Bayesian Probabilistic Methods to Enable Cross-Cutting AI Research in Nuclear Science

Lead Institution
Lawrence Berkeley National Laboratory
1 Cyclotron Road
Berkeley CA 94720

Lead Principal Investigator
P. M. Jacobs
Nuclear Science Division, LBNL
pmjacobs@lbl.gov; 510-688-0055

Administrative Point of Contact
Vance Siebert
Nuclear Science Division, LBNL
vsiebert@lbl.gov; 720-987-4271

Co-principal Investigators:
LBNL: B. Fujikawa, A. Poon, J. Vavrek
Duke University: S. Mak
University of California, Berkeley: Yu.G. Kolomensky, U. Seljak
University of California, San Diego: A. Li
Wayne State University: C. Shen

Additional Information
NOFO number: DE-FOA-0003458
DOE/SC Program Office: Nuclear Physics (NP)
DOE/SC Program Office Technical Contact: Dr. Manouchehr Farkhondeh
PAMS LOI tracking number: LOI-0000049580
Research area: Nuclear Physics

Collaborative Project Outline

Collaborating institutions:
Lawrence Berkeley National Laboratory (PI: Peter Jacobs)
Duke University (PI: Simon Mak)
University of California at Berkeley (PI: Yury Kolomensky)
University of California at San Diego (PI: Aobo Li)
Wayne State University (PI: Chun Shen)

Lead PI: Peter Jacobs, LBNL

This is a Collaborative Application. LBNL is the lead institution. Duke University, the University of California at Berkeley, the University of California at San Diego, and Wayne State University will be funded through sub-awards to LBNL.

Leadership structure: Since the number of collaborators in this project is modest, there is no need for highly-structured management. The leadership structure for decision-making is flat: Jacobs is lead PI for the project, serving as overall project coordinator and project point of contact, but all decisions will be made by consensus with all co-PIs. Each institution also has an institutional PI serving as coordinator and primary point of contact, as follows:

- LBNL: Jacobs is PI. Fujikawa, Poon, and Vavrek are co-PIs, with oversight and coordination responsibilities for neutrinos (Fujikawa, Poon) and Radiological Mapping (Vavrek).
- Duke: Mak is PI, focus is multifidelity and transfer learning.
- UC Berkeley: Kolomensky is PI and Seljak is co-PI for UCB. Focus of activity is two-fold: neutrino experiments (Kolomensky) and gradient-based sampling methods (Seljak).
- UC San Diego: Li is PI. Focus of activity is neutrino experiments.
- Wayne State: Shen is PI. Focus is Quark-Gluon Plasma studies.

Facilities: A description of each institution's facilities, equipment, and resources that will be made available to the team is found in Appendix 2.

Mentoring plan: Detailed description of how students and early-stage researchers will be trained and mentored is found in Appendix 5.

Table 1: Proposed project budget.

Project role	Name	Institution	Year 1 ($K)	Year 2 ($K)	Total($K)
Lead PI	Jacobs	LBNL/UCB	855	881	1,736
Co-PI	Mak	Duke	171	180	351
Co-PI	Li	UCSD	191	222	413
Co-PI	Shen	Wayne State	142	145	288
Total			1,359	1,429	2,787

1 Introduction

Bayes's Theorem is a powerful tool for the quantitative analysis of measurements and numerical calculations in many fields of study, by enabling systematic incorporation of prior knowledge and of correlations and covariance between elements of the measurements and calculations. Probabilistic Bayesian analysis can address "Inverse Problems," in which causal factors are deduced from measurements that are influenced by them, thereby testing model formulations and constraining model parameters. Probabilistic Bayesian analysis can also be applied to challenging computations, such as the emulation of complex simulations and image de–noising. Probabilistic Bayesian calculations are often highly demanding computationally, however, requiring Machine Learning (ML)-based methods for the efficient utilization of practically obtainable resources.

This proposal requests funding to develop and implement general ML-based approaches to Bayesian probabilistic analysis methods, including uncertainty quantification, emulation, inference, and de–noising, for application to a broad range of Nuclear Physics (NP) research areas. These areas include the measurement of the mass and fundamental nature of the neutrino; study of the Quark-Gluon Plasma that filled the early universe; and mapping of natural and anthropogenic radiation environments. The proposal co–PIs are Nuclear Physicists with leading roles in each of these areas, and data scientists developing state–of–the–art ML-based methods for probabilistic Bayesian analysis.

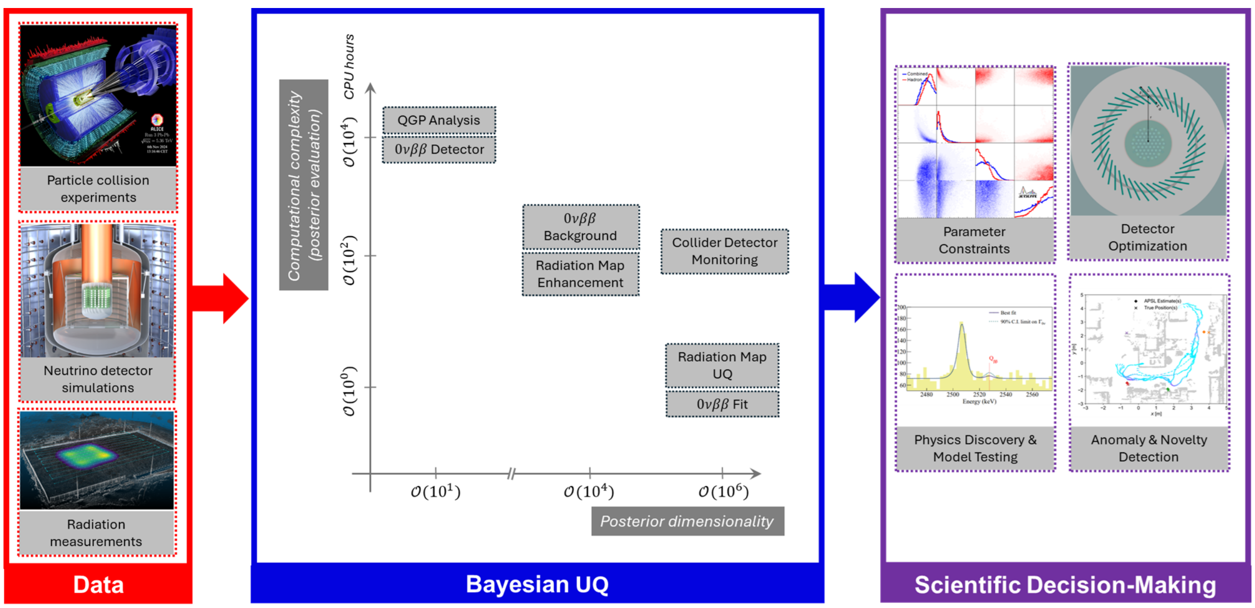


Figure 1: Schematic representation of the key computational requirements for probabilistic Bayesian analysis and UQ of the NP projects considered in the proposal. The left column indicates the various data sources, while the right column specifies the target analyses. The middle box places each project in a two-dimensional space of computational complexity of the forward model (vertical) and posterior dimensionality (horizontal).

This is a renewal proposal, building on the developments of the currently-funded Bayesian Uncertainty Quantification (BUQ) project. In this proposal the current project is labeled "BUQ Phase 1," with the renewal project labeled "BUQ Phase 2." BUQ Phase 1 has focused on a limited set of ML-based Bayesian analysis algorithms to address the inverse problem, and is undertaking a comparative study of their performance in these diverse environments. The BUQ Phase 2 project will extend these methods and explore new approaches, including generative modeling and iterative inference, and additionally explore new approaches to surrogate modeling and emulation to accelerate complex forward model calculations, and to anomaly detection.

The building blocks of Bayesian probabilistic methods include classical algorithms such as Gaussian Processes, Multi-Fidelity Surrogate Modeling, and Adaptive Sampling. The proposed research will use these building blocks to construct novel AI algorithms for challenging Bayesian inference calculations and to carry out computationally expensive simulations. The specification of meaningful, quantitative uncertainties in such calculations is often challenging, and a major focus of the project is the development of methods for that purpose (Bayesian Uncertainty Quantification, or UQ).

	$0\nu\beta\beta$ Detector Optimization	$0\nu\beta\beta$ Background & Fit	QGP Collider Monitoring	QGP Analysis	Radiation Map Enhancement	Radiation Map UQ
Bayesian Transfer Learning				✓	✓	
Bayesian Multi-Fidelity Learning	✓ ○			✓	✓	
Langevin Monte Carlo		✓ ○		✓		✓ ○
Bayesian Manifold Learning	○			○		
Bayesian Optimization	○					○
Boundary-Informed Surrogates	○			○		
Bayesian Image Change Detection			○			

✓: BUQ Phase 1 ○: BUQ Phase 2

Figure 2: Tabulation of BUQ methods and physics projects, indicating the method applied to each project in BUQ Phase 1 and proposed for BUQ Phase 2.

Figure 1 illustrates the computational requirements of the NP projects considered in this proposal, in terms of cost of the forward model for calculating likelihoods and the dimensionality of the posterior parameter space. The various projects differ by several orders of magnitude in these metrics. The proposed research presented in Sect. 3 will explore several probabilistic Bayesian analysis methods to address these challenges, within a common framework. This comprehensive approach promises unique insight into the applicability and performance of such ML-based methods for analyses of widely differing character.

Figure 2 compares BUQ Phase 1 and Phase 2, tabulating the ML-based methods utilized for each physics project in the two phases. In Phase 1, each algorithmic method was applied to more than one physics project, demonstrating their broad applicability and illustrating the BUQ Project approach of their comparative assessment under different operational conditions. The portfolio of methods will be expanded in Phase 2, with application to more than one physics project (except for the last method, whose additional applications are under discussion), thereby extending the cross-cutting approach of the BUQ project for the comparative assessment of the analysis methodologies.

2 Project Objectives and Timeline

2.1 Neutrinos

Underground experiments based on discrete solid-state detectors, such as LEGEND, CUORE, and CUPID, are the workhorse for the search of rare and novel phenomena including neutrinoless double-beta decay (NDBD). If NDBD is observed, crucial insights into the fundamental nature of neutrino mass would be uncovered. Bayesian probabilistic models have demonstrated significant potential in addressing various challenges in neutrino experiments. Bayesian probabilistic models have demonstrated significant potential in addressing various challenges in neutrino experiments.

Meanwhile, there are many other critical yet unsolved challenges in neutrino physics, which could benefit from Bayesian probabilistic models. Solving these challenges could significantly accelerate the science delivery of major results by many different neutrino experiments. Building on our past success in developing the Rare Event Surrogate Model (RESuM) and incorporating advanced sampling techniques into NDBD spectrum fitting, we propose the following key research objectives in phase 2:

- *Bayesian background model for next-generation NDBD experiments*: we propose to enhance RESuM with invertible neural networks and anomaly detection capabilities to create an end-to-end background modeling tool that can efficiently identify and analyze background sources in neutrino experiments, replacing traditional computationally-intensive simulation methods while providing more accurate results with robust uncertainty estimates.
- *Advanced Sampling Techniques in NDBD Bayesian Fit*: we propose to implement advanced gradient-based sampling methods (HMC and MCLMC) combined with domain-specific physics knowledge to overcome computational bottlenecks in analyzing high-dimensional neutrino detector data, enabling more detailed sensitivity studies for neutrinoless double beta decay experiments.
- *Enhance and Benchmark the RESuM Model*: We propose to further enhance the RESuM model with advanced active learning algorithms and additional benchmarking study
- *Precise Spectrum Modeling in KATRIN*: we propose to apply RESuM to KATRIN to model tritium beta decay end-point spectra more efficiently while providing uncertainty estimates needed for neutrino mass measurements.
- *Calibration Source Design in CUPID*: We propose to use RESuM to design calibration source for the CUPID experiment.

2.2 Quark-Gluon Plasma

Quantum Chromodynamics (QCD) describes the strong interactions of quarks and gluons, which form a deconfined state called the Quark-Gluon Plasma (QGP) under extreme conditions. Multi-messenger QGP physics leverages diverse observables—ranging from bulk particle yields to high-momentum jet probes—to study QGP properties comprehensively. The complexity of QGP dynamics, spanning multi-stage collision processes and high-dimensional parameter spaces, necessitates advanced computational methods and robust uncertainty quantification (UQ).

This research proposal addresses these challenges by advancing computational techniques and machine-learning-based approaches for multi-messenger QGP physics. By leveraging innovations in Gaussian Process (GP) modeling, Bayesian methods, and AI-driven data analysis, we aim to enhance the precision and scope of QGP studies, enabling new insights into the behavior of strongly interacting matter under extreme conditions. The following key objectives are included in Phase 2:

- *Deep Heteroskedastic Gaussian Process (GP) Modeling:* Develop and benchmark advanced surrogate models, such as the deep heteroskedastic GP, to interpolate high-dimensional parameter spaces efficiently. This includes optimizing model training with varying statistical and fidelity precisions and applying these methods to datasets from QGP studies for enhanced Bayesian uncertainty quantification (UQ).

- *Boundary-Safe Bayesian Model Selection:* Enhance the accuracy of GP emulators at model prior-space boundaries by incorporating boundary-informed training, ensuring robust performance in physically significant limits.
- *Data-Driven Theory Uncertainty Quantification:* Address theoretical uncertainties in QGP modeling by integrating data-driven UQ methods using GPs to mitigate overfitting and bias in parameter estimations.
- *AI/ML Tools for High-Dimensional Analysis:* Leverage machine learning techniques, such as normalizing flows and manifold learning, to analyze and explore high-dimensional distributions, enabling insights into multi-variable correlations and posterior distributions for QGP analyses.
- *Generative AI Models for QGP:* Develop AI-based generative models to reduce computational costs in simulating heavy-ion collisions while maintaining physical constraints and addressing challenges in uncertainty quantification for these models.
- *Iterative Multi-Messenger Bayesian Analysis:* Implement iterative Bayesian methods to refine parameter space exploration efficiently by sequentially incorporating experimental observables.
- *Collider Monitoring:* Apply Bayesian change detection techniques to monitor and ensure data quality in collider experiments, enhancing fault detection in real-time with improved uncertainty quantification.

