Multi-Fidelity Methods for Fusion Energy

IPAM Long Program, Spring 2026

Victor Artigues, Max-Planck-Institut für Plasmaphysik
Eric Chirtel, University of Arizona
Rory Conlin, University of Maryland
Ionut-Gabriel Farcas, Virginia Polytechnic Institute and State University
Gianluca Geraci, Sandia National Laboratories
Kevin Gill, University of Wisconsin-Madison
Matthew Golden, Los Alamos National Laboratory
Wei Guo, Texas Tech University
Ammar Hakim, Princeton Plasma Physics Laboratory
Jeffrey Hittinger, Lawrence Livermore National Laboratory
George K. Holt, United Kingdom Atomic Energy Authority
Lise-Marie Imbert-Gerard, University of Arizona
Frank Jenko, Max-Planck-Institut für Plasmaphysik / University of Texas at Austin
Boris Kramer, University of California San Diego
Gabriele Merlo, Max-Planck-Institut für Plasmaphysik
Shancong Mou, University of Minnesota Twin Cities
Elizabeth Paul, Columbia University in the City of New York
Uri Shumlak, University of Washington
Phil Travis, Ergodic LLC
Genia Vogman, Lawrence Livermore National Laboratory
Tomo-Hiko Watanabe, Nagoya University
Tim Wildey, Sandia National Laboratories
Adelle Wright, University of Wisconsin-Madison
Ben Zhu, Columbia University in the City of New York

Note: This document was prepared with the assistance of AI tools. These tools were used to help copyedit this white paper. All scientific content reflects the work and review of the named authors.

Table of Contents

1 Executive Summary

The Institute for Pure and Applied Mathematics at UCLA hosted a Long Program on "Multi-Fidelity Methods for Fusion Energy" from March 9 to June 12, 2026. This program brought together researchers from traditionally separate communities – high-fidelity numerical modeling, reduced-order and data-driven surrogate modeling, and multi-fidelity methodology development – to forge new collaborations and accelerate progress toward commercially viable fusion power.

Fusion energy represents one of the most complex scientific and engineering challenges, characterized by multi-scale, multi-physics problems spanning wide ranges of physical and temporal scales. First-principles, high-fidelity simulations require the world's largest supercomputers running for days or weeks to simulate a single plasma realization, severely restricting essential many-query tasks such as design optimization, uncertainty quantification (UQ), and real-time control.

Multi-Fidelity Methods (MFM) systematically leverage models of varying precision, accuracy, resolution, and physical completeness to accelerate convergence and reduce time-to-solution. These approaches have demonstrated computational cost reductions of orders of magnitude across diverse scientific domains. Fusion energy is well-positioned to benefit given its existing ecosystem of models at various fidelities and the urgent need for commercial deployment.

The program addressed critical research areas, beginning with the integration of traditional plasma physics hierarchies with modern data-driven approaches, including proper orthogonal decomposition, dynamic mode decomposition, neural networks, and foundation models. Participants explored verification, validation, and UQ methodologies for trustworthy modeling, differentiable programming, tensor networks, and formal correctness verification. The program examined multi-scale coupling, integrated systems engineering, and Bayesian optimization with high-fidelity metrics for device design. Discussions covered model-based control architectures, sparse-diagnostic state estimation, and controller orchestration for real-time plasma control. Finally, the program investigated AI agents for autonomous simulation, code development, and scientific discovery, demonstrating their potential to transform fusion research workflows.

This program identified opportunities for multi-fidelity methods to transform fusion modeling to achieve robust, computationally tractable frameworks for design, optimization, and control, thereby accelerating the path to commercial fusion energy.

2 Introduction

Fusion energy presents one of the most formidable scientific and engineering challenges of our time, with significant societal implications. The field is characterized by complex, nonlinear, multi-scale, multi-physics problems in which fundamental aspects of the underlying physics remain incompletely understood. Accurate simulation of fusion systems requires capturing the intricate nonlinear interactions among a large number of plasma processes – from radio-frequency heating to small-scale turbulence, large-scale magnetohydrodynamic (MHD) instabilities, neutral-particle physics, and plasma-facing materials – spanning many orders of magnitude in space and time scales. This multi-scale behavior places severe constraints on both theoretical frameworks and computational approaches.

Consequently, both analytical theory and numerical simulation face significant limitations. Even when restricted to fully ionized plasma confinement, first-principles modeling requires the world's largest supercomputers running for days or weeks to simulate a single plasma realization. This expense severely restricts the many-query tasks essential for uncertainty quantification (UQ), design optimization, and real-time control. The fusion community has therefore come to rely on a diverse ensemble of models at various fidelity levels to understand, predict, and design fusion reactor concepts. Historically, however, these models have been developed in relative isolation, with at most simplistic frameworks for combining their complementary strengths.

New opportunities are now emerging to accelerate progress toward commercial viability. The most visible is the rise of data-driven models — from projection-based reduced-order models to deep-learning approaches — increasingly deployed as fast surrogates for computationally expensive sub-models. Beyond individual surrogates lies the more powerful concept of systematically leveraging models of multiple fidelities, which can dramatically reduce time-to-solution for high-fidelity forward modeling. More significantly, rigorous frameworks from statistics and applied mathematics use collections of models at differing fidelities to rapidly estimate sensitivities and uncertainties for design optimization and control. These methods have demonstrated remarkable success across diverse scientific domains, addressing previously cost-prohibitive problems while achieving computational savings of several orders of magnitude.

Fusion energy is particularly well-positioned to benefit: its existing ecosystem of models at various fidelities provides a natural foundation, while the urgency of commercial deployment creates strong motivation for adoption. By exploiting complementary information across models, multi-fidelity approaches have the potential to make UQ, optimization, and control tractable for increasingly realistic fusion applications.

Realizing this potential, however, requires both methodological advances and close collaboration among three communities that have historically operated in relative isolation: high-fidelity modeling (spanning theoretical model development and numerical methods), low-fidelity modeling (reduced models, reduced-order models, and machine learning surrogates), and multi-fidelity methodology development in statistics and applied mathematics.

The importance and timeliness of these topics are underscored by their strong alignment with the recently released *Fusion Science and Technology Roadmap* (FS&T Roadmap) and supporting community reports commissioned by the U.S. Department of Energy (DOE) Office of Science, Office of Fusion Energy Sciences (FES). The Roadmap highlights the crucial role of AI in accelerating fusion research and of cutting-edge methods and algorithms for predictive, trustworthy simulation tools. At the task level, the activities of this program represent concrete steps toward the recommendations of the DOE FES-sponsored TEAMS and STARS workshops, developed concurrently with the Roadmap.

This document summarizes the activities and outcomes of the Long Program "Multi-Fidelity Methods for Fusion Energy," held at the Institute for Pure and Applied Mathematics (IPAM) from March 9 to June 12, 2026. The program brought together researchers from these traditionally separate communities to forge new collaborations, share expertise, and chart a path toward realizing the potential of multi-fidelity approaches in fusion energy. This report presents the open questions and future research directions that emerged from the program, organized into six sections:

- Physics-Based and Data-Driven Models in Fusion
- Multi-Fidelity Methods for Reducing and Characterizing Uncertainties
- Numerics & Computational Acceleration Supporting Multi-Fidelity Methods
- Multi-Fidelity Methods for Fusion Device Design & Optimization
- Multi-Fidelity Methods for Real-Time Control
- Agentic Workflows to Enable Multi-Fidelity Methods for Autonomous Scientific Discovery

It concludes with a discussion of future activities identified by participants to disseminate these advances across multiple communities, further advancing the development and integration of these methods within fusion energy science.

