysis, and visualization, while integrating various libraries using Python. Such workflows are not compatible with monolithic tools that do not operate well within the rest of the scientific computing infrastructure. This need goes beyond traditional metrics (e.g., performance and flexibility) because the creation, deployment, and delivery mechanism of interactive visualization are also key to users' success.

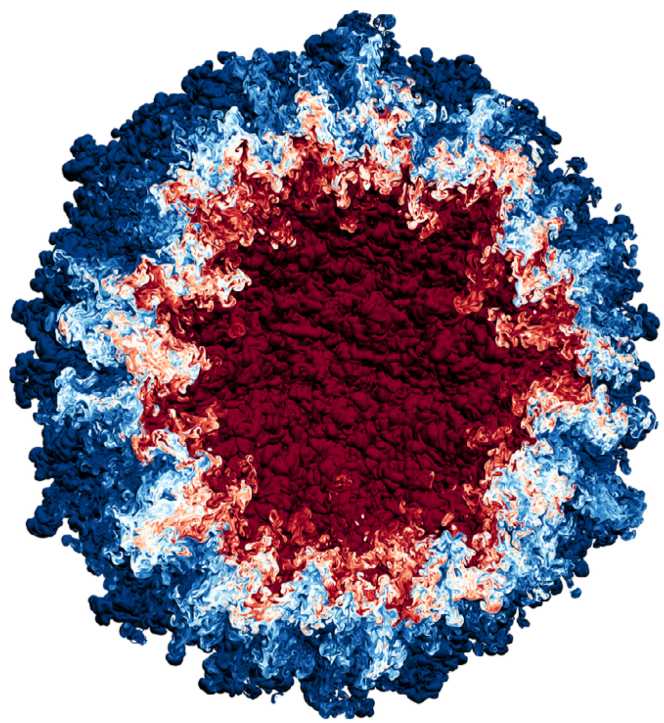


Figure 3.2.2a Visualization of an idealized inertial confinement fusion (ICF) simulation of Rayleigh–Taylor instability with two fluids mixing in a spherical geometry. Rendered with VisIt and data obtained with Ascent.

State of the art

Agility—especially the kind that empowers science domain users during self-driven processes to choose, adopt, experiment, instrument, and share sophisticated visualizations—has not been a focus of past visualization research. The closest equivalents are efforts to provide in situ visualization. In situ processing is a rich space comprising numerous variations [106][107], but techniques are often grouped into three broad categories: in-line (synchronous), in-transit (asynchronous), and hybrid. A major focus of in-line in situ visualization has been on instrumenting a simulation code so that the visualization can use the same resources to process data as it is produced. In-line in situ visualization is possible with Libsim [108] and Catalyst [109] and can be used to instrument a code for VisIt and ParaView, respectively. Ascent [110] and SENSEI [111] can be used to directly instrument simulation codes for in-line in situ visualization. (See Figure 3.2.2a.) Tools such as EPIC [112], Freeprocessing [113], and ICARUS [114] support an in-transit model where the data producer and visualization run on separate resources. Ascent and SENSEI also provide in-transit in situ using ADIOS [94].

The rapid growth of heterogeneous compute nodes, coupled with in situ processing, highlights the need for portability across different architectures. VTK-m [93] provides a portability layer for visualization algorithms. These efforts have demonstrated the benefits of portable algorithm performance across a wide variety of architectures [115]. Nevertheless, despite the goals of adaptability to any simulation code and interoperability, a common trait of existing in situ methods is the heavyweight process of adoption. Usually, not only software integration

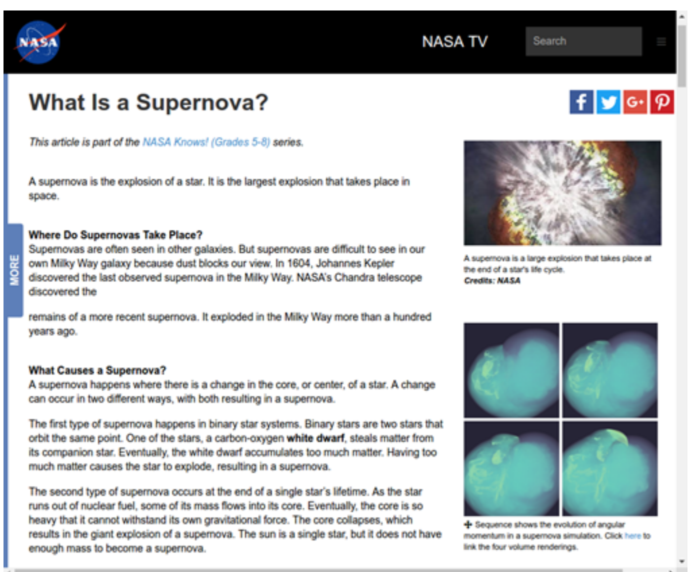


Figure 3.2.2b Interactive visualization can be embedded into a website, as in this example of a "NASA Knows" public educational website featuring four live 3D volume visualizations. Interactive exploration is supported on each of the images. The visualization service runs on a cloud resource and updates the webpage during user exploration [121].

efforts are needed, but the science and visualization teams must also be integrated in the co-design model under which user-centered agility is out of scope. Meanwhile, an industry-proven but new to DOE paradigm that has shown promise is service-oriented architecture (SOA) [116]. At a high level, SOA is characterized by a self-contained black box that provides a well-defined set of features for users. SOA takes several forms, including infrastructure as a service (IaaS) [117], software as a service (SaaS) [118], and microservices [119]. The "as a service (aaS)" paradigm has already been explored in the context of scientific visualization. A set of abstractions for using this paradigm for visualization is described in [120]. Tapestry is a system that can deliver interactive volume renderings of large-scale scientific simulation into the web browser on any device, including laptops, smart phones, and Microsoft Hololens [121]. (See Figure 3.2.2b.) Tapestry uses Amazon Web Services (AWS) where usage costs are very small. As another example, a group of small AWS instances can be organized into a swarm to provide interactive comparative visualization of terabyte-scale turbulent flows [122] from NOAA/NCEP (National Oceanic and Atmospheric Administration/National Centers for Environmental Prediction) models versus an actual observation data repository [123], where the total cost of the AWS instances is also very low. In both cases, the system can elastically scale to support 20–100 concurrent users. Currently, while the microservice model can deliver great interactivity into flexible user devices, the power and feature set of such services are limited in comparison to leading-edge in situ toolchains.

Research directions

The visual analytics and visualization workflows used by DOE scientists are increasingly diverse, presenting a challenge of developing reusable software across multiple domains. This current and future reality, paired with a limited number of visualization researchers and developers, requires software that is less monolithic and consists of modular components that can easily be connected by scientists to suit their needs. Research directions to empower such user-centered agility are considered below.

Categorize motifs in visualization and analysis workflows. Despite significant efforts, a recurring need is to discover and develop commonalities of visualization and analysis patterns across diverse domains. These patterns will provide the necessary abstractions with which to describe and ultimately develop modular visualization systems. Note that the goal is not to develop an all-encompassing standard—a task that has repeatedly failed over decades—but rather to understand which fundamentals are necessary and sufficient to service many application areas and facilitate software and algorithm reuse. Given conceptually complementary modules, connections can be implemented quickly to manifest an agile visualization ecosystem.

Empower distributed collaborations. Many cutting-edge research efforts involve collaborations across distributed facilities, a variety of tools, and diverse teams. A deeper understanding of how these collaborations operate and how both science users and automated analysis tools should best be coupled to reduce friction will significantly accelerate the time to insight. To be successful, visualization tools must be able to span multiple systems, workflows, and modalities. Bespoke solutions are too restrictive; solutions that can be "hot swapped" need to be available.

Develop scalable, responsive, and intelligent visualization systems. While the building blocks discussed above will ensure basic compatibility on a concept level (e.g., data filtering versus data rendering), individual modules must also become inherently more flexible; otherwise, instantiating connections will become too onerous and costly. A premium should be placed on modules that provide rich sets of accessible controls to adapt to new use cases and novel hardware potentially using intelligent feedback mechanisms (see Section 3.3.4) to support agile recombinations.

Expand visualization infrastructure to encompass new modes of interaction. As discussed in Section 3.3.4, a plethora of new display and interface hardware from VR to haptics is coming online. New approaches are needed to efficiently and effectively couple these new interaction paradigms with traditional visualization workflows, thus enabling rapid adoption of these technologies across the DOE complex.

3.2.3 Distributed and Streaming Visualization

Conventionally, visualization is performed post hoc: simulations, experiments, and observations generate and store data, then the data is loaded by the visualization tool for analysis. As discussed above, modern scientific workflows demand a more agile approach where data is accessed and visualized from wherever it was generated to increase the effectiveness of expensive experiments as well as circumvent bottlenecks arising from current storage limitations. This post hoc visualization use case is no longer sufficient. In addition, rapid technological improvements have produced a wide range of available computing hardware for data processing, analysis, and visualization (see Section 3.3.2). These resources can be distributed and include edge computing, supercomputing centers, local clusters, and laptops. Options for interaction and display are rapidly evolving and include VR and AR hardware, mobile devices, and notebook environments such as Jupyter.

Key challenges

Traditionally, visualization is performed post hoc on dedicated workstations or clusters using specialized software tools. Hardware accelerators (GPUs) are often employed to achieve the necessary compute performance because scientific visualization, especially at scale, involves the interactive rendering of numerous graphics primitives. The required hardware acceleration for high performance rendering is now available anywhere on systems ranging from mobile devices (e.g., tablets and phones) to supercomputers. This ubiquity enhances the expectation that visualization software will run everywhere.

Agility and flexibility are increasingly important requirements for visualization software. Scientific campaigns will be controlled by rigid workflow systems that require careful coordination of visualization tasks among the other data processing that occurs. Additionally, scientific campaigns are collaborative in nature and include scientists from multiple disciplines and located around the globe. A central challenge to meet these needs is access to data, which can take many forms: from disk, directly from a data producer (e.g., in situ), from a secondary resource (e.g., in transit), or from remote resources requiring network access (e.g., remote data center, experimental or observational facilities). Such diverse access patterns highlight the limitations of monolithic data visualization tools used in scientific workflows.

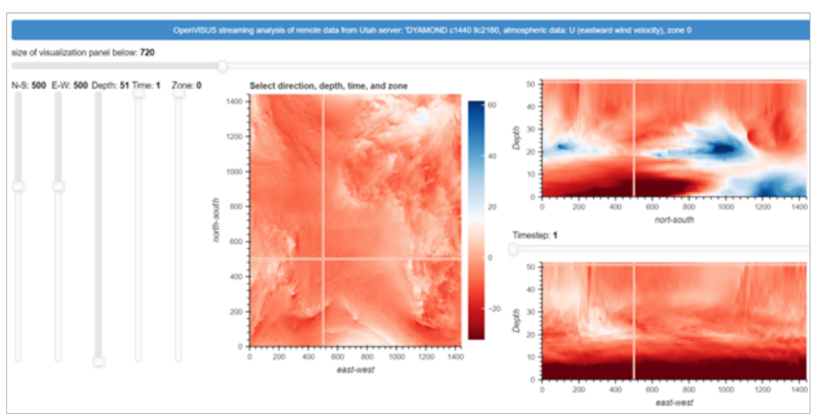


Figure 3.2.3 Interactive visualization of a 200-terabyte scalar field (part of a 4-petabyte climate modeling dataset from NASA) in a Jupyter Notebook streaming the data using the OpenViSUS [126] framework from a remote server on the NASA Advanced Supercomputing infrastructure.

Over the years, a vibrant and active research community has conducted important research and development of in situ [108][124] and streaming [125][126] visualization approaches. (See Figure 3.2.3.) To support current and future scientific needs, use of these techniques needs to become as easy and commonplace as post hoc visualization. The particular types of in situ visualization, as outlined in [107], are numerous and varied, and challenges remain. When paired with the different use cases needed by applications, these challenges only multiply.

State of the art

ParaView [101] and VisIt [127] applications are DOE's visualization workhorses. Development of both software tools began in 2000 to provide next-generation visualization capabilities for large-scale scientific data. Their capabilities include interactive visualization for both quantitative and qualitative analysis, distributed memory parallelism to support extreme-scale data, client/server execution, and state-of-the-art visualization algorithms. Over time, these projects were extended to support scripting through Python to integrate better with broader scientific workflows. These tools depend on VTK [89] for visualization algorithms and rendering. They have also integrated VTK-m, a visualization algorithm library that provides a portable performance implementation across multiple GPUs, and DIY [128], a package of scalable building blocks for data movement tailored to large-scale parallel analysis workloads. VTK, ParaView, and VisIt have been at the epicenter for DOE visualization research over the last two decades, with new algorithms and techniques delivered from numerous research projects. Together, these tools represent almost a human-century effort to provide robust visualization for the DOE community's essential needs.

The visualization community has also developed several domain-specific tools, such as VMD (visualization for molecular dynamics) [129], Tomviz (tomographic visualization of materials) [130], yt (astrophysics and cosmology visualization) [131], and Root (high energy physics visualization) [132]. These tools' modalities vary from interactive visualization (VMD, Tomviz) to scripting environments (yt, Root).

A number of in situ frameworks are also available. Libsim [108] and Catalyst [109] can be used to instrument a code so that VisIt and ParaView, respectively, can be used for synchronous in situ visualization. SENSEI [111] and Ascent [110] use code instrumentation to provide both in situ and in transit visualization. ADIOS [94] uses an abstraction of the I/O layer to provide data access across post-hoc, in situ, and in-transit use cases.

In addition to these specialized scientific visualization tools, the DOE research community leverages some visualization and data analytics libraries, especially those with Python and Jupyter based workflows, including Matplotlib [133], Plotly [134], and Bokeh [135]. These packages enable rich DAV capabilities in conjunction with 3D visualization libraries such as VTK.

The tools described previously are desktop tools and applications. So far, few production-ready visualization frameworks or tools delivered through a Web interface exist. One good example is Trame [136], a framework targeted at developing scientific visualization applications for the Web using only Python. Coupled with vtkWeb [137], which enables the use of VTK, ParaViewWeb [138] (i.e., ParaView over the Web), and vtk.js [139], which provides the rendering subset of VTK written in JavaScript, as well as several state-of-the-art libraries (e.g., Plotly, Matplotlib), Trame takes the first steps towards production visualization on the Web.

Research directions

One key finding from the recent workshop report on in situ data management [140] is the imperative for pervasive in situ processing. To be most effective, visualization must seamlessly fit into the pervasive in situ processing environments of the future. Current solutions that require instrumentation of both the data producer and consumer result in rigid interfaces that would require widespread adoption to become truly pervasive. Such an eventuality is unlikely given the diversity of scientific domains, software stacks, and visualization tools. Research is needed to address these limitations and provide new approaches to in situ visualization.

The distributed nature of future scientific computing ecosystems will result in additional modes of data access. Each of these modes needs to be as available and transparent to visualization tools as post hoc methods currently are. Further, given the diversity of use cases, the access patterns and scale of data will vary. Each domain and scientific team will have a particular type of access, which may dynamically change over the course of a campaign. Research to address these challenges include the following.

Flexibility of visualization across diverse environments, use cases, and scientific domains. For visualization to become more ubiquitous throughout diverse scientific workflows, the latter need to become nimble enough to be placed in arbitrary configurations. The visual analytics and visualization workflows used by DOE scientists are increasingly diverse, which makes developing reusable software across multiple domains a challenge. Thus, the primary research direction in usable and accessible visualization techniques is to establish the commonalities in data access across the domains of these analysis and visualization workflows. The patterns that emerge can serve as a guide for the discovery of new methodologies for data access that can span multiple application use cases, and sources of data. An additional research direction is investigating reusable software development across visualization use cases (i.e., post hoc, in situ, and in-transit) for simulation, experimental, and observational workflows. The needs, constraints, and use cases will significantly vary across different application teams. Understanding the types of optimizations and finding flexible solutions for different combinations of application and visualization use cases will increase the usability of visualization tools. Finally, these software frameworks are anticipated to need complete integration with ever-evolving AI frameworks.

Identify novel ways for collaborative visualization across distributed, streaming environments. Scientific campaigns are generally collaborative efforts across a range of disciplines, including domain, computer, and data sciences. Collaboration among scientists is crucial to the understanding and steering of observational, experimental, and simulation efforts. When coupling these (e.g., simulation and experiment), collaborative visualization is imperative to effective scientific discovery. New techniques, tools, and frameworks for enabling collaborative visualization across distributed resources are needed so that scientists can understand the enormous amounts and speeds of data that will be generated.

Identify novel ways to perform visualization across diverse sets of resources and interactive environments. Mapping a set of visualization tasks onto a set of resources (including distributed resources) is a challenging problem that produces tradeoffs in time, latency, resource requirements, and quality of visualizations, so understanding these tradeoffs and identifying novel ways to meet user requirements will be crucial. Another important research direction is to explore using visualization techniques in multiple computational environments (e.g., desktop, Web, mobile, eXtended reality [XR]) including desktop environments and scripting environments such as Jupyter/JupyterLab. Frameworks compatible with various environments increase accessibility to a diverse audience, including domain scientists and visualization researchers.

3.3 Harnessing Technology Innovations to Accelerate Science through Visualization

To stay current, visualization research must consider and take advantage of technology innovations in other research areas. This includes novel HPC hardware, which has always been a driver of visualization research at DOE, as well emerging computing modalities such as cloud, quantum, or edge computing. Because energy-efficient HPC continues to require increasingly heterogeneous compute environments, efficient DAV on these systems will depend on the ability of future visualization approaches to take advantage of these technologies. On the other side of the spectrum, new hardware for human–computer interaction (HCI) is providing new modalities for visualization, such as consumer VR and AR headsets that present new opportunities as well as significant open challenges. Finally, ML is enabling some disruptive advances (e.g., in dealing models rather than data).

3.3.1 Novel Computing Hardware

Due to a variety of market forces, HPC and computing in general are undergoing a revolution in computing hardware and methodologies. Trends include increasingly specialized computing hardware such as GPUs, ML/AI processors, field-programmable gate arrays (FPGAs), and mobile computing. These new technologies continue to be disruptive forces bringing numerous challenges and opportunities that cannot be ignored if we expect to support discovery in scientific data.

Key challenges

Recent developments in HPC systems have resulted in disruptive changes in computer hardware requiring, in turn, dramatic changes in large-scale parallel codes. Most notably, GPUs have become ubiquitous, both in HPC environments as well as in user's laptops/workstations and at the edge. Trends suggest that we will continue to see regular paradigm shifts as we near the end of Moore's Law [141], and several new hardware technologies have the potential to be such disruptive forces. ML/AI-specific hardware, such as tensor cores (TPUs), half-precision, and neuromorphic chips, is becoming more widespread. Early research has shown this hardware can be applied in other computing scenarios to great effect [142], with potential applications for visualization and analysis. FPGAs may also be on the horizon as features in HPC/cloud computing nodes and already have many applications at the "edge" in sensors and remote instruments.

Ensuring that DOE's visualization capabilities continue to perform through computing paradigm shifts is challenging yet vital. Hardware is a continuously changing landscape for which we must constantly adapt and evolve our visualization algorithms and software. Visualization researchers must reimagine the way DOE scientists work with and experience their data through interactive, collaborative, and intelligent interfaces in concert with advances in diverse DOE-relevant computing modalities, which can include next-generation HPC, edge computing, neuromorphic architectures, novel hardware, quantum computing, and others. Many of these computing modalities originate as target hardware for other applications (e.g., ML/AI), giving rise to the challenges of identifying the best mix of technologies for a given DAV pipeline and mapping the appropriate algorithms to the appropriate hardware.

State of the art

Chris Johnson prophetically identified "efficiently utilizing novel hardware architectures" as a key challenge for visualization research [4]. Indeed, one of the largest software challenges introduced by exascale computing for visualization, as well as other scientific software, is accommodating numerous programming models with each vendor providing their own preferred programming language.

Several software APIs that act as a porting layer between these devices have been created including OpenACC [143], OpenMP with offloading, Thrust [144], SYCL, and Kokkos [145]. These APIs provide abstractions that simplify porting, but leveraging them effectively can still be a challenge. Most of the recent research in utilizing new compute hardware for visualization has leveraged one or more of these porting layers [146]. To date, the most complete visualization library for accelerator processors is VTK-m [93], which can adapt to several of these porting layers for more complete device coverage and contains its own abstraction layer to simplify the development of visualization algorithms [115]. Additionally, several hardware vendors provide hardware-optimized libraries for rendering, an important subset of scientific visualization [147][148]. For example, these improvements have increased the use of interactive ray tracing—once considered too computationally expensive for real-time rendering use—into visualization platforms such as ParaView [101], VisIt [149], VMD [129], Vapor [150], and others.

Note that the majority of research in scientific visualization has focused on leveraging CPUs, GPUs, and related processors. Very little work has addressed FPGAs or other more "exotic" hardware such as neuromorphic chips. Although some porting layers like SYCL have the potential to compile code to processors like FPGAs, the potential is currently unexplored.

Research directions

As computing hardware continues to evolve, diversify, and specialize, DOE researchers must keep an eye on the horizon of new technology. By the time a computational technology becomes widely available, DOE science customers will already be dependent on it, unable to wait for analysis and visualization solutions to catch up. Thus, the DOE visualization community must properly research the use of these disruptive technologies in advance and be ready if and when these processors take hold. The following lines of research should be the most helpful.

Identify novel ways to leverage emerging and existing technologies. Each new form of compute hardware comes with its own benefits, detriments, and idiosyncrasies. Researchers are tasked with learning how to exploit the benefits while avoiding the detriments. Thus, research helps us understand the limits of hardware programming models and helps us push those limits to grow our visualization capabilities. Research should proactively focus on hardware (e.g., TPUs, FPGAs) that may not be widely used in DOE now but could become predominant as HPC and edge systems evolve. More forward-thinking research is necessary for developing technologies that may become viable in the future, such as quantum or neuromorphic computing.

Identify common motifs for DAV along with abstractions for algorithms that encompass a wide range of hardware. A motif is a computational pattern common among a group of software problems, algorithms, or applications. Motifs are useful in examining hardware in the context of the motif so that a discovered solution may be applied to numerous algorithms, rather than rediscovering the same solution multiple times for different algorithms. Motifs can also be used to guide the abstractions created for development on new hardware. Multiple branches of parallel computing have developed useful motifs [151][152]. Taxonomies of visualizations have been proposed [153][154], but they need to address the algorithmic challenges of implementation and, ideally, match with the motifs of other computational domains. From a set of motifs, algorithmic abstractions offer the means for designing platform-portable implementations. VTK-m is an implementation that provides abstractions allowing the efficient design of visualization algorithms across DOE's most prevalent compute hardware [115]. Although effective, these abstractions are insufficient to address visualization needs in upcoming compute hardware. Making broad abstractions that can be applied from FPGA to neuromorphic computing will be challenging and may require completely new approaches.

Provide a unified infrastructure and ecosystem to manage interactive visualization across varied software and hardware. The demand for interactive visualization for remote experiments and simulations is increasing, while the diversity of the hardware infrastructure grows with HPC, edge, cloud, XR, visualization walls, and other features. The goals of visualization overlap greatly with ASCR's research in heterogeneity [155], which focuses mainly on adaptive workflows, and that of edge computing, discussed later in this report. An additional challenge is the requirement to support analysis and visualization in all parts of the ecosystem (see Section 3.2.3). Analysis and visualization do not have the luxury of choosing the preferred hardware and location, and transferring voluminous data among components is often impractical. Visualization and analysis software implementations must therefore be flexible enough to move computation to data, which could reside on a variety of different node types.

3.3.2 Nontraditional Computing

In addition to innovations to compute hardware, mostly in HPC environments, nontraditional computing approaches (e.g., edge and mobile computing, cloud computing, quantum computing for DAV) are becoming more prevalent and thus affect visualization research. This includes edge and mobile computing, cloud computing, and quantum computing for DAV. In particular, for research involving remote facilities and environmental sensing, edge computing will become increasingly important for processing data. However, the term *edge* is quite broad and encompasses sensors and scientific user facilities, networks, and wearable devices. Similarly, mobile devices offer access to data in the field and anywhere in the world and require new visualization approaches. Cloud computing offers resources on demand that in the past have been limited to HPC centers. Finally, DOE is making significant investments in quantum computing due to its potential to revolutionize computing in general. Early research in using quantum computing for visualization and data analysis is necessary to avoid missing future opportunities.

Key challenges

A key challenge driving edge, mobile, and cloud computing is the fact that DOE facilities produce experimental and observational data (EOD) at increasing rates that overwhelm on-site analysis capacity. Thus, there is a need to move analysis to HPC resources or the cloud. Data rates from a large number of experimental facilities and sensors will exceed available network bandwidth, even if 5G networking provides additional capacity, and the growing network bandwidth for data transfer will pose additional pressure on available HPC resources. Data reduction at the edge is one possible approach, but needs to be trusted by scientists [156]. Edge and cloud computing resources consist of heterogeneous hardware and are highly distributed with the possibility of in-network compute resources [157]–[159]. Distributing visualization and analysis computation between edge, network, and HPC center/cloud and developing appropriate compressed data representations suitable for distributed processing remain key challenges. At each location in the data processing pipeline, computing resources are highly heterogeneous (e.g., microcontroller, CPU, GPU, FPGA, custom hardware) and may need to satisfy additional constraints (e.g., cost, power consumption, size, suitable for extreme conditions).

Furthermore, mobile devices are becoming increasingly powerful, having a compute performance that once was only available via supercomputers. Combined with high-resolution displays and intuitive touch interfaces, they provide new opportunities to interact with data anywhere, which is particularly useful in the field. However, visualization needs to consider additional requirements, such as smaller display sizes and real-time data processing. Finally, edge computing systems operate on data in situ—allowing applications to reduce latency, decrease bandwidth, preserve privacy (i.e., process data and discard), and improve resilience (e.g., distributed processing, analysis, and control). Future visualization and analysis tools must support workflows that combine edge-based and centralized-resource analyses, facilitating efficient streaming algorithms, data compression, data decimation, and data transfer among edge resources and between the edge and centralized resources (see Section 3.2). A need also exists to expose meta-analyses of the edge resources themselves to study resource placement, resource efficiency, and related optimization problems.

Another opportunity for visualization to take advantage of the changing computing landscape is through the emergence of ubiquitous cloud resources. DOE science projects are becoming more collaborative, complex, and agile, resulting in a pressing need to make scientific tool sets more flexible, scalable, and available to meet the increasingly diverse and on-demand user tasks. The requirements go beyond scaling and capability, which have been the tenets of mainstream scientific visualization platforms. Recent progress in cloud computing could support this need for faster, more elastic, and more flexible processing to decrease the time to discovery. However, this approach will require significant adaptations of existing visualization infrastructure.

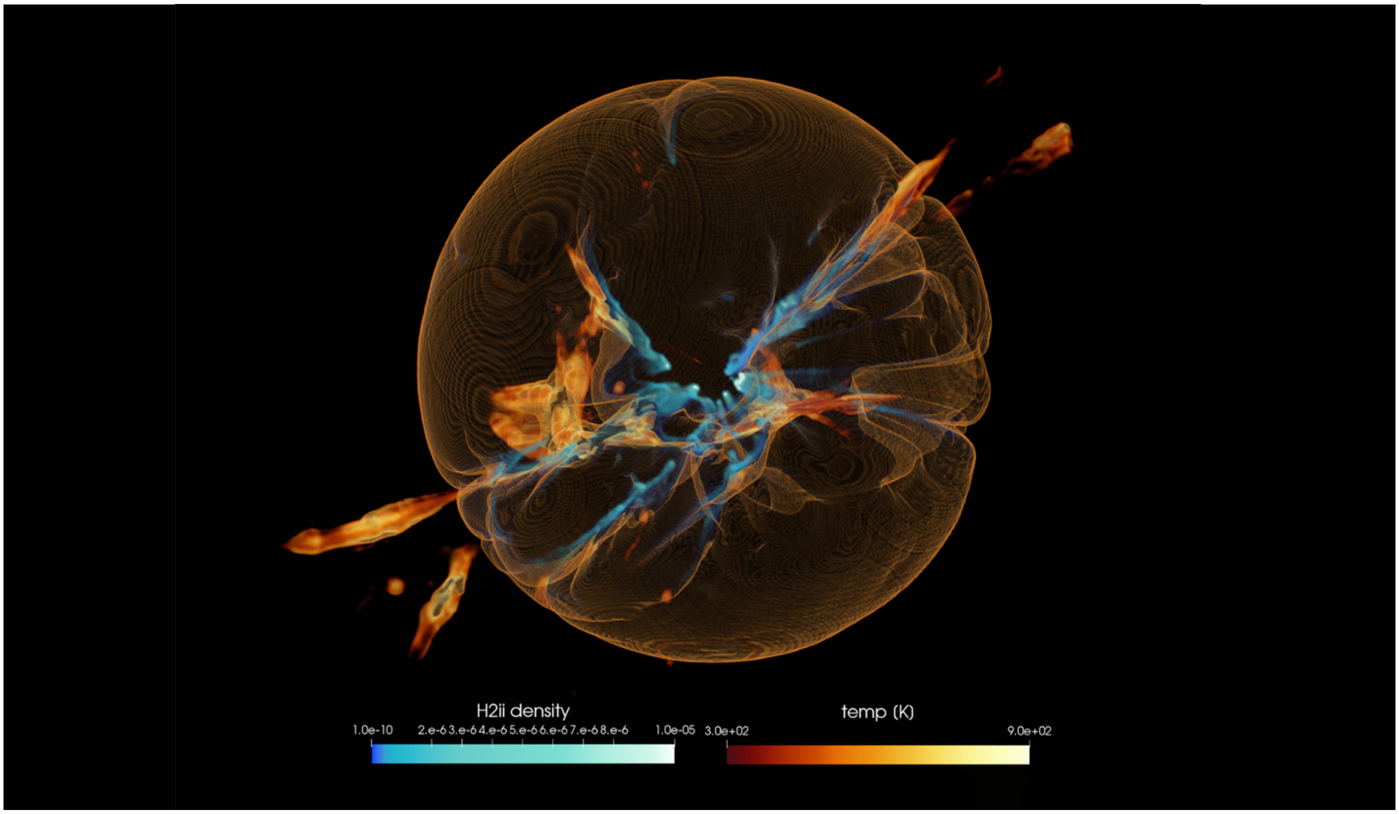


Figure 3.3.2 3D volume renderings of temperature in the wake of the supernova shockwave reveal this seething, turbulent environment in unprecedented detail.