2.3 Radiological Mapping

Radiological mapping seeks to reconstruct spatial-domain radiation intensity distributions from time-domain radiation counts measured as a detector moves about the mapping area. The Applied Nuclear Physics program at LBNL is a world leader in 3D radiation mapping technology, and over the past decade has developed a framework known as radiological scene data fusion (SDF) for combining radiation data streams with contextual data such as lidar and/or visual models of the scene [1–3]. Quantitative reconstruction algorithms are then used to attribute measured radiation counts to spatial elements (pixels or voxels) in the scene, producing a maximum likelihood (or maximum *a posteriori*) estimate of the true radiation distribution. BUQ Phase 1 focused on exploring Bayesian methods for improving the radiation images themselves, as well as producing fast and reliable uncertainty quantification image estimates, in order to produce high-quality, actionable radiation information in near-real-time, which is important for informing human operators where to search or how to avoid high contamination, as well as for driving autonomy algorithms on robotic mapping platforms. To build on our work in phase 1, we intend to explore the following topics in phase 2:

- *MCMC uncertainty quantification*: we propose to implement the new Markov Chain Langevin Monte Carlo (MCLMC) UQ method developed in phase 1 on real radiation mapping data and deploy it on a live mapping system in near-real-time.
- *Data sufficiency*: we propose to explore Bayesian/ML methods to determine whether a dataset is sufficient for a "good quality" reconstruction, as well as for performing or enhancing reconstructions with low-statistics data, either through pre-analysis design optimization or post-analysis processing.

Table 2 shows the timeline and project effort to accomplish the project objectives. The PIs are listed by name, whereas staff and postdoc effort specifies FTE fraction. The student effort is not quantified.

Table 2: Timeline for project objectives. Parentheses in columns 3 and 5 show postdoc effort in fractions of FTE.

Topic	Y1 PI + staff	Y1 PD + students	Y2 PI + staff	Y2 PD + students
Neutrinos				
Bayesian background model	Li, Poon, Fujikawa, Mak	LBNL(0.5), UCSD(0.5), UCSD(students), Duke(students)		
Enhance/Benchmark RESuM			Li, Poon, Fujikawa, Mak	LBNL(0.5), UCSD(0.5), UCSD(students), Duke(students)
Advanced Sampling Techniques	Kolomensky, Poon, Fujikawa, Seljak	LBNL(0.25), UCB(0.5)	Kolomensky, Poon, Fujikawa, Seljak	LBNL(0.25), UCB(0.75)
Spectrum Modeling KATRIN	Poon	LBNL(0.25)	Poon	LBNL(0.25)
Source Design CUPID	Kolomensky	UCB (0.25)	Kolomensky	UCB (0.25)
QGP				
Heteroskedastic GP	Shen, Jacobs, Mak	WSU (0.25), LBNL (0.20), Duke (students)		
Boundary-Safe Model Selection			Shen, Jacobs, Mak	WSU (0.25), LBNL (0.1), Duke (students)
Theory UQ	Shen	WSU (0.25)	Shen	WSU (0.25)
High-dim Analysis	Shen, Seljak, Mak	WSU (0.25), UCB (0.25), Duke (students)		
Generative AI			Shen, Jacobs, Mak	WSU (0.25), LBNL (0.1), Duke (students)
Iterative Multi-Messenger Analysis	Shen, Jacobs	WSU (0.25), LBNL (0.20)	Shen, Jacobs	WSU (0.25), LBNL (0.20)
Collider Monitoring	Jacobs, Mak	LBNL (0.1), Duke (students)	Jacobs, Mak	LBNL (0.1), Duke (students)
Radiation mapping				
Data sufficiency	Vavrek, Mak	LBNL (0.25)	Vavrek, Mak	LBNL (0.25)
Langevin UQ	Vavrek, LBNL staff (0.1)		Vavrek, LBNL staff (0.1)	

3 Progress Report and Proposed Research

3.1 Algorithms

3.1.1 Progress report

The Algorithms working group, which focuses on Bayesian UQ methods and algorithms, has made substantial progress in BUQ Phase 1. This includes the development of algorithms outlined in the initial proposal, as well as innovative Bayesian ML/AI methods that tackle new NP problems arising from collaborative discussions.

The methods proposed for BUQ Phase 1 involve the development of multi-fidelity and transfer learning Bayesian surrogates for NP applications. This work has resulted in three publications on multi-fidelity learning. Ref. [4] presents a new multi-fidelity Gaussian Process (GP) surrogate which leverages graphical dependencies between multi-fidelity forward simulators. Ref. [5] proposes a new conglomerate multi-fidelity GP which leverages multiple fidelity parameters for cost-efficient surrogate modeling. Ref. [6] investigates experimental design algorithms for optimizing multi-fidelity training runs for such models. This work has also resulted in two papers on transfer learning [7, 8] which propose new Bayesian transfer learning models that robustly transfer information from related systems for training surrogates of a costly target simulator. These methods have been applied for emulating QGP observables, with promising preliminary results (Sect. 3.3.1).

Fruitful interdisciplinary discussions within the BUQ project have stimulated additional progress in new Bayesian AI/ML methods and algorithms development. A topic of great interest is the scalability of GPs for massive or complex data [9]. For this, we have published two papers [10, 11], with another in revision [12]. Another direction of interest for NP is active learning [13] – the sequential collection of training data to maximize learning performance. We have published conference proceedings on this topic [14], and our active learning algorithm played a central role in a recent major JETSCAPE publication [15]. Two additional topics of interest are physics-guided ML and Bayesian optimization, discussed further below. We have published four papers on these topics [16–19] and are preparing software for full-scale implementation in BUQ projects.

Collaborative discussions during BUQ Phase 1 have raised interest in Bayesian online monitoring of complex collider data, which has generated several publications. Ref. [20] explores the use of topological analysis (a rising area in ML) for online detection of abrupt changes in data patterns. This work has received prominent awards: the American Statistical Association (ASA) Editor's Choice Collection Award, and first place Student Paper Awards in the INFORMS Section on Data Mining and the ASA Section on Physical & Engineering Sciences. Refs. [21–24] investigate novel Bayesian detection methods for online fault detection and diagnosis with complex high-dimensional data. Preliminary results in BUQ NP projects are promising, as discussed in later sections.

3.1.2 Proposed research

Bayesian Manifold-Embedded Surrogate Models One key bottleneck for timely Bayesian Inference in NP involves the costly nature of a complex forward simulation, which is often required in evaluating the desired posterior distribution, e.g. in our QGP project (Sect. 3.3). A full Bayesian analysis can thus be prohibitively costly, and Bayesian ML with principled UQ is needed to accelerate such a process, enabling incisive analyses with achievable resources.

Bayesian surrogate modeling [25] provides a proven solution. The simulator is first evaluated at a designed set of n parameters, then such data are used to train a probabilistic predictive model which *emulates* the forward simulator efficiently over the parameter space. This trained model (with associated uncertainties) replaces the expensive simulator to accelerate Bayesian Inference. GPs [9] are a popular class of Bayesian surrogates with broad scientific applications [26–30], including in NP studies [31–33]. This includes recent promising work on "deep" GPs that incorporate multiple layers of GPs via its kernel length-scale parameters; see [34, 35] and work [4, 5] by Co-PI Mak. For NP applications, GPs and their deep variants have two key advantages over alternate Bayesian surrogates (e.g., Bayesian neural networks [36, 37], stochastic polynomial chaos [38, 39]). First, GPs provide a flexible predictive model with reliable UQ, supported by Bayesian learning theory [40]; alternate Bayesian surrogates, in contrast, can yield erratic uncertainty quantification with limited training data [41, 42]. Second, GPs provide a closed-form quantification of predictive (epistemic) surrogate uncertainty. This permits a *closed-form*, *easy-to-evaluate* and *differentiable* posterior density, which enables the use of state-of-the-art MCMC samplers for Bayesian Inference (see recent JETSCAPE papers [31–33] on the QGP). We thus leverage GPs and their deep variants below.

Key methodological innovations are needed to harness the full power of GP surrogates for NP, however. One obstacle is, with the costly simulator, only *limited* training data (i.e., small n) can be simulated over the parameter space $\boldsymbol{\Theta} \subseteq \mathbb{R}^d$ for surrogate training, where d is the number of parameters. This poses a significant bottleneck for our NP projects: e.g., for the QGP, with a moderate number of ~ 20 parameters, one may require $\mathcal{O}(10^{20})$ simulation runs over $\boldsymbol{\Theta}$ for accurate GP training [40]. A promising solution is to identify embedded *low-dimensional manifold* structure within the simulator, which can be more easily learned from limited data. Existing work on GPs with such embedded structure includes the single-index GP [43], Gaussian ridge functions [44, 45], and projection pursuit GPs [46,47]; see also related works by Co-PI Mak [26–28,48]. Such surrogates, however, encounter two limitations in NP. Firstly (i), they largely do not quantify *uncertainty* in the learned manifold, resulting in overconfident surrogates that lead to spurious findings. Secondly (ii), they do not account for the intrinsic complexity of heavy-ion measurements and calculations of the QGP, in which an experimental observable can be influenced by *multiple* different physics processes. This may lead to poor surrogates with unreliable uncertainties given limited data, as we show later.

We thus propose a new Bayesian surrogate, the Additive Multi-Index GP (AdMIn-GP), to address limitations (i) and (ii). Let y_i be the observable simulated at parameters $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}$, $i = 1, \cdots, n$. We presume y_i follows the model:

$$y_i = \eta(\boldsymbol{\theta}_i) + \epsilon, \quad \eta(\boldsymbol{\theta}) = \sum_{l=1}^{L} \eta_l(\boldsymbol{\theta}) := \sum_{l=1}^{L} g_l(\mathbf{M}_l \boldsymbol{\theta}), \quad \epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \gamma^2), \quad i = 1, \cdots, n, \tag{1}$$

where $\eta(\boldsymbol{\theta}_i)$ is the simulated observable mean, and ϵ_i models statistical error. Here, each component $\eta_l(\boldsymbol{\theta})$ follows a "ridge" function [49]; it fluctuates only along the active subspace [50] spanned by $\mathbf{M}_l\boldsymbol{\theta}$, where $\mathbf{M}_l \in \mathbb{R}^{p \times d}$, $p \ll d$, is an embedding matrix projecting $\boldsymbol{\theta}$ onto a lower-dimensional manifold. Ridge functions can be justified for simple physical systems [51], and with careful training of $\mathbf{M}_l$ from data, have shown empirical success [50, 52] in modeling simple phenomena with a *single* dominant physics, including in NP [17]. Our model (1) thus accounts for the complex nature of heavy-ion data via multiple ridge functions η_l, with each employing a distinct active subspace via $\mathbf{M}_l$, addressing (ii). We then assign independent GPs on $g_l \sim \mathrm{GP}\{\mu_l, k_l(\cdot, \cdot)\}$, where μ_l and k_l are its mean and covariance kernel; its parameters can estimated from data via maximum likelihood [53].

Recall that existing manifold-embedded GPs lack UQ for manifold estimation (limitation (i)), resulting in overconfident surrogates. We address this in a Bayesian approach by assigning double-exponential shrinkage priors [54] on each entry of $\{\mathbf{M}_l\}_{l=1}^{L}$. Such priors provide shrinkage on the many model parameters, allowing for accurate Bayesian inference with limited data [55]. Using a carefully-constructed variational inference scheme adapted from [56], we can then obtain a closed-form predictive distribution for the response surface η. Thus, the AdMIn-GP not only captures the desired complex multi-physics structure of NP observables with principled UQ, it also provides a *closed-form*, *easy-to-evaluate* and *differentiable* posterior density for efficient Bayesian Inference via state-of-the-art MCMC methods – a key reason for the use of GP surrogates in NP.

Preliminary experiments on the surrogate modeling of a QGP simulation [57] are promising, showing improved probabilistic predictive performance (over 40% reduction in test error) over standard GPs [9], existing manifold-embedded GPs [43,44,47] and a popular probabilistic ML benchmark [58]. The AdMIn-GP also shows excellent coverage for its confidence intervals, highlighting the importance of *Bayesian* manifold learning for reliable surrogates with trustworthy UQ.

During BUQ Phase 2, we will fully develop the AdMIn-GP model for reliable and cost-efficient surrogate modeling. An important direction is the integration of deep GPs to improve surrogate expressiveness. We will adopt a recent promising deep GP formulation [34, 35] (see also work by Co-PI Mak [4, 5]) within the AdMIn-GP, and extend the elliptical slice sampling algorithm in [35] for efficient model training. Another essential development for the BUQ project is the extension of the AdMIn-GP to emulate data with widely varying statistical precision (Sect. 3.3.2). This variation is known in ML as *heteroskedasticity* [25]. To model this, we will adopt the approach in [59], by allowing the noise variance γ^2 to vary over the parameter space, then model its log-variance by a separate GP over $\boldsymbol{\Theta}$. This heteroskedastic extension can then be fit by adapting the Hamiltonian Monte Carlo algorithm in [59]; more on this for the QGP project in Sect. 3.3. Finally, we will develop active learning techniques ([13], also works by Co-PI Mak [29,60]) to select optimal training points. This is a rising area in ML, and we will leverage recent techniques for GPs [25] to improve the AdMIn-GP; see [15] for our preliminary results on the QGP with active learning.

We will then integrate the AdMIn-GP for tackling neutrinos detector optimization problem in Sect. 3.2. There, the goal is to minimize the probability of neutron backgrounds entering the neutrino detector via a careful optimization of neutron moderator design parameters $\boldsymbol{\theta}$. We will leverage Bayesian optimization methods [61] (e.g., the Expected Improvement algorithm [62,63]) to derive *analytical* and *differentiable* acquisition functions for sequential sampling. Further details are provided in Sect. 3.2.

Boundary-Informed Surrogate Models A complementary strategy for improving Bayesian surrogates in problems with limited training data is the integration of known physics on a response surface $\eta(\cdot)$ ("physics-integrated ML" [30, 48, 64–69]). This includes the incorporation of known boundaries of η along the parameter space $\boldsymbol{\Theta}$ [70–72]. For the BUQ QGP project of simulating viscous fluid dynamics, this boundary corresponds to values of shear and bulk viscosity of zero, i.e., the ideal hydrodynamics model, which is computationally much less expensive than viscous hydrodynamics (Sect. 3.3). Such computationally–efficient boundary information can be invaluable for efficient improvement of surrogates, and there is thus a crucial need for the design and modeling of boundary-informed Bayesian surrogates. Our current work on neutrino detector design has shown that incorporating additional physics information into Bayesian probabilistic models significantly enhances both model accuracy and statistical coverage (see Sect. 3.2.1).