3 Physics-Based and Data-Driven Models in Fusion

Understanding, predicting, and designing fusion experiments has led to the development of physics-based and data-driven models. Together, they form an ensemble for MFM, which combine models of differing cost and accuracy whether or not they form a model hierarchy. Controlled fusion experiments span many orders of magnitude in both time and length scales. Coupling across these scales is essential to model the global behavior of experiments and remains an outstanding modeling challenge. Fundamental assumptions define each physics model's validity, and within an experiment the validity domains can vary both in time and space, making whole-device modeling particularly challenging. There are also inherent challenges in handling data, related to both quality and availability.

3.1 Physics-Based Model Hierarchies

Fusion plasmas are examples of multi-scale, multi-physics systems. They comprise multiple charged and neutral species that interact with externally imposed and self-generated electromagnetic fields, while simultaneously undergoing elastic scattering collisions and inelastic reactions that add energy and source new species. No single model resolves this full range of behavior at tractable cost, so the field relies on a hierarchy of physics-based models that trade fidelity for computational economy.

Ordered roughly from highest to lowest fidelity, the physics-based plasma models are the kinetic, gyrokinetic (GK), moment, multi-fluid, and MHD descriptions, followed by quasilinear, analytical, and scaling-law reductions. The dynamics of species and fields are most accurately captured by the multi-species kinetic model coupled to Maxwell's equations, which resolves the full distribution function in phase space and may be collisionless, collisional, or reactive depending on the regime and specific physics of interest. This constitutes the highest-fidelity tier of the hierarchy. Lower-fidelity physics-based models are often obtained from the kinetic description either by taking velocity moments and introducing appropriate closures, or by applying asymptotic reductions tailored to the regime of interest.

Hybrid models are also possible; they combine descriptions of differing fidelity for different species or phase-space directions. Examples include kinetic ions with fluid electrons, and anisotropic closures such as parallel-kinetic, perpendicular-moment models. Such constructions let practitioners place fidelity where the physics demands it while economizing elsewhere. Boundary conditions, sheath formation, plasma and neutral sourcing, plasma-material interactions, direct current drive, and magnetic flux injection introduce additional modeling choices that cut across every tier of the hierarchy.

Beyond reducing the governing equations, fidelity is also managed by restricting the scope of a simulation. Characteristic spatial scales in fusion plasmas span from the electron Debye length to the machine size, and temporal scales from the plasma frequency to the confinement time. Direct numerical simulation of an entire device across all these scales is generally prohibitive, so models are restricted to specific processes (domain separation) or to a limited band of scales (scale separation). Scale separation is a cost-reduction strategy, but it sacrifices fidelity. Related strategies include spectral separation and stage separation, in which a process is decomposed temporally and earlier phases are folded implicitly into initial or boundary conditions.

The physics model hierarchy provides a rich, well-understood spectrum of cost and fidelity. However, its members share a common limitation: each is reached by a priori approximation, and the fidelity attainable at a given cost is fixed by the physics that the reduction chose to keep.

3.2 Data-Driven Reduced Models

Data-driven reduced models relax the physics model constraint, learning a reduced description from data, such as high-fidelity simulation output, experimental measurements, or both, thereby aiming to capture structure that closed-form reductions omit. Their reliance on data makes their reach contingent on data availability and quality, the subject of the following section.

Data-driven methods range from classical projection-based reductions to large deep-learning-based architectures. Classical reductions decompose simulation data into a small set of dominant spatial modes, e.g., proper orthogonal decomposition (POD) and dynamic mode decomposition (DMD). Tensor network methods exploit low-rank structures in high-dimensional phase space to solve continuum kinetic equations directly while retaining essential non-equilibrium effects. Sparse grids can be used to efficiently construct surrogates for the input-to-quantity-of-interest mapping. Operator learning architectures aim to learn mappings between function spaces. Regression-based surrogates include Gaussian processes (GPs) and tabular models. Convolutional and transformer-based networks process plasma data as high-dimensional fields with local and long-range structure. Generative models, including diffusion and flow-matching methods, learn to transport a simple distribution into the data distribution and yield stochastic realizations rather than single point estimates. Foundation models, trained in a self-supervised fashion on broad data, could in principle anchor a generalizable fusion surrogate or digital twin within multi-fidelity workflows, which is contingent on data at a scale the fusion field must work to extend.

Across fusion modeling, these methods appear most often as surrogates that aim to replace or accelerate an expensive computation. Data-driven closures learn subgrid terms directly from simulation or experiment, as with convolutional neural networks (CNN) modeling plasma behavior below the grid scale in magnetic reconnection, or structure-preserving moment closures that reduce the Vlasov and radiative-transfer equations while preserving hyperbolicity.

3.3 Data Availability, Quality, Curation, and Open Sourcing

Data availability, quality, and quantity vary tremendously across fusion concepts, and the gap widens for concepts that operate fewer devices, e.g., alternative concepts and inertial confinement fusion (ICF). Even where data are plentiful, their transferability is limited: results from one device may not translate cleanly to another, even to a device of the same configuration. As a result, data is fragmented by concept, by machine, and by institution.

Several barriers compound this fragmentation. Practically, the data can be expensive to generate, demand large storage, and arrive sparse, noisy, or biased in parameter space, often without the metadata needed to interpret or reuse it. Institutionally, sharing may be restricted by intellectual-property and national-security considerations, by machine-specific formats that complicate combining, and by the reluctance of groups to release data before they have exhausted its scientific value. Codes face parallel obstacles, where availability and reproducibility hinge on version control and documentation that are not always well maintained.

The contrast with other scientific fields is instructive. Large open datasets have repeatedly seen AI-enabled paradigm shifts: the Protein Data Bank underpinned AlphaFold, and the ERA5 reanalysis underpinned GraphCast. Fusion has no cross-device equivalent at the scale that foundation models require, but deliberate effort could build one. Recent efforts to exploit entire datasets for a single machine (MAST), resulting in the creation of TokaMind, a fusion foundation model, have shown promise, but extension to multiple devices is an urgent next step. Emerging resources such as the Integrated Modelling and Analysis Suite (IMAS) as a data standard and FAIR-MAST as an open experimental database offer models, as do calls to open-source new simulation and diagnostic data by default and to treat the codes that produce it as shared infrastructure. Adopting FAIR principles (findable, accessible, interoperable, reusable) as a guiding paradigm, rather than committing to any single framework, offers a flexible standard for making progress. A consortium model with a sharing agreement and standardized formats could reconcile the public, private, and international stakeholders of the field.

Curation itself is increasingly tractable. Many historical barriers, such as poorly recorded data, disparate experimental logs, and missing or inconsistent metadata, stem from labor that was prohibitive to perform by hand. AI agents [Sec. 8] can automate much of this, labeling under-documented data, reconciling heterogeneous notes, and assigning metadata, thereby lowering the barrier to effective reuse of shared data.