Finally, quantum computing also has the potential to significantly impact visualization research. Classical computing encodes information as bits, which can take on a value of either 0 or 1. The qubit is quantum computing's analog to a classical computing bit. A qubit can take on a value of 0 or 1, or a linear combination of values between 0 and 1, due to the quantum physics effect of superposition. This fundamental shift in the compute paradigm will have profound impacts across the hardware-software stack, from how we write and control our software to the core algorithms that support quantum-enabled visualization and science.

State of the art

The traditional workflow for scientists gathering and analyzing data at DOE facilities is to temporarily cache the data at the instrument, then transfer the data across, for example, ESNet to computing centers for analysis, processing, and visualization. However, edge computing shifts this paradigm. Today's scientific endeavor no longer considers the instrument as simply a data source, but rather as an integral part of the new digital continuum. Analysis, processing, visualization, and insight are needed along the entire spectrum of resources from the instrument to the supercomputer. DOE facilities are rapidly embracing this new paradigm and adding edge computing resources that are as diverse as distributed sensors to beamlines at DOE laboratories. Five applications that link scientific instruments and HPC facilities (including Argonne's Advanced Photon Source, the Argonne Leadership Computing Facility, and Stanford's SLAC National Accelerator Laboratory) were studied to understand the patterns and technologies [160]. Researchers found that the experiments varied "significantly in their data rates, flow, and action runtimes, use of heterogeneous resources, and geographically distributed execution." Understanding the patterns they exposed, from online data processing to ML training, will be useful for creating reusable forms and developing solutions that can be shared across domains. Scientific software stacks are evolving; to date, only a handful of prototypes exist to support running AI computation at the edge and to analyze or control the instrument directly. Fewer still analysis and visualization frameworks exist for exploring data being processed at the edge. (See Figure 3.3.2.)

Cloud computing performs parallel computing in an elastic manner. Especially in the consumer Internet world, cloud computing methods (e.g., MapReduce [161] by Google) have achieved success that was previously unfathomable for on-demand use cases. Even in 2004, MapReduce was reported to process terabytes of data daily. Today, petascale cloud computing is routine for Internet companies. Elasticity in cloud computing describes the ability of a system to dynamically adapt to workload changes by provisioning and deprovisioning computing resources [162]. Using elastic parallelism to perform scientific visualization in the cloud is relatively recent. Example systems are Tapestry [163] for time-varying volume rendering and VCI [164] for turbulent flow visualization; when deployed on AWS, such cloud-based visualizations are freely and on-demand available to any users with typical consumer-grade Internet access. Cloud computing also helps to reduce the resource needed in the front-end client, thereby enabling thin clients such as smartphones to access interactive visualizations from a visualization service, also known as a VaaS (visualization as a service). While these examples of public-cloud VaaS deal with datasets in terabytes, research for petascale (and beyond) datasets of petascale has also started in the HPC realm to create visualization services, whether as independent services [120] or in relation to broader data services like Mochi [165][166].

While quantum computing capabilities are still in their infancy, preliminary research is already under way to understand how this revolutionary computational mode can be applied to visual analysis and image processing. For example, Santos et al. have achieved a hybrid classical-quantum Monte Carlo path tracer with the first-known quantum-generated images [167] and an independently proposed quantum-based algorithm for the full ray tracing pipeline [168]. Current quantum devices are extremely limited in their ability to perform computations on arbitrary data, as the data must be transferred to the quantum device and any modifications obtained from the superposition. Santos et al. use the quantum device to perform hemispheric sampling for simulated radiative transfer, leveraging the quantum superposition to calculate multiple sampling results simultaneously. However, this computation mode has unacceptable data transfer costs for nontrivial datasets. Amankwah et al. [169] describe an approach for efficient encoding of pixel data for N-dimensional images, which provides the foundation for new families of quantum methods for image analysis, vision, and ML that work with pixel-based data. Their approach provides optimal circuit implementations that are amenable to compression and that overcome high gate counts associated with previous approaches. Furthermore, recent algorithmic efforts are starting to look at the intersection of ML and quantum computing by attempting to formulate quantum algorithms that perform ML with significant speedup (e.g., polynomial, exponential) over their classical counterparts [170][171]. TensorFlow Quantum [172] is a quantum machine library for rapid prototyping of hybrid quantum–classical ML models with a focus on quantum data.

Research directions

Visualization and analysis algorithms for heterogeneous, distributed compute resources. Novel hardware solutions (e.g., low-power single-board computers with GPUs, FPGAs) provide significant processing power at a low cost while requiring little energy and space. Industry trends in robotics and autonomous driving continue to drive down their costs while improving performance. Compute power is becoming available in networks [173][174]. Using this novel, low-cost hardware presents a unique opportunity to process, filter, and compress data directly where it is produced and avoid wasting network bandwidth as well as computational resources. However, utilizing this hardware effectively requires further research in implementing DAV algorithms on highly heterogeneous architectures (e.g., through the use of abstractions and software frameworks). This research area significantly overlaps with the novel computing hardware discussion of Section 3.3.1 and software for distributed and streaming visualization of Section 3.2.2, but includes research on (1) satisfying power and resource constraints for analysis on site; (2) identifying appropriate data reduction methods [156] and appropriate data formats for distributing computation across edge, network, and HPC center/cloud; (3) leveraging opportunities for using edge compute power to provide real-time feedback during experiments as well as distributed compute power for digital twins at HPC centers; (4) leveraging opportunities of having compute power in the network; (5) taking into account the streaming nature of EOD and developing methods for visualization and analysis of data while it is arriving (e.g., incomplete data, changing data, streaming data); and (6) addressing key challenges and unanswered questions pertaining to use of quantum platforms for visual data analysis and exploration, such as quantum data encoding, effective use of quantum hardware for key algorithmic motifs, and increasing understanding of how quantum platforms can provide an advantage over classical computing in the area of visual data analysis and exploration.

Flexible distribution paradigms. As today's simulations have entered the exascale era, visualization pipelines, whether in situ or post hoc, are becoming very sophisticated. Many existing capabilities have been built on top of traditional programming models such as MPI and OpenMP. These models do not directly translate to elastic parallelism, which is the foundational assumption of cloud computing. Hence, even though early evidence shows that cloud computing can be beneficial to scientific visualization, significant needs have arisen for new research to truly understand how cloud computing can impact user needs at the level required by DOE's visualization tool chain. The following lines of research should be the most helpful: (1) understand and develop ways to leverage elastic parallelism for parallel visualization in general; (2) develop methods for elastic visualization algorithms to effectively interact with diverse data producers; (3) develop methods to forecast, provision, and distribute workloads of parallel visualization so that cloud computing resources can be elastically leveraged; and (4) identify, understand, and enable use cases showing how DOE-scale scientific visualization can be used by teams of scientists in an on-demand fashion.

Visual data analysis and exploration at the edge. One goal of data analysis on an edge platform is to reduce bandwidth, transmitting only the essential observations, compressed data, or the result of data processing (along with some assessment of the uncertainty in the edge data). Usually all data cannot be sent; thus, to ensure important science is not lost, scientists need to visualize and monitor activity at the edge. Visualization tools for monitoring activity at the edge must be developed, and one challenge is how to decide which data to send to the scientist and how to present it. Scientists may want to visualize what led up to an event, so the "process and discard" nature of edge platforms must be adapted to enable this. Detecting and highlighting anomalies [175] can provide insight for scientists for unexpected edge activity. As pointed out in DOE's Advanced Scientific Computing Advisory Committee (ASCAC) subcommittee report on AI [176], pre-programmed triggers are expected to be replaced with "algorithms that can learn and adapt, as well as discover unforeseen or rare, rate-limiting events that would otherwise be lost in compression." Research is required in algorithms that can perform this learning and adaptation. In addition to monitoring the *science* at the edge, the cyberinfrastructure itself must be monitored. Further research is also necessary to develop methods for monitoring complex workflows consisting of compute capabilities in many places (i.e., adaptable visualization to support distributed, diverse scientific workflows of diverse domains). Monitoring distributed workflows includes network visualization for controlling data transfer and dashboards for system health overview.

Exploration at the edge will benefit from advances in hardware (see Section 3.3.1). Mobile devices already provide effective means to interact with data at the edge. Additional opportunities are offered by mixed reality (MR) technology (see Section 3.3.3). Future research is necessary in how to use powerful mobile devices and new MR technology to enable the interaction with data at the edge. This work needs to identify the right MR visualization metaphors/design patterns to make best use of human visual intuition for scientific decision-making. Appropriate visualizations and dashboards can help in controlling experiments, influencing sensor placement, and improving data collection. Furthermore, new ways are necessary to couple experiments and sensors to simulations that are digital twins as well as visualize the output for informing experimenters during experiments.

3.3.3 Hardware for Human–Computer Interaction

While novel computing hardware within DOE has traditionally meant large computing systems that produce data, the scientific information contained within this data is of most interest. Therefore, new hardware that allows scientists to more easily, more intuitively, or in general more effectively explore and understand their data and what it contains is becoming increasingly valuable. In particular, display and interaction modes like VR and haptics may enable innovative ways to experience complex data with a significant potential for new discoveries. (See Figure 3.3.3a.) However, fully exploiting these opportunities requires new tools and techniques in the area of HCI, especially in the context of large-scale science.

Key challenges

As we look towards the future of hardware that facilitates human interaction and understanding, we envision a new era of scientific discoveries where visualizations enable DOE science users to interactively and collaboratively explore their data through novel display and interaction technologies—from commodity head-mounted displays to large-scale immersive environments, from desktop and touch displays to high-resolution display walls, and from haptic interfaces to audio and olfactory devices. The term *novel display and interaction technologies* includes, on the display side, the superset of XR, the superset of AR, MR, VR, and anything in between, high-resolution tiled displays, and tabletop computers. On the interaction side, our definition includes outputs (e.g., display, audio, haptic, olfactory devices) as well as inputs ranging from touch and speech to hand gestures and even head- or eye-tracking devices.

New technologies introduce new HCI challenges, as we rethink and reevaluate how we communicate with the computer and vice versa. The traditional mouse paradigm works well with commodity computers; however, it does not translate well when using novel *display* technologies. In many of these mediums, the DOE scientist is more actively engaged with their data, as they are usually immersed and walking around it. Additionally, novel *interaction* technologies would allow the user to experience their data through other sensory channels and control the simulation using their thoughts, voice, or muscle movements. In other words, the scientific challenge is to devise natural interaction paradigms, allowing scientists to explore, analyze, and make scientific discoveries in novel display technologies.

Like the components of supercomputers and edge devices of today's DOE user community, visualization teams today rely on commodity hardware built for communities with differing interests and goals. DOE users are often concerned with enormous datasets, while many commercial XR (AR, VR, and MR) technologies are resource constrained. The realism needed for their intended audience is faked through texture mapping and similar technologies that are not reasonable solutions for science. The challenge is leveraging the rapidly evolving advanced display space for scientific use cases.

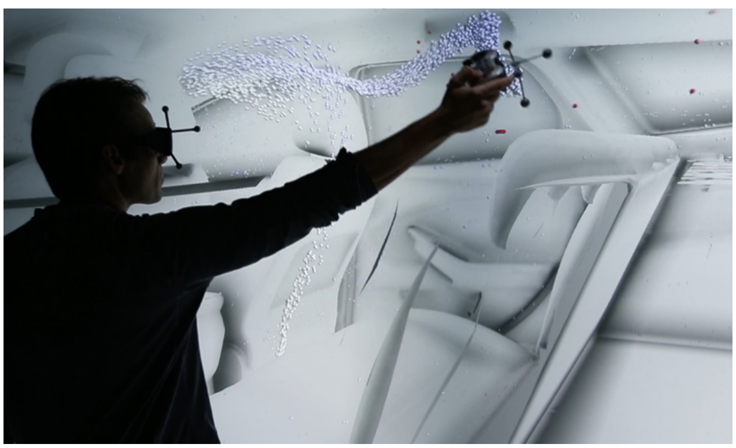


Figure 3.3.3a An engineer interactively exploring complex airflow inside an electric vehicle cabin by directly seeding particles in the flow. Credit: John De La Rosa, National Renewable Energy Laboratory.

Scientists today require more interactive, collaborative tools to analyze their large DOE datasets. Novel display and interaction technologies would allow them to explore their data in unprecedented ways, but a seamless integration of these technologies with DOE scientists' workflow is missing. The technology and tools must allow for a way of accessing and manipulating the data simultaneously while also allowing for feedback from those physically close but not currently immersed in the same environment, as well as potentially those colleagues who are part of the same VR connecting from remote locations. These new paradigms will lead to digital twins of the real experiments, and the implications must be thoroughly researched.

State of the art

Within the context of DOE science, research teams have begun investigating novel display technologies to enhance users' experiences and provide more insight. Raybourn et al. explore the challenges and opportunities for interaction and data display within XR environments, particularly the support for multiple viewpoints and perceptions in multiple spaces [177]. One approach to providing a rich rendering of complex environments is using 360-degree surround-view panoramic images. Marrinan et al. have developed a technique to reduce image artifacts and have begun looking at ways to add interaction [178]. Reipschlager et al. have begun to look at the combination of novel displays, combining AR-based glasses with large-format displays to facilitate data exploration and analysis [179]. Other teams have been researching novel displays to aid analysis and sense-making for over a decade [180], though with little focus on science specific to DOE's Office of Science. State-of-the-art tools are now available from vendors such as Intel's Embree [181] and Nvidia's OptiX [147] and Omniverse, along with tools that DOE has already invested in, such as ParaView and in situ libraries like SENSEI [111] and Alpine [110]. These tools can be coupled with commodity gaming engines (e.g., Unity and Unreal) to simplify and enhance development efforts. Additional understanding is needed about how these tools, when combined with novel displays, can enhance scientific output. (See Figure 3.3.3b.)

Research directions

As discussed above, empowering scientists to interact with their data in fundamentally new ways has the potential to significantly change how we analyze data and how we form and communicate new hypotheses. However, few of the established interaction and visualization paradigms directly translate to new modes of HCI. significant research is needed to bridge this gap and fully realize the potential of novel technology to impact large-scale science.

Rendering large data on novel display technologies. Realistic visualizations of large data on very large displays, domes, VR, and other systems impose significant new constraints and requirements on rendering algorithms substantially differently from the methods used for desktop-based systems. Whether these requirements are due to high pixel counts, the need for constant high frame rates, or low tolerance of imaging artifacts, rendering very large data on such displays will require a new class of techniques that more flexibly and elastically handle multiresolution representations of massive data.

Developing novel user experience components. Novel display technologies do not have the same affordances as a computer monitor, similar to the way that a desktop graphical user interface (GUI) does not translate well on a smartphone, or the way that traditional GUIs will not work well on an XR device or tiled display wall. Likewise, interaction paradigms will differ as the mouse becomes obsolete. Consequently, substantial work in user experience (UX) is needed to identify how users operate these new devices, how we can make them more usable and user-friendly, and how device UX will make users' jobs easier. As solutions are developed, corresponding research must strive to understand the effectiveness of the solutions' impact on DOE's mission science (see Section 3.5.3).

Enabling remote and co-located collaborations. Breakthrough science brings world-renowned experts and unique resources together to solve big, difficult problems. While novel display technologies will enable scientists to immerse themselves in their data literally, these tools currently lack methods for effective collaboration in such an environment. Remote and co-located collaboration infrastructure is needed that will allow two or more scientists to connect, look at, and discuss their data—this includes tools to annotate, filter, explore, analyze, and share findings and data. Furthermore, with continued investment in user facilities across the nation and the ever increasing complexity of these devices, a need exists to interact with experimental science in real-time. Research is also needed to enable changes to be instantly applied and the results visualized, allowing scientists to "teleport" themselves into the micro- or macrocosm of their data independently of where it is being generated.

3.3.4 Machine Learning

As with many scientific fields, ML has been rapidly changing how the field of visualization approaches technical challenges and opportunities, particularly over the last decade. This section specifically focuses on harnessing ML to improve our visualization capabilities for science, rather than on using visualization to evaluate and improve our understanding of ML or coupling ML processes with visualization. Note that ML for data representations and interpretability of ML are discussed in PRD 1, and ML for personalization is discussed in PRD 5.

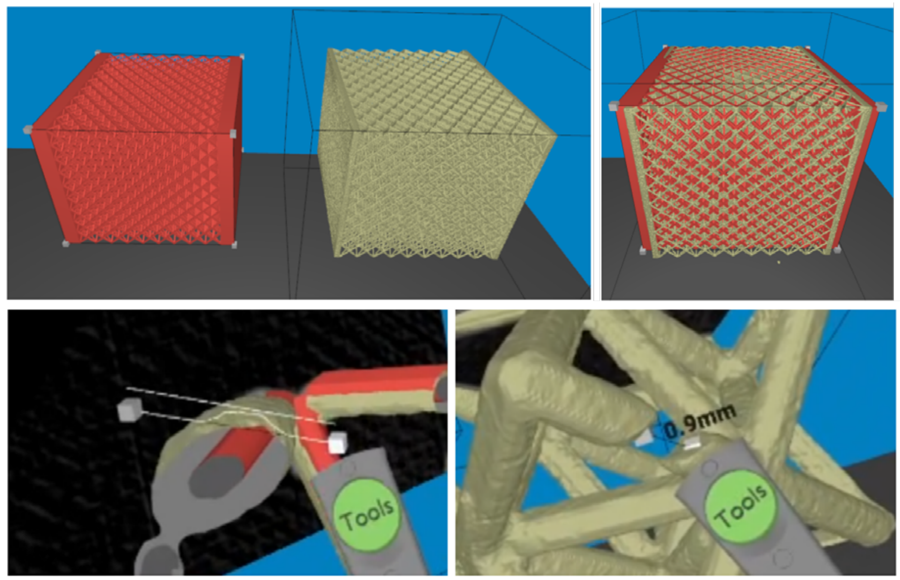


Figure 3.3.3b VR tool to align, compare, and analyze as-designed CAD models (red) with as-built scans (yellow) of additively manufactured parts.

Key challenges

This section focuses on the challenges of using ML in visualization approaches, including how it changes algorithm design and interactions with humans in data analysis pipelines. ML can offer enormous benefits to visualization: improving accuracy and computational speeds, capturing and inferring features that blend human input with computational analysis, and providing statistical models for complex data. Ultimately, ML is helping visualization to keep pace with the massive growth in generated data. The challenges for using ML in data visualization emulate its use in many other subfields of computer science, but can also lead to subtle differences that demand careful consideration of different use cases.

One broad area that requires further investment is applying and developing the most appropriate ML models for visualization applications. This challenge could benefit from a closer collaboration with the ML research community, as many of the primary applications of ML (e.g., computer vision and robotics) reside in a different constraint space than does visualization science. Techniques that work in those applications may not successfully generalize or have the required robustness to work in visualization settings. The demands and expectations for model explainability and interpretability are also different in a visualization use case, particularly when models complement existing data analysis. A second key challenge relates to the availability of data for ML in visualization settings. Often, scientific datasets come with limited supervision/labels and may indeed be a singular instance. Scientific observations and simulations also produce a diverse set of data modalities (e.g., fields, grids, meshes, particles). We further lack benchmarking data for broadly evaluating ML methodologies. Finally, clear computational and mathematical gaps require significant research investments. For example, in an in situ setting, allocated computational resources for an ML model may slow the simulation, or generally using such resources for a deep learning approach may be prohibitively expensive (e.g., if hours of training are required). While the state of the art is demonstrating a number of successful applications of ML, we must also close gaps on the mathematical side to explain how and why these models work (e.g., how best to incorporate physical constraints into a neural network).

State of the art

The visualization research community has used ML in various tasks in visualization pipelines, including ensemble and uncertainty analysis, data and visualization generation, feature extraction, deriving and presenting complex relationships in datasets, and improving performance and scalability in extreme-scale systems. While there are numerous examples of recent work in this area, we point the interested reader to three specific surveys on utilizing ML in visualization, all published and or updated no earlier than 2021 [182]–[184]. To understand parameter spaces in ensemble simulations, researchers developed deep learning–based surrogate visualization models for domain users to explore possible outcomes of different parameters without running expensive simulations [52]. Such deep learning models are also used to emulate part or all of the visualization process in volume visualization [52][185]. Another successful use of ML in visualization is data representation (see Section 3.1.1), and ML algorithms have also demonstrated effectiveness in extracting, characterizing, and analyzing salient geometrical and topological features for visualization [186]. For performance and scalability, researchers have used RL to optimize particle tracing's parallel efficiency, load balance, and communication costs for flow visualization in distributed-memory systems [186][187].

Research directions

Keeping pace with the explosion of ML research requires DOE to make a sustained investment. We outline some of the most promising research avenues below, noting that the best approach might ultimately be to pursue a diverse portfolio of research at the intersection of ML and visualization.

Developing and building trust when ML is applied. Visualization fundamentally helps us interpret complex phenomena. Incorporating a process from ML into a visualization pipeline potentially introduces opportunities for misinterpretation and uncertainty. Additional research is necessary to mitigate this and ensure that we can reap the benefits of using ML without creating additional downstream problems. Potential research directions consist of: (1) developing explainable ML models, for which we can easily interpret what they contributed to the analysis and why; (2) building a representation of model uncertainty via Bayesian inference, targeted at model-based surrogate representations of large-scale data; and (3) further developing visualization methods that are naturally capable of presenting the statistical nature of the model rather than simply treating it as a black-and-white process.

Using ML to enhance our summarization capabilities. While visualization plays a role in efficiently presenting data to analysts, success often relies on careful encoding choices. Meanwhile, ML is extremely promising as a module for extracting features in individual datasets and for computing relationships across multiple datasets. We thus see opportunities for ML in improving human–interface collaboration, with research directions spanning visualization recommendation systems, models of human interaction that facilitate customizable visualization design, and active learning techniques for cost-effective elicitation of domain knowledge in steering ML feature extraction.

Sparse and semi-supervised ML models. While many ML approaches, particularly deep learning, are enormously data hungry, in many DOE applications we may only be visualizing a singular dataset for analysis, or enter an entirely novel problem domain for which data does not exist. Thus, "black swans" can become more common for learning, whereas much of the existing success with ML does not face such data limitations, or can address such limitations with data augmentation. Within the DOE space, ML techniques that support visualization need to address a notion of generalization that differs from convention—one that allows models to rapidly adapt to novel problems, given prior data/knowledge from a set of potentially heterogeneous domains. Research topics ranging from how best to adopt meta learning, few-shot learning, and transfer learning in visualization contexts are promising in addressing these problems.

Understand the tradeoffs of data- versus physics-driven ML. Recent developments in computational science are now utilizing physics-informed neural networks to replace traditional partial differential equation–based science. Directly encoding physical properties is an appealing utilization of ML, for which the visualization community should also seek opportunities to preserve and translate these physical properties into analysis. Physics-informed models for visualization can improve the explanation of ML models, as they satisfy the laws of physics by design, not just approximately via optimization. Knowledge of physics in designing model-based surrogates also has the potential to significantly reduce data requirements, thus making ML models more practical as surrogates, or feature extractors, in visualization applications.

3.4 Improving Equity in Accessing and Engaging with Scientific Data and Processes

The previous sections implicitly, and often explicitly, focus on the challenges the field of visualization faces to support scientific discoveries at DOE. However, scientists are not the only audience of interest to DOE. ASCR has a vested interest, for example, to reach out to future scientists [188], especially those in minorities and underserved institutions. Similarly, DOE's role is not only focused on making discoveries, but also on communicating them to the general public, conveying the resulting policies, and supporting policymakers in their decisions. Reaching these audiences and where possible raising visualization literacy as a whole should be considered an explicit goal of visualization research at DOE. As captured in the well-known phrase "a picture is worth a thousand words," a good illustration is often the most intuitive, concise, and effective way to convey complex information. Moreover, some of the biggest impacts of other agencies such as NASA or the National Institutes of Health have come from inspiring the public's imagination about a Moon landing or communicating the risks of smoking. Additional research opportunities stem from the fact that what constitutes a good illustration depends heavily on the audience and purpose of the visualization.

In the context of this report, acknowledging that science communication is a primary goal, and that visualization in particular has an outsized role to play, raises a number of fundamental challenges. Each audience might require a different visualization. A major challenge remains in understanding how variables from data literacy and prior education to cultural norms and historical biases impact how a visualization is perceived. Simultaneously, determining commonalities, patterns, and cross-cutting approaches will be key to building effective tools. Furthermore, as with any communication strategy, the ultimate purpose matters; instilling trust in a decision might not be the same as explaining the decision, and conveying the risk of long-term consequences might require different strategies than providing near-term decision support. Here, we separate the challenges into two different but related research directions: (1) communicating results, decisions, or information in general to a wide range of audiences; and (2) providing these audiences the means to access relevant information in a meaningful and productive manner.

3.4.1 Improving Scientific Communication and Understanding

Visualization taps into the very best capabilities of our brains. It transforms data that is fundamentally abstract when presented as numbers into something that communicates and illuminates information ranging from the simple to the complex, and draws from science, art, engineering, and technology. The purpose of visualization has long been understood as insight and increased understanding [189], and the process of transforming data into representations, visuals, information, or insight is complex. Designing effective visualizations requires understanding of not only the discipline from which the data was generated, but also knowledge of visual representations, communication, purpose, and intended audience. (See Figure 3.4.1.)

Key challenges

DOE's visualization research has almost exclusively focused on "deriving new insight" as its purpose and "subject matter experts" as its audience. However, the DOE mission space

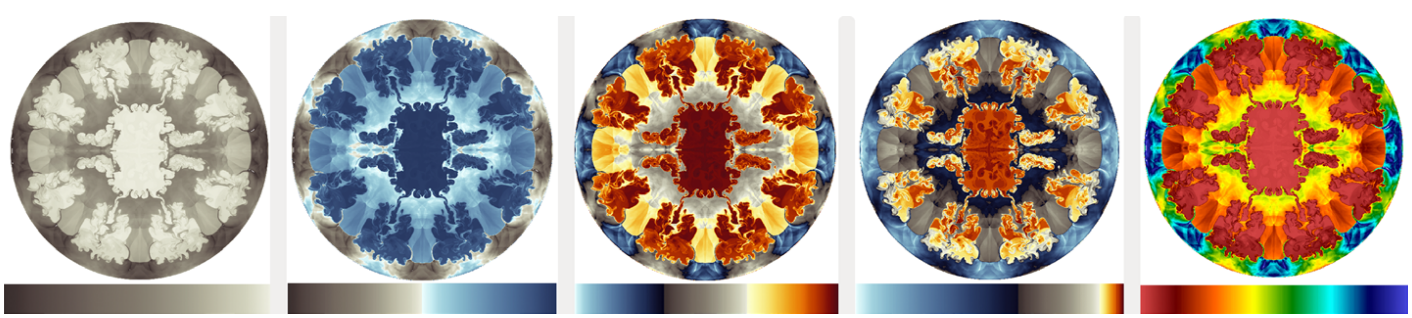


Figure 3.4.1 A system of colormaps ablate to reveal increasingly greater detail within a visualization.

covers a much broader spectrum of both purposes and audiences, and techniques developed for the bleeding edge of science do not necessarily work well to excite future researchers, explain policies, or support decision-making. Addressing more diverse audiences for a wider set of purposes raises new research challenges in scientific visualization in particular and visual communication in general.

The first challenge stems from the fact that visualizations are ultimately interpreted by humans, and each audience will react differently to a given illustration (see Section 3.5.1). A visual metaphor that is intuitive to a scientist might be opaque to a high school student and an educational illustration about wildfires may not help a community leader make zoning decisions. Yet little effort has been spent within DOE to understand what types of visualizations are most appropriate for a given situation, or how one might adapt techniques to become more broadly applicable.

The second challenge in this area relates to trust and transparency. While visualizations can be incredibly instructive, they typically require a complex workflow of collecting and curating the data, aggregating or processing it, and choosing colormaps and representations. For non-experts, these steps are mostly opaque and require users to implicitly trust both the data source and the visualization creator. For any contentious issues, the common lack of trust in either source or creator results in inconvenient conclusions being ignored and allows cognitive bias to thrive. This problem is especially pronounced as visualizations, even more than quantitative statistics, harbor the risk of intentional bias and can be highly misleading.

Even the best visualization can only convey a limited amount of information, especially when considering additional elements that might be required to instill trust and provide transparency. The final challenge requires both communicating where the boundaries are (e.g., how far a given visual metaphor holds) and going beyond individual illustrations to convey information.