Existing work [70,71,73] largely focuses on the *modeling* of boundary-informed GPs. This includes a recent paper by Co-PI Mak on the BdryGP model [72], which integrates boundary information of the form $\{\eta(\boldsymbol{\theta}) : \theta_j = 0\}$ and/or $\{\eta(\boldsymbol{\theta}) : \theta_j = 1\}$, i.e., left and/or right boundaries for a parameter j. One appeal of the BdryGP is, conditioned on training data and boundary information, it provides an analytic predictive distribution, which yields a *closed-form*, *easy-to-evaluate* and *differentiable* posterior density for downstream Bayesian Inference via state-of-the-art MCMC methods. The BdryGP further provides a theoretically well-understood improvement [72] over standard GPs that do not leverage boundary information, enabling improved surrogate predictions and reductions in its uncertainties.

A crucial step for accurate surrogates is the *design* of its training experiment runs over the parameter space [25]. Such experimental design is largely unexplored for boundary-informed GP surrogates, but has potential for considerably improving performance. Existing surrogates in NP use Latin hypercube designs (LHDs; [74]), which may place points on or near known boundaries; this can greatly reduce the information provided on the response surface $\eta(\cdot)$. To address this, we propose a new *boundary maximin* design for selecting simulation runs for boundary-informed GP surrogates. Our design targets two objectives. First, it targets the maximization of the closest two design points; this is known as the "maximin" criterion in statistical learning [75]. Second, it targets the maximization of the closest distance between any design point and a known boundary. This joint optimization ensures minimal overlap of information between design points and known boundaries. Figure 3 shows an optimized $n = 20$-point boundary maximin design for $d = 2$ parameters. Our design can be shown to yield maximum information gain on the response surface $\eta(\cdot)$ under the BdryGP, in the asymptotic sense of [75]. For computation, our design can be optimized via the integration of exchange-type algorithms [25] with particle swarm optimization [76, 77]. Using this, designs with $n = \mathcal{O}(100)$ points for $d = 20$ parameters can be efficiently optimized in seconds. Preliminary experiments on a suite of test functions [78] show that, using the BdryGP surrogate with the proposed designs, one achieves a ten-fold reduction in test prediction error compared to the use of standard GPs (with no boundary information) with LHDs.

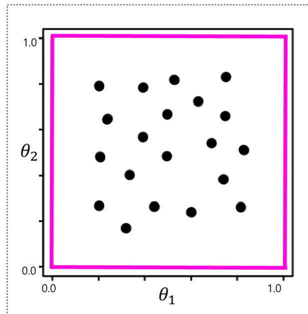


Figure 3: Visualizing the proposed boundary maximin design (known boundaries are marked in pink).

During BUQ Phase 2, we will fully develop the BdryGP and the proposed boundary maximin design for cost-efficient surrogate modeling in our NP projects. One direction is active learning [13], which leverages the trained surrogate with boundary information for selecting subsequent runs; active learning has shown promise for our QGP Bayesian analysis [15]. To do this, we will extend active learning approaches [79, 80] within the BdryGP to provide closed-form acquisition functions for sequential sampling. Model expressiveness will be improved by incorporating the deep GP framework in [34, 35], via multiple GP layers on the BdryGP length-scale parameters. Finally, we will apply the developed BdryGP surrogate framework to BUQ projects (Sects. 3.2 and 3.3).

Bayesian Online Change Detection Another bottleneck for timely Bayesian Inference in NP is the need to make *prompt* decisions with massive complex data. This data can take the form of high-dimensional structured images in NP, and the goal is to use such data to quickly detect, i.e., in an *online* fashion, abrupt system changes. For example, in the ALICE Electromagnetic Calorimeter [81], energy deposition data are recorded many times per second in over 17,000 channels. Such data can be represented as a stream of high-dimensional images [82], and require timely monitoring to ensure data quality with confidence. There is thus an urgent need for an online Bayesian image change detection method that performs prompt anomaly detection with principled uncertainty quantification to guide such monitoring.

Key methodological innovations are needed to tackle such needs for NP. Existing work on online change-point detection are largely frequentist, and do not fully account for uncertainties in either image modeling or change detection. There is a growing literature on Bayesian online change detection [83–85], but such methods do not scale for the large structured images in NP. By ignoring such image structure (e.g., spatial correlations), the online detection of anomalies can be greatly delayed using high-dimensional data [20], which is undesirable for prompt data quality monitoring.

To address these limitations, we propose a new online image Bayesian Change Detection (iBCD) method. Our model builds upon the deep Gaussian Markov Random Field (DGMRF) model in [86], which provides a probabilistic Bayesian framework for expressive image modeling. With this, iBCD leverages a carefully-designed message passing algorithm [87, 88] to efficiently compute the posterior distribution of its run length [85], i.e., the length of time since its last change point. A key appeal of iBCD is its computation of this run length posterior in $\mathcal{O}(\mathrm{pix}^2)$ work at each time step, where pix is the number of image pixels; this offers considerable speed-up over existing Bayesian change detection algorithms [85]. Preliminary experiments on the drift tube detector monitoring set-up in [89] show much quicker change detection performance over the state-of-the-art, with reliable Bayesian UQ to guide downstream decisions.

During BUQ Phase 2, we will fully develop the iBCD for full-scale NP implementation. This includes the integration of GPUs [90] for timely detection with *massive* images. Preliminary experiments with GPUs show that the run length posterior can be computed in seconds for images with $\mathcal{O}(10^6)$ pixels; we will develop this for scalable detection in NP applications. Another direction is the use of iBCD for anomaly *diagnosis*, which targets the root cause of a detected anomaly. We will develop a Bayesian diagnosis framework using iBCD, extending related work by Co-PI Mak [20]; this provides a probabilistic tool that guides subsequent actions to ensure data quality with confidence. Details on its use for collider monitoring are provided in Sect. 3.3. We will also explore the use of the DGMRF later in Sect. 3.4 for inpainting missing pixels in radiological maps.

3.2 Neutrinos

Underground experiments based on discrete solid-state detectors, such as LEGEND, CUORE, and CUPID, deploy large arrays of highly sensitive detectors designed to search for rare and novel phenomena, including neutrinoless double-beta decay (NDBD). If NDBD is observed, crucial insights into the fundamental nature of neutrino mass would be uncovered. Bayesian probabilistic models have demonstrated significant potential in addressing various challenges in neutrino experiments, as evidenced by our past research progress. As detailed in our Progress Report Section 3.2.1, we have developed a Rare Event Surrogate Model (RESuM) for neutrino detector design optimization and built the foundation to incorporate gradient-based samplers into Bayesian spectrum fitting.

Meanwhile, there are many other critical yet unsolved challenges in neutrino physics, which could benefit from Bayesian probabilistic models. Solving these challenges could significantly accelerate the science delivery of major results by many different neutrino experiments. These challenges include (1) background modeling in neutrinoless double beta decay ($0\nu\beta\beta$) experiments; (2) advanced sampling techniques in $0\nu\beta\beta$ Bayesian fits; (3) precise spectrum modeling in direct neutrino mass measurement experiments. Building on our past success, we plan to design end-to-end AI algorithms to solve these challenges by incorporating advanced Bayesian methods, such as invertible variational autoencoders, anomaly detection, and Hamiltonian Monte Carlo sampling. In Section 3.2.2, we will discuss our proposed research which aims to solve each of the challenges in detail (see Figure 5 for an overview of the connections across the proposed projects).

3.2.1 Progress report

Using support from the BUQ Phase 1 project, we developed and applied the Rare Event Surrogate Model (RESuM) [91] to address a specific detector design challenge in the LEGEND experiment. This work has been submitted to The Thirteenth International Conference on Learning Representations (ICLR 2025). The physics motivation is to design a neutron moderator that wraps around the detector array to eliminate one of the primary background sources, $^{77(m)}Ge$ [92, 93]. The neutron moderator can be parameterized by a design parameter vector θ containing five design variables.

RESuM is a surrogate model that maps design parameters θ to the design metric $y = m/N$, where m represents successful instances out of N total simulated instances, given limited access to large N. The model leverages two simulation types: an expensive, accurate High-Fidelity (HF) simulation requiring 170 CPU hours, and a faster, less accurate Low-Fidelity (LF) simulation needing only 0.15 CPU hours. RESuM employs a Conditional Neural Process (CNP) to generate predictive scores μ based on both design parameters θ and instance-specific parameters ϕ, followed by a Multi-Fidelity Gaussian Process (MFGP) that combines LF and HF simulations to estimate HF metrics. Lastly, we used a simple active learning algorithm to optimize the model by selecting new design parameters for additional HF simulations, as shown in Figure 5 (left).

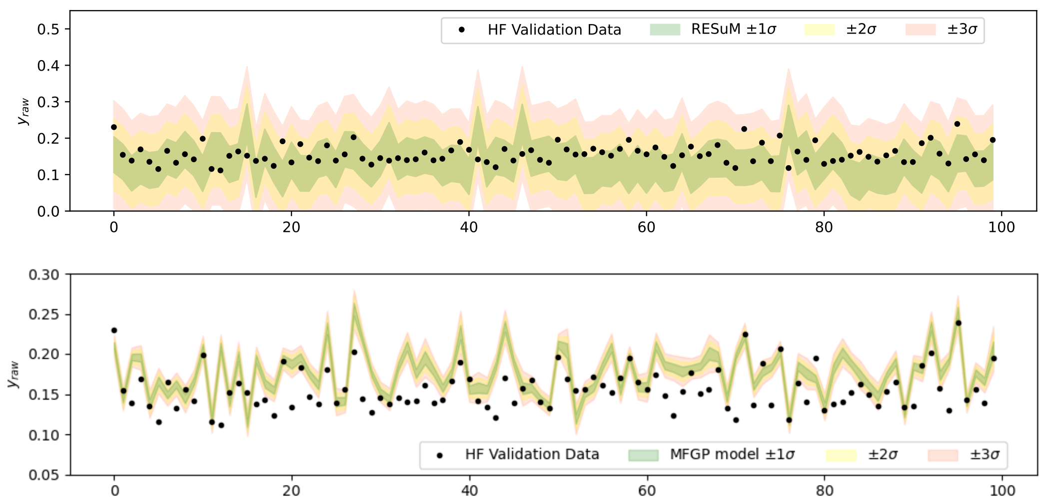


Figure 4: [Top] Statistical coverage of the RESuM model, showing proper coverage is reached for randomly generated out-of-sample simulations. [Bottom] Statistical coverage of a traditional Gaussian Process model trained and validated on the same data, where poor coverage is observed.

Experimental results show that the trained RESuM model successfully identified the optimal design which reduces neutron background by (66.5±3.5)%, while using only 3.3% of the computational resources compared to traditional grid search methods. Given the input design parameter θ, RESuM not only predicts the design metric $\hat{y}$ but also provides an associated uncertainty $\hat{\sigma}$. In [91], additional validation was performed to ensure that $\hat{y} \pm \hat{\sigma}$ is a statistically reliable prediction. This was achieved by generating 100 out-of-sample HF simulations for randomly selected θ values and using RESuM to predict $\hat{y} \pm \hat{\sigma}$. As shown in Figure 4 (top), 69% of the out-of-sample HF simulations fall within the 1-σ band of $\hat{y}$, 95% within the 2-σ band, and 100% within the 3-σ band. These results closely match the expected coverage of a standard normal distribution, which is 68.27%, 95.45%, and 99.73% for 1, 2, and 3 σ, respectively. For comparison, we trained a simple Gaussian Process (GP) surrogate model on the same data and validated it on the same out-of-sample HF simulations. The GP model showed significantly poorer coverage, with 12%, 24%, and 47% coverage at 1, 2, and 3 σ, respectively. These results demonstrate that RESuM can perform robust statistical inference, ensuring that the ground truth is well-covered by the posterior predictive intervals.

On the spectrum fitting front, we have initiated the migration of probabilistic models and analysis infrastructure in the CUORE experiment to ensure compatibility with gradient-based samplers. This process involved evaluating existing Bayesian analysis software for compatibility with experimental workflows while maintaining low-level access for sampler modification and tuning. We identified computational components in the previous analysis model that were incompatible with automatic differentiation for fast gradient computation, and subsequently modified the model's evaluation framework to be fully differentiable.

3.2.2 Proposed research

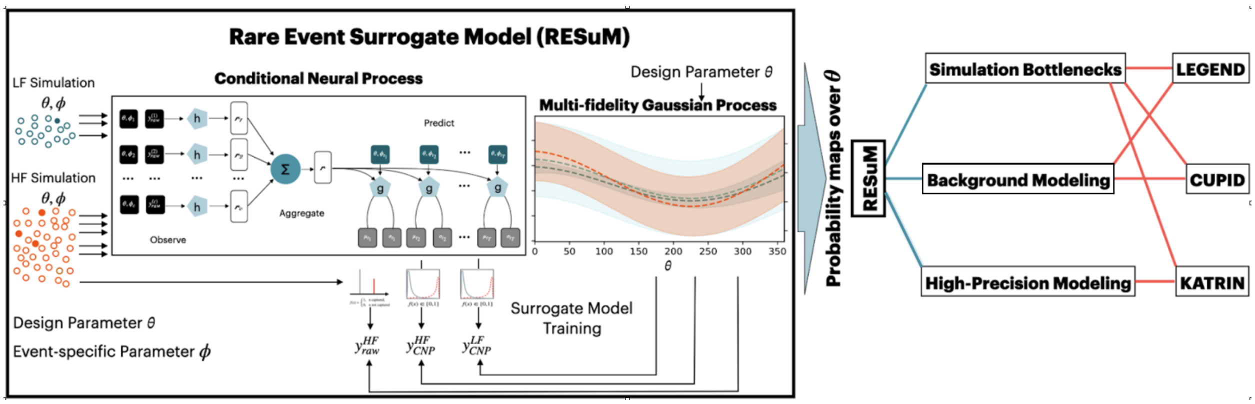


Figure 5: Architecture of the RESuM model and overview of connections across proposed projects.

Bayesian background model for next-generation $0\nu\beta\beta$ experiments A precise understanding of background compositions is crucial for next-generation $0\nu\beta\beta$ experiments, including CUPID and LEGEND. Background sources – natural radioactivity and cosmic-ray interactions – could originate at arbitrary locations within the detector under various conditions, propagate through detector materials, and eventually create signals mimicking those from $0\nu\beta\beta$ in its region of interest (ROI). The traditional method relies on extensive simulations to "grid-search" all possible locations, background types and conditions, which usually comes with high degeneracy. This process is also extremely slow and often requires a prodigious amount of computational power. Currently, both LEGEND and CUPID use this method to model their backgrounds.