3.4 Work Accomplished During Long Program

High-fidelity codes remain crucial in multi-fidelity settings as their computational cost ultimately determines how a given computational budget should be utilized. During the long program the graphics processing unit (GPU) port of GENE was further improved and used to populate a database of high fidelity simulations with the goal of training and comparing different surrogate models for the same downstream task. GPs, conditional flow matching (CFM) and the off-the-shelf AutoGluon library were employed to train an ensemble of models directly predicting the same turbulent fluxes, with the ultimate goal of embedding them in a bi-fidelity framework.

Quasilinear models are well-validated in tokamaks (e.g., TGLF, QuaLiKiz) but are harder to develop for stellarators. Toward closing this gap, we developed multi-mode quasilinear models via dynamic mode decomposition, benchmarking them for the Cyclone Base Case.

To support multi-fidelity workflows and faster scrape-off layer (SOL) simulations, data-driven models capable of handling multiple geometries and scans in fueling, impurity seeding, and transport coefficient profiles were further developed during the long program. Recent tabular foundation models proved performant as surrogates for scalar predictions such as divertor target heat loads, detachment front locations, and midplane densities.

Tensor network-based reduced-order models for efficient simulation of high-dimensional radiative transfer equations were developed.

As a middle fidelity approach to the turbulent transport, complementing the quasilinear model and nonlinear gyrokinetic theory, a reduced-order model of the nonlinear GK simulation was developed by means of the POD-Galerkin projection method, where linear growth, saturation, zonal flow generation, and nonlinear turbulent phases are reasonably reproduced.

Construction of a dataset was initiated for a surrogate model of hot electron generation and laser reflection caused by stimulated Raman scattering. This model will be incorporated into radiation-hydrodynamics codes to replace heuristics of hot electron generation and subsequent fuel preheat.

3.5 Outlook

Optimization and surrogate models are tightly linked: structure-exploiting training improves generalization and reduces costly simulations, while learned components can accelerate solvers by replacing repeated computations. Advancing these capabilities for fusion requires attention to both computational methods and workflow design.

- Embed structure-preserving constraints into surrogate training to maintain stability when coupled to physics solvers.

- Expand GPU-enabled high-fidelity simulation campaigns to populate training databases across key fusion regimes.
- Benchmark competing surrogate approaches on common datasets and tasks to identify regime-dependent best practices.
- Develop multi-fidelity workflows that dynamically select fidelity, quantify uncertainty and trigger high-fidelity calculations when needed [Sec. 8].

Realizing this vision will require robust data infrastructure alongside algorithmic advances. Shared, curated datasets, interoperable standards, open-source codes, benchmarking tasks, and coordination across public, private and international stakeholders are prerequisites for methods that generalize across devices and concepts.

- Adopt FAIR principles as a community-wide standard, guided by resources such as IMAS and FAIR-MAST.
- Establish cross-device benchmark datasets enabling direct comparison of surrogate and reduced-order models.
- Deploy AI agents to automate curation of under-documented and historical data [Sec. 8].
- Maintain version control, documentation, and open-source release of simulation and diagnostic codes by default.
- Scale foundation model initiatives from single-device to multi-device, multi-concept corpora for generalizable digital twin capabilities.

4 Multi-Fidelity Methods for Reducing and Characterizing Uncertainties

Fusion experiments and simulations involve multiple sources of uncertainty that affect outcomes and accuracy of predictions. Experimental uncertainties include instrument or detector noise, calibration errors, finite time resolution, shot-to-shot variability, and limited sample size. Simulation uncertainties include unknown input parameters, model misspecification, random collisions and turbulence, and numerical errors. Taking plasma microturbulence simulations as an example, key input parameters, including density and temperature gradients, collisionality, and magnetic geometry characteristics, often carry uncertainties of several tens of percent. Understanding how these uncertainties propagate to quantities of interest, such as turbulent heat and particle transport, is essential for developing predictive and uncertainty-aware modeling capabilities. More broadly, fusion applications rely on incomplete and imperfect data, and many problems involve high-dimensional parameter spaces. At the same time, the underlying forward models are often computationally expensive. These challenges motivate the development and adoption of systematic verification, validation, and uncertainty quantification (VVUQ) methods, as identified in the FS&T Roadmap, along with multi-fidelity approaches that can efficiently leverage information from models of varying fidelity and cost. Such methods are essential for characterizing model and system properties, assessing predictive capability, and enabling trustworthy predictions for future fusion devices.

4.1 Multi-fidelity Acceleration Strategies for Forward Uncertainty Propagation

Most multi-fidelity sampling-based approaches for forward UQ rely on variance-reduction techniques based on control variates. In this setting, low-fidelity models are used to reduce the variance of Monte Carlo (MC) estimators. Classical approaches such as Multi-Level MC (MLMC) and Multi-Fidelity MC (MFMC) exploit hierarchical model sequences, whereas more general frameworks, including Approximate Control Variates (ACV) and Multi-Level Best Linear Unbiased Estimators (MLBLUE), can accommodate arbitrary model relationships. The efficiency of these methods depends primarily on the computational cost of the low-fidelity models and their statistical relationships with the high-fidelity model and with one another. Multi-fidelity surrogate-based methods seek to learn and exploit structure between low- and high-fidelity models. The underlying assumption is that low-fidelity models capture the dominant features of the input-output mapping, while a limited amount of high-fidelity data is used to correct residual discrepancies. A common challenge across both classes of methods is the optimal allocation of computational resources. For sampling-based approaches, allocation strategies can often be derived analytically from estimator variance and cost models. In contrast, surrogate-based methods typically rely on adaptive sampling, active learning, or greedy refinement strategies to determine where additional high-fidelity information is most valuable.

Although multi-fidelity methodologies have been successfully applied in certain areas pertaining to fusion research (e.g., plasma micro-instability simulations in tokamak configurations), their broad adoption in fusion remains relatively limited. Realizing their full potential will require addressing several challenges, including the automated construction and adaptation of model hierarchies, the treatment of models with disparate parameterizations and outputs, and the integration of fundamentally different simulation paradigms, such as stochastic particle-based methods and deterministic grid-based solvers. More broadly, low-fidelity models should not be designed solely for standalone predictive accuracy. Instead, their construction should explicitly account for their role within a multi-fidelity workflow, balancing computational cost against the information they provide about the high-fidelity model.

4.2 Variance Reduction for Monte Carlo Forward Simulations

Many fusion simulation techniques rely on MC methods, including neutronics, particle-in-cell, and energetic particle simulations. A specific challenge of these applications includes characterizing rare events. For example, in energetic particle simulations, very few particles may impact regions of interest, such as experimental detectors or the first wall, requiring efficient sampling strategies to improve statistics. In neutronics applications, modeling deep-penetration radiation transport similarly requires modeling rare events. In such situations, importance sampling and other variance reduction techniques can reduce the number of required samples to reduce uncertainty in the simulated metrics within a given computational budget. Both importance sampling and variance reduction techniques can leverage multi-fidelity approaches for acceleration. Variance reduction techniques discussed above, including MFMC, have been applied to forward simulations in fusion, including for accelerating the estimation of energetic particle confinement in stellarators, using a data-driven interpolant as a low-fidelity model.