State of the art

The potential of visualizations to communicate ideas has long been recognized [190][191] alongside the need for specialized techniques to create "communication-minded" or "explanatory" visualizations [192][193]. A plethora of books [194] and popular blog posts [195]–[197] are devoted to visual communication of data, yet examples of visualizations explicitly designed to explain—rather than explore or analyze—scientific data remain less common. Some examples are visualizing abstract mathematical [198] or physical [199] concepts as well as visualizing climate change [200]. Yet even simple representations, such as line charts, are not necessarily easy to explain [201] and are the subject of recent research. Furthermore, how to judge the effectiveness of a communication-driven visualization is not necessarily obvious [202]. Successful examples of custom visualization (tools) being highly effective in engaging public audiences have been noted in astronomy [203], history [204], or neuroscience [205], among others. However, as before, these efforts have required significantly different approaches than standard scientific visualization [206]. Finally, a related concept of artistic data visualization [207] uses artistic techniques to more deliberately convey a message (e.g., on climate change) [200]. Nevertheless, despite some successes, the use of visualization for communication in DOE has remained largely unexplored and carries a significant potential to increase outreach, impact, and engagement. The recommendations in this Section consider science communication as a distinct research challenge.

Exploring the general concept of trust has a long history [208], and prior research suggests two related but different notions of trust: relationship-based trust and evidence-based trust. A typical example of the former is our trust in a doctor's opinion based on past interactions and the reputation of their educational institutions. Evidence-based trust, on the other hand, often relies on quantitative data. Some prior work has investigated design criteria to instill trust [209][210], but the challenge remains to evaluate the success. Furthermore, a general goal of increasing trust is not appropriate, which leads to the notion of calibrating trust [211] to provide unbiased information. Ultimately, both notions of trust are intimately connected to transparency—allowing the users to understand the provenance of all data, algorithms, and decisions throughout the visualization process [212][213]. However, while a large body of work focuses on maintaining provenance and reproducibility for visualization tools [214], less clear is guidance on how to collect and represent the corresponding information for the entire data processing pipeline, convey the trustworthiness of a particular algorithm used during the processing, or convey the motive of the creators.

Communication in general has focused on storytelling as a particularly effective way to engage audiences, and the field of visualization is no exception [215]. Similar to exploratory visualizations requiring different approaches than explanatory ones, early research investigated the different design space of storytelling with data [216]. Nevertheless, while aspects of these ideas have certainly influenced the field of visualization as a whole and are commonplace in journalism, only recently has a specific focus on visual storytelling emerged in the research literature [217]. Examples include data stories for COVID-19 [218] and recently proposed general frameworks [219]. How these ideas can be best adapted to DOE science communication and decision-making remains largely unexplored but could have significant impact.

Research directions

Improving scientific communication within an application domain, across science, and to decision-makers and the public will require a renewed investment in building trustworthy, explanatory visualization for a broad range of audiences.

Research the design and impact of explanatory visualizations for diverse audiences. Both the intent and the audience critically matter to the success of a visualization. However, with few exceptions, explanatory visualizations have largely been the domain of popular science publications and journalists. While successful, these outlets do not address all of DOE's needs, audiences, and use cases. Furthermore, the metrics of success in public media can be substantially different from the goals of DOE, and standard approaches to evaluate the impact of a given illustration do not necessarily exist (see Section 3.5.3). Consequently, we need a dedicated research effort focused on a framework to design explanatory visualizations aimed at specific DOE goals (e.g., science engagement, policy explanation, decision support). Furthermore, this effort must investigate how various audiences may require different approaches depending on their background knowledge, data literacy, or experiences, as well as where commonalities emerge to more effectively address these needs.

Develop communication approaches that enable viewers to appropriately calibrate trust in the data and science content. This recommendation goes beyond simply promoting trust—for example, a highly polished data story might appear trustworthy when in fact the underlying data is faulty. Instead, we should better enable viewers to ask questions and approach data sources and analysis processes with healthy skepticism. Significantly more work needs to be done to better understand how building and calibrating trust impacts the design of visualizations and systems. We posit that increasing transparency in several areas will increase and calibrate trust in visualizations: data origins and collection approaches; steps taken to clean, prepare, and transform the data; and any underlying assumptions. Furthermore, wherever possible, visualizations should be self-explanatory both in their provenance and their rationale.

Develop new connections between scientific storytelling and visualizations of the scientific decision-making process. Developing a single, static visualization that is self-explanatory, trustworthy, and contains all necessary provenance information might not be possible in all cases. Instead, reaching a decision or explaining an insight is a process, and using the human predilection for storytelling to convey this process has significant potential. Using scientific storytelling can help to surface the way that people, data, and decisions were combined to achieve some new insight [220]. However, the scientific visualization toolkit for generating visual representations of this decision process has insufficient support. We contend that using the context of provenance and richer visualization approaches based on conveying both results and processes can be transformative in improving science communication to the broader public.

Scientists have substantial training in communicating their results within their particular domains, and our scientific visualization/perceptualization frameworks are built around this goal. However, we must develop approaches to effectively make data storytelling accessible to lay audiences that consider the diversity of user perspectives and literacy (see Section 3.4.2). This effort will require developing new abstractions and approaches that can make storytelling easier to generate for a scientist, while also generating the storytelling context appropriately for non-specialist and public audiences. Methods must be developed to ease the burden of documenting data provenance and context, including ML- and AI-based tools to automate documentation with user interaction research to better understand how different user communities engage with visualizations, as well as associated provenance documentation. Additionally, to have the clearest effect in building trust in scientific results and the scientific process, this research requires exploring representations for scientific visual storytelling and the community-specific contexts required for presenting and specializing those stories.

3.4.2 Increasing Accessibility of Tools and Data

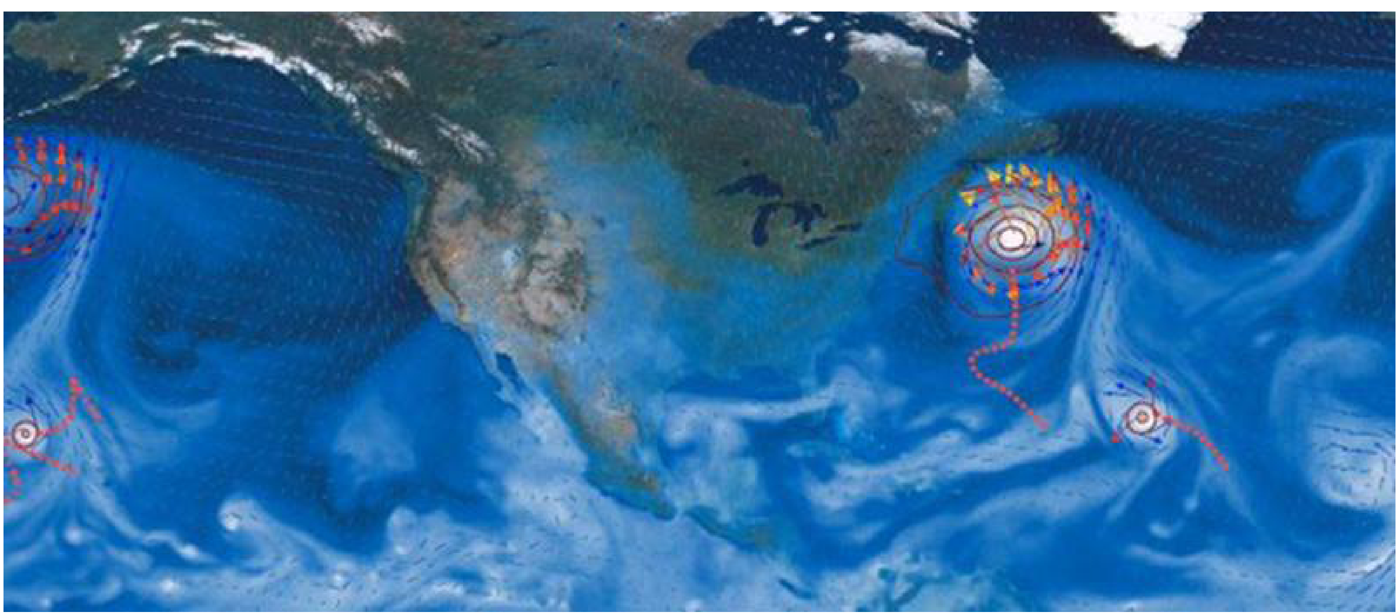
As discussed above, science communication is primarily concerned with conveying curated information to an audience. However, the same approach to visualization puts a significant burden on and affords substantial influence to the creator of a visualization. This outsized role can lead to inherent mistrust, and in some situations (e.g., in decision support), any significant influence by the visualization expert is potentially problematic. Addressing this challenge requires a different paradigm focused on empowering the audience to access, curate, and explore relevant data on their own. DOE scientists are deeply involved in a wide range of applications that create or collect data and/or produce forecasts in areas ranging from climate change and wildfire risk to the impact of fracking and the national power grid. (See Figure 3.4.2a.) Providing audiences outside the traditional user groups access to this wealth of information is crucial to promote an engaged and informed public and, more generally, will increase the trust in and utility of a visualization. Examples range from providing indigenous communities access to environmental data to enabling decision-makers to personally explore data to form their own opinions rather than consuming precompiled results. However, data access by itself is not sufficient. Instead, users require tools to meaningfully engage with this data. Visualizations and visual interfaces are often the primary components of such tools. Furthermore, building broadly applicable and accepted tools without a more diverse workforce is difficult to imagine. Therefore, a crucial step is to engage Minority-Serving Institutions and Historically Black Colleges and Universities in this research [188][221]. Ultimately, this research will lead to a virtuous cycle in which researchers directly engage a much wider audience, which will in turn lead to more exposure and engagement that will ultimately be reflected in the future workforce and the next generation of researchers.

Key challenges

Giving a wide range of audiences meaningful access to DOE data presents three fundamental challenges. First, the data itself must be available alongside the necessary provenance. Active efforts at DOE and elsewhere [222][223] promote the creation of FAIR data (Findable, Accessible, Interoperable, and Reproducible). However, so far, *findable* does not necessarily mean easy to find, nor do *accessible* and *interoperable* mean data is easy to handle. Instead, much of the current infrastructure is centered on text-based searches and assumes significant knowledge of aspects such as naming conventions and downstream tools. To democratize the access and fully realize the intent of FAIR principles requires new dedicated visual interfaces, interactive means to find and subselect data, and visual cues on data origins and provenance. Such visualization front-ends will be especially important to reach nontraditional audiences on new platforms such as mobile devices (see Section 3.2.2).

Given access to large datasets, the next challenge is to enable users to explore this data, draw appropriate conclusions, and communicate results. (See Figure 3.4.2b.) Currently, this process faces significant barriers by requiring the knowledge of multiple tool chains, computational resources, and often significant computer science skills to connect all the pieces. Instead, integrated visualization tools, accessed on mobile devices, and if necessary backed by cloud computing could lower or even eliminate many of these barriers. Apart from the technological challenges (see Section 3.2.2), this integration will also require more intuitive visual interfaces to enable functional access to the available capabilities. Furthermore, these tools must be unbiased and, wherever possible, indicate common pitfalls in creating erroneous or outright misleading results (e.g., in appropriate colormaps, mismatched data scales). Finally, a tradeoff will always exist between how simple an interface can become without

Figure 3.4.2a. The National Energy Research Scientific Computing Center's Toolkit for Extreme Climate Analysis detects and tracks extreme weather events in large climate datasets. This visualization depicts tropical cyclone tracks overlaid on atmospheric flow patterns.



overly limiting the type of visualization and analysis it enables. Ultimately, a highly curated and guided exploration is closer to a communication approach in the sense of Section 3.4.1, rather than enabling flexible data access. The data portal should be flexible enough to allow interested users to progressively engage more deeply with the data and algorithms and expose a path to learn new analysis and visualization skills. This approach will require significant efforts in structuring visualizations to not just inform but to educate a diverse audience.

State of the art

Large data collections are publicly available in a variety of fields such as climate science [224]–[226], biology [227][228], and turbulence [229], and the DOE has long invested in infrastructure to allow the transfer of massive amounts of data across institutions [230]. Data access takes many forms: from the largely text driven search for available downloads in climate archives [226] to individual data collections with code [231] similar to common scientific publications and to highly curated information portals (e.g., for wildfires [232] or energy diversity [233]). Still, finding relevant data remains difficult, especially for non-experts, and only a fraction of the data that exists is available. Some efforts are aimed at enhancing the search process using visualization [234], but more often, navigating large collections of data, as opposed to analyzing the data itself, is a secondary goal [235][236].

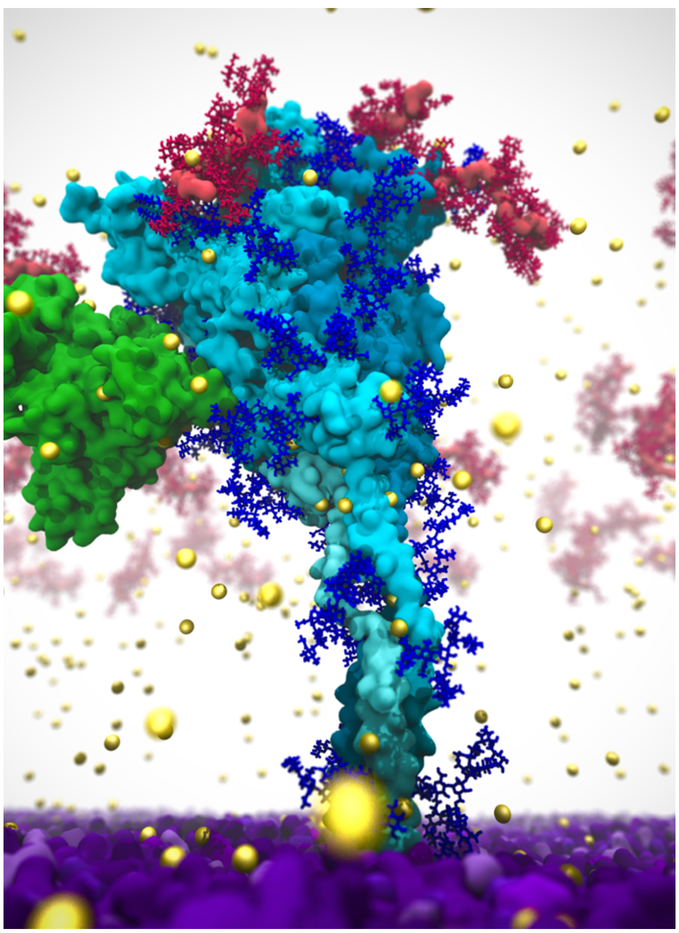


Figure 3.4.2b Visualization of the SARS-CoV-2 virus' spike protein (cyan) surrounded by mucus molecules (red) and calcium ions (yellow) [411].

DOE has also invested significant resources in building openly accessible tools for data analysis [101][127][237][238] that cover a wide range of application domains. However, few of these approaches are novice friendly, and most require both access to computing resources and prior knowledge not common in the general public. Instead, visualization dashboards are used to allow users to explore complex datasets [239][240], but traditionally these are highly focused on particular subjects. Common examples are dashboards in healthcare [241], public services [241]–[243], politics [244], management [245], or the energy grid [246]. Most notably, the COVID-19 pandemic spurred the development of many dashboard-type portals [247][248]. However, with few noticeable exceptions (e.g., in climate [224] or wildfire research [232]), access to DOE science data remains functionally restricted to the corresponding scientific communities. Even where dashboards are available, how they address the above challenges on transparency, accessibility to non-experts, or the desire to guide users in developing expertise is unclear. In fact, evidence suggests that some of the challenges are being actively exploited to promote misinformation [249].

The traditional remedy to intentional or unintentional misuse of any analysis or communication tool is better education in the underlying concepts and potential pitfalls. A promising direction is in developing curated recommender systems [250][251] that lead towards a learning-by-example approach to visualization [252]. Nevertheless, such approaches only shift the need for trust and transparency to the creators of the recommendations. More research is needed to develop tools and approaches that enable both functional data access as well as provide a path towards raising visualization literacy as a whole [253].

Research directions

The ultimate goal of this recommendation is to provide more people not only access to more data, but also the tools and skills necessary to effectively consume this data in the form of intuitive visualizations. This goal leads to three interconnected research directions in the field of visualization.

Develop accessible, transparent, and trustworthy data portals. Massive amounts of data are already publicly available in areas spanning climate simulations, traffic predictions, wildfire risk, census data, and more. The first step in empowering everybody independent of background, resources, or prior knowledge to consume this data is to provide interfaces to find the relevant subsets. Outside of general AI assistants, these interfaces must be visual and interactive, adapted to mobile platforms, and intuitive to use. In this context, visualization can provide an important front-end to the existing efforts on making data FAIR [222] [223] and convey important assumptions, limitations, or other knowledge of the data sources. Transparent information—for example, whether this data is measured or the results of a simulation; how accurate this data is expected to be; whether the data was designed to be predictive or to explore an extreme scenario; and where this data is coming from—must be available in an easy-to-digest manner to promote trust in and discourage misuse of data. Finally, the visualization community must work with data producers to enable an easier path to such active sharing of information.

Lower the barriers to create accurate visualizations from diverse data. Given a data source of interest, new methodologies are needed to allow users to easily express their intent for a desired visualization. Potential avenues are text or voice interfaces, example-driven processes, or other approaches that require minimal prior knowledge. To reach the desired audience, these tools must be mobile enabled and most likely web-centric. Furthermore, in conjunction with the educational focus of the next research direction, an ideal system would be multilayered and enable interested parties to expose successively more complex aspects to gain deeper expertise in consuming data and communicating results.

Develop visualizations for education. Any automatic interface to create visualizations will ultimately be limited due to an inherent tradeoff between complexity and expressiveness. Therefore, new visual interfaces are needed to teach audiences from school children to community organizers the use of more sophisticated approaches. Where Section 3.4.1 discussed tools for visual storytelling, the approach here will focus on teaching how to tell effective stories. Promising directions are the generation of automatic tutorials, gamification of visualization, or intuitive searches of prior examples to emulate. As before, the target audience will play a significant role in the choice of approach, and research opportunities range from how to best incorporate domain concepts to reach application scientists as well as to engage potentially data-illiterate communities. Finally, to encourage productive scientific discourse, any visual curriculum must cover potential pitfalls such as bias, uncertainty, and misinterpretation.

3.5 Developing Intelligent Approaches for Adaptive, Context-Aware Visualization of Scientific Data and AI

A key goal of scientific visualization is to provide important insights for reasoning and decision support. Consider scientific data processed on an exascale computer that must be visualized, inspected, and interactively and collaboratively evaluated by a large, distributed team of experts. The visualization presented might be used in numerous ways by researchers and decision-makers, where conveying the complex information while also considering the cognitive process underlying the use of the visualization are important. (See Figure 3.5.)

Figure 3.5 An image representing future technologies.

Deeper understanding of visual perception, cognition, and reasoning are crucial in developing general-purpose, intelligent visualization tools that are readily customizable to meet the evolving needs of the users (e.g., scientists, general public), team-based collaboration, and downstream tasks. The visualization might be used for hypothesis generation, decision support, interactive ML, and scientific information dissemination. Even though data and visualization have become critical components of scientific exploration and decision-making, methods and strategies that manage perception, cognition, and validation require additional attention. Ensuring that DOE's visualization tools effectively support the user community through computing paradigm and HCI shifts is critical.

3.5.1 Development of Methods and Models for Perception and Cognition

Effective visual representation of complex data is critical for easier understanding, broader communication, better reasoning, and better decision-making. As data and visualization become more embedded in scientific exploration and systemic decision-making processes, demand has grown for visualization tools to support an ever-increasing set of tasks (e.g., exploration, understanding, hypothesis generation, communication, collaboration, decision-making), users (e.g., novice, experts, general public), and scenarios (e.g., high-risk decision-making, rapid response, leisurely navigation). Designing and properly validating generalizable visualization tools to ensure broad and lasting impact is a critical gap. Overwhelming scientific evidence from the visualization, HCI, and psychology communities demonstrates that individual characteristics matter when designing and using visualization tools [254].

Key challenges

Scientific data is often open to wide interpretation, and two users might perceive and interpret it differently. Similarly, two scientific users with different purposes might require specific visual representation geared towards their process (e.g., scientific insight versus decision-making). Addressing a user's individual characteristics when designing visualizations advances accessibility in a number of ways to help adapt the presentation of the scientific data in personal and contextually relevant ways. Scientific users and decision-makers must be presented with visualizations that account for cognition, perception, and bias. Additionally, systems must adapt as a user adapts, and they must learn from a user as the user matures in using tools and systems. However, this research still needs to be developed, and significantly more work is needed to understand how and which data should be collected to minimize disruption in the reasoning process.

State of the art

A growing body of work researching personalized visualizations initially sought to investigate which aspects of a visualization are of interest to an individual user, which representations are more meaningful, and to what extent systems can automatically recommend visual representations [255][256]. This relatively new area of research has demonstrated that individual differences impact how users approach problems and how they use visualizations in that process. Studies have demonstrated that individual differences in experience [257][258], perceptual speed [259][260], spatial ability [261]–[263], and working memory [264][265] can significantly impact the design preferences and effectiveness of visualization tools. Similarly, previous work has looked at the impact of colormaps on cognition [266][267]. Although there has been some advancement in understanding the importance of personalization in the effective utilization of visualization tools, the translation of this knowledge to visualization tools is lacking.

Research directions

Establishing synergistic collaboration between human perception and cognition and the visualization tool presents unique research opportunities for a paradigm shift towards usable and impactful visualization.

Methods to study perception, cognition, and user interactions with visualization tools. Better understanding of human perception and cognition will be necessary to design next-generation visualization tools that can truly improve the human capacity to understand and reason with complex data and AI models. This need includes capturing and sharing the complex informal and formal work practices associated with knowledge discovery from the data through visualization. Understanding and identifying human touchpoints and user interactions with the visualization are also important, requiring deeper investigation into how people use visualization tools and integrate them into end-to-end workflow and data pipelines. As visualization is increasingly used for scientific interpretation and decision-making, visualizations must account for cognition and perception biases and understand what a certain user seeks in the visualization and how they perceive it. The bias needs to be considered not only in terms of automatic visualizations made by the users, but also in the context of any choices the users might make. Our ability to account for perception and cognition related challenges in the context of scientific visualization is limited but critical as we deal with complex data, AI models, and automation.

Methods for personalization and automation. AI provides an opportunity to learn new information about users and how they consume and interact with visual content. Developing intelligent visualization tools and user interfaces that adapt dynamically to the individual's needs and that present options for analysis has the potential to speed up scientific innovations and impact. Creating real-time adaptive, mixed-initiative visualization systems offers new possibilities for understanding complex thought and decision-making more representative of how we interact with visualizations and data, thus giving us tools we need to create better, more informative visualizations. Such adaptation is critical to truly augment every user by leveraging the current context to provide the most useful information in an optimal and efficient manner. With reasoning and decision-making tasks in real-world environments, such adaptive and personalized visualization technologies can reduce visual fatigue, cognitive load, interpretation bias, and decision errors while improving efficiency.

Automation is expected to impact many scientific domains in the form of self-driving and self-guiding experiments, observations, and infrastructure. Automation in scientific visualizations can also improve outcomes as we handle large volumes of data and complex AI models, where interactive visualization might no longer be feasible. For example, real-time analysis of multimodal, high-dimensional, time-series data can be rather complex and computationally intensive. Dynamic dimensionality techniques such as latent discriminant analysis or ML approaches will be necessary to help identify the most informative features to accelerate analysis and help learn about the user's needs and intents for the task at hand. Still, real-time analysis of large multimodal, multidimensional user-interaction data can easily become computationally burdensome, especially when combined with actual visualization of extreme-scale, multimodal, multidimensional scientific data as the primary task. State-of-the-art deep learning models for time-series data can be particularly helpful in inferring users' need for support during interaction or in developing surrogate user models to predict how a specific visualization tool could impact users with specific trait profiles.

3.5.2 Collaborative Visualization

A defining characteristic of DOE research is the need for large, multidisciplinary teams to tackle the grand challenge goals of nuclear physics, material science, or climate science. DOE teams are highly diverse in skills, expertise, and location, making effective collaboration tools a crucial component in solving mission critical problems and in managing the data generated by DOE experiments and observations. In Section 3.5.1, we discussed how individual users might perceive visualizations differently, making collaborative visualization particularly challenging. Collaborative visualization has been defined as a "subset of CSCW [computer-supported cooperative work] applications in which control over parameters or products of the scientific visualization process is shared" [268]. Over the decades, research in this domain has explored the multi-user aspects of shared control and use of tools to augment data-sharing and visualization. While collaborative visualization is a few decades old, comparatively few studies have been conducted on how to improve the process of collaboratively visualizing large volumes of data and scientific phenomena with existing and emerging (e.g., Jupyter) tools.

Key challenges

As science becomes increasingly more complex and interdisciplinary, and as scientists become more distributed among diverse teams, better collaboration tools and processes must anticipate and follow these trends. Overcoming barriers to sharing inherent in the visualization tools themselves (e.g., control of parameter selection, freedom to navigate) is a research goal that will need to be addressed to harness the full capability and promise of social, immersive, and collaborative visualization. (See Figure 3.5.2.) Scientific visualization—whether conducted in co-located settings or experienced through technology-mediated mechanisms such as screen sharing—is a richly collaborative process.

State of the art

Scientists and visualization developers often collaboratively generate visualizations side by side or over diverse networks, devices, shared displays (e.g., screen-sharing via laptops or

Figure 3.5.2 An audience at the Adler Planetarium in Chicago being immersed in a visualization of the Kuiper Belt.

workstations, interactive large-group displays), or VR and AR environments. In many respects, the increase in distributed, technology-mediated collaboration overall during the COVID-19 pandemic highlighted some of the challenges to the process of collaboration itself that are characteristic of distributed teams working together to visualize large volumes of data residing on HPC systems. The unplanned and imposed nature of the pandemic [269] exposed several shortcomings in the scientific community that present unique opportunities to expand research in distributed, collaborative visualization. We learned that synchronous video conferencing collaboration tools alone are inadequate substitutes for side-by-side, co-located visualization. Synchronous communication tools could be accompanied by collaborative workspaces that allow real-time, persistent interactions such as note-taking and clustering. We also learned that typical scientific workflows involving the need to access data residing on remote systems (e.g., supercomputers) while using disparate visualization tools could be improved to provide more seamless analysis experiences, perhaps by leveraging in situ methods and exploring enhancements to the analysis environment.

In collaborative sense-making, many visualization tools have been developed for use by a single user [270] and therefore lack metadata to assist in visualization software reuse. Existing tools do not provide mechanisms by which control can be transferred to other remote participants. Most commodity tools used today (e.g., Teams, Zoom, Bluejeans) do not provide the ability to capture simultaneous remote interactions. Past tools such as Lawrence Berkeley National Laboratory's real-time Video Conferencing Tool could be coupled and extended to provide scalable remote sharing of custom VTK applications [271]. Furthermore, most visualization tools do not offer the ability to share out state by default. Given the open-source nature of many standard visualization tools (e.g., VTK, ParaView, VisIt) and their ability to save a given state, one could imagine an easy adaption to provide sharing/collaboration capabilities. For example, ParaView has had this capability, but whether it still exists in the current codebase is unclear. To truly enable collaborative visualization, the underlying technology must be in place to provide a standard set of capabilities to all users.

Research directions

Methods to support collaborative visualization. Scientists are now used to seamless, live online collaboration through tools like Google Docs, and have come to expect that collaborative visualization will also function similarly. Collaborative visualization requires each participant to have a consistent view of a shared state that may include arbitrarily complex external resources such as local files and the in-memory state. The ability to capture and synchronize across multiple participants is nontrivial, especially as we handle large data volumes and complex AI models. Additionally, tools such as Jupyter are increasingly used by scientists for collaborative analytics and visualization, and supporting collaborative visualization in these environments is imperative.

A collaborative environment necessitates ensuring consistency and reproducibility. Each step of the visualization should be consistent, reproducible, and validated across the actions of all participants. If a given visualization were rerun, the same result should be generated irrespective of who performed what, when, and where. Ensuring reproducibility in collaborative visualization requires several levels of provenance collecting and testing: unit testing, testing the integration and workflow pipeline, verifying that the images produced remain the same [272], verifying that interactions preserve expected behavior [273], and ensuring that error and uncertainty can be understood and reported [274].

Visualization technologies to support collaboration. More research is needed on immersive visualization and cognitive context in collaborative visualization—namely, the extent to which existing single-user visualization tools and immersive technologies such as MR, VR, and AR will facilitate result interpretation, creativity, and productivity. We will need immersive VR, AR, or cross-reality environments that incentivize reflection, active simultaneous participation, persistent recall of information, scaffolded assistance, and cultural cues and that support multiple points of view and perceptions [177]. Finally, the context-aware tools and methods of technology-mediated communication must evolve to address cognitive bias in collaborative and team decision-making, particularly when data, simulations, and models are poorly visualized and decision-makers lack adequate interaction options or the interaction is hampered by technology designs that may impede best practices of collaborative visualization. We anticipate the impact of this research to radically improve collaborative discovery through visualization, and lead to development of better approaches for engagement among scientists, the public and policymakers at all levels.

3.5.3 Evaluation Methods and Metrics

Visualization tools are created with the goals of helping humans generate insights and communicate with others. However, the ultimate outcomes are difficult to measure objectively and can vary widely across user groups and even individuals. Representations that are intuitive to a subject matter expert might be opaque to a decision-maker, and a successful popular science illustration might feel imprecise and misleading to a scientist. Nevertheless, for any field to make progress requires the ability to measure the impact of a new idea and to objectively determine whether a new solution is more (or less) effective than prior art. Since interpreting visualizations is inherently subjective, the appropriate metrics and methods to evaluate them are elusive.