The Rare Event Surrogate Model (RESuM) is a statistically robust, uncertainty-aware deep learning model designed to accelerate computationally intensive tasks in rare-event physics. As demonstrated in the manuscript, RESuM offers two significant advantages over traditional, resource-intensive Monte Carlo methods (even with biasing): (1) it achieves valid results with up to 97% less computational power, and (2) each of its predictions is accompanied by a statistically robust measure of uncertainty. As the creators of RESuM, we aim to refine and expand its capabilities, enabling it to address a wider range of challenges in neutrinoless double beta decay ($0\nu\beta\beta$) and direct neutrino-mass measurement experiments. Specifically, we propose enhancements to support high-dimensional objective functions and advanced active learning methods, extending RESuM's reach to solve pivotal issues in the LEGEND and CUPID ($0\nu\beta\beta$) experiments. Additionally, we plan to leverage the upgraded RESuM model for high-precision modeling in the KATRIN experiment, where progress is currently constrained by computational limitations. The success of this proposal would position RESuM as a scalable solution not only for neutrino experiments but also for broader scientific applications that require extensive simulations.

The success of RESuM shows that Bayesian modeling can address such optimization problems efficiently by incorporating additional prior information to break the degeneracy (in RESuM, we use CNP to break the degeneracy between 1 (background created) and 0 (background not created)). However, from an AI perspective, background modeling is considered an inverse problem: where we would like to surrogate the relationship from background composition to the energy spectrum, but the goal is to understand background composition. This requires us to "inversely redesign" the RESuM model with invertible neural networks. The development involves two steps. (1) Using the RESuM model to generate a detection probability map, which maps different spatial distributions of various background sources into their signal rate in the energy spectrum. This can be done via the Forward RESuM model described in the detector design challenge, the key difference is that now the surrogate model output is an entire energy spectrum instead of a single design metric. Doing this with the RESuM model supports more accurate source inferences while reducing false positives and computational cost. (2) We will design an Invertible RESuM model, where the CNP is replaced by an invertible neural network, e.g., a Variational Auto Encoder or Normalizing Flow. This would allow us to infer the detector configuration and source properties that best explain the observed data. (3) We will develop anomaly detection techniques over the detector spectrum to detect unusual background excess, such as asymmetric peak shape or small background excess.

Combining these developments, we will have an end-to-end background modeling tool: when a background excess is found by anomaly detection (3), we can leverage the RESuM-generated detection probability map (1) to analyze these bins with excess, and eventually use the invertible model (2) to identify possible location and the background types. This approach allows us to automatically detect unknown background sources, without relying on expensive and often degenerate simulations. Given that both LEGEND and CUPID use a similar approach for background modeling, this end-to-end AI algorithm can provide a fast and efficient way to identify backgrounds in both CUPID and LEGEND. Moreover, as a predecessor of LEGEND, the Majorana Demonstrator experiment also uses a similar method to perform background modeling. The success of this end-to-end AI algorithm could provide an independent way to validate the accuracy and provide hints for the currently ongoing MJD background modeling effort.

Advanced Sampling Techniques in NDBD Bayesian fits Analyses for these NDBD experiments are demanding in that they require high spatio-temporal granularity and specificity in modeling detector response, which is necessary to maximize the experiments' sensitivity. They require a large number of the so-called nuisance parameters – numbering in the hundreds to tens of thousands – that are crucial for accurately calibrating and understanding the behavior of the discrete solid-state detectors. These parameters need to be precisely accounted for, demanding statistical methods capable of handling this high-dimensional problem space.

To tackle these challenges, Bayesian computational techniques, such as Markov Chain Monte Carlo (MCMC), provide a powerful framework that allows experiments to robustly incorporate and quantify the impact of nuisance parameter uncertainty into statistical models. These methods sample across the entire landscape of parameters, offering a more comprehensive view of the problem as compared to point-wise estimators such as maximum likelihood methods. However, the current MCMC approaches used by the experiments encounter significant computational bottlenecks when dealing with such high-dimensional problems, leading to computational inefficiencies. The computational requirements result in longer model inference times and a reduction in the robustness of the results, which becomes particularly problematic in searches for new physics, such as NDBD. These searches require inference models to be run many thousands of times, often using synthetic data to cross-check results, which exacerbates the computational burden.

To overcome these limitations, modern state-of-the-art Bayesian samplers, such as Hamiltonian Monte Carlo (HMC) and Microcanonical Langevin Monte Carlo (MCLMC), offer a promising pathway. These methods significantly enhance computational efficiency by leveraging the gradients of the posterior distribution to guide the sampling process more effectively. Calculating these gradients is made possible through computational techniques and frameworks for automatic differentiation, which have been instrumental in the machine learning community for training artificial intelligence models. By incorporating gradient-based methods into the sampling process, these techniques improve both the speed and precision of Bayesian Inference, making them ideal candidates for the high-dimensional challenges faced in NDBD experiments.

The primary goal of this work is to develop, integrate, and deploy these gradient-based samplers into physics analyses conducted within CUORE, CUPID, and LEGEND. The first major target will be the application of these samplers to experimental searches for NDBD. Following this, the focus will shift to detector background modeling, which typically spans a broader range of detected energies and involves data selections different from those used in NDBD searches. Another analysis avenue is to search for other rare phenomena, such as double-beta decay to excited nuclear states. These phenomena introduce an additional layer of complexity, as they require the simultaneous modeling of multiple detector channels, further expanding the dimensionality of the problem.

Moreover, this work will explore ways to further reduce the computational bottlenecks encountered in inference models by merging HMC and MCLMC samplers with domain-specific physical input related to the experiments' detection techniques, geometric layouts, and methodologies. While these gradient-based algorithms offer great improvement potential over current methods, any model-agnostic "black-box" sampling technique will inevitably saturate in computational performance. These algorithms typically comprise an optimization stage, followed by an inference stage. The optimization stage runs at lesser efficiency to self-optimize sampler hyperparameters, which improves performance during the subsequent inference stage. Incorporation of domain-specific knowledge can be used to pre-tune the optimization stage with greater fidelity, or even modify the sampling algorithm altogether to include awareness of problem-specific particularities.

The improvements in computational efficiency and reduced overhead achieved through these methods will enable the experiments to perform more detailed studies quantifying the impact of systematic uncertainties on the discovery sensitivity to NDBD. Currently, performing such detailed sensitivity profiling is too computationally expensive compared to existing techniques. The outcome of such analyses could help inform the design and operational strategies of both current and future experiments, ensuring that they maximize their potential to discover new physics.

Enhance and Benchmark the RESuM Model Despite its success, the RESuM model has several limitations that we plan to address in our proposed research. A major bottleneck is the relatively simple active learning algorithm currently employed. We aim to enhance the active learning process by implementing the Expected Improvement method (EI) described in the Algorithms proposed research (Sect. 3.1.2). Furthermore, we intend to benchmark RESuM against the other proposed surrogate models in Sect. 3.1.2. This benchmarking study can inspire the development of a more powerful Bayesian surrogate model that combines the strengths of different approaches, such as Conditional Neural Processes, Bayesian manifold learning, and boundary-informed surrogates. Given the prevalence of rare event challenges in neutrino experiments, this enhanced RESuM model could be applied to various other challenges in neutrino physics, as described in the following two subsections.

Precise Spectrum Modeling in KATRIN Modeling the end-point spectrum for tritium beta decay is crucial for direct neutrino mass measurement, as different distortions of the end-point spectrum could correspond to different neutrino mass. Currently, the KATRIN experiment employs a fully-connected neural network to achieve this task. While effective, this simple network relies on large training datasets and expensive numerical computations, while it only outputs a deterministic value instead of a quantity with its associated uncertainty.

The RESuM model leverages Gaussian Process (GP) regression as an efficient probabilistic interpolator with less training data. This is achieved by its multi-fidelity nature which combines high- and low-fidelity data sources, reducing computational demands while maintaining accuracy. The two advantages of the RESuM model make it a perfect fit for modeling the end-point spectrum in KATRIN: (1) it requires much less data to train, which could allow us to explore a broader range of spectrum shape correction; (2) its ability to provide uncertainty estimates and enhance confidence in predictions near the endpoint, where spectral precision is essential for neutrino mass sensitivity. Therefore, we propose to adopt the RESuM model to KATRIN spectrum modeling to improve the current end-point spectrum modeling.

Calibration Source Design in CUPID Designing calibration sources is crucial for $0\nu\beta\beta$ experiments to understand detector response, which significantly impacts the $0\nu\beta\beta$ sensitivity. A common challenge in calibration source design stems from the rare event nature of these experiments: due to the ultra-pure environment required for $0\nu\beta\beta$ detection, a large number of calibration events must be simulated to observe even a single calibration event in the signal region of interest. The RESuM model effectively addresses this challenge, as demonstrated in the LEGEND experiment, where it efficiently and robustly emulates rare event design metrics to guide detector design. Building on this success, we propose to apply RESuM to CUPID's calibration source design, particularly to accelerate computationally intensive simulations such as phonon propagation throughout the detector.

3.3 Quark-Gluon Plasma

3.3.1 Progress report

Bayesian inference for RHIC BES data During BUQ phase 1, open-source Bayesian analysis tools from the JETSCAPE and BAND Collaborations were applied to pre-generated RHIC Beam Energy Scan training data. Three types of Gaussian Process (GP) emulators were assessed using these data, quantifying their accuracy via closure tests [94]. Three different MCMC implementations were also assessed via closure tests. With optimal choices of the GP emulator and the UCB-developed MCMC algorithm (pocoMC [95]), a systematic Bayesian inference study of QGP properties at finite baryon densities was carried out using RHIC Beam Energy Scan data [96] (PRC Editor's Suggestion). This analysis provides robust constraints on QGP transport properties and (3+1)D relativistic nuclear dynamics. A sensitivity analysis elucidated how experimental observables respond to specific model parameters, providing new insight into phenomenological modeling of heavy-ion collisions.

Bayesian inference for jet quenching data Cost-efficient methods for Bayesian inference of jet quenching data have been assessed using JETSCAPE simulations [97]. Preliminary characterization of heteroskedastic GPs (Sect. 3.1.2) shows improved performance relative to traditional GPs. The pocoMC algorithm [95] is likewise being explored using JETSCAPE jet quenching data, to characterize its performance for Bayesian inference in this area.

Comprehensive Bayesian inference of heavy–ion data A unique, comprehensive Bayesian inference study of QGP properties has been initiated, incorporating both soft and hard sector observables. A key objective is to investigate the sensitivity of hard–sector observables to QGP evolution, and assess whether hard–sector measurements can constrain bulk QGP properties; this is a central question in the field, which has previously not been addressed systematically. The soft and hard sector evolution are generated using the JETSCAPE framework and integrated into a Bayesian inference workflow. These calculations are computationally demanding, and multi-fidelity approaches (Sect. 3.1.1) are being explored to enhance simulation efficiency and scalability.

Generative ML A generative model based on Gaussian Process Regression for nuclear matter equations of state at high temperatures and densities is being developed [98]. It will be used to generate large–scale training simulations for a new Bayesian inference analysis constraining the QCD equation using heavy-ion data.

Software infrastructure Software packages and simulation data from Refs. [94,96,98] are publicly available on Zenodo [99–101] to ensure long-term software stewardship and result reproducibility.

3.3.2 Proposed research

Deep heteroskedastic GP Bayesian inference of QGP dynamics requires running a Monte–Carlo model framework (e.g. JETSCAPE [102]) with $10^3 - 10^7$ calculations at each point $\boldsymbol{\theta}_i$ in the model parameter space. Due to the high computational cost of such calculations, an efficient surrogate model, such as a pre–trained GP emulator, is needed to cover the model prior space efficiently.

A primary goal of this proposal is to explore model emulation approaches that optimize learning performance at fixed numerical cost. Section 3.1.2 proposes a new approach to surrogate modeling, the deep heteroskedastic GP, which is trained on simulations with varying statistical precision. This approach complements our current effort based on surrogates with varying fidelity precision, providing a comprehensive framework for Bayesian UQ of QGP emulation.

In BUQ Phase 2, we will first apply this approach to existing training datasets from the previous JETSCAPE and RHIC BES projects. This will benchmark the deep heteroskedastic GP against other open–source GP models, such as PCGP and PCSK from the BAND Collaboration [103,104]. It will then be applied to a new set of training data, generated in the 23-dimensional model parameter space, to study constraints on QGP properties using the bulk and high p_T observables together (Sect. 3.3.1) - a high-dimensional multi-messenger analysis.

Bayesian model selection with boundary-safe model emulation Bayesian model selection, based on comparing statistical evidence between different models, is a powerful method to determine whether the underlying theoretical model has redundant parameters. In QGP analyses with complex models, it is important to determine whether specific model elements are relevant to the experimental measurements [33]. For example, ideal fluid dynamics is the limit of viscous hydrodynamics for zero shear and bulk viscosities. We define ideal hydrodynamics as the prior space boundary of full hydrodynamics because the QGP viscosity must be non-negative.

A second model element to consider in this regard is color coherence in the parton shower. JETSCAPE models color–coherence effects as a virtuality-dependent modification factor $f(Q^2) \in [0,1]$ of the QGP transport coefficient $\hat{q}$ [105], where Q^2 is the parton virtuality. The color–incoherent limit is defined at the boundary of the model parameter space, $f(Q^2) = 1$. Since Bayesian QGP analyses require fast surrogate GP models, trained GP emulators should be accurate at the model prior–space boundary. However, as the dimension of the model parameter space increases, the model prior space boundary may be outside the region covered by training data whose design points are generated by Latin hypercube designs [106,107]. This traditional design approach will therefore behave poorly for boundary-informed surrogate modeling, as outlined in Sect. 3.1.2.

In BUQ Phase 2, we will apply the boundary-informed surrogate modeling framework proposed in Sect. 3.1.2, to incorporate boundary information at certain physically significant limits (outlined above), to ensure accurate prediction and UQ at such limits. We will then explore the improvement of such boundary-informed surrogates over existing GP emulators, and its impact on QGP Bayesian analysis.

Data-driven theory uncertainty quantification in Bayesian inference Theoretical modeling of QGP dynamics is multi-stage and complex [102]. Because the model incorporates physics at multiple length scales, it is difficult to quantify its theory uncertainties. Bayesian inference analyses that do not treat theoretical uncertainties systematically result in overfitting and bias in posterior distributions of QGP properties. For example, Bayesian inference of bulk–sector observables exhibits tensions in QGP bulk viscosity [33, 108, 109]. Tension is also observed in a jet quenching analysis incorporating a broad transverse–momentum range [97].