Importance sampling techniques leverage a biasing distribution to reduce the variance. Rather than sampling from the physical distribution of interest (e.g., the fusion birth distribution), samples are taken from a biasing distribution strategically chosen to reduce variance. Some methods, such as the cross-entropy method, iteratively solve for the optimal biasing distribution. Multi-fidelity importance sampling techniques use low-fidelity surrogate models to facilitate the construction of a biasing distribution for the high-fidelity model, such as using a low-fidelity model within the cross-entropy optimization loop.

4.3 Multi-Fidelity Methods for Inference and Inverse Problems

Inverse problems are widely utilized across computational science and engineering to inform model inputs, quantify uncertainties, and test assumptions. However, experimental data is often noisy due to both turbulence and measurement errors in fusion devices, and characterizing the impact of this uncertainty is critical if one seeks to make credible predictions.

Bayesian methods are widely regarded as the standard approach for probabilistic inference. They make it possible to combine expert or user knowledge, expressed through expert-informed prior distributions, with experimental or observational data. In the context of analysis of fusion diagnostics, the prior could represent physical constraints or information from other diagnostics. This provides a principled way to characterize epistemic uncertainty, which reflects incomplete knowledge of the true data-generating inputs. For characterizing aleatoric, or intrinsically random, uncertainty, hierarchical Bayesian methods are often used. These methods solve a parameterized inference problem where the distribution over model inputs is given by seeking the best match within a parameterized family of distributions.

Bayesian methods have become increasingly important in plasma physics and fusion research, particularly through integrated data analysis frameworks and data assimilation techniques that combine experimental measurements with physics-based models. Nevertheless, their application to high-fidelity models, such as gyrokinetic simulations of tokamaks and stellarators, remains challenging. The primary difficulties include the extreme computational cost of high-fidelity forward models, the poor scaling of Markov Chain Monte Carlo (MCMC) methods in high-dimensional parameter spaces, and the presence of model-form errors. For hierarchical Bayesian approaches, additional challenges arise from potential misspecification of the statistical model, limited robustness to model discrepancies, and the substantial computational burden associated with repeated model evaluations. These challenges present a significant opportunity for multi-fidelity methodologies to accelerate Bayesian inference while maintaining rigorous uncertainty estimates. However, the literature on MFM for inverse UQ remains considerably smaller than that for forward UQ. Multi-fidelity MCMC methods, such as delayed acceptance schemes, are now well established. More recently, multi-fidelity data assimilation approaches, including filtering and particle-based methods for state and parameter estimation, have emerged as promising techniques that leverage models and data sources of varying fidelity to improve computational efficiency while maintaining estimation accuracy.

4.4 Work Accomplished During Long Program

A collaboration between Virginia Tech and University of Colorado Boulder explored sparse-grid-based UQ and surrogate modeling for divertor heat-load predictions in tokamaks. In the CU Boulder modeling framework several inputs (e.g., blob width, amplitude, and spacing) are uncertain and require systematic analysis. Leveraging sparse-grid methodologies we obtained promising preliminary results, demonstrating efficient UQ and accurate surrogate construction with valuable insights into the relative importance of different model parameters. Incorporation of adaptive sparse-grid refinement and multi-fidelity methodologies will further improve efficiency and scalability. This collaboration is expected to lead to at least one joint publication, potentially as part of the special issue associated with the program.

As an alternative to hierarchical Bayesian approaches, one can utilize a measure-theoretic approach for characterizing aleatoric uncertainties. While such approaches are nonparametric and can be shown to converge to the true data-generating distribution under certain assumptions, most of the numerical demonstrations in the literature utilize suboptimal sampling strategies, such as importance sampling. While visiting IPAM, a new sampling methodology for measure-theoretic inversion was inspired and developed based on a multi-fidelity delayed acceptance algorithm. This new approach resulted in a 10x efficiency gain when used as a preconditioning strategy for an MCMC sampler. The results were presented at the Society for Industrial and Applied Mathematics (SIAM) Conference on Uncertainty Quantification in late March 2026.

4.5 Outlook

Systematically characterizing uncertainties is critical for assessing predictive capability, establishing confidence in simulations, and enabling trustworthy predictions for future fusion devices operating beyond currently accessible experimental regimes. To this end we advocate

- Experimental measurements of distribution functions on fusion platforms would provide critically needed validation for models that incorporate high-fidelity kinetic descriptions.
- Broader integration of VVUQ into fusion modeling workflows is essential for improving predictions, addressing model-form uncertainty, the impact of simplifying assumptions, and numerical errors arising from discretizations and solver approximations.
- Develop techniques that efficiently handle high-dimensional parameter spaces arising from complex magnetic geometries, profile uncertainties, and integrated modeling workflows (e.g., analysis of stellarator coil manufacturing and installation errors on plasma physics metrics).
- Develop MFM to enable moving beyond moment estimation toward approximating densities (cumulative distribution function (CDF), probability density function (PDF), etc.).

5 Numerics and Computational Acceleration Supporting Multi-Fidelity Methods

Simulation tools remain critical for fusion research and for probing physics that experiments cannot currently access. State-of-the-art numerical methods should be adopted and leveraged to best integrate and favor the development of multi-fidelity frameworks. Key questions in this area include how modern numerical tools can accelerate physics-based simulations to provide more and better data; how, especially in the generative and agentic AI era, we can ensure that our codes produce trustworthy results; and how other modern enabling technologies, such as differentiable programming, can enhance our multi-fidelity and broader modeling capabilities.

5.1 Accelerating High-Fidelity Simulations

Direct simulations of high-fidelity kinetic equations are often prohibitively expensive due to the high dimensionality of phase space. Tensor network methods exploit the inherent low-rank structure of solutions and operators. This enables efficient compression and substantially reduces computational and storage costs, thus mitigating the curse of dimensionality. Furthermore, tensor networks provide a natural framework for multi-fidelity modeling via tuning the representation complexity (e.g., ranks or bond dimensions) rather than changes to the underlying physics. As a result, tensor-network models can serve as efficient surrogates and accelerators for high-fidelity simulations.

Borrowing from classical fluid dynamics, large eddy simulations are a viable technique to accelerate higher-fidelity turbulence solvers, requiring one to adopt proper subgrid models. The difficulty lies in the different nature of turbulence in magnetized plasmas compared to regular fluids, such that existing models cannot be directly applied. Ad-hoc closure terms need to be developed, and machine learning methods can help learn them directly from simulations or experimental data.

Across all fidelity levels, a common computational challenge is the iterative solution of large implicit systems. Surrogate models can reduce time to solution by generating data-driven initial guesses that capture subtle correlations in these high-dimensional problems, while neural networks can provide learned nonlinear preconditioners that map variables into a space favoring faster convergence. Modern GPUs further support truncated datatypes such as FP8 and FP16: although too imprecise for many scientific applications on their own, they offer a promising route to sketching matrices or powering cheap surrogates that seed initial guesses for implicit solves.