Key challenges

Some of the most frequently stated goals of visualization are new insight, better understanding, and clear communication. These abstract and qualitative concepts are difficult to distill into concrete observables. Furthermore, they are broad terms that combine a plethora of more specific aims such as, usability, aesthetics, and engagement. For any specific use case, the relative importance of these goals might vary by target audience. For example, should a data analysis tool be aesthetically pleasing or as precise as possible? How might this calculus change if an unattractive tool hampers adoption? Once the primary goals are determined, the next challenge is to define metrics for how well the goals have been achieved. For some tasks, creating quantitative metrics is straightforward (e.g., the time or number of clicks required to achieve a predetermined outcome), while other metrics (e.g., time to insight) are anecdotal at best.

The final challenge is how to evaluate the chosen metrics in practice, which remains an open problem in the larger community; DOE-specific environments are creating additional complexities. One aspect is the need for reproducibility to ensure a fair comparison. Even creating a comparable environment in terms of monitor size, input devices, and other factors might not be straightforward. The problem becomes even more pronounced for high-level decision-makers who need to make timely decisions based on the visualizations. Collectively, determining in the near term if a visualization approach is effective is another challenge.

State of the art

Demonstrating that tools lead to human scientific insight requires such tools be used by scientists who have limited time to devote to visualization testing and are often hesitant to switch from their existing tools [275]. Previous work has looked at usability heuristics for scientific visualization [276][277], which evaluate visualization tools on specific dimensions of usability [276]–[279]. However, we do not yet understand many metrics in the evaluation of visualization tools, and as visualizations are developed for people with considerable expertise, the pool suitable for human-centric evaluation is small. This challenge is compounded in collaborative scenarios that should be validated by multiple groups of users. Validating and evaluating visualization methods and results is an open area of research.

Research directions

The two key open challenges in enabling quantitative evaluations of visualization approaches for DOE are (1) development of appropriate metrics and an understanding of when to apply them; and (2) new methods to reliably assess these metrics in the DOE context.

Informative metrics to guide the development of visualization tools. New research is needed to define a comprehensive set of metrics to evaluate the entire breadth of DOE relevant use cases discussed in Sections 3.1 to 3.4. These metrics might build from diverse communities including qualitative and quantitative metrics and account for DOE-specific metrics (e.g., reliability or uncertainty quantification). Furthermore, many metrics lead to invariable tradeoffs (e.g., between simplicity and generality), and not all are equally important in all cases. Currently, no clear guidelines exist for what types of DOE applications prioritize which metrics and how to make this determination. Using a systematic approach to developing an informative set of metrics is important.

Reliable and practical methods to evaluate visualization tools. Assuming one is given a set of metrics to evaluate a new approach, new techniques are required to assess these metrics. In particular, a number of DOE-specific challenges arise with existing approaches. The first challenge is that relevant user groups often have highly specific needs and skills, creating a very small population that is not readily accessible. Not only do small numbers of potential users make traditional evaluations anecdotal at best, but they also suffer from potential bias as users and visualization are often long-term collaborators. One possible direction to address this challenge is to carefully analyze commonalities between use cases and to develop techniques that evaluate common properties across applications while managing the increased heterogeneity of users. The second challenge is how to record quantitative data in the highly time-constrained, potentially security-sensitive, and often unique environments of interest to DOE. Evaluation approaches that require hours of dedicated testing will create significant barriers to adoption, as might tracking software. Furthermore, the evaluation software itself might need to run on nonstandard hardware (e.g., HPC systems or mobile platforms), which creates practical challenges. Finally, DOE is engaged in many scientific grand challenges with ambitious targets and multidecade visions. However, this long-term perspective often leads to somewhat intangible goals like "promoting new insights" and abstract needs such as "trust." These ideas are key to the overall mission but difficult to quantify. New approaches are needed to reliably consider such concepts.

Appendix A - Workshop Summary

In the fall of 2021, ASCR portfolio program manager Margaret Lentz approached Peer-Timo Bremer from Lawrence Livermore National Laboratory (LLNL) and Gina Tourassi from Oak Ridge National Laboratory (ORNL) to organize a workshop aimed at identifying priority research directions for data visualization at DOE. The charge was to bring together the visualization community at DOE and at interested academic institutions to identify a 10-year outlook on upcoming challenges and opportunities. The team assembled a steering committee that consisted of laboratory scientists and long-term academic partners:

1. Peer-Timo Bremer: LLNL (co-chair)
2. Georgia Tourassi: ORNL (co-chair)
3. Wes Bethel: San Francisco State University / Lawrence Berkeley National Laboratory
4. Kelly Gaither: Texas Advanced Computing Center
5. Valerio Pascucci: University of Utah
6. Wei Xu: Brookhaven National Laboratory

With the support of the DOE Early Career researchers in visualization funded at the time—Kathrine Isaacs (University of Utah), Joshua Levine (University of Arizona), and Bei Wang (University of Utah)—the committee developed an initial charge document outlining the workshop's scope and a list of four areas of discussion [280]. The document and a call for community input in the form of white papers were subsequently published [281] online. The committee selected 66 submissions as part of the workshop material, with 23 slated for live presentations [282]. The first authors of all accepted white papers alongside a diverse set of researchers were invited to participate in a workshop over two-and-a-half days in January 18–20, 2021. The event was held online due to COVID-19 restrictions.

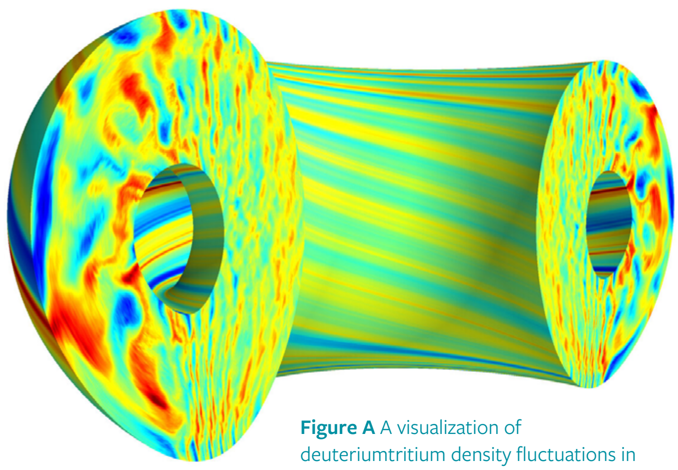


Figure A A visualization of deuteriumtritium density fluctuations in a tokamak driven by turbulence. Areas of red are representative of high density and areas of blue are representative of low density.

Back to TOC

A.1 Participants

The final list of participants including keynote speakers is given here.

James Ahrens
Los Alamos National Laboratory

Ilkay Altintas
University of California, San Diego

Dustin Arendt
Pacific Northwest National Laboratory

Tushar Athawale
Oak Ridge National Laboratory

Zhe Bai
Lawrence Berkeley National Laboratory

Divya Banesh
Los Alamos National Laboratory

Jeffrey Baumes
Kitware, Inc.

Paul Bayer
U.S. Department of Energy, Office of Biological and Environmental Research

Pete Beckman
Argonne National Laboratory

Janine Bennett
Sandia National Laboratories

Brian Benscoter
US DOE, EESSD

Matthew Berger
Vanderbilt University

Wes Bethel
Lawrence Berkeley National Laboratory

Jean Luca Bez
Lawrence Berkeley National Laboratory

Harsh Bhatia
Lawrence Livermore National Laboratory

Ayan Biswas
Los Alamos National Laboratory

Laura Biven
NIH

Alexander Bock
Linköping University

Mark Bolstad
Sandia National Laboratories

Peer-Timo Bremer
Lawrence Livermore National Laboratory

Ben Brown
DOE ASCR

David Brown
Brookhaven National Laboratory

Joshua Brown
Oak Ridge National Laboratory

Paul Brunk
University of Georgia

Reuben Budiardja
Oak Ridge National Laboratory

Roxana Bujack
Los Alamos National Laboratory

Suren Byna
Lawrence Berkeley National Laboratory

Tara Camacho-Lopez
Sandia National Laboratories

Carla Campos
Pennsylvania State University

Franck Cappello
Argonne National laboratory

Richard Carlson
DOE/SC ASCR

Brad Carvey
Sandia National Laboratories

Thomas Caswell
Brookhaven National Laboratory

Junghoon Chae
Oak Ridge National Laboratory

Iris Chang
SLAC National Accelerator Laboratory

Aashish Chaudhary
Kitware, Inc.

Jieyang Chen
Oak Ridge National Laboratory

Jian Chen
Ohio State University

Mandy Chen
University of Chicago

Siu Wun Cheung
Lawrence Livermore National Laboratory

Hank Childs
University of Oregon

Shreyas Cholia
Lawrence Berkeley National Laboratory

Housen Chu
Lawrence Berkeley National Laboratory

Eric Church
HEP

Valentino Constantinou
NASA

Joseph Cottam
Pacific Northwest National Laboratory

Nina Crosby Walton
Washington University in St.Louis

Biprateep Dey
University of Pittsburgh

George Djorgovski
California Institute of Technology

Bin Dong
Lawrence Berkeley National Laboratory

Soumya Dutta
Los Alamos National Laboratory

Kel Elkins
NASA

Alex Endert
Georgia Institute of Technology

David Etim
NNSA

Rebecca Faust
Virginia Tech

Eliot Feibush
Princeton Plasma Physics Laboratory

Shaw Feng
National Institute of Standards and Technology

Nicola Ferrier
Argonne National Laboratory

Hal Finkel
DOE/SC/ASCR

Kelly Gaither
University of Texas at Austin

Gerald Geernaert
DOE

Berk Geveci
Kitware, Inc.

Devarshi Ghoshal
Lawrence Berkeley National Laboratory

Maria Glenski
Pacific Northwest National Laboratory

Stuart Gluck
DOE Office of Science, AAAS Science
& Technology Policy Fellow

Andy Gotz
European Synchrotron Radiation Facility

Stephanie Granger
California Institute of Technology

Pascal Grosset
Los Alamos National Laboratory

Kenny Gruchalla
National Renewable Energy Laboratory

Qiang Guan
Kent State University

Dan Gunter
Lawrence Berkeley National Laboratory

Hanqi Guo
Argonne National Laboratory

Attila Gyulassy
University of Utah

Mengjiao Han
University of Utah

Kyle Harrington
Oak Ridge National Laboratory

Cyrus Harrison
Lawrence Livermore National Laboratory

Gerd Heber
The HDF Group

Barbara Helland
DOE/Advanced Scientific Computing Research

Chris Henze
NASA

Benjamin Hernandez
Oak Ridge National Laboratory

Gavin Hester
Purdue University

Justin Hnilo
UD DOE

Thuc Hoang
NNSA

Aaron Holder
DOE-BES

Anthony Hoogs
Kitware, Inc.

Susannah Howieson
DOE Office of Science

Brian Hu
Kitware, Inc.

Jian Huang
University of Tennessee

Joseph Insley
Argonne National Laboratory

Katherine Isaacs
University of Arizona

Amber Jenkins
NASA

Sumit Kumar Jha
University of Texas at San Antonio

Yi Jiang
Argonne National Laboratory

Graham Johnson
National Renewable Energy Laboratory

Chris Johnson
University of Utah

Maureen Johnson-Leon
University of Texas at Austin

Daniel Keefe
University of Minnesota

Shane Kelly
Washington State University

Jin Ryong Kim
University of Texas at Dallas

Janet Knowles
Argonne National Laboratory

Tzanio Kolev
Lawrence Livermore National laboratory

Helen-Nicole Kostis
NASA

Boris Kovalerchuk
Central Washington University

Hari Krishnan
Lawrence Berkeley National Laboratory

Olga Kuchar
Oak Ridge National Laboratory

Harriet Kung
DOE Office of Science

Dimitri Kusnezov
DOE

David Laidlaw
Brown University

Tsengdar Lee
NASA

Jason Leigh
University of Hawaii at Manoa

Joshua Levine
University of Arizona

Zhimin Li
University of Utah

Xin Liang
Missouri S&T

Peter Lindstrom
Lawrence Livermore National Laboratory

Zhengchun Liu
Argonne National laboratory

Shusen Liu
Lawrence Livermore National Laboratory

Mark Livingston
Naval Research Laboratory

Li-Ta Lo
Los Alamos National Laboratory

Kwan-Liu Ma
University of California, Davis

Ross Maciejewski
Arizona State University

Ramana Madupu
Office of Biological and Environmental Research (BER)

Valerio Mariani
SLAC National Accelerator Laboratory

Laurentiu Marinovici
Pacific Northwest National Laboratory

Thomas Marrinan
Argonne National Laboratory

Donnie Mason
Brookhaven National Laboratory

Victor Mateevitsi
Argonne National Laboratory

Alex May
Oak Ridge National Laboratory

Harshitha Menon
Lawrence Livermore National Laboratory

Haichao Miao
Lawrence Livermore National Laboratory

David Miller
NIH

Samantha Molnar
National Renewable Energy Laboratory

Kenneth Moreland
Oak Ridge National Laboratory

Nathan Morrical
University of Utah

Paul Navratil
University of Texas at Austin

Jon Nelson
NASA

Van Ngo
Oak Ridge National Laboratory

Cassandra Nickles
NASA

Chris North
Virginia Tech

Stan Nowak
Simon Fraser University

Benjamin Nuernberger
NASA

Patrick O'Leary
Kitware, Inc.

Leigh Orf
University of Wisconsin

Michael Papka
Argonne National Laboratory

Konstantinos Parasyris
Lawrence Livermore National Laboratory

Dilworth Parkinson
Lawrence Berkeley National Laboratory

Valerio Pascucci
University of Utah

Paolo Patelli
Los Alamos National Laboratory

Talita Perciano
Lawrence Berkeley National Laboratory

Lori Perkins
NASA

Tom Peterka
Argonne National Laboratory

Steve Petruzza
Utah State University

Kelly Pierce
University of Texas at Austin

Robinson Pino
Department of Energy

Kristi Potter
National Renewable Energy Laboratory

Martin Pratt
Pacific Northwest National Laboratory

David Pugmire
Oak Ridge National Laboratory

Jesus Pulido
Los Alamos National Laboratory

Lavanya Ramakrishnan
Lawrence Berkeley National Laboratory

Arvind Ramanathan
Argonne National Laboratory

Silvio Rizzi
Argonne National Laboratory

Carolina Roe-Raymond
Princeton University

David Rogers
Los Alamos National Laboratory

Jen Rogers
University of Utah

Oliver Ruebel
Lawrence Berkeley National Laboratory

Sonia Sachs
DOE ASCR

Francesca Samsel
University of Texas at Austin

Jibonananda Sanyal
Oak Ridge National Laboratory

Han-Wei Shen
Ohio State University

Benjamin Shu
Brookhaven National Laboratory

Claudio Silva
New York University

Horst D. Simon
Lawrence Berkeley National Laboratory

Yue-Ting Siu
San Francisco State University

Andrew Solis
University of Texas at Austin

Jessica Spayd
Johns Hopkins University

Bill Spotz
Office of Science/ASCR

Rohit Srivastava
Oak Ridge National Laboratory

David Stevens
Lawrence Livermore National Laboratory

Hendrik Strobelt
IBM Research

Simon Su
National Institute of Standards and Technology

Mark SubbaRao
NASA

Sreenivas Rangan Sukumar
Cray - HPE

Brian Summa
Tulane University

Melissa Swift
Pacific Northwest National Laboratory

Nur Afsa Syeda
Washington State University

Dingwen Tao
Washington State University

Aniket Tekawade
Argonne National Laboratory

Chris Tennant
Thomas Jefferson National Accelerator Facility

Jordan Thomas
DOE (AAAS STPF)

Melanie Tory
Northeastern University

Georgia Tourassi
Oak Ridge National Laboratory

Dani Ushizima
Lawrence Berkeley National Laboratory

Donald Vasco
Lawrence Berkeley National Laboratory

Carlos Verdoza
Pacific Northwest National Laboratory

Lucas Villa Real
IBM Research

Venkatram Vishwanath
Argonne National Laboratory

Svitlana Volkova
Pacific Northwest National Laboratory

Emily Wall
Emory University

Amy Walton
National Science Foundation

Qianwen Wang
Harvard University

Bei Wang Phillips
University of Utah

Gunther Weber
Lawrence Berkeley National Laboratory

Richard Witt
DOE Office of Science, Office of Nuclear Physics

Matthew Wolf
Oak Ridge National Laboratory

John Wu
Lawrence Berkeley National Laboratory

Wei Xu
Brookhaven National Laboratory

Haowen Xu
Oak Ridge National Laboratory

Yalong Yang
Virginia Tech

Anders Ynnerman
Linköping University

Ilmi Yoon
San Francisco State University

Chengzhu (Jill) Zhang
Lawrence Livermore National Laboratory

Eugene Zhang
Oregon State University

Yue Zhang
Oregon State University

Zhiming Zhang
Yahoo

Back to TOC

A.2 Agenda

The final agenda included two keynotes, two rounds of lightning talks, and breakout groups on nine subtopics led by two pre-appointed session chairs. The agenda below highlights blocks that were open to the public in green and blocks that were restricted to invited participants in orange.

Day 1: January 18, 2022

Time (EST)	Topic
12:00 - 12:15	Opening Remarks & Introduction - Margaret Lentz, DOE/ASCR
12:15 - 1:15	Keynote - Anders Ynnerman, Linkoping University, Sweden
1:15–1:25	Break
1:25–2:00	Lightning Talks • **Thomas Caswell**, Visualization of Structured Data • **Eugene Zhang**, Tensor Field Visualization: Challenges and Opportunities • **Kenneth Moreland**, The Importance of Scientific Visualization on Novel Hardware • **Nathan Morrical**, Leveraging Ray Tracing Coprocessors to Advance Visualization for Science • **Victor Mateevitsi,** Novel Display Technologies for Accelerating Scientific Discoveries • **Kenneth Moreland**, The Exploitation of Data Reduction for Visualization • **Soumya Dutta**, Model-Based Visual Analytics of Big Data: Challenges and Opportunities • **Maria Glenski,** Beyond Communication, Visualization for Hypothesis Generation (and Testing) • **Talita Perciano**, The Role of Complex Data Visualization Tools to Support Current and Future Machine Learning for Science Methods • **Brian Hu**, Operationalizing Explainable AI (XAI): Challenges and Opportunities for Saliency Maps • **Shusen Liu**, Visual Interpretation of Complex Systems Needs Causal Reasoning at Concept Level • **Oliver Ruebel**, The Role of FAIR Methods in Visualization Towards Trusted Decision Making • **Kelly Pierce**, Equity-Focused Data Context Documentation to Promote Ethical Data Reuse
2:00–2:15	Break

Back to TOC

2:15–3:15	**Breakout Groups A: Grand Challenge Problems** 1. Multivariate and Multimodal Data (R. Bujack, H.-W. Shen, session chairs) 2. Novel Technologies in Visualization (M. Papka, D. Keefe, session chairs) 3. Extreme-Scale Data (J. Ahrens, V. Pascucci, session chairs) 4. Interpretability of Complex Systems (M. Berger, B. Hu, session chairs) 5. Equity in Access to Science (K. Gaither, M. Tory, session chairs)
3:15–3:30	Break
3:30–4:30	**Breakout Groups A: Road to Solutions** 1. Multivariate and Multimodal Data (R. Bujack, H.-W. Shen, session chairs) 2. Novel Technologies in Visualization (M. Papka, D. Keefe, session chairs) 3. Extreme-Scale Data (J. Ahrens, V. Pascucci, session chairs) 4. Interpretability of Complex Systems (M. Berger, B. Hu, session chairs) 5. Equity in Access to Science (K. Gaither, M. Tory, session chairs)
4:30–5:00	Report out from breakouts

Day 2: January 19, 2022

Time (EST)	Topic
12:00 - 12:15	Opening Remarks and Logistics - Margaret Lentz, DOE/ASCR
12:15 - 1:15	Keynote - Jackie Chen, Sandia National Laboratory, United States
1:15–1:25	Break

[Back to TOC]

1:25–2:00	**Lightning Talks** • **Kristi Potter**, Actionable Uncertainty Visualization • **Hanqu Guo**, Intelligent Visual Analytics for Ensemble Simulations • **Nicola Ferrier**, Visualization for Scientific Applications of Edge Computing • **Lavanya Ramakrishnan**, Interactive Visualization in Scientific Workflows • **Berk Geveci**, Visualization Workflows for Temporal Data from Post Hoc, In Situ, Experimental and Observational Use Cases • **Valerio Mariani**, Remote Real-Time Data Visualization for Fast Decision-Making and Experiment Steering at Light Sources • **Laura Matzen**, Research Challenges for Visualizations Involving State Uncertainty • **Mark Livingston**, Measuring Correct Interpretation of Graphs, Figures, and Tables • **Roxana Bujack**, Vector Field Segmentation for Data Analysis • **Janine Bennett**, Quantum Information Technologies: A New Frontier for Visualization
2:00–2:15	Break
2:15–3:15	**Breakout Groups B: Grand Challenge Problems** 1. Uncertainty Visualization and Ensembles (C. Johnson, K. Potter, session chairs) 2. Visualization at the Edge (H. Krishnan, C. Silva, session chairs) 3. Human Factors and Usability (A. Endert, C. North, session chairs) 4. High-Dimensional Data (P.-T. Bremer, B. Wang, session chairs)
3:15–3:30	Break
3:30–4:30	**Breakout Groups B: Road to Solutions** 1. Uncertainty Visualization and Ensembles (C. Johnson, K. Potter, session chairs) 2. Visualization at the Edge (H. Krishnan, C. Silva, session chairs) 3. Human Factors and Usability (A. Endert, C. North, session chairs) 4. High-Dimensional Data (P.-T. Bremer, B. Wang, session chairs)
4:30–5:00	Report out from breakouts

A.3 Documentation

In a joint discussion during the last day of the workshop, the community suggested five top-level priority research directions to DOE:

1. Advancing Theory and Techniques for Visualization to Support the Analysis and Understanding of Complex Scientific Data
2. Introducing Interoperable and Adaptable Visualization to Support Diverse Scientific Workflows Across All Scales
3. Harnessing Technology Innovations to Accelerate Science through Visualization
4. Improving Equity in Accessing and Engaging with Scientific Data and Processes
5. Developing Intelligent Approaches for Adaptive, Context-Aware Visualization of Scientific Data and AI

Each direction included a list of open challenges, opportunities, and potential future impacts. Shortly after the workshop, the committee together with the breakout session chairs expanded this initial list into a two-page summary brochure [283], which forms the basis for this report.

Appendix B - Pre-Workshop Document

Visualization for Scientific Discovery, Decision-Making & Communication

orau.gov/ASCR_DataVisWS

Organizing Committee:

- Peer-Timo Bremer, Lawrence Livermore National Laboratory
- Georgia Tourassi, Oak Ridge National Laboratory
- Wes Bethel, Lawrence Berkeley National Laboratory
- Kelly Gaither, Texas Advanced Computing Center
- Valerio Pascucci, University of Utah
- Wei Xu, Brookhaven National Laboratory

DOE Points of Contact:

- Margaret Lentz, DOE, Advanced Scientific Computing Research
- Hal Finkel, DOE, Advanced Scientific Computing Research

B.1 Introduction

Visualization—the use of visual elements to explore data, form hypotheses, or convey conclusions—has always been an integral part of the scientific process. Starting from an initial exploration of new data to illustrating outcomes to the general public, visualization is one of the most intuitive and powerful ways of communication. This is especially true in the team based, cross-discipline environment of the many cutting edge, large scale projects funded by the Department of Energy (DOE). The Advanced Scientific Computing Research (ASCR) program in particular, has long supported visualization, a highly effective means of exploring data and communicating results. Intuitive visualizations represent a significant force multiplier, connecting scientists across domains, with their stakeholders, and ultimately to policy makers and the public writ large.

Despite significant efforts from many communities, visualization in practice often remains limited to a handful of common techniques; the majority of which are restricted to showing spatial distributions of individual variables or statistical summaries of more complex data. These limitations can be due to a lack of scaling of existing techniques, the lack of easily accessible tools, or that for various types of data there may not exist a straightforward visual encoding (e.g., high-dimensional and multimodal data).

The purpose of this workshop will be to bring together visualization experts from DOE and the broader stakeholder community to better understand: (a) the current state of the art; (b) identify future visualization needs of the scientific community and gaps in current capabilities; and (c) emerging technologies that will aid in visualization and communication. Furthermore, the workshop will deliberately expand the target audience and potential impact beyond prior efforts to include factors relevant to decision making and visualization including human factors, cognition, interpretation, and evaluation. Additionally, we are intentionally elevating discussion of science communication and interaction to discuss issues related to access, bias, inclusion, and usability.

The workshop will have four themes. The first focuses on visualizing data such as, non-scalar spatial data (multivariate data, spectra, tensors, etc.), data defined in high dimensions (phase spaces, probability distributions, parameter spaces, etc.), or abstract data types (nuclear cross-sections, graphs, facility data, etc.). The second theme concentrates on the role visualization plays in decision support for applications ranging from situational awareness and time sensitive applications to strategic planning. This includes interconnected thrusts on how to represent uncertainty, how to instill confidence in complex decision tools, and how to take human factors, such as, cognitive biases and limitations into account when designing and evaluating tools. The third theme focuses on the interplay between new technologies and visualization approaches. One aspect will be the opportunities for visualization to support and enhance emerging capabilities, such as, providing visualizations on edge devices like experimental diagnostics or distributed sensor networks. Another aspect will be how new technologies, such as virtual reality or machine learning, can open new frontiers in visualization research. The final theme recognizes the crucial role visualization can play in communicating ideas, results, and predictions to the general public, domain scientists and decision makers based on the knowledge created and curated by the DOE. Whether this means understanding climate predictions, mapping flood plains, or enabling citizen science in general, visualization can provide a crucial gateway to data, analysis, and decisions.

B.2 History

B.2.1 Visualization to Advance DOE's Science Mission

ASCR as well as the SciDAC (Scientific Discovery through Advanced Computing) program have a long-standing history of leading-edge visualization R&D in support of the DOE science mission. The SciDAC2 Visualization and Analytics Center for Enabling Technologies (VACET) [284] made petascale-capable, production quality visualization a reality on ASCR supercomputing science user facilities, and by extension, benefited the worldwide scientific community. VACET pursued advances in visualization software, such as enhancing the scalability of the the VisIt parallel visualization application [285] for accommodating extreme-scale scientific data exceeding a trillion mesh cells [286] or a trillion particles [287] in size, and application of new methods deployed in software infrastructure to key science problems, such as the feature-based analysis of combustion simulation output [288] or using visualization for comparative analysis of cosmology codes run with different parameters [289].

The SciDAC3 Scalable Data Management, Analysis, and Visualization (SDAV) Institute [290] consisted of three integrated focus areas for data management, analysis, and visualization. Like VACET, SDAV facilitated advances in a number of key software tools like scalable visualization applications (VisIt, ParaView), and libraries like VTK-m, a library for platform portable visualization algorithms and data structures [93], and DIY, a library for building block-based parallel applications [291]. These tools were applied to a number of different science application areas, such as visualizing the solar wind in plasma physics models [292], visualizing the output from high-resolution atmospheric model, CAM5, on 64K cores of ALCF's Intrepid supercomputer [293], and integrating geometric analysis methods into cosmological simulations [294].

In the SciDAC4/5 RAPIDS2 Institute for Computer Science, Data, and Artificial Intelligence, these themes continue with an emphasis on both software implementation of methods and their application to DOE mission science problems. In RAPIDS, there is a continued emphasis on in situ processing for analysis and visualization due to the widening gap between our ability to compute values and our ability to store them for later analysis (e.g., ParaView/Catalyst [295], VisIt/Libsim [296], SENSEI [297]). There are libraries for implementing data-parallel visualization (VTK-m [298]) and analysis (DIY [299]), building graph algorithms in the language of linear algebra (GraphBLAS [300]), parallel computation of Delaunay and Voronoi tessellations (Tess [300][301]). Some example applications include using in situ visualization and code coupling for CFD code design space optimization [302] and in situ analysis and visualization of fusion simulations [303].

B.2.2 Previous workshops that Involve Data Management, Analysis, and Visualization

There is a rich history of community workshops focusing on research challenges in visualization, data management, and data analysis. Some of these have focused on the research challenges resulting from an evolving computational landscape, characterized by a deepening memory and storage hierarchy combined with increasing concurrency and heterogeneity in hardware, and how the advances that benefit simulation science also will benefit data-intensive workloads [304]. Related issues include the need to rethink algorithm design as we have evolved from vector to MPP to now heterogeneous architectures characterized by increasing node- or chip-level concurrency [305][306]. Several point out the rapidly growing heterogeneity in the software ecosystem and the opportunities for leveraging advances and tools from industry and academia, particularly in data-intensive regimes like AI/ML [11][307].

In response to the well established trend of growth in volume and veracity of scientific data, some workshops have examined topics of adding processing at various stages in the storage hierarchy as a way to reduce I/O and also to enhance the value and usefulness of scientific data [12]. Others focused on reducing I/O loads by performing as much processing as possible while data is still in memory, an idea that has application to computational as well as experimental sciences [11][307][308].

Several workshops examine topics in the area of the convergence of computing and data, particularly in the regime of experimental and observational sciences, where advances in computational technology, such as use of digital twins run on remotely located HPC facilities, can help to optimize and control experiments in real time so as to produce better scientific data.

In this same vein, the data plays a central role in sustaining the R&D of data-centric methods, such as Artificial Intelligence (AI) and Machine Learning (ML) [309], where having a corpus of high-quality, curated reference data can advance many different fields, including visualization and analysis. In this problem space, the issue of data models and their impact on toolchains is particularly important for the design and implementation of multi-stage data-intensive processing pipelines and workflows [310].