In BUQ phase 2, we will apply a data-driven approach to quantitative theory UQ using the following GP framework [110–112],

$$y_{\mathrm{exp}}(p_T) = y_{\mathrm{model}}(p_T, \boldsymbol{\theta}_\star) + \delta_{\mathrm{GP}}(p_T) + \varepsilon(p_T), \tag{2}$$

where $y_{\mathrm{model}}(p_T, \boldsymbol{\theta}_\star)$ is the model result at the optimal parameter set $\boldsymbol{\theta}_\star$, and $\varepsilon(p_T)$ denotes the experimental uncertainties. The term $\delta_{\mathrm{GP}}(p_T)$ represents its theoretical uncertainty, which we model by the GP $\delta_{\mathrm{GP}}(p_T) \sim \mathrm{GP}\{\mu(p_T), k(p_{T,i}, p_{T,j})\}$. The parametric forms of the mean and covariance kernel of the GP require only qualitative guidance from the underlying theory, which reduces bias from uncontrolled theory uncertainty estimation. The values of $\delta_{\mathrm{GP}}(p_T)$ are constrained by the model, providing a data-driven approach to theory UQ. We will first apply this framework to a limited set of experimental observables, where tension between the model and data has been observed [97] and assess the impact of the $\delta_{\mathrm{GP}}(p_T)$ term on the posterior distribution of model parameters. We will also analyze the functional dependence of $\delta_{\mathrm{GP}}(p_T)$ to guide the theory.

AI/ML-based tools for high–dimensional distributions Our proposed project deals with probability distributions with high dimension and of unknown analytic form. The standard Scatterplot matrices [113] provide limited information about marginal parameter distributions and their pairwise correlations. Analyzing non-trivial multi-variable correlations of the posterior distribution beyond two dimensions is yet more challenging. However, the exploration of multi-variable correlations hidden in the high-dimension posterior distribution of a QGP analysis can provide crucial physics insight. For example, kinetic and strongly–coupled holographic theories predict different temperature behavior for the ratio of QGP bulk and shear viscosity, ζ/η [114–116]. Analyzing this ratio as a function of temperature from the posterior distribution requires examining the correlation of eight model parameters.

Normalizing Flows provide a powerful framework for learning complex distributions from posterior samples [117]. In BUQ Phase 2, we will extend this approach to analyze $\mathcal{O}(20)$-dimensional distributions of QGP models. We will explore its performance and compare with different machine learning approaches, such as the stochastic diffusion method. Once these ML models are trained, they will serve as fast generators, enabling efficient exploration of the high-dimensional parameter space. We will develop analysis tools to compare high-dimensional posterior distributions from different MCMC algorithms, notably pocoMC vs. Micro-canonical LMC [118]. We will develop tools for auto-detecting multivariable correlations in the posterior distributions and effective clustering algorithms for multimodal distribution in high dimensions, which go beyond computing the standard Kullback–Leibler divergence [119].

AI/ML-based generative models for QGP: Due to the high computational cost of simulating heavy-ion collisions, the development of AI/ML-based generative models for the QGP is an important avenue of exploration. We will explore models which respect the physical constraints of collision dynamics. The consequent reduction in computing time will enable physics studies that would otherwise not be possible, such as a combined Bayesian inference analysis of low (soft) and high momentum (hard) observables. This approach will likewise benefit experimental analyses that require simulations for data corrections but are limited by computational costs. A key challenge is the principled quantification of uncertainty in generative AI models, which must be resolved in order to be deployed in Bayesian calibrations and experimental analysis. We will explore several approaches to UQ, guided and validated by large-scale, realistic simulations. Once this UQ challenge has been addressed, generative AI models can provide unique insights into heavy-ion collisions, enabling precision physics analyses that were previously computationally prohibitive.

Iterative multi-messenger Bayesian analysis In Bayesian QGP analyses, theory uncertainties vary widely among observables, and it is impractical to include all heavy-ion measurements in one fully comprehensive Bayesian inference analysis. An iterative Bayesian analysis could reveal the constraining power of specific experimental measurements on the model's posterior distribution. Iterative Bayesian QGP analysis requires the GP training described in Sect. 3.1.2 for new observables. In BUQ Phase 2, we will explore Bayesian history matching [71, 120] to optimize the model prior range at every iteration, thereby reducing computational costs of additional simulations for new observables. We will utilize a normalizing flow-based generator in Sect. 3.3.2 to first learn the high-dimensional posterior distribution from previous iterations, then use it as the prior distribution at a new iteration. Calculations will be based on the JETSCAPE framework. Building upon the results in Ref. [97], we will sequentially include new observables from jet substructure, boson-tagged jets, and a joint calibration using measurements from bulk and high-p_T sectors together.

Collider monitoring Online change detection (Sect. 3.1.2) has promising applications for data monitoring and quality assurance of collider experiments, e.g., for the ALICE Electromagnetic Calorimeter [81], which records energy deposition data more than 30000 times per second in 17664 channels. To ensure high quality data, faulty electronics need to be promptly identified. We will explore the application of the proposed iBCD method (Sect. 3.1.2) to identify such faulty channels in real ALICE data. A distinct advantage of this approach over the more conventional methods currently in use is that it provides efficient Bayesian UQ, enabling improved quantification of data quality. This is particularly important to enable a rapid determination of which channels to exclude for subsequent physics analysis and provide guidance to the running experiment about which electronic channels require maintenance.

3.4 Radiological Mapping

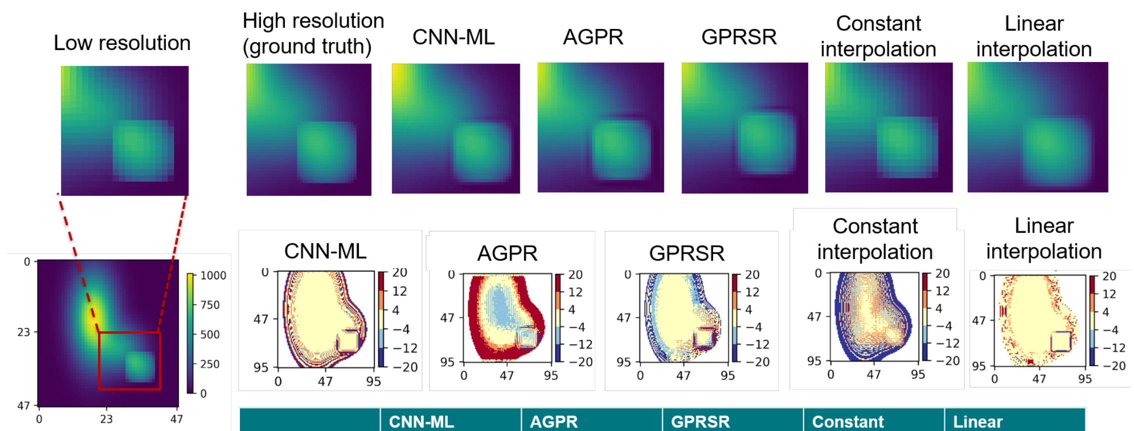
3.4.1 Progress report

During BUQ Phase 1, we studied several relevant ML algorithms under the framework of BUQ for application in radiological mapping.

1) Due to the high dimensionality of the radiation map image space, for example $\mathcal{O}(10^4) - \mathcal{O}(10^6)$, it can be computationally expensive to reconstruct the high-fidelity radiation image. Yet, real-time image reconstruction is necessary for timely decision-making processes, especially if variations of the images need to be repeatedly recomputed for UQ calculations. ML upscaling or "super-resolution" algorithms were therefore of interest for accelerating radiation image reconstructions by performing

the reconstruction at a lower fidelity and using a trained model to upscale the low-resolution image to a higher fidelity.

We identified in literature and tested ML algorithms for the acceleration of radiation image reconstruction via image upscaling. These include Convolutional Neutral Network based algorithms [121], as well as Bayesian algorithms that can provide uncertainty quantification [122]. The performance of these algorithms was compared, with insights gained of their suitability on radiation image upscaling. In general we found that the ML algorithms reach but do not exceed the performance of classical upscaling, e.g., by linear interpolation, likely because the level of fine-grained detail in radiation images is low due to detector angular resolution—see Fig. 6. This work was presented as an oral presentation at the 2024 IEEE NSS/MIC/RTSD conference [123], and we are preparing a manuscript on our systematic comparisons of upscaling methods.



	CNN-ML	AGPR	GPRSR	Constant interpolation	Linear interpolation
SSIM	**0.994**	0.989	0.991	0.948	0.995

Figure 6: Different algorithms tested for radiation image upscaling on an example problem. CNN-ML: a convolutional neutral network based ML model. AGPR, GPRSR: Gaussian Process Regression based models. Constant, linear interpolation: non-ML algorithms. The percentage deviation maps and SSIM (structural similarity index measure) are used to quantify the model upscaling performance.

2) The reconstructed radiation image can have missing regions due to various artifacts in the image domain (e.g., a ground elevation model) prior to reconstruction. Such missing regions create ridges of artificially-high radiation intensities, since activity that would normally be attributed to those missing regions gets pushed to nearby pixels/voxels instead. ML algorithms can be used to fill in (i.e., inpaint) these missing ground surface regions with estimated values, and provide uncertainty estimates on the inpainted values.

We have identified and tested a Bayesian ML model, the Deep Gaussian Markov Random Field (DGMRF) [124] from Section 3.1.2, that can be used for radiation image inpainting. We developed synthetic ground height maps, artificially removed sections of the maps to simulate measurement artifacts, and then used the DGMRF model to inpaint the missing values. The DGMRF model produced promising performance, quantified by high SSIM scores with respect to the true height map values, and it also provides uncertainty on the inpainted values—see Fig. 7.

3) An uncertainty map provides critical information that complements the radiation intensity map. The most widely-deployed radiation image reconstruction algorithm, Maximum Likelihood Expectation Maximization (ML-EM), provides no uncertainty information of the radiation image data by default. The exact error propagation of ML-EM, known as Iterative Bayesian Unfolding

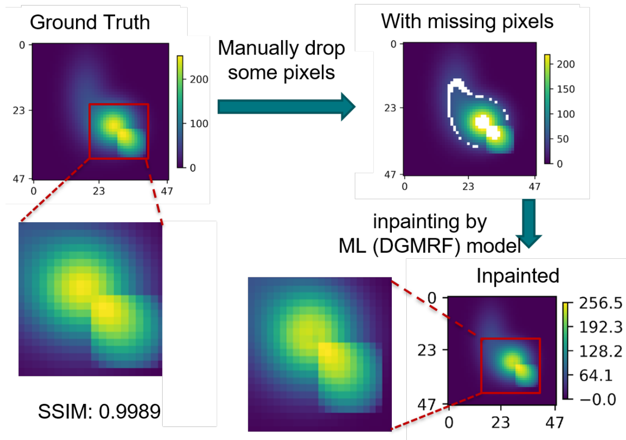


Figure 7: DGMRF algorithm for radiation image inpainting

(IBU) [125], is extremely computationally expensive, scaling as (at least) N^5, where N is the number of image pixels/voxels. Exact UQ via IBU is therefore unsuitable for real-time UQ of radiation images, and fast approximate methods are required instead.

We have tested a Markov Chain Monte Carlo MCMC sampler developed by the UC Berkeley sub-group, Microcanonical Langevin Monte Carlo (MCLMC) [126, 127], with a focus on evaluation of computation time and reconstruction accuracy. The MCLMC is able to reconstruct a radiation intensity and uncertainty map in less than 20 s for dimensionality $N \sim 10^4$, which significantly outperforms other MCMC samplers—see Fig. 8. In particular the MCLMC vastly improves upon the state-of-the-art Hamiltonian Monte Carlo (HMC) times of ∼1 hour found prior to this project. The radiation map and its associated uncertainty map agrees well with expected correct results. Such performance makes online radiation map reconstruction with uncertainty values feasible. We have organized the code in a online repository to make it easy to use and readily available to be interfaced with existing LBNL radiation map data processing tools. Finally, we are preparing a manuscript comparing the performance of these various UQ methods, MCMC-based and otherwise.

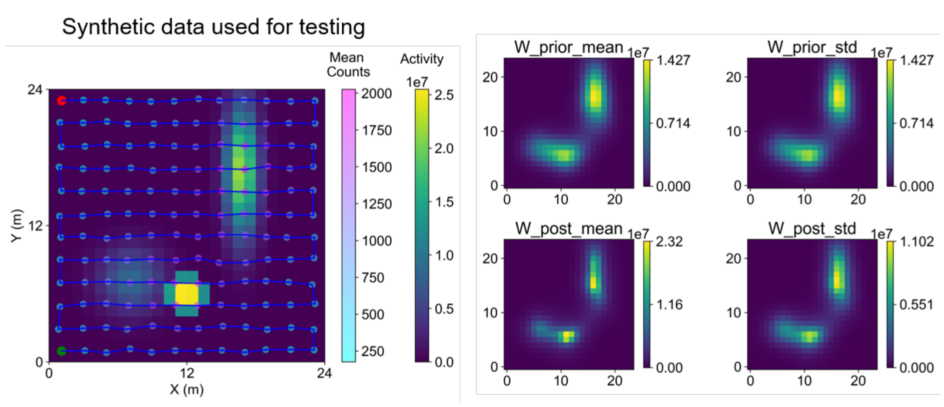


Figure 8: Overview of the MCLMC radiation image UQ results. Left: the synthetic ground truth distribution. Upper center and upper right: the ML-EM reconstruction result used as the mean and standard deviation of a Gaussian prior for the MCLMC. Lower center and lower right: the mean and standard deviation (UQ estimate) from the MCLMC.

3.4.2 Proposed research

Operationalizing MCLMC uncertainty quantification In phase 2 of this project, we intend to further operationalize the MCLMC UQ method developed in phase 1, in particular by deploying it on a live radiation mapping system such as LBNL's NG-LAMP [128], and demonstrating its real-time usage. Consideration will be given to computational performance, power usage, and how to present the UQ results to the system operator in an intuitive and informative way. One possibility for a demonstration of this technology is to leverage an upcoming multi-lab, multi-project, wide-area distributed sources measurement campaign that is being planned for 2025 or 2026 under the auspices of DOE NNSA NA-22. The fast but accurate nature of the MCLMC UQ results will also enable path planning methods based on the UQ map, e.g., moving in the direction of the largest relative uncertainty pixel. Operationalizing MCLMC will therefore benefit several other LBNL radiation mapping projects that are increasingly focused on real-time autonomous mapping.