5.2 Differentiable Programming and Simulations

The ubiquitous success of deep learning has co-evolved with the development of autodifferentiation frameworks (e.g., JAX, PyTorch) optimized for high-throughput tensor operations. By writing scientific software with these packages, the scientific computing community inherits this substantial development investment: such codes benefit from GPU/TPU (tensor processing unit) acceleration, are approachable to a wider audience through the Python ecosystem, and expose derivative information through autodifferentiation. Derivatives are critical to both optimization and sensitivity analysis, and can often be obtained in a few lines of code from a differentiable simulation. This enables sophisticated schemes such as Adam, Newton-Krylov, and Krylov-Levenberg-Marquardt directly within scientific codes. Coupling to neural networks is similarly streamlined, since solvers and networks reside in the same software frameworks. This allows neural network components to be embedded directly within the timestepping loop, and enables gradients to propagate through the solver, aiding the development of derivative-informed surrogate models such as derivative-informed neural operators (DINO).

5.3 Building Trustable Numerical Methods Via Formal Certificates of Correctness

Numerical methods must provably preserve the mathematical structure of the partial differential equations (PDE) themselves as well as physical correctness properties, such as conservation laws. Constructing provable correct numerical methods for fusion simulations is possible via special Domain Specific Languages (DSLs) that encode, in an information-dense manner, the underlying structure of the equations. Once constructed, the DSL can generate: (1) formal proofs in the Lean language that the numerical scheme satisfies certain properties, and (2) the C code that implements the scheme. Initial work demonstrating this approach was performed for the compressible Euler equations and the MHD equations. Producing production quality tools for this, however, requires significant further research and concomitant investments.

5.4 Work Accomplished During Long Program

iGENE was converted from TensorFlow to JAX adding missing physics modules and multi-GPU support. New code capability includes a large eddy simulation model based on the Germano identity. The possibility of using implicit time schemes to overcome the Courant–Friedrichs–Lewy (CFL) constraint, as well as non-uniform fast Fourier transforms (FFTs), was also explored.

A database of thousands of traveling waves in 2D Kolmogorov flow was constructed using a multi-fidelity convergence scheme. The flow fields were converged at low precision, Krylov dimension, and spatial resolution before being finely converged in double precision. A denoising diffusion probabilistic model (DDPM) was fine-tuned on the above Kolmogorov flow data and used to generate plausible novel equilibria. Work continues to introduce equivariance into the unconditional generative model to enforce boundary symmetry, which may improve the likelihood of generating physical solutions.

Random subsampling was compared against a POD projection as a ROM coordinate system for 2D turbulence. While POD is provably optimal for reconstructing the state, random subsampling had competitive performance in fitting the time derivative of the chaotic flow, especially when the reduced degrees of freedom were O(100).

Tensor network methods have been developed for the Boltzmann collision operator, achieving substantial reductions in computational and storage costs while maintaining high accuracy.

Using lagradept, a differentiable ICF rad-hydro solver, laser pulse shapes were reparametrized using a neural network and optimized to maximize the Lawson criterion. Use of the Adam optimizer enabled rapid exploration of possible pulse shapes, with exploration speed following a power law in network width.

5.5 Outlook

Accelerating existent codes and exploiting state-of-the-art computational techniques remains a crucial element even in a multi-fidelity setting. To this end we advocate

- Write or port codes using differentiable frameworks. Such codes would be inherently platform-agnostic, providing gradients at a fraction of the cost of traditional approaches, and naturally suited for integration of AI/ML pipelines.
- Leverage the hierarchy of floating precision datatypes available on modern GPUs, trading fidelity for speed as needed.
- Exploit tensor network methods to overcome the curse of dimensionality typical of high-fidelity kinetic simulations. The Vlasov–Maxwell system is particularly amenable to tensor-network representations, also providing a systematic framework for constructing reduced-order multi-fidelity models.
- Explore construction of new asymptotic hierarchical kinetic equations suitable for magnetized plasmas (for example, the Parallel-Kinetic, Perpendicular-Moment models) to reduce computational cost yet retaining parallel kinetic dynamics.

6 Multi-Fidelity Methods for Fusion Device Design and Optimization

Any potential fusion reactor requires large-scale optimization to balance competing physics, engineering, and economic objectives, which inherently rely on models of varying fidelity. Historically, these models and their surrogates were utilized in a decoupled manner. However, optimization algorithms naturally exploit surrogate errors, demanding extensive high-fidelity validation. Relying on decoupled, low-fidelity surrogates often yields overly optimistic designs that fail under rigorous testing, or overly pessimistic ones that discard promising configurations because the reduced models miss critical physical features. This highlights the need for new research: improving the fidelity of existing models, directly incorporating high-fidelity simulations into optimization, and developing robust surrogates for objectives that remain cost prohibitive. These challenges and resulting methodologies were central themes explored during Workshop 3: Fusion Device Design & Engineering and its associated seminars.

6.1 The Need for Multi-Scale and Multi-Physics Models for Design

Device design rests on predictive modeling, which is severely complicated by the multi-scale, multi-physics character of fusion plasmas. In magnetically confined plasma, resolving turbulence-driven core transport requires expensive simulations, while core-edge integration requires self-consistent coupling across vastly different domains. The wide range of scales makes multi-fidelity approaches promising.

In the core, turbulent transport is governed by cross-scale interactions that may not be recovered by summing single-scale results. The highest-fidelity method (direct numerical simulation spanning both ion and electron scales) requires ~10^{10} grid points, making data-driven approaches challenging. One possible route with lower fidelity is ion scale simulations with an effective diffusion model of electron-scale turbulence, though estimating the diffusion coefficient without high-fidelity input remains an open question.

Core-edge integration typically uses a multi-fidelity chain: a GK or quasilinear core model, a reduced pedestal model, and a fluid plasma model coupled to a kinetic neutral model in the divertor. Such integrated simulations are today's workhorse for tokamaks but are often time-consuming, may lack self-consistency, and are usually unsuited to routine operations or optimization. An immediate priority is self-consistent coupling for rapid prediction, plus validation of the coupling schemes themselves. An emerging pathway involves coupling 4D GK (i.e., approximate perpendicular dynamics as a cross-field diffusion process) edge/divertor models to adequate neutral models.

6.2 Reactor Physics Design and Optimization

Optimization is fundamental to fusion device design, driving everything from shaping tokamak divertors and inertial fusion energy (IFE) targets to navigating the high-dimensional landscape of stellarator configurations. Because fusion plasmas are inherently multi-scale and multi-physics, objective functions require integrated simulation frameworks such as FREDA or FUSE that couple core plasma, edge dynamics, and neutronics. Historically, optimization relied on simplified, physics-based proxies, such as using effective helical ripple to estimate confinement. However, direct optimization reveals these traditional proxies can be overly restrictive. Consequently, incorporating true high-fidelity metrics—like direct energetic particle confinement or GK turbulence evaluations—is necessary for accurate reactor design.

Given the immense computational cost of these high-fidelity evaluations, MFM make this deep integration tractable. Techniques like Bayesian optimization utilize Gaussian process surrogates trained on high-fidelity samples, while methods like co-Kriging efficiently fuse data across multiple fidelity levels. In ICF, for example, multi-fidelity Bayesian methods and deep learning surrogates have successfully optimized complex target designs at the National Ignition Facility. By incorporating simulated and experimental information, these tools provide probabilistic predictions for achieving ignition, facilitate pre-shot validation, and accurately model experimental shot-to-shot variations.