Multiple reports point to challenges in the area of data management, analysis, and visualization where the onus of managing data and software used for analysis and visualization is the responsibility of the individual user, and how many programs would benefit from a more centralized approach to managing data and software R&D and deployment [311] [312]. One workshop, from an earlier period, was forward looking in this regard and highlighted the scientific and programmatic benefits of a large, coordinated data-centric effort on the scale of the SciDAC program to focus on data-related issues [313]. Noticeably absent from the list of prior efforts are workshops and reports centered on visualization, rather than on related topics, such as, data management, I/O, data reduction, etc. This workshop will address this gap by focusing primarily on visualization and including long neglected topics such as human cognition and science communication.

B.3 Scope

As discussed above, DOE has focused much of its research efforts in visualization on the support of the many scientific applications within the Office of Science. As part of the SciDAC program and the ASCR research portfolio this has led to many successful collaborations and has given rise to two visualization related Exascale Computing Project (ECP) projects tasked with providing the DOE community with reliable and scalable tools on future exascale platforms. However, the focus on (relatively) short term applications has also resulted in an emphasis on practical challenges in data management, in situ computations, and software integration and a comparative scarcity of fundamentally new visualization techniques and research directions. Over the last decades these trends have resulted in a sophisticated and scalable set of core tools and techniques addressing many day to day challenges. The goal of this workshop is to develop a research strategy aimed at expanding these core capabilities to include emerging applications, address existing gaps, and expand the reach of visualization beyond specific domain challenges. This will include a critical evaluation of past practices and new approaches to assess future advances.

The workshop comes at a time where data in all forms and shapes has become a ubiquitous and highly prized commodity, whether it comes from predictive simulations, scientific instruments, or sensor networks of any kind. Currently, only a small fraction of this data is amenable to straight forward visualization. Yet illustrations are typically the first step to understanding a new challenge, and are used for validation and communication throughout the lifetime of a project. Consequently, there exist tremendous opportunities in developing new techniques for currently underserved applications and challenging data types that can have an outsized impact on the overall outcome. Furthermore, the emergence of machine learning techniques, new technologies such as virtual and augmented reality, and the continued increase in processing power across all scales can open entirely new research directions and solution spaces. Finally, as an increasing number of decisions, scientific and otherwise, are based on data, both the importance of visualization and the size of its target audience has the potential to increase dramatically.

The workshop will collect feedback from the broader visualization community to identify research requirements and priorities for the next decade in developing new visualization capabilities that address and anticipate the entire breadth of DOE mission challenges from protecting critical infrastructure to accelerating scientific discovery across many domains such as climate science, material science, and high energy physics. The anticipated outcome is a collection of Priority Research Directions (PRDs) for DOE to support in order to create the foundation for a generation of highly impactful visualization tools and techniques.

B.4 Discussion Topics

In preliminary discussions the workshop committee has identified nine focus areas, grouped into four general themes that have been used to create the workshop agenda and recruit session leads for the respective breakouts. This section briefly recaps the different areas and provides a preliminary description of the respective technical challenges. However, this list should not be considered complete or final and refining topic areas and organizing them into meaningful research directions is one of the declared objectives of the workshop. Following a brief introduction this section will provide an overview of the current state of the art and open challenges to provide context and stimulate discussions.

B.4.1 Visualizing Complex Data

Typically, the most straight forward data to visualize are spatial distributions of single scalar values. One can either map the value to an additional visual channel, i.e., color and opacity which leads to volume rendering type visualizations [314], or extract subsets with specific values, which leads to isosurface type visualizations [315]. However, there exist a vast range of data for which such simple approaches do not apply. In particular, the data might defy a simple mapping. For example, what if there exist multiple values at each location or each value represents a more complex concept (e.g., a vector, tensor, image, spectrum). Additionally, data is now coming from a diverse range of sources, in a diverse range of formats and types, each of which may contain complex data representations that do not easily map to existing visual analysis methods. Significant challenges remain in this multimodal, multivariate space. Another axis of complexity is data defined in dimensions higher than three, for which there exist no simple spatial representation even for something as (apparently) simple as a high-dimensional scalar function. While both challenges are related and in fact often occur jointly, considering them as the first two topics provides a convenient split into finding new visual encodings for nonscalar data and representing domains beyond 3D physical space.

B.4.2 Multivariate & Multimodal Data

Scientific data are becoming increasingly complex. They may arise from different data sources (multimodel) or consist of different attributes (multivariate), often coupled with spatiotemporal elements. We formalize our setting of complex data, in terms of scalar fields, vector fields, tensor fields, and multi-fields. For example, a combustion simulation produces scalar fields that represent physical measurements such as temperature and pressure, as well as vector fields that model turbulent flows. A molecular dynamics simulation generates tensor fields capturing a material's strain and stress. The multi-scale and multi-physics nature of such simulations also gives rise to multi-fields, a collection of fields with different modalities on the same domain.

The study of multimode and multivariate data play an important role in advancing scientific understanding, from oceanography to astrophysics, chemistry to meteorology, and nuclear engineering to molecular dynamics. The analysis and visualization of such data is considered as one of the top challenges in scientific visualization [4]; see [316], [317], and [45] for surveys. Our ability to provide usable visualization and analysis methods and tools that compensate for the wide range of sources, modalities and content of data is key to advancing science. Integrating and interpreting diverse data coming from images and video, text, audio, sensor streams, instruments, and simulations is a significant research challenge that our community must address. Core challenges include representation, translation, alignment, fusion and co-learning or the ability to transfer knowledge between modalities [318].

Despite recent advances in the visualization of scalar fields [319], vector fields [320], and tensor fields [321], there are a number of research challenges surrounding the study of multivariate data, including data representation, comparison, integration, and fusion. Accounting for data that is both multimodal and multivariate mandates that we address challenges in this intersection space. For instance, how can multiple data attributes and their interrelation be encoded in a coherent representation? How robust is such a representation with respect to missing data and outliers? How can simulation results generated on different types of grids be intermixed using a common frame of reference? How can different attributes from different models be compared to each other? How can we design visualizations to minimize artifacts, during the rendering and registration of multiple data modalities? What are the next frontiers for interactive methods for multimodal and multivariate data?

B.4.3 High-Dimensional Data

With the ever-increasing amount of available computing resources, our ability to collect and generate a wide variety of large, complex, high-dimensional datasets continues to grow. High-dimensional datasets show up in numerous fields of study, such as economy, biology, chemistry, political science, astronomy, and physics, to name a few. Their wide availability, increasing size, and complexity continuously lead to new challenges and opportunities for their effective visualization. The physical limitations of the display devices and our visual system prevent the direct display and instantaneous recognition of structures with higher dimensions than two or three. In the past decades, a variety of approaches have been introduced to visually convey high-dimensional structural information by utilizing low-dimensional projections or abstractions: from dimensionality reduction to visual encoding, and from quantitative analysis to interactive exploration.

Many surveys have focused on different aspects of high-dimensional data visualization, such as parallel coordinates [66][322][323], quality measures [40], clutter reduction [41], visual data mining [324]–[326], and interactive techniques [327]. High-dimensional aspects of scientific data have also been investigated within the surveys [45] [328], while [316][329][330] focus on the various aspects of visual encoding techniques for multivariate data. The most recent survey [46] was also associated with an interactive website organizing most of the relevant literature [331].

Despite this impressive body of literature available, effective and intuitive graphical representation of high-dimensional data remains one of the most difficult open challenges in visualization. More research is needed in all aspects of this field, including interactive exploration of high-dimensional spaces, development of new visual metaphors, qualitative and quantitative representations, and domain-agnostic and domain-specific approaches.

B.4.2 Supporting Trusted Decision Making

The second theme of visualization techniques is focused on one the key use cases of visualization: providing decision makers the necessary insight and situational awareness to plan the next action, whether that is an immediate emergency response or a multi-year strategic plan. To enable trusted decision making requires three fundamental capabilities that define the three focus areas. First, is the ability to convey uncertainties in data, analysis, or predictions. By definition, difficult decisions accept multiple viable solutions and without proper care a single illustration might imply an unsupported confidence. This leads to the second focus area aimed at understanding the impacts of human peculiarities and shortcomings. Cognitive biases, limited visual perception for detail, and many other factors may result in a visualization being interpreted quite differently by different audiences and potentially very differently from its intended effect. Therefore, understanding the interplay between visualization design and its ultimate impact as well as how to properly evaluate the effect of a visualization are key in providing decision support. The final focus area acknowledges the fact that decisions are increasingly supported by and aimed at highly complex systems that are incomprehensible in their entirety. Developing techniques to interpret such systems is key to provide confidence and inspire trust in the results.

B.4.2.1 Uncertainty Visualization and Ensembles

Uncertainty visualization is a rapidly developing field with broad impacts in scientific discovery, science communication, and data-driven decision-making. While point estimates are a common quantity to consult in reasoning and decision-making, deviations from point estimates are troublesome. For example, trust in scientific content may erode as events contrary to a point estimate occur. Consumers of data-driven content are increasingly seeking information regarding how their point estimates may vary.

A recent survey by Kamal et al. [332] provides a comprehensive review of state-of-the-art uncertainty visualization approaches based on the data type and attributes they utilize along with their advantages and limitations. The authors also discuss popular evaluation methodologies for uncertainty visualization that are either empirical or theoretical. In a different survey by Padilla et al. [333], the authors suggest two broad categories of uncertainty visualization: graphical annotations, which may show properties of a distribution, such as the mean or confidence intervals, and visual encodings, which seek to evoke a sensation of uncertainty and do so by using visual channels, mapping uncertainty properties to visual characteristics, such as size, alpha or arrangement. With ensemble approaches becoming more widely adopted, there are many visualization efforts to capture the uncertainty of ensemble data. For example, accurately interpreting complex ensemble data plays a crucial role in high-risk decision-making situations or for understanding complex physical phenomena. Wang et al. [334] provided a comprehensive review of ensemble visualization techniques and a taxonomy covering multiple facets and dimensions of the ensemble data, including visualization of ensemble uncertainty.

While uncertainty visualization has seen large gains over the past few years, there are still valuable research directions. As extreme scale data, ensemble data and machine learning become increasingly popular in the DOE mission space and beyond, uncertainty visualization becomes even more challenging. Past work suggests the need for application-specific techniques and robust ways to test how uncertainty visualization methods may help their target user complete tasks of interest [66]. The breakout session will discuss challenges and opportunities on the topic to identify novel research directions with broad impact to DOE mission space.

B.4.2.2 Human Factors and Usability

Human-computer interaction (HCI), particularly in scientific visualization, is a rapidly growing field, driven not only by increases in data availability and size but also by a wide array of applications and modalities in which visualization can make an impact. Recent advances in our understanding of perceptual learning and reasoning processes are crucial in developing impactful visualization tools that adapt to the evolving needs of the target users. Furthermore, as data and visualization become more embedded in systemic decision-making processes, it is important that visualization tools which stakeholders use are properly validated. There exist valuable opportunities to develop human-centric design and evaluation guidelines, particularly those which may elevate minority viewpoints.

Three recent studies highlight that emphasis on human factors and usability are as critical as technological innovation for broad and lasting impact. Offenwanger et. al. [335] suggest that bias in the gender representation of HCI study participants can call into question a study's generalizability, and provide evidence of underrepresentation of women in HCI studies. Bergstrom et. al. [336] identify that few virtual reality (VR) studies establish guidelines on how to conduct user studies, and standards developed for evaluating 2D interaction may not apply. They propose recommendations and provide a checklist for future VR studies. Mamykina et. al. [337] describe two case studies of research projects that attempt to scale HCI research beyond small evaluation studies. They propose four design considerations for large-scale studies: designing for longevity, diversity, adoption, and abandonment. They suggest great importance on implementation and deployment, and building long-term relationships with user communities.

As the DOE scientific community experiences explosive growth in data volumes and dimensions, new human-centered visualization design frameworks and validation protocols are needed to capture context that goes beyond traditional user characteristics and account for cognitive and perceptual factors, domain expertise and experience, and factors related to the task at hand. There is plenty of evidence from medical imaging that one-size-does-not fit all when it comes to visualization tools, small scale usability studies in artificial settings do not translate well in the general user population, and universally optimized designs often have detrimental effects to individual users. How can future technologies effectively leverage user interactions during visual explorations to further inform the user model on the user's expertise, experience, and specific perceptual and cognitive needs?

B.4.2.3 Interpretability of Complex Systems

The rapid adoption and evolution of artificial intelligence/ machine learning (AI/ML) technologies in scientific research and automation has revolutionarily enlarged the spectrum of methodologies for scientific discovery. However, due to the high non-linearity and complexity of AI models, more widespread adoption is hindered by a lack of understanding of these black-box models in their internal working mechanisms and decision-making process. Explainable AI (XAI) has emerged to provide interpretation, increase transparency and establish trust targeting in all stages of AI model lifecycle, including model development, training, validation, and the practical use. This technique is essential for scientific users to have AI enabled solutions to deal with the increased volume and complexity of scientific data and problems and significantly increase the automation and reduce the time and cost of experiments.

XAI has grown exponentially for over a half decade since the seminal work introduced in 2014. Both AI and visual analytics (VA) communities have contributed to the state-of-the-art XAI with different preferences. The AI community aims to develop alternative interpretable models or use local/global explanations to explain one type of model behaviors, e.g., highlighting relevant input features to a model's decision [338]–[340]. Instead, VA community excels at building an interactive visualization system that presents and connects multiple network components or learned features and supports drill-down study through user interactions to understand a specific network type or application [47][341][342]. There are also evaluation works to discuss the effectiveness, robustness, sensitivity and satisfaction of the XAI methods [338][343][344].

There are a few major challenges affecting the adoption of XAI in science. First, scientific data has its unique complexity and heterogeneity compared to natural images that most XAI methods are designed and demonstrated with. Potential data biases, theoretical approximations, a mixture of multi-modal and multivariate attributes associated with data exist that must be addressed before feeding them to the models. Second, domain scientists are with distinguished knowledge to guide throughout model development, evaluation and actual use. A pathway of direct knowledge injection and human intervention is yet to establish beyond current focus on either presenting low-level computational results or generating high-level explanations. Third, thorough evaluation and utility of XAI methods in scientific applications are critical but insufficiently investigated that consider the scenarios with different targeted user groups (e.g., AI developers, domain experts), explanation tasks, and the correct code of conduct to adopt these methods.

The variety of modern scientific domains and experiment and simulation data of DOE national laboratories present a unique environment for this multidisciplinary and collaborative research, which will advance both ML and visualization techniques in industry and academia.

B.4.3 New Technological Frontiers

The third theme aims to look forward, to provide context for research and development that will be necessary to enable next-generation decision makers that are tasked with illuminating ever larger grand challenge science problems. Key problems must be addressed to facilitate scientific understanding at the largest scales, tackling problems that require multidisciplinary teams, deep expertise in multiple disciplines and infrastructure capable of responding to these problems as science evolves and actionable decision making is an imperative. We introduce these in three topic areas, the first of which is data and technology at the edge which provides an overview of modern data acquisition and science driven analysis. The second topic area discusses novel technologies for visualization including novel interaction techniques, novel display modalities and the evolution of novel decision making ecosystems. Our final focus area provides an overview of the issues that must be addressed with extreme scale data.

B.4.3.1 Data and Technology at the Edge

Our world is becoming increasingly connected with evermore need for near real time data coming from instruments, sensors, networks, and wearables, etc. Furthermore, data-driven science has become a legitimate fourth pillar of science, and has in fact provided the underpinnings for the other three pillars: theory, experimentation, and computation. While data is being generated at a faster and faster rate, resulting in ever larger stores, it is of little use without the ability to translate this data into information, insight, and actionable decisions [345].

As a scientific community, particularly at scale, we are familiar with a centralized concept of data management and analysis that moves data close to computation. However, we are entering an age in which more diverse types of data are being collected, curated and made public in the interest of furthering research and discovery. Coalescing heterogeneous, geographically distributed digital assets for actionable decision making became a critical need during the COVID-19 pandemic. While the primary mechanism for access was through crude data repository access systems like GitHub, the availability of these disparate datasets gave rise to a large body of data-driven research, demonstrating the value of integrating these digital assets into the models that drive prediction and subsequent decision-making [346].

Cross-platform mechanisms for data access and analysis are now available that account for scale, heterogeneity, provenance, and the need for blending or linking with other digital assets [347][348]. While easy access to high quality, trusted data is a necessity, the ability to analyze these assets in the context of the given science is just as, if not more important. Statistical analysis and machine learning are critical tools for evidenced based decision making, informed by forecasting and predictions [349]. Furthermore, interpretability of these results is a necessary step towards actionable decision making in all scientific disciplines [349]–[353].

B.4.3.2 Novel Technology for Visualization

With respect to technology, change is the only constant. Display systems have evolved dramatically over the past decade. Resolution has increased; form-factors have diversified; all of which have given rise to a re-imagining of displays in decision making environments [354][355]. Virtual reality and augmented reality have evolved from purely experimental and gaming platforms to being put into practice in educational settings, medicine, behavioral health, and command and control environments [356]–[359]. One of the key enablers is the ability to easily and to the extent possible, naturally interact with digital assets and collaborators, whether remote or in-person.

Significant research has been conducted in the novel interaction space, providing the basis for future work [360]. Interacting with and controlling 3D widgets has been largely under explored, particularly in the context of aiding understanding [361][362]. Natural language processing (NLP) and toolkits like NLTK [363], Stanford CoreNLP [364], and NER [365] have been used to aid developers to perform tasks using speech. However, Natural language interfaces (NLIs) for visualization have emerged to allow users to interact with data via data-related queries that generate visualizations [366]–[369]. Additionally, toolkits are now available that reduce the barrier for building natural language interface systems for visualization and data analysis [370][371].

As these novel technologies mature, researchers are investigating novel ecosystems comprising a multitude of display and interaction modalities. Display ecologies that link spatially aware heterogeneous displays have been shown to effectively enable users to search for, organize and synthesize information, effectively acting as a single display environment from the user's perspective [372]. Combining multiple displays, novel interaction mechanisms, and augmented and virtual reality is a growing area in which research has been conducted in the feasibility of these ecosystems for broad use [373][374] and the usability of these ecosystems for visual analysis and decision making [375]–[377].

B.4.3.3 Extreme Scale Data

The Department of Energy (DOE) Office of Science (SC) operates dozens of national science user facilities (SUFs) that span many disciplines [378]. These facilities include accelerators, colliders, supercomputers, light sources, and neutron sources, as well as facilities for studying the nanoworld, genomes, the environment, the atmosphere, and the cosmos. Each of these facilities generates vast amounts of scientific data, and thanks to advances in technology, the size, rate, and complexity of this data is rapidly increasing.

Simulation, experiment, and observation are producing data at unprecedented rates and quantities, to the point that supercomputers are needed to analyze the data and draw meaningful scientific conclusions. On the experimental science side, the drivers of increasing data volume and velocity are a combination of factors. One is the increase in the resolution and readback rates of the instrument sensors and detectors. Readback rates for the sensors at LCLS-II are expected to increase by 4 orders of magnitude in the period between 2016 and 2025 [379]. Over the same period of time, annual data volume from the CMS experiment at LHC is expected to increase about 3 orders of magnitude from ~5 PB to ~197 PB [379][380]. On the computational science side, increase in data volume and velocity result from an increase in computational capacity at SUFs as the underlying systems continue to evolve and grow in concurrency, heterogeneity, and complexity. Science drivers include needing to run codes in minutes rather than weeks for effective experiment planning or real-time optimization, among others [381].

Simulated datasets produced in cosmology, earth system science, and fusion research are large and complex and require post-processing to find important effects like extreme weather events, the formation of structure in the early universe, and key plasma instabilities that affect fusion reactor performance. On the experimental/ observational side, the volume of data produced is rising rapidly. For example, the current LHC output is on the order of 10 PB/year and is expected to rise to more than 150 PB/year in less than a decade, with a need to store exabytes' worth of data permanently. Projects like CMB-S4 (cosmic microwave background) and the Large Synoptic Survey Telescope (LSST) (optical) will gather tens to hundreds of PBs' worth of data [381]. From summaries of projected growth rates across the SUFs, we see a near future where individual facilities, of which there are dozens, are each generating collections of data in the range of tens to hundreds of petabytes per year. These projections suggest, when integrating across the entire program, that these science user facilities will be soon collectively acquiring exabytes of data per year. Affordable data storage, effective data access, distribution, and curation, and meaningful analysis are key challenges that these facilities increasingly face [382].

Beyond data volume and velocity, other trends include evolution in how scientific research is conducted that present challenges to established patterns. There is an increasing convergence of computing and data, where computational methods like machine learning are brought to bear on datasets like results of previous experiments so as to improve the accuracy or usefulness of a new experiment [382], and visual data exploration and analysis often plays a key role in this process [140]. Science projects are increasingly diverse in terms of geographic distribution of participants and resources. An increasingly common motif is where data is collected at an instrument where it undergoes initial processing, then is moved over the network to a central facility for additional processing and storage, then made accessible to a broad community of collaborators [382].

Another ongoing trend is the increasing heterogeneity in the software landscape. While applications like VisIt and ParaView are staples for HPC-based uses, there is a rapid growth in useful tools from academia, industry, and consortia, particularly those distributed as Python packages like Matplotlib [383]. There is a long-standing interest in being able to leverage software from diverse sources in scientific HPC applications, including those for visualization, analysis, data management, and learning [140][382].

DOE has funded R&D in several topical areas that are motivated by these trends. Scalable visualization applications like VisIt [285] and ParaView [384] routinely run at high concurrency on modern supercomputing platforms, and libraries like VTK-m provide the ability for developers to create platform-portable, custom visualization applications that run in shared-memory parallel fashion on a number of different modern architectures [93][384].

A useful approach for visualizing and analyzing very large data is to focus processing to a subset of data on the premise that for any given scientific inquiry, the portion of the dataset contributing to the answer is quite small compared to the size of the full-resolution data. The concept of query-driven visualization refers to focusing processing on a subset of data that is scientifically meaningful as expressed through a compound set of multidimensional- or multivariate-range queries [385][386]. A different approach for subset selection is to focus processing on features of interest, where features may be topological [387], geometric [388], or statistical [388][389].

Others have pursued approaches that involve compression and multi-resolution representations of data. Error-bounded lossy compression has been identified as one solution and has been tested for many use-cases: reducing streaming intensity (instruments), reducing storage and memory footprints, accelerating computation and accelerating data access and transfer [390]. Different approaches include SZ [391], orthogonal (or not) block transforms like ZFP [24], those based on multiresolution and hierarchical basis functions like wavelet transforms [392], singular value decomposition [393], and multi-level approaches [394] .

Finally, another approach for working with very large data is to do analysis or visualization on data while it is still resident in memory as it is generated. This pattern is known as *in situ* processing, and the idea is to avoid the widening gap between our ability to generate data and our ability to write it to storage for later post hoc analysis or visualization [140] [395]. There is a diversity in the ways this processing motif may be implemented: in some cases, data moves from one set of ranks to another for processing, while in other cases the data does not move and is processed in place. A recent community document enumerates the taxonomy and terminology for these different configurations and modalities [107]. Such an approach has been shown to significantly reduce I/O demands by computing data extracts, such as rendered images from visualization or data subsets, from full spatiotemporal resolution data [396]. Since both data reduction and in situ processing were recently the focus of dedicated workshops able to cover them much more extensively they are considered out of scope for the current visualization focused effort.

B.4.4 Equity in Access to Science

There is general agreement that visualization plays a critical role in the development and dissemination in many scientific disciplines. The modern age of information and data availability provides the promise for solutions to larger, more impactful problems. However, significant disparities in access to data, computation, and research exist. The visualization community has an opportunity and a responsibility to engage in research that specifically addresses issues contributing to barriers to or lack of access to visual representations, analytical reasoning and decision making methods and tools. In addition to availability, access refers to issues related to bias, inclusion, and the general consideration of the person or persons in the visualization process.

Taking advantage of the highest bandwidth channel in the human brain, visualization provides a powerful means for synthesizing data and information and communicating complex phenomena to diverse populations. For visualization to become truly inclusive, we must grow our body of visualization research that adopts a “do no harm” mantra. This necessarily dictates that research must be done to measure and prevent introducing bias in both data and visualization, to reflect equity awareness for race, ethnicity, gender, and disability, and to include cultural competency. While the visualization community has researched cognitive bias during the visualization process [397]–[399], there has been little work for strategies to mitigate bias [400], an area that has potential for considerable growth. As part of the Visualization for Communication workshop at IEEE Visualization 2020, work was presented that identified eight areas in which researchers and analysts could be more inclusive with their visualizations [401].

There is a growing body of work investigating personalized visualizations, initially seeking to investigate which aspects of a visualization are of interest, what representations are more meaningful, and to what extent systems can automatically recommend visual representations based on an individual’s preferences [255][256]. Recently, a research agenda has been proposed that addresses the need to consider individual differences in the design process. Moreover, evaluation research needs to be expanded to further explore whether there is a signal that suggests personality is encoded in a user’s response during interaction, reasoning and decision-making [402].

For visualization to be truly accessible by the masses, methods, tools and systems must be ubiquitous. Usability and interpretability are key to adoption and there is a large body of literature dedicated to the study and application of visualization in diverse domains, including works similar to [403]–[408]. While visual analytics emerged over two decades ago in response to the growing need to examine the intersection of interactive visualization, computational analysis and analytical reasoning [409], there is potential for growth to examine the computational analysis space, and to a greater degree, analytical reasoning. Including research from other, more mature fields studying decision making, and developing evaluation methodologies that measure and validate decision making as a part of the visualization process is key to adoption and usability from diverse communities [409][410]. Substantial advances in this space will allow visualization to play a pivotal role in *democratizing access to science* outside the traditional boundaries of highly specialized audiences.

Appendix C - Acronyms and Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
ACORN	Adaptive coordinate networks for neural scene representation
ADIOS	Adaptable Input Output System
AI	Artificial intelligence
AMM	Adaptive multilinear meshes
API	Application programming interface
AR	Augmented reality
ASCAC	Advanced Scientific Computing Advisory Committee
ASCI	Accelerated Strategic Computing Initiative
ASCR	Office of Advanced Scientific Computing Research
AWS	Amazon Web Services
CPU	Central processing unit
CSCW	Computer-supported cooperative work
DAV	Data analysis and visualization
DOE	Department of Energy
EOD	Experimental and observational data
FAIR	Findable, Accessible, Interoperable, and Reproducible.
FPGA	Field-programmable gate array
GPU	Graphics processing unit
GUI	Graphical user interface
HCI	Human–computer interaction
HDF5	Hierarchical data format version 5

HPC	High performance computing
I/O	Input/output
IaaS	IInfrastructure as a service.
LLNL	Lawrence Livermore National Laboratory
MFEM	Modular Finite Element Methods software
ML	Machine learning
MR	Mixed reality
NASA	National Aeronautics and Space Administration
NCEP	National Centers for Environmental Prediction
NeRF	Neural radiance fields
netCDF	network Common Data Form
NOAA	National Oceanic and Atmospheric Administration
ONNX	Open Neural Network Exchange
ORNL	Oak Ridge National Laboratory
PRD	Priority research direction
RL	Reinforcement learning
SaaS	Software as a service
SciDAC	Office of Science Scientific Discovery through Advanced Computing
SOA	Service-oriented architecture
TPU	Tensor processing unit
UC	University of California
UX	User experience
VaaS	Visualization as a service
VACET	Visualization and Analytics Center for Enabling Technologies
VCI	Visualization Cloud Instances

VMD	Visual molecular dynamics
VR	Virtual reality
VTK	Visualization Toolkit
XAI	eXplainable artificial intelligence
XDMF	eXtensible Data Model and Format.
XR	eXtended reality

Appendix D - Bibliography

[1] O. Benesch, Leonardo Da Vinci and the Beginning of Scientific Drawing. 1943.

[2] B. H. McCormick, “Visualization in scientific computing,” SIGBIO newsl., vol. 10, no. 1, pp. 15–21, Mar. 1988.

[3] L. J. Rosenblum, “Research issues in scientific visualization,” IEEE Comput. Graph. Appl., vol. 14, no. 2, pp. 61–63, Mar. 1994.

[4] C. Johnson, “Top scientific visualization research problems,” IEEE Comput. Graph. Appl., vol. 24, no. 4, pp. 13–17, Jul. 2004.

[5] A. R. Larzelere II, “The History of the Accelerated Strategic Computing Initiative (ASCI).” Accessed: Feb. 17, 2022. [Online]. Available: https://asc.llnl.gov/sites/asc/files/2020-09/DeliveringInsightASCI_0.pdf

[6] E. J. Nathan Galli, “The Visualization and Analytics Center for Enabling Technologies (VACET).” www.vacet.org (accessed Feb. 17, 2022).

[7] D. Silver, “SciDAC Institute for Ultra-Scale Visualization: Activity Recognition for Ultra-Scale Visualization.” 2014. doi: 10.2172/1162059.

[8] “RAPIDS2.” https://rapids.lbl.gov/ (accessed Feb. 17, 2022).

[9] T. Peterka et al., “ASCR workshop on in situ data management: Enabling scientific discovery from diverse data sources,” USDOE Office of Science (SC) (United States), Feb. 2019. doi: 10.2172/1493245.

[10] E. W. Bethel, E. Wes Bethel, and M. G. Eds, “Report of the DOE Workshop on Management, Analysis, and Visualization of Experimental and Observational data – The Convergence of Data and Computing.” 2016. doi: 10.2172/1525145.