Low-dose radiation imaging and data sufficiency A major outstanding question in radiation mapping is: how much data is necessary for a "good quality" reconstruction? Image quality is generally improved with more data, but collecting more data requires time and battery life and increases dose to human operators. For example, [3, Fig. 14] by co-PI Vavrek et al. shows the effect of raster line spacing (and therefore total measurement time) on image quality. In this particular example, the image quality metrics suggest that the original measurement was far above some abstract "data sufficiency criterion" and that a measurement half as long would have been sufficient.

However, data sufficiency for radiation imaging remains more an art than a science. As such, in phase 2, we will explore Bayesian/ML methods for determining data sufficiency and thus enabling efficient and low-dose radiation imaging. We will first investigate whether there exist any (possibly) Bayesian methods for quantifying data sufficiency, looking in particular at the system matrix in the radiation mapping linear model.

Then, for imaging cases near- or below-criterion, we will investigate methods to produce a "good quality" image with less-than-sufficient data. We will look to the wealth of ML image de-noising techniques already developed in the medical imaging domain (e.g. for PET and SPECT imaging) to retain high image quality while reducing dose to the patient. In radiation mapping by analogy such techniques could allow human and/or robotic detector operators to move faster through an area while retaining high image quality, reducing dose and/or increasing coverage. The GPP and Bayesian optimization algorithms of the Duke sub-group provide a useful starting point; there are other methods of interest such as the Reader bootstrap [129]. Matrix approximation techniques such as randomized singular value decomposition may also allow for the learning of low-dimensional structure in the model system matrix, which can be leveraged in de-noising. Finally, we may leverage our phase 1 work by converting the image upscaling framework to an image de-noising framework, i.e., training the ML model to go from a reconstructed image to the ground truth image rather than from a low-fidelity to a high-fidelity reconstructed image.

Appendix 1: References

[1] K. Vetter, R. Barnowski, J. W. Cates, A. Haefner, T. H. Joshi, R. Pavlovsky, and B. J. Quiter, "Advances in nuclear radiation sensing: Enabling 3-d gamma-ray vision," *Sensors* **19** no. 11, (2019) 2541.

[2] D. Hellfeld, M. S. Bandstra, J. R. Vavrek, D. L. Gunter, J. C. Curtis, M. Salathe, R. Pavlovsky, V. Negut, P. J. Barton, J. W. Cates, *et al.*, "Free-moving quantitative gamma-ray imaging," *Scientific reports* **11** no. 1, (2021) 1–14.

[3] J. R. Vavrek, J. Lee, M. Salathe, M. S. Bandstra, D. Hellfeld, B. J. Quiter, and T. H. Joshi, "Surrogate distributed radiological sources iii: quantitative distributed source reconstructions," *arXiv preprint arXiv:2412.02926* (2024) .

[4] Y. Ji, S. Mak, D. Soeder, J. F. Paquet, and S. A. Bass, "A graphical multi-fidelity Gaussian process model, with application to emulation of expensive computer simulations," *Technometrics* **66** no. 2, (2024) 267–281.

[5] Y. Ji, H. S. Yuchi, D. Soeder, J.-F. Paquet, S. A. Bass, V. R. Joseph, C. F. Wu, and S. Mak, "Conglomerate multi-fidelity Gaussian process modeling, with application to heavy-ion collisions," *SIAM/ASA Journal of Uncertainty Quantification* **12** (2024) 473–502.

[6] C. L. Sung, Y. Ji, S. Mak, W. Wang, and T. Tang, "Stacking designs: designing experiments for multi-fidelity modeling with confidence," *SIAM/ASA Journal on Uncertainty Quantification* **12** no. 1, (2024) 157–181.

[7] S. R. Narayanan, Z. Sun, S. Yang, J. Miller, S. Mak, K. Kim, and C. B. Kweon, "Local-transfer gaussian process (ltgp) learning for multi-fuel diesel engines," in *Proceedings of the 2025 AIAA Science and Technology (SciTech) Forum and Exposition*. 2025.

[8] X. Wang, S. Mak, J. Miller, and J. Wu, "Local transfer learning Gaussian process modeling, with applications to surrogate modeling of expensive computer simulators," *arXiv preprint arXiv:2410.12690* (2024) .

[9] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[10] T. Tang, S. Mak, and D. Dunson, "Hierarchical shrinkage Gaussian processes: applications to computer code emulation and dynamical system recovery," *SIAM/ASA Journal on Uncertainty Quantification* **12** no. 4, (2024) 1085–1112.

[11] K. Li, M. Balakirsky, and S. Mak, "Trigonometric quadrature fourier features for scalable Gaussian process regression," in *2024 International Conference on Artificial Intelligence and Statistics*. 2024.

[12] K. Li and S. Mak, "ProSpar-GP: scalable Gaussian process modeling with massive non-stationary datasets," *arXiv preprint arXiv:2311.08752* (2023) .

[13] W.-N. Hsu and H.-T. Lin, "Active learning by learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29. 2015.

[14] D. Song, S. Mak, and C. F. Wu, "ACE: Active learning for causal inference with expensive experiments," in *29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining: Workshop on Causal Inference and Machine Learning*. 2023.

[15] R. Ehlers, Y. Chen, J. Mulligan, Y. Ji, A. Kumar, S. Mak, P. Jacobs, A. Majumder, A. Angerami, R. Arora, *et al.*, "Bayesian inference analysis of jet quenching using inclusive jet and hadron suppression measurements," *arXiv preprint arXiv:2408.08247* (2024) .

[16] S. R. Narayanan, Y. Ji, H. D. Sapra, C.-B. M. Kweon, K. S. Kim, Z. Sun, S. Kokjohn, S. Mak, and S. Yang, "A misfire-integrated Gaussian process (MInt-GP) emulator for energy-assisted compression ignition (EACI) engines with varying cetane number jet fuels," *International Journal of Engine Research* (2024) 14680874241229514.

[17] K. Li, S. Mak, J.-F. Paquet, and S. A. Bass, "Additive Multi-Index Gaussian process modeling, with application to multi-physics surrogate modeling of the quark-gluon plasma," *arXiv preprint arXiv:2306.07299* (2023) .

[18] Z. Chen, S. Mak, and C. F. J. Wu, "A hierarchical expected improvement method for Bayesian optimization," *Journal of the American Statistical Association* **119** no. 546, (2024) 1619–1632.

[19] J. J. Miller, S. Mak, B. Sun, S. R. Narayanan, S. Yang, Z. Sun, K. S. Kim, and C.-B. M. Kweon, "Diverse Expected Improvement (DEI): Diverse Bayesian Optimization of Expensive Computer Simulators," *arXiv preprint arXiv:2410.01196* (2024) .

[20] X. Zheng, S. Mak, L. Xie, and Y. Xie, "Percept: a new online change-point detection method using topological data analysis," *Technometrics* **65** no. 2, (2023) 162–178.

[21] X. Zheng, S. Mak, L. Xie, and Y. Xie, "eRPCA: Robust Principal Component Analysis for Exponential Family Distributions," *Statistical Analysis and Data Mining: The ASA Data Science Journal* **17** no. 2, (2024) e11670.

[22] Y. Ji, R. Lekivetz, S. Mak, and J. Morgan, "BayesFlo: Bayesian Fault Localization for Software Testing," in *2023 IEEE 23rd International Conference on Software Quality, Reliability, and Security Companion (QRS-C)*, pp. 865–866, IEEE. 2023.

[23] Y. Ji, S. Mak, R. Lekivetz, and J. Morgan, "MaLT: Machine-Learning-Guided Test Case Design and Fault Localization of Complex Software Systems," in *2024 22nd ACM-IEEE International Symposium on Formal Methods and Models for System Design (MEMOCODE)*, pp. 58–62, IEEE. 2024.

[24] Y. Ji, S. Mak, R. Lekivetz, and J. Morgan, "Bayesflo: Bayesian fault localization of complex software systems," *arXiv preprint arXiv:2403.08079* (2024) .

[25] R. B. Gramacy, *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. CRC Press, 2020.

[26] S. Mak, C.-L. Sung, X. Wang, S.-T. Yeh, Y.-H. Chang, V. R. Joseph, V. Yang, and C. F. J. Wu, "An efficient surrogate model for emulation and physics extraction of large eddy simulations," *Journal of the American Statistical Association* **113** no. 524, (2018) 1443–1456.

[27] S.-T. Yeh, X. Wang, C.-L. Sung, Mak, S, Y.-H. Chang, L. Zhang, C. J. Wu, and V. Yang, "Common proper orthogonal decomposition-based spatiotemporal emulator for design exploration," *AIAA Journal* **56** no. 6, (2018) 2429–2442.

[28] Y.-H. Chang, X. Wang, L. Zhang, Y. Li, Mak, S, C.-F. J. Wu, and V. Yang, "Reduced-order modeling for complex flow emulation by common kernel-smoothed proper orthogonal decomposition," *AIAA Journal* **59** no. 9, (2021) 1–13.

[29] J. Chen, Mak, S, V. R. Joseph, and C. Zhang, "Adaptive design for Gaussian process regression under censoring," *Annals of Applied Statistics* (2021) . Accepted.

[30] J. Chen, Mak, S, V. R. Joseph, and C. Zhang, "Function-on-function kriging, with applications to three-dimensional printing of aortic tissues," *Technometrics* **63** (2020) 384–395.

[31] **JETSCAPE** Collaboration, S. Cao *et al.*, "Determining the jet transport coefficient $\hat{q}$ from inclusive hadron suppression measurements using Bayesian parameter estimation," *Phys. Rev. C* **104** no. 2, (2021) 024905, `arXiv:2102.11337 [nucl-th]`.

[32] **JETSCAPE** Collaboration, D. Everett *et al.*, "Phenomenological constraints on the transport properties of QCD matter with data-driven model averaging," *Phys. Rev. Lett.* **126** no. 24, (2021) 242301, `arXiv:2010.03928 [hep-ph]`.

[33] **JETSCAPE** Collaboration, D. Everett *et al.*, "Multisystem Bayesian constraints on the transport coefficients of QCD matter," *Phys. Rev. C* **103** no. 5, (2021) 054904, `arXiv:2011.01430 [hep-ph]`.

[34] A. Damianou and N. D. Lawrence, "Deep Gaussian processes," in *Artificial Intelligence and Statistics*. 2013.

[35] A. Sauer, A. Cooper, and R. B. Gramacy, "Vecchia-approximated deep Gaussian processes for computer experiments," *Journal of Computational and Graphical Statistics* **32** no. 3, (2023) 824–837.

[36] C.-C. Chang and L. Zhou, "Neural network emulation of inverse dynamics for a magnetorheological damper," *Journal of Structural Engineering* **128** no. 2, (2002) 231–239.

[37] V. M. Krasnopolsky, M. S. Fox-Rabinovitz, and D. V. Chalikov, "New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of longwave radiation in a climate model," *Monthly Weather Review* **133** no. 5, (2005) 1370–1383.

[38] D. Xiu, *Numerical Methods for Stochastic Computations*. Princeton University Press, 2010.

[39] J. P. Mattern, K. Fennel, and M. Dowd, "Estimating time-dependent parameters for a biological ocean model using an emulator approach," *Journal of Marine Systems* **96** (2012) 32–47.

[40] A. W. van der Vaart and J. H. van Zanten, "Rates of contraction of posterior distributions based on Gaussian process priors," *The Annals of Statistics* **36** no. 3, (2008) 1435 – 1463. `https://doi.org/10.1214/009053607000000613`.

[41] B. Adlam, J. Lee, L. Xiao, J. Pennington, and J. Snoek, "Exploring the uncertainty properties of neural networks' implicit priors in the infinite-width limit," *arXiv preprint arXiv:2010.07355* (2020) .

[42] J. Yao, W. Pan, S. Ghosh, and F. Doshi-Velez, "Quality of uncertainty quantification for Bayesian neural network inference," *arXiv preprint arXiv:1906.09686* (2019) .

[43] R. B. Gramacy and H. Lian, "Gaussian process single-index models as emulators for computer experiments," *Technometrics* **54** no. 1, (2012) 30–41.

[44] P. Seshadri, S. Yuchi, and G. T. Parks, "Dimension reduction via Gaussian ridge functions," *SIAM/ASA Journal on Uncertainty Quantification* **7** no. 4, (2019) 1301–1322.

[45] R. Tripathy, I. Bilionis, and M. Gonzalez, "Gaussian processes with built-in dimensionality reduction: Applications to high-dimensional uncertainty propagation," *Journal of Computational Physics* **321** (2016) 191–223.

[46] I. Delbridge, D. Bindel, and A. G. Wilson, "Randomly projected additive Gaussian processes for regression," *Proceedings of the 37th International Conference on Machine Learning* **119** (13–18 jul, 2020) 2453–2463.

[47] G. Chen and R. Tuo, "Projection pursuit Gaussian process regression," *IISE Transactions* (2022) . To appear.

[48] R. Zhang, S. Mak, and D. Dunson, "Gaussian process subspace prediction for model reduction," *SIAM Journal on Scientific Computing* **44** no. 3, (2022) A1428–A1449.

[49] A. Pinkus, *Ridge Functions*, vol. 205. Cambridge University Press, 2015.

[50] P. G. Constantine, B. Zaharatos, and M. Campanelli, "Discovering an active subspace in a single-diode solar cell model," *Statistical Analysis and Data Mining: The ASA Data Science Journal* **8** no. 5, (2015) 264–273.

[51] E. Buckingham, "On physically similar systems; illustrations of the use of dimensional equations," *Physical Review* **4** no. 4, (1914) 345.

[52] P. G. Constantine, E. Dow, and Q. Wang, "Active subspace methods in theory and practice: applications to kriging surfaces," *SIAM Journal on Scientific Computing* **36** no. 4, (2014) A1500–A1524.

[53] G. Casella and R. Berger, *Statistical Inference*. CRC press, 2024.

[54] R. F. MacLehose and D. B. Dunson, "Bayesian semiparametric multiple shrinkage," *Biometrics* **66** no. 2, (2010) 455–462.

[55] T. Park and G. Casella, "The Bayesian lasso," *Journal of the American Statistical Association* **103** no. 482, (2008) 681–686.

[56] M. Titsias and N. D. Lawrence, "Bayesian Gaussian process latent variable model," in *International Conference on Artificial Intelligence and Statistics*, pp. 844–851. 13–15 may, 2010. https://proceedings.mlr.press/v9/titsias10a.html.

[57] D. Everett, W. Ke, J.-F. Paquet, G. Vujanovic, S. Bass, L. Du, C. Gale, M. Heffernan, U. Heinz, D. Liyanage, *et al.*, "Multisystem Bayesian constraints on the transport coefficients of QCD matter," *Physical Review C* **103** no. 5, (2021) 054904.