Furthermore, robust reactor design must account for real-world uncertainties, such as manufacturing tolerances and coil installation errors. To manage this, Monte Carlo methods are embedded within the optimization loop to evaluate the distribution of quantities of interest due to a distribution of potential errors. By leveraging multi-fidelity MC-based methods [Sec. 4], designers can obtain significantly more accurate estimates for a given number of samples. When objective functions themselves rely on statistical simulations, such as in fast ion confinement analysis, these multi-fidelity variance reduction techniques ensure the final design is both physically optimal and practically realizable.

6.3 Integrated Systems Engineering and Device Design

Fusion design has made substantial progress through advances in plasma physics, and emerging integrated modeling frameworks are now enabling earlier consideration of engineering constraints, facility requirements, and economic objectives. Recent integrated design studies, including Pareto-front analyses of fusion reactor concepts, highlight the value of a systems-engineering approach that jointly balances plasma performance, engineering constraints, power-handling requirements, safety margins, and cost objectives.

However, directly incorporating high-fidelity engineering analysis into early-stage design optimization remains computationally challenging. MFM provide a practical pathway by using reduced-order models and lower-fidelity metrics to represent key constraints—such as coil stresses, structural integrity, power exhaust limits, and thermal or radiation-induced material degradation—without the bottleneck of full-scale simulation. As discussed in [Sec. 7], this integrated perspective also extends to diagnostics and actuators, where low-fidelity sensor placement methods or high-fidelity predictive control simulations can be incorporated into the design loop to improve controllability and observability from the outset.

The long-term success of fusion power plants requires system-level integration of engineering design, operational decision making, and economic viability over the full plant lifetime. This calls for advanced monitoring, diagnostics, and predictive maintenance systems that combine real-time sensor data with physics-based models and digital twins to assess plant health, anticipate plasma disruptions, and detect abnormal behavior. These methods must also incorporate component- and material-level degradation physics to support timely intervention, maintenance planning, and controlled shutdowns, while informing manufacturing and supply-chain planning for critical components.

6.4 Work Accomplished During Long Program

A concept of dynamic coupling for ion- and electron-orbit-scale turbulence was proposed as a middle-fidelity method for handling multi-scale turbulence, and a feasibility study was initiated to investigate this model.

Additionally, advancements were made on a new multi-fidelity Bayesian optimization approach for solving low-dimensional optimization problems, where models of various fidelities can provide function evaluations and/or gradient information. This work emerged from a collaboration between core members and attendees of the fourth workshop. The workshop also facilitated broader discussions with other researchers in Bayesian optimization, helping to contextualize this research within the state-of-the-art. These findings will be presented at the SIAM Annual Meeting on July 6–10, 2026.

6.5 Outlook

- Develop and apply multi-fidelity techniques to stellarator optimization, specifically targeting turbulent transport and fast particle confinement. Existing levels of fidelity can be combined to make complex optimization more tractable.

- Develop large-scale algorithms that leverage multi-scale and multi-fidelity structures to reduce computational cost while preserving convergence guarantees, stability, and reliability.
- Continue development of multi-scale coupling for electron- and ion-scale turbulence, potentially steering new directions for proxy models used in optimization.
- Create fast, effective core-edge integration techniques and advanced higher-fidelity (e.g., GK) edge and divertor models for realistic predictions of plasma confinement and plasma-structure interactions, along with reduced order models suitable for optimization.

7 Multi-Fidelity Methods for Real-Time Control

Effective control of plasma, whether through real-time adjustments in magnetic confinement fusion or inter-shot tuning in ICF, is essential to achieving net-positive fusion energy, safe operation, anomaly detection and avoidance, and machine protection. Good control design is necessary to reach or exceed the performance foreseen at the design stage. Historically, fusion experiments usually take many years after first plasma to reach their peak performance as understanding of the machine and its control matures. It is therefore essential to consider control early in the design phase to ensure adequate diagnostics and actuators are included.

While different fusion concepts demand tailored diagnostic and control solutions, reactor-grade devices share critical challenges: sparse and noisy diagnostic data, the transition from single-task to multi-objective integrated control, orchestration of hundreds of controllers, and long latencies. In addition, regulators ultimately will have to approve future fusion power plants and certify machine safety: pure data-driven control methods could be difficult to verify, and MFM can play a key role, as discussed in Workshop 4: Multi-Fidelity Methods to Enable Robust Optimization and Real-Time Control of Fusion Processes.

7.1 Diagnostics, Actuators and Effect on Control

The effectiveness of control and state estimation is constrained by hardware for actuation and diagnostics. Furthermore, their number, type, and availability in current physical devices and future fusion power plants are severely limited.

On the diagnostic side, the plasma state is high-dimensional and largely interior, yet most measurements are sparse, peripheral, and noisy. Current diagnostics are dominated by boundary magnetic-field measurements, integrated values, and spatiotemporally localized internal parameters. Reconstructing a full interior profile of temperature, density, flow, current, and field throughout the volume from such limited observations is an ill-posed inverse problem: distinct internal states can produce nearly identical measurements. This is precisely the regime in which model-based state estimation is most valuable, supplying the prior structure that closes the gap between what is measured and what must be known.

On the actuator side, the available means of pushing the plasma toward a target state are comparatively few and slow. Magnetic-field coils, external radio-frequency heating, and neutral-beam injection constitute a common set, and their response times are often long relative to transients that control must suppress. Other fusion concepts have far fewer actuators. Control authority is therefore limited not only in degrees of freedom but in bandwidth: an actuator that cannot act faster than the dynamics it targets exercises little authority over them. Characterizing which states are reachable with a given actuator set, and at what speed, is also central to control design.

The diagnostic and actuator requirements diverge sharply between experimental test devices and power-producing reactors. An experimental device can be extensively diagnosed, with access designed into the machine and components serviced whenever necessary. A reactor cannot. The nuclear environment degrades sensors and actuators through neutron flux and activation, restricts physical access, and makes it hard to perform frequent maintenance. Diagnostics that are routine on today's experiments may be unavailable, unreliable, or prohibitively short-lived in a reactor, and the actuator set may likewise be constrained by survivability and reliability, rather than performance alone. Control strategies developed on experimental devices must therefore anticipate a future in which the plasma is observed and actuated through a sparser set of interfaces, which places a further premium on model-based state estimation to extract maximal information from minimal measurements.

7.2 Control and Plasma State Estimation

There are a wide variety of control strategies commonly used in plasma experiments, from feedforward to feedback, model based to model free, frequency-based and state-space, and more. Model free frequency-domain methods such as Proportional–Integral–Derivative (PID) have historically been the most widely used, but struggle with multiple inputs and outputs, or complex nonlinear dynamics which are common in fusion plasmas. This, and the ability to include notions of optimality of the controls, motivates the use of model-based control using either physics based or data driven models, such as model-predictive control, robust nonlinear control, or reinforcement learning. State-space control of fusion plasma in real time requires: (1) observability: the plasma state must first be reconstructed from available measurements through state estimation techniques, (2) controllability: the state must be driven toward a more desirable one through the actuators, (3) the control algorithm provides the connective tissue between the current measurements (or state), the desired state, and activation of actuators.

A key consideration for model-based control is understanding the fidelity required of models given the target task and the relevant control timescales. Some control tasks may only need very approximate models, while others likely need high predictive accuracy. Model-based control has appealing features, and MFM can play an enabling role in model acceleration to enable real-time deployment.