[11] R. Gerber et al., “Crosscut report: Exascale requirements reviews, march 9–10, 2017 – Tysons Corner, Virginia. An office of science review sponsored by: Advanced scientific computing research, basic energy sciences, biological and environmental research, fusion energy sciences, high energy physics, nuclear physics,” Office of Scientific and Technical Information (OSTI), Jan. 2018. doi: 10.2172/1417653.

[12] R. Ross et al., “Storage systems and input/output: Organizing, storing, and accessing data for scientific discovery. Report for the DOE ASCR workshop on storage systems and I/O. [full workshop report],” USDOE Office of Science (SC) (United States), Sep. 2018. doi: 10.2172/1491994.

[13] T. Peterka, D. Morozov, and C. Phillips, “High-Performance Computation of Distributed-Memory Parallel 3D Voronoi and Delaunay Tessellation,” SC14: International Conference for High Performance Computing, Networking, Storage and Analysis. 2014. doi: 10.1109/sc.2014.86.

[14] J. Woodring, J. Ahrens, J. Figg, J. Wendelberger, S. Habib, and K. Heitmann, “In-situ Sampling of a Large-Scale Particle Simulation for Interactive Visualization and Analysis,” Computer Graphics Forum, vol. 30, no. 3. pp. 1151–1160, 2011. doi: 10.1111/j.1467-8659.2011.01964.x.

[15] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, “Implicit Neural Representations with Periodic Activation Functions,” Adv. Neural Inf. Process. Syst., vol. 33, pp. 7462–7473, 2020.

[16] “Website.” https://doi.org/10.48550/arXiv.2105.02788

[17] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis,” Computer Vision – ECCV 2020. pp. 405–421, 2020. doi: 10.1007/978-3-030-58452-8_24.

[18] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, “Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields,” 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 2021. doi: 10.1109/iccv48922.2021.00580.

[19] “Website.” https://doi.org/10.48550/arXiv.2111.11426

[20] M. Soler, M. Plainchault, B. Conche, and J. Tierny, “Topologically Controlled Lossy Compression,”

2018 IEEE Pacific Visualization Symposium (PacificVis). 2018. doi: 10.1109/pacificvis.2018.00015.

[21] X. Liang et al., “Toward Feature-Preserving 2D and 3D Vector Field Compression,” 2020 IEEE Pacific Visualization Symposium (PacificVis). 2020. doi: 10.1109/pacificvis48177.2020.6431.

[22] M. Ainsworth, O. Tugluk, B. Whitney, and S. Klasky, “Multilevel Techniques for Compression and Reduction of Scientific Data-Quantitative Control of Accuracy in Derived Quantities,” SIAM Journal on Scientific Computing, vol. 41, no. 4. pp. A2146–A2171, 2019. doi: 10.1137/18m1208885.

[23] S. Di and F. Cappello, “Fast Error-Bounded Lossy HPC Data Compression with SZ,” 2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS). 2016. doi: 10.1109/ipdps.2016.11.

[24] P. Lindstrom, “Fixed-Rate Compressed Floating-Point Arrays,” IEEE Trans. Vis. Comput. Graph., vol. 20, no. 12, pp. 2674–2683, Dec. 2014.

[25] R. Ballester-Ripoll, P. Lindstrom, and R. Pajarola, “TTHRESH: Tensor Compression for Multidimensional Visual Data,” IEEE Trans. Vis. Comput. Graph., vol. 26, no. 9, pp. 2891–2903, Sep. 2020.

[26] S. Li, N. Marsaglia, C. Garth, J. Woodring, J. Clyne, and H. Childs, “Data Reduction Techniques for Simulation, Visualization and Data Analysis,” Computer Graphics Forum, vol. 37, no. 6. pp. 422–447, 2018. doi: 10.1111/cgf.13336.

[27] H. Bhatia, D. Hoang, N. Morrical, V. Pascucci, P.-T. Bremer, and P. Lindstrom, “AMM: Adaptive Multilinear Meshes,” IEEE Trans. Vis. Comput. Graph., vol. 28, no. 6, pp. 2350–2363, Jun. 2022.

[28] L. J. Gosink, C. Garth, J. C. Anderson, E. W. Bethel, and K. I. Joy, “An Application of Multivariate Statistical Analysis for Query-Driven Visualization,” IEEE Transactions on Visualization and Computer Graphics, vol. 17, no. 3. pp. 264–275, 2011. doi: 10.1109/tvcg.2010.80.

[29] S. Hazarika, A. Biswas, and H.-W. Shen, “Uncertainty Visualization Using Copula-Based Analysis in Mixed Distribution Models,” IEEE Trans. Vis. Comput. Graph., vol. 24, no. 1, pp. 934–943, Jan. 2018.

[30] X. He, Y. Tao, Q. Wang, and H. Lin, “Biclusters Based Visual Exploration of Multivariate Scientific Data,” 2018 IEEE Scientific Visualization Conference (SciVis). 2018. doi: 10.1109/scivis.2018.8823605.

[31] X. He, Y. Tao, Q. Wang, and H. Lin, “Multivariate spatial data visualization: a survey,” Journal of Visualization, vol. 22, no. 5. pp. 897–912, 2019. doi: 10.1007/s12650-019-00584-3.

[32] D. Schneider, A. Wiebel, H. Carr, M. Hlawitschka, and G. Scheuermann, “Interactive comparison of scalar fields based on largest contours with applications to flow visualization,” IEEE Trans. Vis. Comput. Graph., vol. 14, no. 6, pp. 1475–1482, Nov-Dec 2008.

[33] H. Carr and D. Duke, “Joint Contour Nets,” IEEE Trans. Vis. Comput. Graph., vol. 20, no. 8, pp. 1100–1113, Aug. 2014.

[34] H. Edelsbrunner, J. Harer, V. Natarajan, and V. Pascucci, “Local and global comparison of continuous functions,” IEEE Visualization 2004. doi: 10.1109/visual.2004.68.

[35] H. Bhatia, B. Wang, G. Norgard, V. Pascucci, and P.-T. Bremer, “Local, smooth, and consistent Jacobi set simplification,” Computational Geometry, vol. 48, no. 4. pp. 311–332, 2015. doi: 10.1016/j.comgeo.2014.10.009.

[36] P. Klacansky, J. Tierny, H. Carr, and Zhao Geng, “Fast and Exact Fiber Surfaces for Tetrahedral Meshes,” IEEE Trans. Vis. Comput. Graph., vol. 23, no. 7, pp. 1782–1795, Jul. 2017.

[37] J. Tierny and H. Carr, “Jacobi Fiber Surfaces for Bivariate Reeb Space Computation,” IEEE Trans. Vis. Comput. Graph., vol. 23, no. 1, pp. 960–969, Jan. 2017.

[38] A. Inselberg, Parallel Coordinates: Visual Multidimensional Geometry and Its Applications. Springer, 2017.

[39] J. Heinrich and D. Weiskopf, “State of the art of parallel coordinates.” The Eurographics Association, 2012. doi: 10.2312/conf/EG2013/stars/095-116.

[40] E. Bertini, A. Tatu, and D. Keim, “Quality metrics in high-dimensional data visualization: an overview and systematization,” IEEE Trans. Vis. Comput. Graph., vol. 17, no. 12, pp. 2203–2212, Dec. 2011.

[41] G. Ellis and A. Dix, "A taxonomy of clutter reduction for information visualisation," IEEE Trans. Vis. Comput. Graph., vol. 13, no. 6, pp. 1216–1223, Nov. 2007.

[42] D. A. Keim, "Information visualization and visual data mining," IEEE Transactions on Visualization and Computer Graphics, vol. 8, no. 1. pp. 1–8, 2002. doi: 10.1109/2945.981847.

[43] M. C. F. de Oliveira, M. C. F. de Oliveira, and H. Levkowitz, "From visual data exploration to visual data mining: A survey," IEEE Transactions on Visualization and Computer Graphics, vol. 9, no. 3. pp. 378–394, 2003. doi: 10.1109/tvcg.2003.1207445.

[44] A. Buja, D. Cook, and D. F. Swayne, "Interactive High-Dimensional Data Visualization," Journal of Computational and Graphical Statistics, vol. 5, no. 1. p. 78, 1996. doi: 10.2307/1390754.

[45] J. Kehrer and H. Hauser, "Visualization and visual analysis of multifaceted scientific data: a survey," IEEE Trans. Vis. Comput. Graph., vol. 19, no. 3, pp. 495–513, Mar. 2013.

[46] S. Liu, D. Maljovec, B. Wang, P.-T. Bremer, and V. Pascucci, "Visualizing High-Dimensional Data: Advances in the Past Decade," IEEE Trans. Vis. Comput. Graph., vol. 23, no. 3, pp. 1249–1268, Mar. 2017.

[47] F. M. Hohman, M. Kahng, R. Pienta, and D. H. Chau, "Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers," IEEE Trans. Vis. Comput. Graph., Jun. 2018, doi: 10.1109/TVCG.2018.2843369.

[48] J. Wang, S. Hazarika, C. Li, and H.-W. Shen, "Visualization and Visual Analysis of Ensemble Data: A Survey," IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 9. pp. 2853–2872, 2019. doi: 10.1109/tvcg.2018.2853721.

[49] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," Neurocomputing, vol. 312. pp. 135–153, 2018. doi: 10.1016/j.neucom.2018.05.083.

[50] R. T. Whitaker, M. Mirzargar, and R. M. Kirby, "Contour boxplots: a method for characterizing uncertainty in feature sets from simulation ensembles," IEEE Trans. Vis. Comput. Graph., vol. 19, no. 12, pp. 2713–2722, Dec. 2013.

[51] M. Sedlmair, C. Heinzl, S. Bruckner, H. Piringer, and T. Möller, "Visual Parameter Space Analysis: A Conceptual Framework," IEEE Trans. Vis. Comput. Graph., vol. 20, no. 12, pp. 2161–2170, Dec. 2014.

[52] W. He et al., "InSituNet: Deep Image Synthesis for Parameter Space Exploration of Ensemble Simulations," IEEE Trans. Vis. Comput. Graph., vol. 26, no. 1, pp. 23–33, Jan. 2020.

[53] S. Liu, J. A. Levine, P.-T. Bremer, and V. Pascucci, "Gaussian mixture model based volume visualization," in IEEE Symposium on Large Data Analysis and Visualization (LDAV), Oct. 2012, pp. 73–77.

[54] L. Gosink et al., "Characterizing and visualizing predictive uncertainty in numerical ensembles through Bayesian model averaging," IEEE Trans. Vis. Comput. Graph., vol. 19, no. 12, pp. 2703–2712, Dec. 2013.

[55] W. He, X. Liu, H.-W. Shen, S. M. Collis, and J. J. Helmus, "Range likelihood tree: A compact and effective representation for visual exploration of uncertain data sets," in 2017 IEEE Pacific Visualization Symposium (PacificVis), Apr. 2017, pp. 151–160.

[56] "UFLOW: visualizing uncertainty in fluid flow." https://doi.org/10.1109/VISUAL.1996.568116 (accessed Apr. 13, 2022).

[57] C. R. Johnson and A. R. Sanderson, "A Next Step: Visualizing Errors and Uncertainty," IEEE Computer Graphics and Applications, vol. 23, no. 5. pp. 6–10, 2003. doi: 10.1109/mcg.2003.1231171.

[58] K. Potter, P. Rosen, and C. R. Johnson, "From Quantification to Visualization: A Taxonomy of Uncertainty Visualization Approaches," IFIP Adv. Inf. Commun. Technol., vol. 377, pp. 226–249, 2012.

[59] A. Kamal et al., "Recent advances and challenges in uncertainty visualization: a survey," Journal of Visualization, vol. 24, no. 5. pp. 861–890, 2021. doi: 10.1007/s12650-021-00755-1.

[60] D. Weiskopf, "Uncertainty Visualization: Concepts, Methods, and Applications in Biological Data Visualization," Frontiers in Bioinformatics, vol. 2. 2022. doi: 10.3389/fbinf.2022.793819.

[61] G. Ristovski, T. Preusser, H. K. Hahn, and L. Linsen, "Uncertainty in medical visualization: Towards a taxonomy," Computers & Graphics, vol. 39. pp. 60–73, 2014. doi: 10.1016/j.cag.2013.10.015.

[62] T. M. Athawale, K. A. Johnson, C. R. Butson, and C. R. Johnson, "A statistical framework for quantification and visualisation of positional uncertainty in deep brain stimulation electrodes," Comput Methods Biomech Biomed Eng Imaging Vis, vol. 7, no. 4, pp. 438–449, 2019.

[63] K. Potter et al., "Ensemble-Vis: A Framework for the Statistical Visualization of Ensemble Data," 2009 IEEE International Conference on Data Mining Workshops. 2009. doi: 10.1109/icdmw.2009.55.

[64] F. Raith, G. Scheuermann, and C. Gillmann, "Uncertainty-aware detection and visualization of ocean eddies in ensemble flow fields - A case study of the Red Sea." The Eurographics Association, 2021. doi: 10.2312/ENVIRVIS.20211080.

[65] J. Sanyal, S. Zhang, G. Bhattacharya, P. Amburn, and R. J. Moorhead, "A user study to compare four uncertainty visualization methods for 1D and 2D datasets," IEEE Trans. Vis. Comput. Graph., vol. 15, no. 6, pp. 1209–1218, Nov. 2009.

[66] J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay, "In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation," IEEE Trans. Vis. Comput. Graph., Sep. 2018, doi: 10.1109/TVCG.2018.2864889.

[67] G.-P. Bonneau et al., "Overview and State-of-the-Art of Uncertainty Visualization," Mathematics and Visualization. pp. 3–27, 2014. doi: 10.1007/978-1-4471-6497-5_1.

[68] M. Brooks, "Beyond quantum supremacy: the hunt for useful quantum computers," Nature, vol. 574, no. 7776. pp. 19–21, 2019. doi: 10.1038/d41586-019-02936-3.

[69] W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K.-R. Müller, Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Springer Nature, 2019.

[70] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in Computer Vision – ECCV 2014, Cham: Springer International Publishing, 2014, pp. 818–833.

[71] N. Kokhlikyan et al., "Captum: A unified and generic model interpretability library for PyTorch," 2020, doi: 10.48550/ARXIV.2009.07896.

[72] B. Hu, P. Tunison, B. Vasu, N. Menon, R. Collins, and A. Hoogs, "XAITK : The explainable AI toolkit," Applied AI Letters, vol. 2, no. 4, Dec. 2021, doi: 10.1002/ail2.40.

[73] B. Kim et al., "Interpretability beyond feature attribution: Quantitative testing with Concept Activation Vectors (TCAV)," 2017, doi: 10.48550/ARXIV.1711.11279.

[74] D. Bau, B. Zhou, A. Khosla, A. Oliva, and A. Torralba, "Network Dissection: Quantifying interpretability of deep visual representations," 2017, doi: 10.48550/ARXIV.1704.05796.

[75] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nat Mach Intell, vol. 1, no. 5, pp. 206–215, May 2019.

[76] F. Doshi-Velez and B. Kim, "Towards A rigorous science of interpretable machine learning," 2017, doi: 10.48550/ARXIV.1702.08608.

[77] J. Choo and S. Liu, "Visual Analytics for Explainable Deep Learning," IEEE Computer Graphics and Applications, vol. 38, no. 4. pp. 84–92, 2018. doi: 10.1109/mcg.2018.042731661.

[78] X. Huang et al., "Interactive Visual Study of Multiple Attributes Learning Model of X-Ray Scattering Images," IEEE Trans. Vis. Comput. Graph., vol. 27, no. 2, pp. 1312–1321, Feb. 2021.

[79] T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, "Measuring the capabilities of quantum computers," Nature Physics, vol. 18, no. 1. pp. 75–79, 2022. doi: 10.1038/s41567-021-01409-7.

[80] R. Blume-Kohout and K. C. Young, "A volumetric framework for quantum computer benchmarks," Quantum, vol. 4. p. 362, 2020. doi: 10.22331/q-2020-11-15-362.

[81] M. Benedetti, D. Garcia-Pintos, O. Perdomo, V. Leyton-Ortega, Y. Nam, and A. Perdomo-Ortiz, "A generative modeling approach for benchmarking and training shallow quantum circuits," npj Quantum Information, vol. 5, no. 1. 2019. doi: 10.1038/s41534-019-0157-8.

[82] F. Arute et al., “Quantum supremacy using a programmable superconducting processor,” Nature, vol. 574, no. 7779, pp. 505–510, Oct. 2019.

[83] M. Sasakura, S. Taniuchi, and K. Iwata, “A Proposal of Visualization System for Understanding Quantum Algorithms,” 2019 23rd International Conference in Information Visualization – Part II. 2019. doi: 10.1109/iv-2.2019.00015.

[84] S. Lin, J. Hao, and L. Sun, “QuFlow: Visualizing Parameter Flow in Quantum Circuits for Understanding Quantum Computation,” 2018 IEEE Scientific Visualization Conference (SciVis). 2018. doi: 10.1109/scivis.2018.8823602.

[85] S. Ruan, Y. Wang, W. Jiang, Y. Mao, and Q. Guan, “VACSEN: A visualization approach for noise awareness in quantum computing,” IEEE Trans. Vis. Comput. Graph., pp. 1–11, 2022.

[86] “IBM Quantum Composer,” IBM Quantum. https://quantum-computing.ibm.com/composer/docs/iqx/ (accessed Dec. 20, 2022).

[87] “Cirq,” Google Quantum AI. https://quantumai.google/cirq (accessed Dec. 20, 2022).

[88] J. Mackinlay, “Automating the design of graphical presentations of relational information,” ACM Trans. Graph., vol. 5, no. 2, pp. 110–141, Apr. 1986.

[89] W. Schroeder, K. Martin, and B. Lorensen, The Visualization Toolkit: An Object-oriented Approach to 3D Graphics. Ingram, 2006.

[90] A. Huebl et al., Openpmd 1.1.0: Base paths for mesh- and particle-only files and updated attributes. Zenodo, 2018. doi: 10.5281/zenodo.1167843.

[91] L. J. Clarke and E. Mark, “Enhancements to the eXtensible Data Model and Format (XDMF),” 2007 DoD High Performance Computing Modernization Program Users Group Conference. 2007. doi: 10.1109/hpcmp-ugc.2007.30.

[92] D. Pugmire et al., “Fides: A General Purpose Data Model Library for Streaming Data,” Lecture Notes in Computer Science. pp. 495–507, 2021. doi: 10.1007/978-3-030-90539-2_34.

[93] K. Moreland et al., “VTK-m: Accelerating the Visualization Toolkit for Massively Threaded Architectures,” IEEE Comput. Graph. Appl., vol. 36, no. 3, pp. 48–58, May 2016.

[94] W. F. Godoy et al., “ADIOS 2: The Adaptable Input Output System. A framework for high-performance data management,” SoftwareX, vol. 12, no. 100561, p. 100561, Jul. 2020.

[95] “Conduit.” https://llnl-conduit.readthedocs.io/en/latest/conduit.html (accessed Dec. 20, 2022).

[96] “Blueprint.” https://llnl-conduit.readthedocs.io/en/latest/blueprint.html,

[97] “ONNX.” https://onnx.ai/ (accessed Dec. 20, 2022).

[98] M. J. Dixon, M. Curtis, D. Michelson, J. Hardin, K. Kehoe, and S. Haimov, “CfRadial2 data file format: CF2 NetCDF format for RADAR and LIDAR data in radial coordinates - v2.0.” UCAR/NCAR, 2019. doi: 10.5065/FY2K-X587.

[99] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer, “Vega-Lite: A Grammar of Interactive Graphics,” IEEE Trans. Vis. Comput. Graph., vol. 23, no. 1, pp. 341–350, Jan. 2017.

[100] A. Santos et al., “Visus,” in Proceedings of the Workshop on Human-In-the-Loop Data Analytics - HILDA’19, Amsterdam, Netherlands, 2019. doi: 10.1145/3328519.3329134.

[101] J. Ahrens, B. Geveci, and C. Law, “Paraview: An end-user tool for large data visualization,” The visualization handbook, 2005, [Online]. Available: https://datascience.dsscale.org/wp-content/uploads/2016/06/ParaView.pdf

[102] R. Anderson et al., “The multiphysics on advanced platforms project,” Lawrence Livermore National Laboratory (LLNL), Nov. 2020. doi: 10.2172/1724326.

[103] G. Mengaldo et al., “Industry-Relevant Implicit Large-Eddy Simulation of a High-Performance Road Car via Spectral/hp Element Methods,” SIAM Review, vol. 63, no. 4. pp. 723–755, 2021. doi: 10.1137/20m1345359.

[104] T. Kolev et al., “Efficient exascale discretizations: High-order finite element methods,”

Int. J. High Perform. Comput. Appl., vol. 35, no. 6, pp. 527–552, Nov. 2021.

[105] "FMS: FMS Documentation." https://codedocs.xyz/CEED/FMS/ (accessed Dec. 20, 2022).

[106] A. C. Bauer et al., "In SituMethods, Infrastructures, and Applications on High Performance Computing Platforms," Computer Graphics Forum, vol. 35, no. 3. pp. 577–597, 2016. doi: 10.1111/cgf.12930.

[107] H. Childs et al., "A terminology for in situ visualization and analysis systems," Int. J. High Perform. Comput. Appl., vol. 34, no. 6, pp. 676–691, Nov. 2020.

[108] Whitlock, Brad and Favre, Jean M. and Meredith, Jeremy S., "Parallel in situ coupling of simulation with a fully featured visualization system," Proceedings of the 11th Eurographics, doi: 10.2312/egpgv.egpgv11.101-109/101-109.

[109] U. Ayachit et al., "ParaView catalyst," in Proceedings of the First Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization, Austin TX USA, Nov. 2015. doi: 10.1145/2828612.2828624.

[110] M. Larsen et al., "The ALPINE In Situ Infrastructure," Proceedings of the In Situ Infrastructures on Enabling Extreme-Scale Analysis and Visualization. 2017. doi: 10.1145/3144769.3144778.

[111] U. Ayachit et al., "The SENSEI Generic In Situ Interface," in 2016 Second Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization (ISAV), Salt Lake City, UT, USA, Nov. 2016. doi: 10.1109/isav.2016.013.

[112] E. P. Duque et al., "EPIC - An Extract Plug-In Components Toolkit for In-Situ Data Extracts Architecture," 22nd AIAA Computational Fluid Dynamics Conference. 2015. doi: 10.2514/6.2015-3410.

[113] T. Fogal, F. Proch, A. Schiewe, O. Hasemann, A. Kempf, and J. Krüger, "processing: Transparent visualization via data interception," Eurographics Symp Parallel Graph Vis, vol. 2014, pp. 49–56, 2014.

[114] M. Rivi, L. Calori, G. Muscianisi, and V. Slavnic, "In-situ visualization: State-of-the-art and some use cases." Zenodo, 2012. doi: 10.5281/ZENODO.6303483.

[115] K. Moreland et al., "Minimizing development costs for efficient many-core visualization using MCD3," Parallel Computing, vol. 108. p. 102834, 2021. doi: 10.1016/j.parco.2021.102834.

[116] M. Lian, "Introduction to Service Oriented Architecture." https://home.cs.colorado.edu/~kena/classes/5828/s12/presentation-materials/lianming.pdf (accessed Dec. 20, 2022).

[117] A. Amies, H. Sluiman, Q. G. Tong, and G. N. Liu, Developing and Hosting Applications on the Cloud. IBM Press, 2012.

[118] O. D. H. Hang, "Software as a Service," Software as a Service. https://home.cs.colorado.edu/~kena/classes/5828/s12/presentation-materials/dibieogheneovohanghaojie.pdf

[119] N. Dragoni et al., "Microservices: Yesterday, Today, and Tomorrow," Present and Ulterior Software Engineering. pp. 195–216, 2017. doi: 10.1007/978-3-319-67425-4_12.

[120] D. Pugmire et al., "Visualization as a Service for Scientific Data," in Communications in Computer and Information Science, Cham: Springer International Publishing, 2020, pp. 157–174.

[121] M. Raji, J. Duncan, T. Hobson, and J. Huang, "Dataless Sharing of Interactive Visualization," IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 9. pp. 3656–3669, 2021. doi: 10.1109/tvcg.2020.2984708.

[122] T. Hobson, O. Yildiz, B. Nicolae, J. Huang, and T. Peterka, "Shared-Memory Communication for Containerized Workflows," 2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid). 2021. doi: 10.1109/ccgrid51090.2021.00022.

[123] S. Saha, S. Moorthi, X. Wu, and J. Wang, "NCEP climate forecast system version 2 (CFSv2) 6-hourly products," Data Archive at ..., 2011, [Online]. Available: https://scholar.google.ca/scholar?cluster=16864460308326032258&hl=en&as_sdt=0,5&sciodt=0,5

[124] "The Paraview Catalyst Users Guide." https://www.paraview.org/files/catalyst/docs/ParaViewCatalystUsersGuide_v2.pdf (accessed Dec. 20, 2022).

[125] V. Pascucci, "ViSUS: Visualization Streams for Ultimate Scalability." 2005. doi: 10.2172/15014677.

[126] “The ViSUS Visualization Framework,” High Performance Visualization. pp. 439–452, 2012. doi: 10.1201/b12985-32.

[127] “VisIt: An End-User Tool for Visualizing and Analyzing Very,” High Performance Visualization. pp. 395–410, 2012. doi: 10.1201/b12985-29.

[128] D. Morozov and T. Peterka, “Block-Parallel Data Analysis with DIY2.” 2017. doi: 10.2172/1377403.

[129] W. Humphrey, A. Dalke, and K. Schulten, “VMD: visual molecular dynamics,” J. Mol. Graph., vol. 14, no. 1, pp. 33–8, 27–8, Feb. 1996.

[130] Y. Jiang et al., “tomviz: Providing Advanced Electron Tomography by Streamlining Alignment, Reconstruction, and 3D Visualization,” Microscopy and Microanalysis, vol. 23, no. S1. pp. 222–223, 2017. doi: 10.1017/s1431927617001799.

[131] M. J. Turk et al., “yt: A MULTI-CODE ANALYSIS TOOLKIT FOR ASTROPHYSICAL SIMULATION DATA,” The Astrophysical Journal Supplement Series, vol. 192, no. 1. p. 9, 2011. doi: 10.1088/0067-0049/192/1/9.

[132] R. Brun and F. Rademakers, “ROOT — An object oriented data analysis framework,” Nucl. Instrum. Methods Phys. Res. A, vol. 389, no. 1–2, pp. 81–86, Apr. 1997.

[133] J. D. Hunter, “Matplotlib: A 2D Graphics Environment,” Computing in Science & Engineering, vol. 9, no. 3. pp. 90–95, 2007. doi: 10.1109/mcse.2007.55.

[134] “Plotly: Low-Code Data App Development.” https://plot.ly (accessed Dec. 20, 2022).

[135] “Bokeh.” http://www.bokeh.pydata.org

[136] Kitware Inc, “trame,” trame, Dec. 18, 2022. https://kitware.github.io/trame/index.html (accessed Dec. 20, 2022).

[137] “vtkWeb.” https://vtk.org/flavors/#web (accessed Dec. 20, 2022).

[138] “Solutions.” https://www.paraview.org/solutions/ (accessed Dec. 20, 2022).

[139] Kitware Inc, “vtk.js,” vtk.js, Dec. 14, 2022. https://kitware.github.io/vtk-js/index.html (accessed Dec. 20, 2022).

[140] T. Peterka et al., “Priority research directions for in situ data management: Enabling scientific discovery from diverse data sources,” Int. J. High Perform. Comput. Appl., vol. 34, no. 4, pp. 409–427, Jul. 2020.

[141] B. Hendrickson et al., “ASCR@40: Highlights and impacts of ASCR’s programs,” USDOE Office of Science (SC) (United States), Jun. 2020. doi: 10.2172/1631812.

[142] T. Kurth et al., “Exascale Deep Learning for Climate Analytics,” in SC ’18: Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis, Nov. 2018.

[143] S. Chandrasekaran and G. Juckeland, OpenACC for Programmers: Concepts and Strategies. Addison-Wesley Professional, 2017.

[144] N. Bell and J. Hoberock, “Chapter 26 - Thrust: A Productivity-Oriented Library for CUDA,” in GPU Computing Gems Jade Edition, W.-M. W. Hwu, Ed. Boston: Morgan Kaufmann, 2012, pp. 359–371.

[145] H. Carter Edwards, C. R. Trott, and D. Sunderland, “Kokkos: Enabling manycore performance portability through polymorphic memory access patterns,” J. Parallel Distrib. Comput., vol. 74, no. 12, pp. 3202–3216, Dec. 2014.

[146] C. Sewell et al., “The SDAV Software Frameworks for Visualization and Analysis on Next-Generation Multi-Core and Many-Core Architectures,” in 2012 SC Companion: High Performance Computing, Networking Storage and Analysis, Nov. 2012, pp. 206–214.

[147] S. G. Parker et al., “GPU ray tracing,” Commun. ACM, vol. 56, no. 5, pp. 93–101, May 2013.

[148] A. Knoll, G. P. Johnson, and J. Meng, “Path Tracing RBF Particle Volumes,” in Ray Tracing Gems II: Next Generation Real-Time Rendering with DXR, Vulkan, and OptiX, A. Marrs, P. Shirley, and I. Wald, Eds. Berkeley, CA: Apress, 2021, pp. 713–723.

[149] H. Childs et al., “VisIt: An End-User Tool For Visualizing and Analyzing Very Large Data,” in High Performance Visualization, Chapman and Hall/CRC, 2012, pp. 395–410.