[58] H. A. Chipman, E. I. George, and R. E. McCulloch, "BART: Bayesian additive regression trees," *The Annals of Applied Statistics* **4** no. 1, (2010) 266 – 298. https://doi.org/10.1214/09-AOAS285.

[59] M. Heinonen, H. Mannerström, J. Rousu, S. Kaski, and H. Lähdesmäki, "Non-stationary Gaussian process regression with Hamiltonian Monte Carlo," in *Artificial Intelligence and Statistics*, pp. 732–740, PMLR. 2016.

[60] Mak, S, H. S. Yuchi, and Y. Xie, "Information-guided sampling for low-rank matrix completion," in *ICML Workshop on Information Theoretic Methods for Rigorous, Responsible, and Reliable Machine Learning*. 2021.

[61] P. I. Frazier, "A tutorial on Bayesian optimization," *arXiv preprint arXiv:1807.02811* (2018) .

[62] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization* **13** no. 4, (1998) 455–492.

[63] A. D. Bull, "Convergence rates of efficient global optimization algorithms," *Journal of Machine Learning Research* **12** (2011) 2879–2904.

[64] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova, and C. Zhong, "Interpretable machine learning: Fundamental principles and 10 grand challenges," *Statistic Surveys* **16** (2022) 1–85.

[65] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics* **378** (2019) 686–707.

[66] M. W. Wheeler, D. B. Dunson, S. P. Pandalai, B. A. Baker, and A. H. Herring, "Mechanistic hierarchical Gaussian processes," *Journal of the American Statistical Association* **109** no. 507, (2014) 894–904.

[67] J. Chen, Z. Chen, C. Zhang, and C. Jeff Wu, "APIK: Active physics-informed kriging model with partial differential equations," *SIAM/ASA Journal on Uncertainty Quantification* **10** no. 1, (2022) 481–506.

[68] S. Golchi, D. R. Bingham, H. Chipman, and D. A. Campbell, "Monotone emulation of computer experiments," *SIAM/ASA Journal on Uncertainty Quantification* **3** no. 1, (2015) 370–392.

[69] X. Wang and J. O. Berger, "Estimating shape constrained functions using Gaussian processes," *SIAM/ASA Journal on Uncertainty Quantification* **4** no. 1, (2016) 1–25.

[70] M. H. Y. Tan, "Gaussian process modeling with boundary information," *Statistica Sinica* (2018) 621–648.

[71] I. Vernon, J. Liu, M. Goldstein, J. Rowe, J. Topping, and K. Lindsey, "Bayesian uncertainty analysis for complex systems biology models: emulation, global parameter searches and evaluation of gene functions," *BMC systems biology* **12** (2018) 1–29.

[72] L. Ding, S. Mak, and C. F. Wu, "BdryGP: a new Gaussian process model for incorporating boundary information," *arXiv preprint arXiv:1908.08868* (2019) .

[73] S. E. Jackson and I. Vernon, "Efficient emulation of computer models utilising multiple known boundaries of differing dimension," *Bayesian Analysis* **18** no. 1, (2023) 165–191.

[74] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics* **42** no. 1, (2000) 55–61.

[75] M. E. Johnson, L. M. Moore, and D. Ylvisaker, "Minimax and maximin distance designs," *Journal of Statistical Planning and Inference* **26** no. 2, (1990) 131–148.

[76] R. Eberhart and J. Kennedy, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, Citeseer. 1995.

[77] S. Mak and V. R. Joseph, "Minimax and minimax projection designs using clustering," *Journal of Computational and Graphical Statistics* **27** no. 1, (2018) 166–178.

[78] S. Surjanovic and D. Bingham, "Virtual Library of Simulation Experiments: Test Functions and Datasets,".

[79] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of Artificial Intelligence Research* **4** (1996) 129–145.

[80] D. J. MacKay, "Information-based objective functions for active data selection," *Neural computation* **4** no. 4, (1992) 590–604.

[81] **ALICE** Collaboration, S. Acharya *et al.*, "Performance of the ALICE Electromagnetic Calorimeter," *JINST* **18** no. 08, (2023) P08007, `arXiv:2209.04216 [physics.ins-det]`.

[82] R. Ehlers and J. Mulligan, "ALICE Overwatch: Online monitoring and data quality assurance using HLT data," in *EPJ Web of Conferences*, vol. 214, p. 01038, EDP Sciences. 2019.

[83] M. West, "Bayesian model monitoring," *Journal of the Royal Statistical Society. Series B (Methodological)* **48** no. 1, (1986) 70–78.

[84] P. Fearnhead and Z. Liu, "On-line inference for multiple changepoint problems," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **69** no. 4, (2007) 589–605.

[85] R. P. Adams and D. J. MacKay, "Bayesian online changepoint detection," *arXiv preprint arXiv:0710.3742* (2007) .

[86] P. Sidén and F. Lindsten, "Deep gaussian markov random fields," in *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020.

[87] J. S. Yedidia, "Message-Passing Algorithms for Inference and Optimization: "Belief Propagation" and "Divide and Concur"," *Journal of Statistical Physics* **145** (2011) 860–890.

[88] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *2010 IEEE Information Theory Workshop on Information Theory*, pp. 1–5, IEEE. 2010.

[89] G. Grosso, N. Lai, M. Letizia, J. Pazzini, M. Rando, L. Rosasco, A. Wulzer, and M. Zanetti, "Fast kernel methods for data quality monitoring as a goodness-of-fit test," *Machine Learning: Science and Technology* **4** no. 3, (2023) 035029.

[90] O. Zachariadis, N. Satpute, J. Gómez-Luna, and J. Olivares, "Accelerating sparse matrix–matrix multiplication with GPU tensor cores," *Computers & Electrical Engineering* **88** (2020) 106848.

[91] A.-K. Schuetz, A. W. P. Poon, and A. Li, "RESuM: Rare Event Surrogate Model for Physics Detector Design," arXiv:2410.03873 [physics.ins-det].

[92] LEGEND-Collaboration *et al.*, "Legend-1000 preconceptual design report," 2021. https://arxiv.org/abs/2107.11462.

[93] C. Wiesinger, L. Pandola, and S. Schönert, "Virtual depth by active background suppression: revisiting the cosmic muon induced background of gerda phase II," *Eur. Phys. J. C* **78** no. 7, (2018) . https://doi.org/10.1140%2Fepjc%2Fs10052-018-6079-3.

[94] H. Roch, S. A. Jahan, and C. Shen, "Model emulation and closure tests for (3+1)D relativistic heavy-ion collisions," *Phys. Rev. C* **110** no. 4, (2024) 044904, arXiv:2405.12019 [nucl-th].

[95] M. Karamanis, D. Nabergoj, F. Beutler, J. A. Peacock, and U. Seljak, "pocoMC: A Python package for accelerated Bayesian inference in astronomy and cosmology," *J. Open Source Softw.* **7** no. 79, (2022) 4634, arXiv:2207.05660 [astro-ph.IM].

[96] S. A. Jahan, H. Roch, and C. Shen, "Bayesian analysis of (3+1)D relativistic nuclear dynamics with the RHIC beam energy scan data," *Phys. Rev. C* **110** no. 5, (2024) 054905, arXiv:2408.00537 [nucl-th].

[97] **JETSCAPE** Collaboration, R. Ehlers *et al.*, "Bayesian Inference analysis of jet quenching using inclusive jet and hadron suppression measurements," arXiv:2408.08247 [hep-ph].

[98] J. Gong, H. Roch, and C. Shen, "A Gaussian Process Generative Model for QCD Equation of State," arXiv:2410.22160 [nucl-th].

[99] H. Roch and S. A. Jahan, "Hendrik1704/gpbayestools-hic: v1.1.0," July, 2024. https://doi.org/10.5281/zenodo.12807892.

[100] S. A. Jahan, H. Roch, and C. Shen, "Bayesian analysis of (3+1)d relativistic nuclear dynamics with the rhic beam energy scan data," July, 2024. https://doi.org/10.5281/zenodo.12807556.

[101] J. Gong, H. Roch, and C. Shen, "Hendrik1704/gp_eos_generator: v1.0.0," Oct., 2024. https://doi.org/10.5281/zenodo.14008389.

[102] J. H. Putschke *et al.*, "The JETSCAPE framework," arXiv:1903.07706 [nucl-th].

[103] M. Plumlee, O. Sürer, S. M. Wild, and M. Y.-H. Chan, "surmise 0.2.1 users manual," Tech. Rep. Version 0.2.1, NAISE, 2023. https://surmise.readthedocs.io.

[104] K. Beyer *et al.*, "BANDFramework: An open-source framework for Bayesian analysis of nuclear dynamics," Tech. Rep. Version 0.3.0, 2023. https://github.com/bandframework/bandframework.

[105] **JETSCAPE** Collaboration, A. Kumar *et al.*, "Inclusive jet and hadron suppression in a multistage approach," *Phys. Rev. C* **107** no. 3, (2023) 034911, arXiv:2204.01163 [hep-ph].

[106] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics* **21** no. 2, (1979) 239–245. http://www.jstor.org/stable/1268522.

[107] R. L. Iman, J. M. Davenport, and D. K. Zeigler, "Latin hypercube sampling (program user's guide). [lhc, in fortran]," tech. rep., Sandia Labs., Albuquerque, NM (USA), 01, 1980. https://www.osti.gov/biblio/5571631.

[108] G. Nijs, W. van der Schee, U. Gürsoy, and R. Snellings, "Bayesian analysis of heavy ion collisions with the heavy ion computational framework Trajectum," *Phys. Rev. C* **103** no. 5, (2021) 054909, arXiv:2010.15134 [nucl-th].

[109] D. Liyanage, O. Sürer, M. Plumlee, S. M. Wild, and U. Heinz, "Bayesian calibration of viscous anisotropic hydrodynamic simulations of heavy-ion collisions," *Phys. Rev. C* **108** no. 5, (2023) 054905, arXiv:2302.14184 [nucl-th].

[110] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society Series B: Statistical Methodology* **63** no. 3, (01, 2002) 425–464, https://academic.oup.com/jrsssb/article-pdf/63/3/425/49590547/jrsssb_63_3_425.pdf. https://doi.org/10.1111/1467-9868.00294.

[111] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne, "Combining field data and computer simulations for calibration and prediction," *SIAM Journal on Scientific Computing* **26** no. 2, (2004) 448–466, https://doi.org/10.1137/S1064827503426693. https://doi.org/10.1137/S1064827503426693.

[112] J. Brynjarsdóttir and A. O'Hagan, "Learning about physical parameters: the importance of model discrepancy," *Inverse Problems* **30** no. 11, (Oct, 2014) 114007. https://dx.doi.org/10.1088/0266-5611/30/11/114007.

[113] D. Foreman-Mackey, "corner.py: Scatterplot matrices in python," *Journal of Open Source Software* **1** no. 2, (2016) 24. https://doi.org/10.21105/joss.00024.

[114] S. Weinberg, "Entropy generation and the survival of protogalaxies in an expanding universe," *Astrophys. J.* **168** (1971) 175.

[115] G. S. Denicol, S. Jeon, and C. Gale, "Transport Coefficients of Bulk Viscous Pressure in the 14-moment approximation," *Phys. Rev. C* **90** no. 2, (2014) 024912, arXiv:1403.0962 [nucl-th].

[116] S. I. Finazzo, R. Rougemont, H. Marrochio, and J. Noronha, "Hydrodynamic transport coefficients for the non-conformal quark-gluon plasma from holography," *JHEP* **02** (2015) 051, arXiv:1412.2968 [hep-ph].

[117] Y. Yamauchi, L. Buskirk, P. Giuliani, and K. Godbey, "Normalizing Flows for Bayesian Posteriors: Reproducibility and Deployment," arXiv:2310.04635 [nucl-th].

[118] J. Robnik, G. B. De Luca, E. Silverstein, and U. Seljak, "Microcanonical Hamiltonian Monte Carlo," arXiv:2212.08549 [stat.CO].

[119] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics* **22** no. 1, (1951) 79 – 86. https://doi.org/10.1214/aoms/1177729694.

[120] R. G. Bower, M. Goldstein, and I. Vernon, "Galaxy formation: a Bayesian uncertainty analysis," *Bayesian Analysis* **5** no. 4, (2010) 619 – 669. https://doi.org/10.1214/10-BA524.

[121] K. Prajapati, V. Chudasama, H. Patel, A. Sarvaiya, K. P. Upla, K. Raja, R. Ramachandra, and C. Busch, "Channel split convolutional neural network (chasnet) for thermal image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4368–4377. 2021.

[122] H. He and W.-C. Siu, "Single image super-resolution using gaussian process regression," in *CVPR 2011*, pp. 449–456, IEEE. 2011.

[123] L. Pan and J. Vavrek, "Acceleration of radiological mapping by image super resolution," in *2024 IEEE Nuclear Science Symposium (NSS), Medical Imaging Conference (MIC) and Room Temperature Semiconductor Detector Conference (RTSD)*, pp. 1–2, IEEE. 2024.

[124] P. Sidén and F. Lindsten, "Deep gaussian markov random fields," in *International conference on machine learning*, pp. 8916–8926, PMLR. 2020.

[125] G. D'Agostini, "Improved iterative bayesian unfolding," *arXiv preprint arXiv:1010.0632* (2010) .

[126] J. Robnik and U. Seljak, "Fluctuation without dissipation: Microcanonical langevin monte carlo," *arXiv preprint arXiv:2303.18221* (2023) .

[127] J. Robnik, G. B. De Luca, E. Silverstein, and U. Seljak, "Microcanonical hamiltonian monte carlo," *arXiv preprint arXiv:2212.08549* (2022) .

[128] R. Pavlovsky, J. Cates, W. Vanderlip, T. Joshi, A. Haefner, E. Suzuki, R. Barnowski, V. Negut, A. Moran, K. Vetter, *et al.*, "3d gamma-ray and neutron mapping in real-time with the localization and mapping platform from unmanned aerial systems and man-portable configurations," *arXiv preprint arXiv:1908.06114* (2019) .

[129] A. J. Reader and S. Ellis, "Bootstrap-optimised regularised image reconstruction for emission tomography," *IEEE transactions on medical imaging* **39** no. 6, (2020) 2163–2175.

[130] "HEPData.". https://www.hepdata.net/.

[131] "Zenodo.". https://zenodo.org/.