7.3 Reduced-Order Models and System-Level Control

With regulation and certification being key hurdles to fusion power plant (FPP) approval, model-based controllers provide a viable avenue, given the rich control theory on robustness and stability of such controllers. Yet current roadblocks for model-predictive control are that the physics-based models are too slow, inaccurate, or not robust, leading to online software crashes and premature optimization termination.

Reduced-order models can capture complex nonlinear dynamics, as demonstrated in a variety of applications, and often have better fidelity than simplified-physics or coarse-grid surrogate models requiring the same low computational cost. However, the generation of control-aware nonlinear reduced models is underexplored for applications as complex as fusion plasmas. The development of such efficient and control-aware reduced-order models capable of predicting the full plasma state is key to model-based control and requires significant research.

Current fusion plasma machines operate with tens to hundreds of controllers to achieve plasma shaping, instability detection and avoidance, profile and burn control, current drive, and many other tasks. Orchestration and management of competing objectives for tens to hundreds of controllers on a fusion power plant is necessary for high-performing machines. Currently, controllers are often operating independently, at times leading to unexpected shutoff and yielding suboptimal operation. While this may be acceptable for scientific experiments, it is not for fusion power plants. These systems require both centralized and localized control coordination to increase the top-level objective of stable and long-term power output, while staying in a safe regime for the plant.

7.4 Work Accomplished During Long Program:

Over the course of the long program, open control problems in magnetic confinement fusion research were reviewed, the outcomes of which are presented in the previous subtopics. Methods were also explored to enable gradient-enhanced training of surrogate models for PDEs that serve as the basis for models used to develop tokamak plasma shape control algorithms. These methods are envisioned to improve the performance and training efficiency of surrogate models used in multi-fidelity plasma control schemes.

7.5 Outlook

Control will play a critical role to make future fusion power plants economical and safe. Based on this long program, we recognize that additional research is needed in the following areas:

- Adding control considerations into the design stage. This is a common issue across industries, from automotive to aerospace to industrial plants. Since control is often considered a software task, it is unfortunately common practice to pass a final design to a control team and let them do magic. This is suboptimal. Control and design should synchronize early, especially for systems as complex as fusion power plants. Novel methods for high-dimensional control co-design that take into account transient dynamics and rare events (e.g., edge localized modes in tokamaks) are needed to achieve this goal.
- Control-aware surrogate models are needed to enable more efficient and provably robust real-time control and nonlinear state estimation for fusion power plants. This requires control engineers working with plasma scientists to achieve interpretable model forms.
- Next-generation nonlinear state estimation must bridge the gap between sparse, noisy and intermittent measurements and real-time state reconstruction, e.g., integrating state-of-the-art generative AI and scientific machine learning with traditional Kalman filtering may allow rapid reconstruction of full plasma states under tight operational constraints.

8 Agentic Workflows to Enable Multi-Fidelity Methods for Autonomous Scientific Discovery

AI agents are systems coupling reasoning models (which are typically based on large language models) with tool use (such as file system access, command and code execution, web browsing, API calls, and version control), which receive iterative feedback from the environment with which they interact. Agents are now used in high-performance computing (HPC) and computational physics workflows. By autonomously executing simulations and developing surrogates, agents meaningfully accelerate integration of MFM into existing workflows.

8.1 Agentic Workflows

Agent tooling evolved rapidly over the course of the program culminating in this document, progressing from an early curiosity to production-ready systems. The emergence of mature agentic tools has been highly disruptive, reshaping research practices and opening new avenues for exploration in a short period. In the following, we report on our experiences using agentic AI for fusion research and offer our perspectives on the applicability of agents to MFM in the field.

Current usage of agents, while still in the early stages and largely anecdotal, includes lifecycle support tasks: compiling codes on heterogeneous architectures; managing simulation campaigns, including systematic input deck variation and validation, and monitoring of logs and simulation state; resource management and scheduling; and debugging at build and run time.

Agents further assist with synchronizing state between systems; authoring new code (differentiable solvers, transport solvers, machine learning pipelines, optimization experiments, and their integration into existing frameworks); comprehending legacy codebases; producing documentation; and conducting literature review to identify promising methods.

8.2 Agentic Research

Agentic research is still in its early stages, although multiple attempts have been made across the broader machine learning community. Most successful experiences so far have come from closely supervised agentic behavior on well-defined research tasks. For example, code-based migration from central processing unit (CPU) to GPU is a relatively concrete task where agents can be useful, though even this setting still presents multiple challenges.

For more exploratory research, such as fast iteration on research ideas, agent behavior still requires frequent monitoring and steering. A major challenge is the wide bandwidth needed for researchers to fully digest, evaluate, and understand agent-generated suggestions and proposals before making meaningful decisions and progress. Blindly trusting agent recommendations often leads to inefficient exploration rather than productive research acceleration.

8.3 Work Accomplished During Long Program

Over the duration of this long program, agents have been extensively used to deploy, run, debug, and port codes on HPC systems. Agents have fully automated the deployment of the GK code GENE on several HPC systems. Using purpose-built skills, agents are responsible for cloning repositories, building production environments, compiling, and running tests. Furthermore, agents were used to identify performance bottlenecks, develop and tune GPU kernels, and autonomously debug code. Agents produced a refactored and performance-portable port of GENE-3D, and an existing differentiable GK solver, iGENE, was ported to JAX. Agents were also used to port Gkeyll's CUDA code to AMD GPUs. For laser-plasma interaction studies, agents launched OSIRIS on Perlmutter given only coarse objectives, autonomously preparing and validating input decks, estimating resource usage, and queuing and monitoring jobs. Agents also generated a comprehensive wrapper integrating OSIRIS into a common input/output framework. Finally, agents developed surrogate models of plasma transport codes by rapidly experimenting on models guided by quantitative metrics under resource constraints, substantially improving on baseline models. Generally, agents dramatically reduce cognitive overhead in all aspects of HPC simulation.

8.4 Outlook

Agentic AI has advanced rapidly over the past year and is likely to become even more capable, reliable, and autonomous. Its impact on the work within this program should not be underestimated. In particular, agentic approaches to code development and research have significantly expanded researchers' ability to test new ideas, iterate on experimental workflows, and enhance the robustness and functionality of existing codes.

Agentic workflows are closely aligned with MFM: they substantially lower the barrier to adopting MFM by making transitions between fidelities far more seamless. We envision their development and deployment along a hierarchy of increasing autonomy:

- Attendant agents: aid researchers – we are here.
    - Navigates model choices, data requirements, and trade-offs between cost and accuracy within multi-fidelity workflows.
- Semi-supervised agents: researcher maintains oversight.
    - Selects appropriate fidelity levels and plans computational campaigns.
    - Determines what additional data or simulations are needed.
- Self-supervised agents: given an objective but largely autonomous.
    - Selects, coordinates, and adapts MFM in response to the evolving demands of the problem.