[150] S. Li, S. Jaroszynski, S. Pearse, L. Orf, and J. Clyne, "VAPOR: A Visualization Package Tailored to Analyze Simulation Data in Earth System Science," Atmosphere , vol. 10, no. 9, p. 488, Aug. 2019.

[151] P. Colella, "Defining software requirements for scientific computing," presented at the DARPA High Productivity Computing Program Meeting, Fairfax, VA, 2004.

[152] K. Yelick et al., "The parallelism motifs of genomic data analysis," Philos. Trans. A Math. Phys. Eng. Sci., vol. 378, no. 2166, p. 20190394, Mar. 2020.

[153] M. Tory and T. Moller, "Rethinking Visualization: A High-Level Taxonomy," in IEEE Symposium on Information Visualization, Oct. 2004, pp. 151–158.

[154] K. Moreland, B. Geveci, K.-L. Ma, and R. Maynard, "A classification of scientific visualization algorithms for massive threading," in Proceedings of the 8th International Workshop on Ultrascale Visualization, Denver, Colorado, Nov. 2013, pp. 1–10.

[155] J. S. Vetter et al., "Extreme heterogeneity 2018 - productive computational science in the era of extreme heterogeneity: Report for DOE ASCR workshop on extreme heterogeneity," USDOE Office of Science (SC) (United States), Dec. 2018. doi: 10.2172/1473756.

[156] S. Klasky, J. Thayer, and H. Najm, "Data Reduction for Science: Brochure from the Advanced Scientific Computing Research Workshop." 2021. doi: 10.2172/1770192.

[157] Q.-V. Pham et al., "A Survey of Multi-Access Edge Computing in 5G and Beyond: Fundamentals, Technology Integration, and State-of-the-Art," IEEE Access, vol. 8. pp. 116974–117017, 2020. doi: 10.1109/access.2020.3001277.

[158] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5G Evolution: A View on 5G Cellular Technology Beyond 3GPP Release 15," IEEE Access, vol. 7. pp. 127639–127651, 2019. doi: 10.1109/access.2019.2939938.

[159] Z. Liu, B. Mah, Y. Kumar, C. Guok, and R. Cziva, "Programmable Per-Packet Network Telemetry," Proceedings of the 2021 on Systems and Network Telemetry and Analytics. 2020. doi: 10.1145/3452411.3464443.

[160] R. Vescovi et al., "Linking Scientific Instruments and HPC: Patterns, Technologies, Experiences," SSRN Electronic Journal. doi: 10.2139/ssrn.4141629.

[161] J. Dean and S. Ghemawat, "MapReduce," Commun. ACM, vol. 51, no. 1, pp. 107–113, Jan. 2008.

[162] M. Caballer, C. de Alfonso, F. Alvarruiz, and G. Moltó, "EC3: Elastic cloud computing cluster," J. Comput. System Sci., vol. 79, no. 8, pp. 1341–1351, Dec. 2013.

[163] M. Raji, A. Hota, T. Hobson, and J. Huang, "Scientific Visualization as a Microservice," IEEE Trans. Vis. Comput. Graph., vol. 26, no. 4, pp. 1760–1774, Apr. 2020.

[164] T. Hobson, J. Hammer, P. P. Provins, and J. Huang, "Interactive visualization of large turbulent flow as a cloud service," IEEE trans. cloud comput., pp. 1–1, 2021.

[165] M. Dorier, Z. Wang, U. Ayachit, S. Snyder, R. Ross, and M. Parashar, "Colza: Enabling Elastic In Situ Visualization for High-performance Computing Simulations," 2022 IEEE International Parallel and Distributed Processing Symposium (IPDPS). 2022. doi: 10.1109/ipdps53621.2022.00059.

[166] R. B. Ross et al., "Mochi: Composing data services for high-performance computing environments," J. Comput. Sci. Technol., vol. 35, no. 1, pp. 121–144, Jan. 2020.

[167] L. P. Santos, T. Bashford-Rogers, J. Barbosa, and P. Navrátil, "Towards quantum ray tracing," 2022, doi: 10.48550/ARXIV.2204.12797.

[168] X. Lu and H. Lin, "A Framework for Quantum Ray Tracing," arXiv [quant-ph], Mar. 29, 2022. [Online]. Available: http://arxiv.org/abs/2203.15451

[169] M. G. Amankwah, D. Camps, E. W. Bethel, R. Van Beeumen, and T. Perciano, "Quantum pixel representations and compression for N-dimensional images," 2021, doi: 10.48550/ARXIV.2110.04405.

[170] A. Sornsaeng, N. Dangniam, P. Palittapongarnpim, and T. Chotibut, "Quantum diffusion map

for nonlinear dimensionality reduction," Phys. Rev. A, vol. 104, no. 5, p. 052410, Nov. 2021.

[171] S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum algorithms for supervised and unsupervised machine learning," arXiv [quant-ph], Jul. 01, 2013. [Online]. Available: http://arxiv.org/abs/1307.0411

[172] M. Broughton et al., "TensorFlow Quantum: A Software Framework for Quantum Machine Learning," arXiv [quant-ph], Mar. 06, 2020. [Online]. Available: http://arxiv.org/abs/2003.02989

[173] P. Bressana, N. Zilberman, D. Vucinic, and R. Soulé, "Trading latency for compute in the network," in Proceedings of the Workshop on Network Application Integration/CoDesign, Virtual Event USA, Aug. 2020. doi: 10.1145/3405672.3405807.

[174] Y. Tokusashi, H. Matsutani, and N. Zilberman, "LaKe: The power of in-network computing," in 2018 International Conference on ReConFigurable Computing and FPGAs (ReConFig), Cancun, Mexico, Dec. 2018. doi: 10.1109/reconfig.2018.8641696.

[175] D. Hendrycks, M. Mazeika, and T. Dietterich, "Deep Anomaly Detection with Outlier Exposure," Dec. 2018, doi: 10.48550/arXiv.1812.04606.

[176] T. Hey, "Opportunities and challenges from artificial intelligence and machine learning for the advancement of science, technology, and the office of science missions," USDOE Office of Science (SC) (United States), Sep. 2020. doi: 10.2172/1734848.

[177] E. M. Raybourn, W. A. Stubblefield, M. Trumbo, A. Jones, J. Whetzel, and N. Fabian, "Information Design for XR Immersive Environments: Challenges and Opportunities," in Virtual, Augmented and Mixed Reality. Multimodal Interaction, Cham: Springer International Publishing, 2019, pp. 153–164.

[178] T. Marrinan and M. E. Papka, "Real-Time Omnidirectional Stereo Rendering: Generating 360° Surround-View Panoramic Images for Comfortable Immersive Viewing," IEEE Trans. Vis. Comput. Graph., vol. 27, no. 5, pp. 2587–2596, May 2021.

[179] P. Reipschlager, T. Flemisch, and R. Dachselt, "Personal Augmented Reality for Information Visualization on Large Interactive Displays," IEEE Trans. Vis. Comput. Graph., vol. 27, no. 2, pp. 1182–1192, Feb. 2021.

[180] C. Andrews, A. Endert, and C. North, "Space to think," in Proceedings of the 28th international conference on Human factors in computing systems - CHI '10, Atlanta, Georgia, USA, 2010. doi: 10.1145/1753326.1753336.

[181] I. Wald, S. Woop, C. Benthin, G. S. Johnson, and M. Ernst, "Embree: a kernel framework for efficient CPU ray tracing," ACM Trans. Graph., vol. 33, no. 4, pp. 1–8, Jul. 2014.

[182] Q. Wang, Z. Chen, Y. Wang, and H. Qu, "A Survey on ML4VIS: Applying Machine Learning Advances to Data Visualization," IEEE Trans. Vis. Comput. Graph., vol. 28, no. 12, pp. 5134–5153, Dec. 2022.

[183] C. Wang and J. Han, "DL4SciVis: A State-of-the-Art Survey on Deep Learning for Scientific Visualization," IEEE Trans. Vis. Comput. Graph., vol. PP, Apr. 2022, doi: 10.1109/TVCG.2022.3167896.

[184] A. Wu et al., "AI4VIS: Survey on Artificial Intelligence Approaches for Data Visualization," IEEE Trans. Vis. Comput. Graph., vol. 28, no. 12, pp. 5049–5070, Dec. 2022.

[185] S. Weiss and R. Westermann, "Differentiable Direct Volume Rendering," IEEE Trans. Vis. Comput. Graph., vol. 28, no. 1, pp. 562–572, Jan. 2022.

[186] Y. Qin, B. T. Fasy, C. Wenk, and B. Summa, "A Domain-Oblivious Approach for Learning Concise Representations of Filtered Topological Spaces for Clustering," IEEE Trans. Vis. Comput. Graph., vol. 28, no. 1, pp. 302–312, Jan. 2022.

[187] J. Xu, H. Guo, H.-W. Shen, M. Raj, S. W. Wurster, and T. Peterka, "Reinforcement Learning for Load-balanced Parallel Particle Tracing," IEEE Trans. Vis. Comput. Graph., vol. PP, Feb. 2022, doi: 10.1109/TVCG.2022.3148745.

[188] "ADVANCED SCIENTIFIC COMPUTING RESEARCH - REACHING A NEW ENERGY SCIENCES WORKFORCE (ASCR-RENEW)." https://science.osti.gov/grants/FOAs/-/media/grants/pdf/foas/2022/DE-FOA-0002767.pdf (accessed Dec. 20, 2022).

[189] M. Card, Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann, 1999.

[190] F. B. Viegas and M. Wattenberg, "Communication-minded visualization: A call to action

[Technical forum],” IBM Systems Journal, vol. 45, no. 4. pp. 801–812, 2006. doi: 10.1147/sj.454.0801.

[191] Z. Pousman, J. Stasko, and M. Mateas, “Casual information visualization: depictions of data in everyday life,” IEEE Trans. Vis. Comput. Graph., vol. 13, no. 6, pp. 1145–1152, Nov-Dec 2007.

[192] R. Kosara and E. Bertini, “Reflecting on the design criteria for explanatory visualizations.” https://niiv.njitvis.com/assets/tvcg16.pdf (accessed Dec. 21, 2022).

[193] R. Kosara, “Presentation-Oriented Visualization Techniques,” IEEE Comput. Graph. Appl., vol. 36, no. 1, pp. 80–85, Jan-Feb 2016.

[194] S. B. Barnes, An Introduction to Visual Communication: From Cave Art to Second Life. Peter Lang Pub Incorporated, 2011.

[195] “The Three Elements of Successful Data Visualizations,” Harvard Business Review, Apr. 19, 2013. https://hbr.org/2013/04/the-three-elements-of-successf (accessed Dec. 20, 2022).

[196] S. Glibova, “Visual Communication for Data Scientists,” Analytics Vidhya, Jul. 18, 2021. https://medium.com/analytics-vidhya/visual-communication-for-data-scientists-2dc3e9e4fb3 (accessed Dec. 20, 2022).

[197] E. Murray, “How Data Visualization Supports Communication,” Forbes, Jan. 28, 2019. https://www.forbes.com/sites/evamurray/2019/01/28/how-data-visualization-supports-communication/ (accessed Dec. 20, 2022).

[198] J. J. van Wijk and A. M. Cohen, “Visualization of Seifert surfaces,” IEEE Trans. Vis. Comput. Graph., vol. 12, no. 4, pp. 485–496, Jul-Aug 2006.

[199] D. Weiskopf et al., “Explanatory and illustrative visualization of special and general relativity,” IEEE Trans. Vis. Comput. Graph., vol. 12, no. 4, pp. 522–534, Jul-Aug 2006.

[200] U. Hahn and P. Berkers, “Visualizing climate change: an exploratory study of the effectiveness of artistic information visualizations,” World Art, vol. 11, no. 1, pp. 95–119, Jan. 2021.

[201] C. Yau, M. Karimzadeh, C. Surakitbanharn, N. Elmqvist, and D. S. Ebert, “Bridging the data analysis communication gap utilizing a three- component summarized line graph,” Comput. Graph. Forum, vol. 38, no. 3, pp. 375–386, Jun. 2019.

[202] E. Adar and E. Lee, “Communicative Visualizations as a Learning Problem,” Sep. 2020, doi: 10.48550/arXiv.2009.07095.

[203] F. Lan et al., “Visualization in Astrophysics: Developing New Methods, Discovering Our Universe, and Educating the Earth,” Computer Graphics Forum, vol. 40, no. 3. pp. 635–663, 2021. doi: 10.1111/cgf.14332.

[204] A. Ynnerman, T. Rydell, D. Antoine, D. Hughes, A. Persson, and P. Ljung, “Interactive visualization of 3d scanned mummies at public venues,” Commun. ACM, vol. 59, no. 12, pp. 72–81, Dec. 2016.

[205] “A walk inside the brain at The Leonardo Museum using Virtual Reality.” https://dataintensivescience.org/?p=1 (accessed Dec. 20, 2022).

[206] M. Böttinger, H.-N. Kostis, M. Velez-Rojas, P. Rheingans, and A. Ynnerman, “Reflections on visualization for broad audiences,” in Foundations of Data Visualization, Cham: Springer International Publishing, 2020, pp. 297–305.

[207] F. B. Viégas and M. Wattenberg, “Artistic Data Visualization: Beyond Visual Analytics,” Online Communities and Social Computing. pp. 182–191, 2007. doi: 10.1007/978-3-540-73257-0_21.

[208] Trust in organizations : frontiers of theory and research. Thousand Oaks, Calif.: Sage Publications.

[209] A. Dasgupta et al., “Familiarity Vs Trust: A Comparative Study of Domain Scientists’ Trust in Visual Analytics and Conventional Analysis Methods,” IEEE Trans. Vis. Comput. Graph., vol. 23, no. 1, pp. 271–280, Jan. 2017.

[210] J. Ooge and K. Verbert, “Trust in Prediction Models: a Mixed-Methods Pilot Study on the Impact of Domain Expertise,” 2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX). 2021. doi: 10.1109/trex53765.2021.00007.

[211] W. Han and H.-J. Schulz, “Beyond Trust Building — Calibrating Trust in Visual Analytics,” 2020 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX). 2020. doi: 10.1109/trex51495.2020.00006.

[212] T. L. Weissgerber et al., “Reveal, Don’t Conceal: Transforming Data Visualization to Improve Transparency,” Circulation, vol. 140, no. 18, pp. 1506–1518, Oct. 2019.

[213] B. Plale and S. L. Harrell, “Transparency and Reproducibility Practice in Large-Scale Computational Science: A Preface to the Special Section,” IEEE Trans. Parallel Distrib. Syst., vol. 32, no. 11, pp. 2607–2608, Nov. 2021.

[214] C. Silva, J. Freire, E. Santos, and E. Anderson, “Provenance-Enabled Data Exploration and Visualization with VisTrails,” 2010 23RD SIBGRAPI - Conference on Graphics, Patterns and Images Tutorials. 2010. doi: 10.1109/sibgrapi-t.2010.9.

[215] E. Segel and J. Heer, “Narrative visualization: telling stories with data,” IEEE Trans. Vis. Comput. Graph., vol. 16, no. 6, pp. 1139–1148, Nov. 2010.

[216] J. Hullman and N. Diakopoulos, “Visualization rhetoric: framing effects in narrative visualization,” IEEE Trans. Vis. Comput. Graph., vol. 17, no. 12, pp. 2231–2240, Dec. 2011.

[217] N. H. Riche, C. Hurter, N. Diakopoulos, and S. Carpendale, Data-Driven Storytelling. AK Peters Visualization Series, 2018.

[218] Y. Sakamoto, S. Sallam, A. Salo, J. Leboe-McGowan, and P. Irani, “Persuasive Data Storytelling with a Data Video during Covid-19 Infodemic: Affective Pathway to Influence the Users’ Perception about Contact Tracing Apps in less than 6 Minutes,” 2022 IEEE 15th Pacific Visualization Symposium (PacificVis). 2022. doi: 10.1109/pacificvis53943.2022.00028.

[219] Y. Zhang, M. Reynolds, A. Lugmayr, K. Damjanov, and G. M. Hassan, “A visual data storytelling framework,” Informatics (MDPI), vol. 9, no. 4, p. 73, Sep. 2022.

[220] S. Gratzl, A. Lex, N. Gehlenborg, N. Cosgrove, and M. Streit, “From Visual Exploration to Storytelling and Back Again,” Comput. Graph. Forum, vol. 35, no. 3, pp. 491–500, Jun. 2016.

[221] “Funding for Accelerated, Inclusi.,” Nov. 07, 2022. https://science.osti.gov/Initiatives/FAIR (accessed Dec. 20, 2022).

[222] “Department of Energy Announces $8.5 Million for FAIR Data to Advance Artificial Intelligence for Science,” Energy.gov. https://www.energy.gov/articles/department-energy-announces-85-million-fair-data-advance-artificial-intelligence-science (accessed Dec. 20, 2022).

[223] M. D. Wilkinson et al., “The FAIR Guiding Principles for scientific data management and stewardship,” Sci Data, vol. 3, p. 160018, Mar. 2016.

[224] National Centers for Environmental Information (NCEI), “Climate Data Online.” https://www.ncei.noaa.gov/cdo-web/ (accessed Dec. 20, 2022).

[225] “Climate,” Data.gov. https://www.data.gov/climate/ (accessed Dec. 20, 2022).

[226] “Website.” https://esgf-node.llnl.gov/projects/esgf-llnl/ https://esgf.llnl.gov/

[227] “Home - GEO - NCBI.” https://www.ncbi.nlm.nih.gov/geo/ (accessed Dec. 20, 2022).

[228] “Website.” https://www.rcsb.org/

[229] “Johns Hopkins Turbulence Databases (JHTDB).” http://turbulence.pha.jhu.edu/ (accessed Dec. 20, 2022).

[230] “Home,” ESnet. https://www.es.net/ (accessed Dec. 20, 2022).

[231] “Open Data Initiative.” https://data-science.llnl.gov/open-data-initiative (accessed Dec. 20, 2022).

[232] “WIFIRE Firemap.” https://firemap.sdsc.edu/ (accessed Dec. 20, 2022).

[233] “Energy Justice Dashboard (BETA),” Energy.gov. https://www.energy.gov/diversity/energy-justice-dashboard-beta (accessed Dec. 20, 2022).

[234] E. Clarkson, K. Desai, and J. Foley, “ResultMaps: visualization for search interfaces,” IEEE Trans. Vis. Comput. Graph., vol. 15, no. 6, pp. 1057–1064, Nov-Dec 2009.

[235] W. de Leeuw, P. J. Verschure, and R. van Liere, “Isualization and analysis of large data collections: a case study

applied to confocal microscopy data," IEEE Trans. Vis. Comput. Graph., vol. 12, no. 5, pp. 1251–1258, Sep-Oct 2006.

[236] B. Kim, B. Lee, S. Knoblach, E. Hoffman, and J. Seo, "GeneShelf: a web-based visual interface for large gene expression time-series data repositories," IEEE Trans. Vis. Comput. Graph., vol. 15, no. 6, pp. 905–912, Nov-Dec 2009.

[237] "CDAT." https://cdat.llnl.gov/ (accessed Dec. 20, 2022).

[238] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, "MDAnalysis: a toolkit for the analysis of molecular dynamics simulations," J. Comput. Chem., vol. 32, no. 10, pp. 2319–2327, Jul. 2011.

[239] B. Bach et al., "Dashboard Design Patterns," IEEE Trans. Vis. Comput. Graph., vol. PP, Sep. 2022, doi: 10.1109/TVCG.2022.3209448.

[240] S. Few, Information Dashboard Design: The Effective Visual Communication of Data. O'Reilly Media, Inc., 2006.

[241] M. Elshehaly et al., "QualDash: Adaptable Generation of Visualisation Dashboards for Healthcare Quality Improvement," IEEE Trans. Vis. Comput. Graph., vol. 27, no. 2, pp. 689–699, Feb. 2021.

[242] R. Kitchin, T. P. Lauriault, and G. McArdle, "Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards," Reg. Stud. Reg. Sci., vol. 2, no. 1, pp. 6–28, Jan. 2015.

[243] D. Lee, J. Ricardo, A. Felix, S. He, D. Offenhuber, and C. Ratti, "CityEye: Real-time visual dashboard for managing urban services and citizen feedback loops." http://senseable.mit.edu/papers/pdf/20150707_Lee_etal_CityeyeRealtime_CUPUM.pdf (accessed Dec. 21, 2022).

[244] J. C. M. Serrano, O. Papakyriakopoulos, M. Shahrezaye, and S. Hegelich, "The Political Dashboard: A Tool for Online Political Transparency," Proceedings of the International AAAI Conference on Web and Social Media, vol. 14. pp. 983–985, 2020. doi: 10.1609/icwsm.v14i1.7371.

[245] O. M. Yigitbasioglu and O. Velcu, "A review of dashboards in performance management: Implications for design and research," International Journal of Accounting Information Systems, vol. 13, no. 1. pp. 41–59, 2012. doi: 10.1016/j.accinf.2011.08.002.

[246] S. Goodwin, Y. Dkhissi, Q. Wu, B. Moyle, K. Freidin, and A. Liebman, "Informed dashboard designs for microgrid electricity market operators," in Proceedings of the Twelfth ACM International Conference on Future Energy Systems, Virtual Event Italy, Jun. 2021. doi: 10.1145/3447555.3466604.

[247] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," Lancet Infect. Dis., vol. 20, no. 5, pp. 533–534, May 2020.

[248] Y. Zhang, Y. Sun, L. Padilla, S. Barua, E. Bertini, and A. G. Parker, "Mapping the Landscape of COVID-19 Crisis Visualizations." doi: 10.31219/osf.io/kd3y9.

[249] C. Lee, T. Yang, G. D. Inchoco, G. M. Jones, and A. Satyanarayan, "Viral visualizations: How Coronavirus skeptics use orthodox data practices to promote unorthodox science online," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama Japan, May 2021. doi: 10.1145/3411764.3445211.

[250] H. Li, Y. Wang, S. Zhang, Y. Song, and H. Qu, "KG4Vis: A Knowledge Graph-Based Approach for Visualization Recommendation," IEEE Trans. Vis. Comput. Graph., vol. 28, no. 1, pp. 195–205, Jan. 2022.

[251] K. Hu, M. A. Bakker, S. Li, T. Kraska, and C. Hidalgo, "VizML," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow Scotland Uk, May 2019. doi: 10.1145/3290605.3300358.

[252] P. Ruchikachorn and K. Mueller, "Learning Visualizations by Analogy: Promoting Visual Literacy through Visualization Morphing," IEEE Trans. Vis. Comput. Graph., vol. 21, no. 9, pp. 1028–1044, Sep. 2015.

[253] E. E. Firat, A. Joshi, and R. S. Laramee, "Interactive visualization literacy: The state-of-the-art," Inf. Vis., vol. 21, no. 3, pp. 285–310, Jul. 2022.

[254] Z. Liu, R. Jordan Crouser, and A. Ottley, "Survey on Individual Differences in Visualization," Computer Graphics Forum, vol. 39, no. 3. pp. 693–712, 2020. doi: 10.1111/cgf.14033.

[255] B. Mutlu, E. Veas, and C. Trattner, "VizRec," ACM Trans. Interact. Intell. Syst., vol. 6, no. 4, pp. 1–39, Dec. 2016.

[256] P. Gmelch et al., "The case for low-cost, personalized visualization for enhancing natural hazard preparedness," ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., vol. XLIV–M–2–2020, pp. 37–44, Nov. 2020.

[257] W. Dou, D. H. Jeong, F. Stukes, W. Ribarsky, H. R. Lipford, and R. Chang, "Recovering Reasoning Processes from User Interactions," IEEE Computer Graphics and Applications, vol. 29, no. 3. pp. 52–61, 2009. doi: 10.1109/mcg.2009.49.

[258] S. Lee, S.-H. Kim, Y.-H. Hung, H. Lam, Y.-A. Kang, and J. S. Yi, "How do People Make Sense of Unfamiliar Visualizations?: A Grounded Model of Novice's Information Visualization Sensemaking," IEEE Trans. Vis. Comput. Graph., vol. 22, no. 1, pp. 499–508, Jan. 2016.

[259] C. Conati and H. Maclaren, "Exploring the role of individual differences in information visualization," Proceedings of the working conference on Advanced visual interfaces - AVI '08. 2008. doi: 10.1145/1385569.1385602.

[260] D. Toker, C. Conati, G. Carenini, and M. Haraty, "Towards Adaptive Information Visualization: On the Influence of User Characteristics," User Modeling, Adaptation, and Personalization. pp. 274–285, 2012. doi: 10.1007/978-3-642-31454-4_23.

[261] C. A. Cohen and M. Hegarty, "Individual differences in use of external visualisations to perform an internal visualisation task," Applied Cognitive Psychology, vol. 21, no. 6. pp. 701–711, 2007. doi: 10.1002/acp.1344.

[262] A. Ottley et al., "Improving Bayesian Reasoning: The Effects of Phrasing, Visualization, and Spatial Ability," IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1. pp. 529–538, 2016. doi: 10.1109/tvcg.2015.2467758.

[263] M. C. Velez, D. Silver, and M. Tremaine, "Understanding Visualization through Spatial Ability Differences," IEEE Visualization 2005 - (VIS'05). doi: 10.1109/vis.2005.108.

[264] C. G. Healey and J. T. Enns, "Attention and visual memory in visualization and computer graphics," IEEE Trans. Vis. Comput. Graph., vol. 18, no. 7, pp. 1170–1188, Jul. 2012.

[265] C. Conati, G. Carenini, E. Hoque, B. Steichen, and D. Toker, "Evaluating the Impact of User Characteristics and Different Layouts on an Interactive Visualization for Decision Making," Computer Graphics Forum, vol. 33, no. 3. pp. 371–380, 2014. doi: 10.1111/cgf.12393.

[266] F. Samsel, S. Klaassen, and D. H. Rogers, "ColorMoves: Real-time Interactive Colormap Construction for Scientific Visualization," IEEE Comput. Graph. Appl., vol. 38, no. 1, pp. 20–29, Jan. 2018.

[267] F. Samsel, L. Bartram, and A. Bares, "Art, affect and color: Creating engaging expressive scientific visualization," in 2018 IEEE VIS Arts Program (VISAP), Berlin, Germany, Oct. 2018. doi: 10.1109/visap45312.2018.9046053.

[268] G. Johnson and T. Todd Elvins, "Introduction to collaborative visualization," ACM SIGGRAPH Computer Graphics, vol. 32, no. 2. pp. 8–11, 1998. doi: 10.1145/282037.282039.

[269] E. Raybourn, "Why We Need Strategies for Working Remotely: the Exascale Computing Project (ECP) Panel Series," Proposed for presentation at the SC20 held November 17 - October 17, 2020 in Virtual Event, Virtual Event, US. 2020. doi: 10.2172/1825833.

[270] C. Mouton, K. Sons, and I. Grimstead, "Collaborative visualization: current systems and future trends," in Proceedings of the 16th International Conference on 3D Web Technology, Paris, France, Jun. 2011, pp. 101–110.

[271] Remote Rendering Using Vtk and Vic. 2000.

[272] "The Front-End Testing of Data Visualizations," Paperpile. https://paperpile.com/app/p/c8a22dc8-41a0-0ec9-be11-20f90c11b63e (accessed Apr. 20, 2022).

[273] M. Stegmaier, D. Engel, J. Olbrich, T. Ropinski, and M. Tichy, "Property-Based Testing for Visualization Development," 2021, [Online]. Available: http://diglib.eg.org/handle/10.2312/visgap20211087

[274] R. M. Kirby and C. T. Silva, "The need for verifiable visualization," IEEE Comput. Graph. Appl., vol. 28, no. 5, pp. 78–83, Sep. 2008.

[275] Y. L. Wong, K. Madhavan, and N. Elmqvist, "Towards characterizing domain experts

as a user group," in 2018 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV), Berlin, Germany, Oct. 2018. doi: 10.1109/beliv.2018.8634026.

[276] S. Swaid, M. Maat, H. Krishnan, D. Ghoshal, and L. Ramakrishnan, "Usability Heuristic Evaluation of Scientific Data Analysis and Visualization Tools," Advances in Usability and User Experience, pp. 471–482, 2018.

[277] Y. C. Ye, F. Sauer, K.-L. Ma, K. Aditya, and J. Chen, "A User-Centered Design Study in Scientific Visualization Targeting Domain Experts," IEEE Trans. Vis. Comput. Graph., vol. 26, no. 6, pp. 2192–2203, Jun. 2020.

[278] T. Munzner, "A nested model for visualization design and validation," IEEE Trans. Vis. Comput. Graph., vol. 15, no. 6, pp. 921–928, Nov-Dec 2009.

[279] Maneesh Agrawala University of California, Berkeley, Wilmot Li Adobe Systems Inc. , San Francisco, CA, and Floraine Berthouzoz University of California, Berkeley, "Design principles for visual communication," Commun. ACM, Apr. 2011, doi: 10.1145/1924421.1924439.

[280] Peer-Timo Bremer, Georgia Tourassi, Wes Bethel, Kelly Gaither, Valerio Pascucci, Wei Xu, "Visualization for Scientific Discovery, Decision-Making & Communication." https://custom.cvent.com/DCBD4ADAAD004096B1E4AD96F3C8049E/files/event/fc8f3c090e35418982b417f2d3e4e73e/26a8f1a6239947b798de042ba2c5b793.pdf (accessed Dec. 20, 2022).

[281] P.-T. Bremer, G. Tourassi, W. Bethel, K. Gaither, V. Pascucci, and W. Xu, "Position Papers for the ASCR Workshop on Visualization for Scientific Discovery, Decision-Making, and Communication." 2022. doi: 10.2172/1843572.

[282] "Agenda - ASCR Workshop on Visualization for Scientific Discovery, Decision-Making, & Communication." https://web.cvent.com/event/fc8f3c09-0e35-4189-82b4-17f2d3e4e73e/websitePage:645d57e4-75eb-4769-b2c0-f201a0bfc6ce (accessed Dec. 20, 2022).

[283] P.-T. Bremer, G. Tourassi, W. Bethel, K. Gaither, V. Pascucci, and W. Xu, "Visualization for Scientific Discovery, Decision-Making, and Communication." 2022. doi: 10.2172/1845708.

[284] N. Galli and E. Jorgensen, "The Visualization and Analytics Center for Enabling Technologies (VACET)." www.vacet.org (accessed Jan. 08, 2022).

[285] H. Childs et al., "VisIt: An End-User Tool For Visualizing and Analyzing Very Large Data," in High Performance Visualization—Enabling Extreme-Scale Scientific Insight, E. W. Bethel, H. Childs, and C. Hansen, Eds. CRC Press, 2012, pp. 357–372.

[286] H. Childs et al., "Extreme scaling of production visualization software on diverse architectures," IEEE Comput. Graph. Appl., vol. 30, no. 3, pp. 22–31, May 2010.

[287] S. Byna et al., "Parallel I/O, analysis, and visualization of a trillion particle simulation," in SC '12: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, Nov. 2012, pp. 1–12.

[288] P.-T. Bremer, G. H. Weber, V. Pascucci, M. Day, and J. B. Bell, "Analyzing and tracking burning structures in lean premixed hydrogen flames," IEEE Trans. Vis. Comput. Graph., vol. 16, no. 2, pp. 248–260, Mar. 2010.

[289] E. W. Anderson, J. P. Ahrens, K. Heitmann, S. Habib, and C. T. Silva, "Provenance in Comparative Analysis: A Study in Cosmology," Computing in Science Engineering, vol. 10, no. 3, pp. 30–37, May 2008.

[290] "Website." https://sdm.lbl.gov/sdav/

[291] D. Morozov and T. Peterka, "Block-parallel data analysis with DIY2," in 2016 IEEE 6th Symposium on Large Data Analysis and Visualization (LDAV), Oct. 2016, pp. 29–36.

[292] H. Karimabadi et al., "Coherent structures, intermittent turbulence, and dissipation in high-temperature plasmas," Phys. Plasmas, vol. 20, no. 1, p. 012303, Jan. 2013.

[293] "Website." https://www.alcf.anl.gov/news/new-groundbreaking-global-climate-visualization-debuts-92nd-american-meteorological-society

[294] T. Peterka et al., "Meshing the Universe: Integrating Analysis in Cosmological Simulations," in 2012 SC Companion: High Performance Computing, Networking Storage and Analysis, Nov. 2012, pp. 186–195.

[295] "In situ." https://www.paraview.org/in-situ/ (accessed Jan. 08, 2022).

[296] "VisIt-tutorial-in-situ - VisItusers.org." https://visitusers.org/index.php?title=VisIt-tutorial-in-situ (accessed Jan. 08, 2022).

[297] "sensei." https://sensei-insitu.org/index.html (accessed Jan. 08, 2022).

[298] "VTK-m." https://m.vtk.org/ (accessed Jan. 08, 2022).

[299] diatomic, "GitHub - diatomic/diy: data-parallel out-of-core library." https://github.com/diatomic/diy (accessed Jan. 08, 2022).

[300] "Website." https://graphblas.org/

[301] "Website." https://github.com/diatomic/tess2

[302] F. Newberry, C. Wetterer-Nelson, J. A. Evans, A. Doostan, and K. E. Jansen, "Software Tools to Enable Immersive Simulation," May 2021. doi: 10.31224/osf.io/mrf5d.

[303] M. Kim et al., "In Situ Analysis and Visualization of Fusion Simulations: Lessons Learned," in High Performance Computing, 2018, pp. 230–242.

[304] N. G. Author, "Synergistic challenges in data-intensive science and exascale computing. Summary report of the advanced scientific computing advisory committee (ASCAC) subcommittee, march 2013," USDOE Office of Science (SC) (United States), Mar. 2013. doi: 10.2172/1471113.

[305] "[No title]." https://science.osti.gov/-/media/ascr/pdf/program-documents/docs/Exascale-ASCR-Analysis.pdf (accessed Jan. 08, 2022).

[306] "[No title]." https://science.osti.gov/-/media/ascr/pdf/program-documents/docs/Crosscutting_grand_challenges.pdf (accessed Jan. 08, 2022).

[307] T. Peterka et al., "ASCR workshop on in situ data management: Enabling scientific discovery from diverse data sources," USDOE Office of Science (SC) (United States), Feb. 2019. doi: 10.2172/1493245.

[308] "[No title]." https://science.osti.gov/-/media/ascr/pdf/program-documents/docs/Exascale-ASCR-Analysis.pdf (accessed Jan. 08, 2022).

[309] K. Fagnan, Y. Nashed, G. Perdue, D. Ratner, A. Shankar, and S. Yoo, "Data and models: A framework for advancing AI in science," USDOE Office of Science (SC) (United States), Dec. 2019. doi: 10.2172/1579323.

[310] E. Deelman et al., "The future of scientific workflows," Int. J. High Perform. Comput. Appl., vol. 32, no. 1, pp. 159–175, Jan. 2018.

[311] K. Kleese van Dam, A. Shoshani, and C. Plata, "Data crosscutting requirements review," Office of Scientific and Technical Information (OSTI), Apr. 2013. doi: 10.2172/1226419.

[312] E. W. Bethel and M. G. Eds, "Report of the DOE workshop on management, analysis, and visualization of experimental and observational data – the convergence of data and computing," Lawrence Berkeley National Laboratory, May 2016. doi: 10.2172/1525145.

[313] R. P. Mount, "The Office of Science Data-Management Challenge," Office of Scientific and Technical Information (OSTI), Oct. 2005. doi: 10.2172/878079.

[314] C. Xu, G. Sun, and R. Liang, "A survey of volume visualization techniques for feature enhancement," Vis. inform., vol. 5, no. 3, pp. 70–81, Sep. 2021.

[315] T. S. Newman and H. Yi, "A survey of the marching cubes algorithm," Comput. Graph., vol. 30, no. 5, pp. 854–879, Oct. 2006.

[316] P. C. Wong and R. D. Bergeron, "30 years of multidimensional multivariate visualization," Scientific Visualization, 1994, [Online]. Available: http://wwwx.cs.unc.edu/~taylorr/Comp715/papers/Wong97_30_years_of_multidimensional_multivariate_visualization.pdf

[317] R. Fuchs and H. Hauser, "Visualization of multi-variate scientific data," Comput. Graph. Forum, vol. 28, no. 6, pp. 1670–1690, Sep. 2009.

[318] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 2, pp. 423–443, Feb. 2019.

[319] L. Yan et al., "Scalar Field Comparison with Topological Descriptors: Properties and Applications for Scientific Visualization," arXiv [cs.HC], Jun. 01, 2021. [Online]. Available: http://arxiv.org/abs/2106.00157

[320] R. Bujack, L. Yan, I. Hotz, C. Garth, and B. Wang, "State of the art in time- dependent flow topology: Interpreting physical meaningfulness through mathematical properties," Comput. Graph. Forum, vol. 39, no. 3, pp. 811–835, Jun. 2020.

[321] C. Hergl et al., "Visualization of tensor fields in mechanics," Comput. Graph. Forum, vol. 40, no. 6, pp. 135–161, Sep. 2021.

[322] A. Inselberg, "Parallel Coordinates: Interactive Visualisation for High Dimensions," in Trends in Interactive Visualization: State-of-the-Art Survey, R. Liere, T. Adriaansen, and E. Zudilova-Seinstra, Eds. London: Springer London, 2009, pp. 49–78.

[323] J. Heinrich and D. Weiskopf, "State of the Art of Parallel Coordinates," in Eurographics (State of the Art Reports), 2013, pp. 95–116.

[324] M. C. Ferreira de Oliveira and H. Levkowitz, "From visual data exploration to visual data mining: a survey," IEEE Trans. Vis. Comput. Graph., vol. 9, no. 3, pp. 378–394, Jul. 2003.

[325] P. E. Hoffman and G. G. Grinstein, "A survey of visualizations for high-dimensional data mining," in Information visualization in data mining and knowledge discovery, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 47–82.

[326] D. A. Keim, "Information visualization and visual data mining," IEEE Trans. Vis. Comput. Graph., vol. 8, no. 1, pp. 1–8, Jan. 2002.

[327] A. Buja, D. Cook, and D. F. Swayne, "Interactive High-Dimensional Data Visualization," J. Comput. Graph. Stat., vol. 5, no. 1, pp. 78–99, Mar. 1996.

[328] R. Bürger and H. Hauser, "Visualization of multi-variate scientific data," in Eurographics (State of the Art Reports), 2007, pp. 117–134.

[329] W. W.-Y. Chan, "A survey on multivariate data visualization," Department of Computer Science and Engineering. Hong Kong University of Science and Technology, vol. 8, no. 6, pp. 1–29, 2006.

[330] T. Munzner, Visualization Analysis and Design. CRC Press, 2014.

[331] "[No title]." http://www.sci.utah.edu/~shusenl/highDimSurvey/website/ (accessed Jan. 08, 2022).

[332] A. Kamal et al., "Recent advances and challenges in uncertainty visualization: a survey," J. Vis. , May 2021, doi: 10.1007/s12650-021-00755-1.

[333] L. Padilla, M. Kay, and J. Hullman, "Uncertainty visualization," 2020, [Online]. Available: https://psyarxiv.com/ebd6r/download?format=pdf

[334] J. Wang, S. Hazarika, C. Li, and H.-W. Shen, "Visualization and Visual Analysis of Ensemble Data: A Survey," IEEE Trans. Vis. Comput. Graph., vol. 25, no. 9, pp. 2853–2872, Sep. 2019.

[335] A. Offenwanger, A. J. Milligan, M. Chang, J. Bullard, and D. Yoon, "Diagnosing Bias in the Gender Representation of HCI Research Participants: How it Happens and Where We Are," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, New York, NY, USA: Association for Computing Machinery, 2021, pp. 1–18.

[336] J. Bergström, T.-S. Dalsgaard, J. Alexander, and K. Hornbæk, "How to evaluate object selection and manipulation in VR? Guidelines from 20 years of studies," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama Japan, May 2021. doi: 10.1145/3411764.3445193.

[337] L. Mamykina et al., “Scaling Up HCI Research: from Clinical Trials to Deployment in the Wild,” in Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, New York, NY, USA: Association for Computing Machinery, 2021, pp. 1–6.

[338] H. Lakkaraju, J. Adebayo, and S. Singh, “Explaining machine learning predictions: State-of-the-art, challenges, and opportunities,” NeurIPS Tutorial, 2020.

[339] X.-H. Li et al., “A survey of data-driven and knowledge-aware explainable ai,” IEEE Trans. Knowl. Data Eng., 2020, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9050829/?casa_token=gQARnw l2LfwAAAAA:1FJTrm9P8CrwJDlBeaN-olfolnbMeaBxJZRE3XNgUsJWdRkKvRCDIcd6cGlmzLMyGSJzdoo

[340] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K.-R. Müller, “Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications,” Proc. IEEE, vol. 109, no. 3, pp. 247–278, Mar. 2021.

[341] J. Choo and S. Liu, “Visual Analytics for Explainable Deep Learning,” IEEE Comput. Graph. Appl., vol. 38, no. 4, pp. 84–92, 2018.

[342] D. Sacha, M. Kraus, D. A. Keim, and M. Chen, “VIS4ML: An Ontology for Visual Analytics Assisted Machine Learning,” IEEE Trans. Vis. Comput. Graph., Aug. 2018, doi: 10.1109/TVCG.2018.2864838.

[343] F. Doshi-Velez and B. Kim, “Towards A Rigorous Science of Interpretable Machine Learning,” arXiv [stat.ML], Feb. 28, 2017. [Online]. Available: http://arxiv.org/abs/1702.08608

[344] S. Mohseni, N. Zarei, and E. D. Ragan, “A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems,” ACM Trans. Interact. Intell. Syst., vol. 11, no. 3–4, pp. 1–45, Aug. 2021.

[345] M. L. Nelson, “Data-driven science: A new paradigm,” Educause Review, 2009, [Online]. Available: https://scholar.google.ca/scholar?cluster=9959993130008725069&hl=en&as_sdt=0,5&sciodt=0,5

[346] X. Zuo, Y. Chen, L. Ohno-Machado, and H. Xu, “How do we share data in COVID-19 research? A systematic review of COVID-19 datasets in PubMed Central Articles,” Brief. Bioinform., vol. 22, no. 2, pp. 800–811, Mar. 2021.

[347] D. Agrawal et al., “RHEEM: enabling cross-platform data processing: may the big data be with you!,” Proceedings VLDB Endowment, vol. 11, no. 11, pp. 1414–1427, Jul. 2018.

[348] J. Traub, Z. Kaoudi, J.-A. Quiané-Ruiz, and V. Markl, “Agora: Bringing Together Datasets, Algorithms, Models and More in a Unified Ecosystem [Vision],” SIGMOD Rec., vol. 49, no. 4, pp. 6–11, Mar. 2021.

[349] S. L. Brunton and J. Nathan Kutz, Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control. Cambridge University Press, 2019.

[350] X. Yan et al., “A Spatial-Temporal Interpretable Deep Learning Model for improving interpretability and predictive accuracy of satellite-based PM2.5,” Environ. Pollut., vol. 273, p. 116459, Jan. 2021.

[351] G. Iadarola, F. Martinelli, F. Mercaldo, and A. Santone, “Towards an interpretable deep learning model for mobile malware detection and family identification,” Comput. Secur., vol. 105, p. 102198, Jun. 2021.

[352] G. Singh and K.-C. Yow, “An Interpretable Deep Learning Model for Covid-19 Detection With Chest <italic>X</italic>-Ray Images,” IEEE Access, vol. 9, pp. 85198–85208, 2021.

[353] K.-L. Ma, “Intelligent Visualization Interfaces,” in 26th International Conference on Intelligent User Interfaces, New York, NY, USA: Association for Computing Machinery, 2021, pp. 2–3.

[354] B. Memarian, C. Koulas, and B. Fisher, “Novel Display Design for Spatial Assessment in Virtual Environments,” Proc. Hum. Fact. Ergon. Soc. Annu. Meet., vol. 65, no. 1, pp. 1437–1442, Sep. 2021.

[355] L. Bu, C.-H. Chen, K. K. H. Ng, P. Zheng, G. Dong, and H. Liu, “A user-centric design approach for smart product-service systems using virtual reality: A case study,” J. Clean. Prod., vol. 280, p. 124413, Jan. 2021.

[356] D. Hamilton, J. McKechnie, E. Edgerton, and C. Wilson, “Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design,” Journal of Computers in Education, vol. 8, no. 1, pp. 1–32, Mar. 2021.

[357] P. Araiza-Alba, T. Keane, W. S. Chen, and J. Kaufman, “Immersive virtual reality as a tool to learn problem-solving skills,” Comput. Educ., vol. 164, p. 104121, Apr. 2021.

[358] R. Q. Mao et al., “Immersive Virtual Reality for Surgical Training: A Systematic Review,” J. Surg. Res., vol. 268, pp. 40–58, Dec. 2021.

[359] M. Mirzaei, P. Kán, and H. Kaufmann, “Head Up Visualization of Spatial Sound Sources in Virtual Reality for Deaf and Hard-of-Hearing People,” in 2021 IEEE Virtual Reality and 3D User Interfaces (VR), Mar. 2021, pp. 582–587.

[360] B. Lee, P. Isenberg, N. H. Riche, and S. Carpendale, “Beyond Mouse and Keyboard: Expanding Design Considerations for Information Visualization Interactions,” IEEE Trans. Vis. Comput. Graph., vol. 18, no. 12, pp. 2689–2698, Dec. 2012.

[361] L. Besançon, A. Ynnerman, D. F. Keefe, L. Yu, and T. Isenberg, “The state of the art of spatial interfaces for 3D visualization,” Comput. Graph. Forum, vol. 40, no. 1, pp. 293–326, Feb. 2021.

[362] P. Goffin, T. Blascheck, P. Isenberg, and W. Willett, “Interaction Techniques for Visual Exploration Using Embedded Word-Scale Visualizations,” in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–13.

[363] E. Loper and S. Bird, “NLTK: The Natural Language Toolkit,” in Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics, 2002, pp. 63–70.

[364] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, “The Stanford CoreNLP natural language processing toolkit,” in Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, 2014, pp. 55–60.

[365] J. R. Finkel, T. Grenager, and C. D. Manning, “Incorporating non-local information into information extraction systems by gibbs sampling,” in Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL’05), 2005, pp. 363–370.

[366] E. Hoque, V. Setlur, M. Tory, and I. Dykeman, “Applying Pragmatics Principles for Interaction with Visual Analytics,” IEEE Trans. Vis. Comput. Graph., vol. 24, no. 1, pp. 309–318, Jan. 2018.

[367] V. Setlur, M. Tory, and A. Djalali, “Inferencing underspecified natural language utterances in visual analysis,” in Proceedings of the 24th International Conference on Intelligent User Interfaces, Marina del Ray, California, Mar. 2019, pp. 40–51.

[368] V. Setlur, S. E. Battersby, M. Tory, R. Gossweiler, and A. X. Chang, “Eviza: A Natural Language Interface for Visual Analysis,” in Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, Japan, Oct. 2016, pp. 365–377.

[369] A. Srinivasan, B. Lee, N. Henry Riche, S. M. Drucker, and K. Hinckley, “InChorus: Designing Consistent Multimodal Interactions for Data Visualization on Tablet Devices,” in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–13.

[370] R. Sicat et al., “DXR: A Toolkit for Building Immersive Data Visualizations,” IEEE Trans. Vis. Comput. Graph., vol. 25, no. 1, pp. 715–725, Jan. 2019.

[371] A. Narechania, A. Srinivasan, and J. Stasko, “NL4DV: A Toolkit for Generating Analytic Specifications for Data Visualization from Natural Language Queries,” IEEE Trans. Vis. Comput. Graph., vol. 27, no. 2, pp. 369–379, Feb. 2021.

[372] H. Chung and C. North, “SAViL: cross-display visual links for sensemaking in display ecologies,” Pers. Ubiquit. Comput., vol. 22, no. 2, pp. 409–431, Apr. 2018.

[373] R. Langner, M. Satkowski, W. Büschel, and R. Dachselt, “MARVIS: Combining Mobile Devices and Augmented Reality for Visual Data Analysis,” in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, New York, NY, USA: Association for Computing Machinery, 2021, pp. 1–17.

[374] P. Reipschläger and R. Dachselt, “DesignAR: Immersive 3D-Modeling Combining Augmented Reality with Interactive Displays,” in Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces, Daejeon, Republic of Korea, Nov. 2019, pp. 29–41.

[375] T. Sun, Y. Ye, I. Fujishiro, and K.-L. Ma, “Collaborative Visual Analysis with Multi-level Information Sharing Using a

Wall-Size Display and See-Through HMDs," in 2019 IEEE Pacific Visualization Symposium (PacificVis), Apr. 2019, pp. 11–20.

[376] J. Zagermann, U. Pfeil, P. von Bauer, D. Fink, and H. Reiterer, "'It's in my other hand!' Studying the Interplay of Interaction Techniques and Multi-Tablet Activities," in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu HI USA, Apr. 2020. doi: 10.1145/3313831.3376540.

[377] R. Skarbez, N. F. Polys, J. T. Ogle, C. North, and D. A. Bowman, "Immersive Analytics: Theory and Research Agenda," Front Robot AI, vol. 6, p. 82, Sep. 2019.

[378] "Office of Science User Facilities." https://www.energy.gov/science/science-innovation/office-science-user-facilities (accessed Jan. 08, 2022).

[379] "DOE Exascale Requirements Review." https://exascaleage.org/bes/ (accessed Jan. 08, 2022).

[380] "DOE Exascale Requirements Review." https://exascaleage.org/hep/ (accessed Jan. 08, 2022).

[381] R. Gerber et al., "Crosscut report: Exascale requirements reviews, march 9--10, 2017--tysons corner, virginia. an office of science review sponsored by: Advanced scientific computing research, basic energy sciences, biological and environmental research, fusion energy sciences, high energy physics, nuclear physics," Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States); Argonne ..., 2018. [Online]. Available: https://www.osti.gov/biblio/1417653

[382] E. W. Bethel, M. Greenwald, K. K. Van Dam, M. Parashar, S. M. Wild, and H. S. Wiley, "Management, analysis, and visualization of experimental and observational data—the convergence of data and computing," in 2016 IEEE 12th International Conference on e-Science (e-Science), 2016, pp. 213–222.

[383] "Website." https://matplotlib.org/

[384] J. Ahrens, B. Geveci, and C. Law, "Paraview: An end-user tool for large data visualization," The visualization handbook, vol. 717, no. 8, 2005, [Online]. Available: https://datascience.dsscale.org/wp-content/uploads/2016/06/ParaView.pdf

[385] O. Rübel, E. W. Bethel, Prabhat, Wu, and Keshang, "Query-Driven Visualization and Analysis," in High Performance Visualization—Enabling Extreme-Scale Scientific Insight, E. W. Bethel, H. Childs, and C. Hansen, Eds. CRC Press, 2012, pp. 117–144.

[386] K. Stockinger, J. Shalf, K. Wu, and E. W. Bethel, "Query-driven visualization of large data sets," in VIS 05. IEEE Visualization, 2005., Oct. 2005, pp. 167–174.

[387] G. H. Weber et al., "Topology-based Visualization and Analysis of High-dimensional Data and Time-varying Data at the Extreme Scale," presented at the DOE Exascale Research Conference, Portland, OR, 2012.

[388] M. Kerber, D. Morozov, and A. Nigmetov, "Geometry Helps to Compare Persistence Diagrams," ACM J. Exp. Algorithmics, vol. 22, pp. 1–20, Sep. 2017.

[389] O. Rübel, C. G. R. Geddes, M. Chen, E. Cormier-Michel, and E. W. Bethel, "Feature-based analysis of plasma-based particle acceleration data," IEEE Trans. Vis. Comput. Graph., vol. 20, no. 2, pp. 196–210, Feb. 2014.

[390] F. Cappello, S. Di, and A. M. Gok, "Fulfilling the Promises of Lossy Compression for Scientific Applications," in Driving Scientific and Engineering Discoveries Through the Convergence of HPC, Big Data and AI, 2020, pp. 99–116.

[391] S. Di and F. Cappello, "Fast Error-Bounded Lossy HPC Data Compression with SZ," in 2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS), May 2016, pp. 730–739.

[392] J. Clyne, P. Mininni, A. Norton, and M. Rast, "Interactive desktop analysis of high resolution simulations: application to turbulent plume dynamics and current sheet formation," New J. Phys., vol. 9, no. 8, p. 301, Aug. 2007.

[393] G. Ballard, A. Klinvex, and T. G. Kolda, "TuckerMPI: A Parallel C++/MPI Software Package for Large-scale Data Compression via the Tucker Tensor Decomposition," ACM Trans. Math. Softw., vol. 46, no. 2, pp. 1–31, Jun. 2020.

[394] M. Ainsworth, O. Tugluk, B. Whitney, and S. Klasky, "Multilevel techniques for compression and reduction of scientific data—the univariate case," Comput. Vis. Sci., vol. 19, no. 5–6, pp. 65–76, Dec. 2018.

[395] A. C. Bauer et al., "In Situ Methods, infrastructures, and applications on high performance computing platforms," Comput. Graph. Forum, vol. 35, no. 3, pp. 577–597, Jun. 2016.

[396] U. Ayachit et al., "Performance Analysis, Design Considerations, and Applications of Extreme-Scale In Situ Infrastructures," in SC '16: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, Nov. 2016, pp. 921–932.

[397] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic, "A Task-Based Taxonomy of Cognitive Biases for Information Visualization," IEEE Trans. Vis. Comput. Graph., vol. 26, no. 2, pp. 1413–1432, Feb. 2020.

[398] E. Wall, L. M. Blaha, L. Franklin, and A. Endert, "Warning, Bias May Occur: A Proposed Approach to Detecting Cognitive Bias in Interactive Visual Analytics," 2017 IEEE Conference on Visual Analytics Science and Technology (VAST). 2017. doi: 10.1109/vast.2017.8585669.

[399] T. Zuk and S. Carpendale, "Visualization of Uncertainty and Reasoning," Smart Graphics. pp. 164–177. doi: 10.1007/978-3-540-73214-3_15.

[400] E. Dimara, G. Bailly, A. Bezerianos, and S. Franconeri, "Mitigating the Attraction Effect with Visualizations," IEEE Trans. Vis. Comput. Graph., Aug. 2018, doi: 10.1109/TVCG.2018.2865233.

[401] J. Schwabish and A. Feng, "Applying racial equity awareness in data visualization," Aug. 27, 2020. doi: 10.31219/osf.io/x8tbw.

[402] A. Ottley, Adaptive and Personalized Visualization. Morgan & Claypool Publishers, 2020.

[403] C. Stoiber, C. Walchshofer, F. Grassinger, H. Stitz, M. Streit, and W. Aigner, "Design and comparative evaluation of visualization onboarding methods," in The 14th International Symposium on Visual Information Communication and Interaction, Potsdam Germany, Sep. 2021. doi: 10.1145/3481549.3481558.

[404] U. Schlegel and D. A. Keim, "Time series model attribution visualizations as explanations," in 2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX), New Orleans, LA, USA, Oct. 2021. doi: 10.1109/trex53765.2021.00010.

[405] Q. Wang, Z. Li, S. Fu, W. Cui, and H. Qu, "Narvis: Authoring Narrative Slideshows for Introducing Data Visualization Designs," IEEE Trans. Vis. Comput. Graph., Aug. 2018, doi: 10.1109/TVCG.2018.2865232.

[406] L. South, M. Schwab, N. Beauchamp, L. Wang, J. Wihbey, and M. A. Borkin, "DebateVis: Visualizing political debates for non-expert users," in 2020 IEEE Visualization Conference (VIS), Salt Lake City, UT, USA, Oct. 2020. doi: 10.1109/vis47514.2020.00055.

[407] S. Afzal, S. Ghani, H. C. Jenkins-Smith, D. S. Ebert, M. Hadwiger, and I. Hoteit, "A visual analytics based decision making environment for COVID-19 modeling and visualization," in 2020 IEEE Visualization Conference (VIS), Salt Lake City, UT, USA, Oct. 2020. doi: 10.1109/vis47514.2020.00024.

[408] T. McDonald et al., "Improving the Usability of Virtual Reality Neuron Tracing with Topological Elements," IEEE Trans. Vis. Comput. Graph., vol. 27, no. 2, pp. 744–754, Feb. 2021.

[409] J. J. Thomas, Illuminating the Path: The Research and Development Agenda for Visual Analytics. 2005.

[410] E. Dimara and J. Stasko, "A Critical Reflection on Visualization Research: Where Do Decision Making Tasks Hide?," IEEE Trans. Vis. Comput. Graph., vol. 28, no. 1, pp. 1128–1138, Jan. 2022.

[411] A. Dommer et al., "#COVIDisAirborne: AI-enabled multiscale computational microscopy of delta SARS-CoV-2 in a respiratory aerosol," Int J High Perform Comput Appl., vol. 37, no. 1, pp. 28–44, Jan. 2023.

[412] P.-T. Bremer et al., "Analyzing and Tracking Burning Structures in Lean Premixed Hydrogen Flames," IEEE Trans. Vis. Comput. Graph., vol. 16, no. 2, pp. 248–260, Mar-Apr 2010, doi: 10.1109/TVCG.2009.69.

[413] D. Laney et al., "Understanding the Structure of the Turbulent Mixing Layer in Hydrodynamic Instabilities," IEEE Trans. Vis. Comput. Graph., vol. 12, no. 05, pp. 1053–1060, 2006. doi: 10.1109/TVCG.2006.186.

[414] P. Klacansky et al., "Virtual Inspection of Additively Manufactured Parts," 2022 IEEE Visualization Symposium (PacificVis), 2022, pp. 81–90, doi: 10.1109/PacificVis53943.2022.00017.

Appendix E - Image Credits

Cover	Lawrence Livermore National Laboratory; see also [412][413]
1.1	Workshop committee
2	VisIt team (visit-dav.github.io/visit-website/index.html)
3.1.1	MFEM team (mfem.org/gallery)
3.1.2a	Lawrence Livermore National Laboratory; see also [412][413]
3.1.2b	See [36]
3.15	Kenny Gruchalla, National Renewable Energy Laboratory
3.2.2a	Ascent team (data-science.llnl.gov/latest/news/successful-simulation-visualization-coupling-proves-power-sierra)
3.2.2b	Jian Huang, University of Tennessee; see also [121]
3.2.3	ViSUS team (visus.org)
3.3.2	Joseph Smidt, Brandon Wiggins, and Francesca Samsel, Texas Advanced Computing Center
3.3.3a	John De La Rosa, National Renewable Energy Laboratory
3.3.3b	See [414]
3.4.1	Visualization by Francesca Samsel at Texas Advanced Computing Center/University of Texas at Austin; combustion data from Matt Larsen at Lawrence Livermore National Laboratory
3.4.2a	National Energy Research Scientific Computing Center
3.4.2b	See [411]; created by Lorenzo Casalino, Amaro Lab, UC San Diego, for the #COVIDisAirborne Team including DOE Laboratories
3.5	Adobe Stock
3.5.2	AdlerPlanetarium.org
A	Emily Belli, General Atomics