9 Future Directions for Multi-Fidelity Methods in Fusion

9.1 The Emergence of Digital Twins for Fusion Systems

A central long-term goal of fusion system modeling is the development of digital twins, that is, computational representations of current and future fusion power plants or their subsystems. Following a frequently cited National Academies of Sciences, Engineering, and Medicine (NASEM) definition, a digital twin is more than a simulation: it integrates verified and validated numerical models, experimental observations, machine learning, and UQ into a predictive framework that is continuously informed by, and can inform decisions about, its physical counterpart. Importantly, the virtual representation need only be as detailed as required for its intended purpose. By enabling virtual exploration of operating scenarios and design choices, digital twins have the potential to accelerate the design-test-refine cycle and support more informed decision making.

Realizing such capabilities for fusion requires an ecosystem of models spanning a range of fidelities. High-fidelity tools, particularly nonlinear GK simulations, have matured substantially over the past two decades and have undergone extensive validation against experiments. These models provide the predictive accuracy needed to investigate regimes that remain inaccessible experimentally. At the same time, advances in whole-device modeling are increasingly enabling the coupling of core and edge physics, while engineering models continue to capture more detailed aspects of reactor design and operation. However, the computational cost of high-fidelity models limits their applicability in many-query settings such as optimization, UQ, and real-time control. As a result, they must be complemented by lower-fidelity models within workflows that strategically balance accuracy and computational cost.

Two research directions closely aligned with the themes of this long program are particularly important for advancing credible fusion digital twins. The first is VVUQ, which provides the foundation for assessing predictive capability and establishing confidence in simulation-based decisions. In addition to uncertainties in input parameters, VVUQ must address model-form uncertainty, simplifying assumptions, numerical approximations, and uncertainties associated with experimental measurements. Bayesian approaches have emerged as a powerful framework for these tasks, although further integration with plasma theory and physics-based modeling remains an important area of research. The second is multi-fidelity modeling, which seeks to combine information from models of varying cost and accuracy rather than treating them in isolation. By leveraging the speed of lower-fidelity models together with the predictive power of higher-fidelity simulations, MFM offer a promising pathway toward scalable, uncertainty-aware workflows. Together, VVUQ and multi-fidelity methodologies provide key ingredients for the development of credible and actionable digital twins for fusion systems.

9.2 Alignment with the U.S. DOE Research Priorities

The emergence of AI tools has the potential to drastically transform how the U.S. DOE and other departments in the U.S. federal government approach scientific discovery, energy innovation, and national security. Fusion energy is uniquely positioned to take advantage of this disruptive technology to accelerate the transition of fusion energy from experiments in highly specialized laboratories to providing energy on the national electrical grid.

In the spring of 2026, the DOE announced the Genesis Mission National Science and Technology Challenges Request for Applications (RFA), which is a competitive call for proposals seeking to accelerate scientific discovery and workflows using novel AI models and frameworks. In response to this opportunity, several new teams were formed among the IPAM Long Program participants, taking advantage of this rare opportunity to collaborate on multi-disciplinary proposals while co-located with experts in fusion energy modeling, reduced-order and surrogate modeling, and VVUQ. Several proposals were initiated and submitted to this call based on conversations and collaborations at IPAM. Regardless of the outcome of these proposals in this particular call, these efforts have accelerated collaboration, produced tangible outcomes and promising research directions that will shape future work.

In addition, the FS&T Roadmap, which is intended to target actions and milestones through mid-2030, highlighted a set of critical priority research opportunities related to the use of AI for fusion, including the development of surrogate and reduced-order models with data from HPC codes; accelerating numerical algorithms with AI/ML methods; developing AI-enabled digital twins; and exploring the use of AI/ML methods for optimizing the design of facilities and experiments. Each of these topics were key themes in the activities and discussions at this IPAM Long Program.

9.3 Dissemination via Publications and Conference Contributions

At the conclusion of the long program, the participants attending the retreat at Lake Arrowhead Lodge proposed and discussed a number of follow-on activities to communicate the ideas and opportunities identified in the program:

- All participants of the Long Program on *Multi-Fidelity Methods for Fusion Energy* are invited to submit a paper to a special issue of *Plasma Physics and Controlled Fusion (PPCF)*. The submission deadline is currently December 31, 2026. Given the format of the meeting, we are soliciting several article types:
  - Tutorial and Review articles will be solicited from speakers during the tutorials week
  - Perspective articles will be solicited from participants of the culminating week
  - Standard research articles and Letters will be solicited from all workshop participants
- Several opportunities exist to communicate the key ideas emerging from the Long Program to the wider scientific community. These include:
  - Propose a tutorial on *Multi-Fidelity Methods for Fusion Energy* at the American Physical Society Division of Plasma Physics (APS-DPP) annual meeting (Nov 2027 or Nov 2028).
  - Propose a mini-symposium on *Multi-Fidelity Methods for Fusion Energy* at SIAM Computational Science & Engineering 2027 (SIAM CSE27), which will occur in Pittsburgh, PA in February 2027. Deadline for submission: Jul 27, 2026.
  - Propose a mini-symposium on *Multi-Fidelity Methods for Fusion Energy* at the Platform for Advanced Scientific Computing (PASC 2027), which will occur in Switzerland in June 2027. Deadline for full proposal: Jan 2027.
  - Propose a mini-symposium on *Multi-fidelity UQ Methods for Fusion Energy* at the SIAM Conference on Uncertainty Quantification, which will occur in March of 2028.

10 Glossary of Abbreviations

ACV	Approximate Control Variates
AI	Artificial Intelligence
APS-DPP	American Physical Society, Division of Plasma Physics
CDF	Cumulative Distribution Function
CFL	Courant–Friedrichs–Lewy (condition)
CFM	Conditional Flow Matching
CNN	Convolutional Neural Network
DDPM	Denoising Diffusion Probabilistic Model
DINO	Derivative-Informed Neural Operators
DMD	Dynamic Mode Decomposition
DOE	U.S. Department of Energy
DSL	Domain Specific Language
FAIR	Findable, Accessible, Interoperable, Reusable
FES	Office of Fusion Energy Sciences
FFT	Fast Fourier Transform
FP8 / FP16	8-bit / 16-bit floating-point datatypes
FPP	Fusion Power Plant
FS&T	Fusion Science and Technology
GK	Gyrokinetic
GP	Gaussian Process
GPU / CPU / TPU	Graphics / Central / Tensor Processing Unit
HPC	High-Performance Computing
ICF	Inertial Confinement Fusion
IFE	Inertial Fusion Energy
IMAS	Integrated Modelling and Analysis Suite
IPAM	Institute for Pure and Applied Mathematics
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MFM	Multi-Fidelity Methods
MFMC	Multi-Fidelity Monte Carlo
MHD	Magnetohydrodynamic(s)
ML	Machine Learning
MLBLUE	Multi-Level Best Linear Unbiased Estimators
MLMC	Multi-Level Monte Carlo
NASEM	National Academies of Sciences, Engineering, and Medicine
PASC	Platform for Advanced Scientific Computing
PDE	Partial Differential Equation
PDF	Probability Density Function
PID	Proportional–Integral–Derivative (control)
POD	Proper Orthogonal Decomposition
PPCF	Plasma Physics and Controlled Fusion (journal)
RFA	Request for Applications
ROM	Reduced-Order Model
SIAM	Society for Industrial and Applied Mathematics
SIAM CSE	SIAM Conference on Computational Science and Engineering
SOL	Scrape-Off Layer
UQ	Uncertainty Quantification
VVUQ	Verification, Validation, and Uncertainty Quantification