Appendix 2: Facilities

LBNL and UCB

The National Energy Research Scientific Computing Center (NERSC) at LBNL hosts the Perlmutter supercomputer. Perlmutter is an HPE Cray system based on the Shasta platform. The HPE Cray Shasta system integrates NVIDIA A100 GPUs, AMD "Milan" EPYC CPUs, a novel HPE Slingshot high-speed network, and a 35-petabyte FLASH scratch file system. It comprises of 3,072 CPU-only and 1,792 GPU-accelerated nodes. The PIs will apply for a NERSC computing and storage allocation to cover those aspects of the project that are best provided by NERSC computing.

Additional computing resources are provided on the Berkeley campus HPC cluster (Savio) for UC Berkeley faculty. LBNL also hosts the Lawrencium cluster [71], a general-purpose HPC facility comprising multiple generations of linux-based multi-core processors plus associated infrastructure, most notably Infiniband interconnects. The architecture of this facility maps well onto some of the projects in this proposal, in particular the large-scale computing needs of Bayesian Inference for Quark-Gluon Plasma studies.

Duke University:

The Duke University Department of Statistical Science maintains a near state of the art network of approximately fifty single-and dual-processor x86 and x86-64-based Linux workstations, approximately a dozen Windows PCs in a Samba network environment, and a range of networked monochrome and color Postscript printers for its faculty, Ph.D. students, and staff. Rack-mounted servers offer file, e-mail, web, and authentication service. A RAID storage server facility offers something under one terabyte of disk capacity, backed up daily to an LTO tape changer. The software environment includes a wide array of scientific programming tools, including the GNU suite of libraries, compilers, and development tools, a range of scientific and statistical computing environments such as Matlab, Maple, Mathematica, S-Plus, R, OpenBUGS, etc. An MPI-based parallel computing environment is provided that is consistent with the HPC environment at Duke's Computational Science, Engineering and Medicine (CSEM) facility, to aid investigators in prototyping and debugging parallel computer code. The computing environment is maintained by a full-time systems manager and systems programmer.

The backbone and other university-level infrastructure needs of the University are maintained by a central IT organization, the Office of Information Technology (OIT). OIT is responsible for the operation, testing, support, and engineering of the campus-wide data, voice, and video communications infrastructure. This includes the design and subsequent implementation of structured wiring and switching systems, enterprise-level servers, including Domain Name Server (DNS) and Dynamic Host Configuration Protocol (DHCP) servers, routing systems, and wireless systems.

Duke University's high-speed backbone, DukeNet, provides researchers, staff, faculty and students with a robust, redundant conduit for data. The backbone consists of Cisco routers with redundant 10 gigabit ethernet links. Most buildings on campus are wired with Category 5 cabling and have 10M/100M Ethernet ports supplied to each desktop. Servers and high speed research workstations can be provided with gigabit or ten gigabit ethernet ports as needed. Building networks connect to the backbone via dual gigabit or 10 gigabit ethernet uplinks.

The Duke Shared Cluster Resource (DSCR) facility maintains a shared computational cluster facility of over 600 machines (1152 processors) to which we have access. The processors range from 2.8GHz to 3.6GHz. While the cluster must be shared by the entire community at Duke, it provides a useful resource for computational science.

The University also maintains a campus-wide AFS file system infrastructure with terabytes of storage; a campus-wide electronic mail infrastructure supporting over 35,000 mailboxes and handling in excess of a million messages a day; a server-based file service, authentication services; directory services; web service; and name service and other network services.

Wayne State University

HPC Grid: The current HPC Grid consists of 502 servers running CentOS 7 and OpenHPC components with approximately 9,500 cores and 70 GPUs of various vintages. It has grown over the years based upon funding from multiple sources such as grants, investments by C&IT, and investments by colleges and the office of the Vice President for Research. As such, it has a broad range of configurations. The challenge of managing this complex environment is made possible by software tools such as xCAT (installation and configuration), PBSPro (scheduling) and Xymon and BigBrother (monitoring). These same tools will simplify and speed installation of the new cluster. Our HPC services are documented in detail at https://tech.wayne.edu/hpc. For example, a complete list of nodes with their associated GPU cards can be found at https://tech.wayne.edu/hpc/nodes.

Networking: C&IT operates a campus network with a dual 10G backbone (soon to be upgraded to 40G or 100G depending upon bids) with the main hub at the central computing facility and a secondary hub located on the west end of campus for redundancy. Connections to buildings are currently 1G with upgrades planned for 10G and higher where needed. WSU is a member of the Merit Network which provides connectivity to the Internet and to Internet2. WSU also has a 10G Science DMZ to four locations on campus with a link to Starlight in Chicago for access to national and global networks such as ESNet and LHCOne.

UCSD:

San Diego Supercomputer Center: SDSC hosts an array of five substantial High-Performance Computing (HPC) systems, including one specifically optimized for modern Machine Learning, utilizing Intel/Habana processors. Of these, the Expanse system—a Dell cluster initiated in 2020—stands out with its 13 SDSC Scalable Compute Units (SSCUs). This configuration includes 56 standard nodes powered by AMD EPYC (Rome) processors, alongside four GPU nodes equipped with Nvidia V100 GPUs, all interconnected through a 100 GB/s HDR InfiniBand. Expanse is also integrated with the Open Science Grid for high-throughput computing, supporting vast numbers of single-core jobs, and features a low-latency Mellanox HDR InfiniBand network, ideal for medium-scale parallel tasks requiring several thousand cores.

As an SDSC Faculty Fellow, the PI is entitled to 5,000 NVIDIA A10 equivalent GPU computing hours annually over three years, extending until 2027. This resource allocation supports escalating research demands, and the SDSC staff offers collaborative and intellectual engagement opportunities. Additionally, the PI benefits from 100 TB of storage space provided by SDSC, facilitating data storage and open data sharing initiatives.

Appendix 3: Equipment

No additional equipment will be required for this project.

Appendix 4: Data Management Plan

Data sources: The research supported by this proposal entails the application of novel analysis methods to experimental data and model calculations in several topical areas within the NP research portfolio. These data are generated primarily by large experimental and theoretical collaborations. Some relevant data have been published in refereed journals and are publicly available, and their usage does not require special consideration.

The PIs of this proposal include NP Domain Scientists who collaborate on these experiments, and their activities within this project require access to raw data and detailed experimental and calculational information that is unavailable to researchers who are not collaboration members. Such an approach is essential for this project, which cannot be carried out based solely on publicly available data and the high-level information provided in publications. However, in each collaboration, the usage and publication of data is governed by formal rules which specify how such internal analyses can be presented in conferences, public talks, and publications. The PIs will establish agreements with each collaboration and project whose data we will utilize, to ensure full conformity to their rules of data usage. These agreements will likewise conform to all DOE requirements concerning the usage of unpublished data. Publication of analyses based on such data will be handled on a case-by-case basis, consistent with these agreements.

Data Management: Effective data management enables developers and end users to quickly and efficiently read in simulation and experimental data, analyze them, and store the results. Since there are multiple sources of data that will be used in this project, Data Management requires a hybrid approach. For experiment-internal data, the data management approach of each collaboration will be followed. For multi-experiment analysis of published data, existing data management solutions that are in widespread usage in the community, such as HEPData [130] and Zenodo [131], will be applied.

Computing resources: As noted above, access to detailed experimental and calculational information is required for effective utilization of raw data from large experimental collaborations. This may necessitate analysing data in the native environment of each experiment, which requires access by participants in this project to the the computing facilities of the experiments themselves. Large physics experiments commonly have mechanisms that enable joint work with non-collaborators on specific technical projects, which grant the non-collaborators access to collaboration resources. We will utilize these mechanisms as needed. Several PIs of this proposal play leading roles in the relevant experiments, and we expect that such arrangements will not be burdensome to establish.

Data Archiving and Distribution: Data will be archived and made available publicly in accordance with best scientific practices and with the specific objective that all published results should be reproducible. To facilitate this objective, datasets generated and managed by this project will be registered through OSTI's DOE Data ID Service and made available through NERSC data portals, or equivalent. Data generated by the project and used in publications, including any reduced data, the results of inference algorithms, and all data used to generate charts and graphs, will be archived by the authors and made available publicly.

Software: The goal of this proposal is to develop general solutions that are broadly applicable to multiple projects. However, the NP projects in the proposal developed independently, using different computing languages and tool sets. This barrier must be overcome for the project to be successful, and new software interfaces will therefore have to be developed. Interfaces between C++, Python, and Julia will be provided. This will not only serve the NP projects in this proposal but also applications in the wider community.

Similarly, usage of high performance computing clusters requires additional interfaces for the new tools. Containers such as Singularity, Docker, and the NERSC-specific Shifter environment will be implemented. Conflicts may arise from parallelization, for instance when experiment native code uses the same parallelization methods as the Bayesian UQ and they interfere on the computing cluster. Collaboration with computer scientists will address these issues and provide solutions for the NP projects as well as for the wider community.

Standard software management tools such as GitHub will be employed.

Publications and Presentations: Publications and presentations will be reviewed before release in accordance with requirements established by DOE and by the home institutions of the authors. Publications and presentations will be archived by the project in PDF format as well as a source format, e.g., LaTeX, Word or Powerpoint. Publications and presentations will also be archived and made available through OSTI Open Archives Initiative and through the authors' home institutions. Pre-prints will be archived through http://arxiv.org.

Appendix 5: Promoting Inclusive and Equitable Research (PIER) Plan

This proposal is a multi-institutional collaboration between Lawrence Berkeley National Laboratory, Duke University, Wayne State University, the University of California at Berkeley, and the University of California at San Diego. Each institution has a strong commitment to the principles of IDEA (Inclusion, Diversity, Equity and Accountability) and has integrated them fully into their work culture and managerial practices.

Collaboration environment

The proposed collaborative effort will involve personnel with a very wide range of career experience and professional level: career faculty and laboratory career staff, term staff, postdocs, graduate students, and undergraduates. Much of the daily research work will be carried out by collaborators who do not have a permanent position, and whose career needs must be taken carefully into account to ensure their future success. Giving positions of responsibility to the early-career collaborators and promoting their work to the community provides the foundation for a welcoming, constructive collaboration culture.

The proposed project is highly multi-disciplinary, spanning multiple sub-fields of Nuclear Physics as well as areas of Data Science. Members of the collaboration therefore come from different communities with different expertise, culture, and modes of working. As noted in the proposal, the essential motivation to carry out such a multi-disciplinary project is to find and develop new connections between these sub-areas: the whole may then be greater than the sum of its parts. However, building such connections will require deliberate effort. Regular meetings (currently on a monthly basis) are held to discuss research progress. These meetings are chaired by early career scientists and provide them opportunities to speak on their research in a low-pressure, propitious environment, thereby preparing them for future opportunities to present publicly.

While postdocs and graduate students are often focused largely on their specific research, participation in this project will require early-career researchers to take a broader view of physics and data science than they would otherwise have, which is beneficial for their education and next career steps. However, many collaborating postdocs and students will in practice be dividing their time between an NP experiment and this project. The mentoring of a postdoc or student in such a circumstance has the additional challenge of ensuring that both projects get the appropriate level of attention, and that the student or postdoc is not pulled into too large and time-consuming a project in one area at the expense of the other. It is the collaboration's joint responsibility to ensure that postdocs and students have sufficient space and time to carry out their research, without feeling undue pressure from internal competition (this, unfortunately, happens in large experimental collaborations). Finding the right balance for each early-career researcher in the collaboration will require care and attention by collaboration members on a case-by-case basis; there is no general approach that works for everyone.

The collaboration is also multi-institutional, comprising one National Laboratory (LBNL) and four universities (Duke, Wayne State, UC San Diego, UC Berkeley). This provides the opportunities for extended research visits of early-career collaborators to different institutions, which will likewise broaden their horizons.

Leadership training

Leadership training is best done by providing the opportunity to lead. The career stage at which this opportunity is most crucial is at the level of postdoc and term staff, where an early-career person has developed experience and judgment and is looking for her next career step in academia or industry with larger leadership responsibility. There are several opportunities for leadership in the collaboration: organizing and running regular collaboration meetings; giving prominent conference talks on behalf of the collaboration; being a lead or corresponding author on a collaboration paper; serving on the editorial committee for other papers, etc. Collaborators at a vulnerable career stage will be given the highest priority for these leadership tasks, according to their interests and needs.

Formal Mentoring:

Each of our institutions provides a formal mentoring program and other resources for students and postdocs. We describe them here:

- **LBNL/UCB:** LBNL provides a Physical Sciences Area (PSA) mentoring program which partners a mentee with a mentor from a different field, to provide a perspective that is different than their group and supervisory chain. This program is not restricted to early-career mentees. The Berkeley Lab Postdoc Association at LBNL and Berkeley Postdoctoral Association at UCB offer the opportunity for postdocs to meet mentors and peers. We will encourage the postdocs to connect with both the PSA mentoring program as well as their respective postdoc associations. More senior members of the project are likewise encouraged to participate in the PSA Mentoring program. The LBNL co-PIs in this proposal have mentored postdoctoral fellows in grant writing and involved them in proposal reviews (with permission from DOE NP), so that they are more prepared for the typical tasks in their next career step.

- **Duke:** Duke has numerous formal mentoring programs geared towards training graduate students and preparing them for future endeavors. The Emerging Leaders Institute provides an 8-week program that helps graduate students and postdocs develop competencies in communication, self-awareness, professional adaptability, interdisciplinary teamwork, and leadership, leading to a certificate of completion. Duke offers a Professional Development Series, which features one-off talks, workshops, and events that help students identify and develop transferable skills to prepare them for careers in academia, industry, government, nonprofit, and entrepreneurship. The Graduate School at Duke also provides writing support such as academic courses, writing spaces, writing consultations, online resources, and additional support for international students.

- **Wayne State:** The postdoctoral fellow will get training in the professional skills that will aid them in succeeding in their future endeavors and reaching their full potential. The WSU Individual Development Plan (IDP) is an ideal framework to assist the postdoc scholars in developing their career plans and trajectories. By taking advantage of the Graduate and Postdoctoral Professional Development (GPPD) seminar series and the Grant Writing Seminars, the postdoc fellows will improve their abstract writing, job search, and presentation skills. The professional development programs such as Academic Leadership Academy, Broadening Experience in Scientific Training, reBUILDetroit, and the Postdoctoral to Faculty Transition Fellowship program can further boost their leaderships in the fields.

- **UCSD:** UCSD offers a variety of training resources for postdocs to enhance their professional development and prepare them for future career. The Office of Postdoctoral Scholar Affairs provides essential training, including a New Postdoc Orientation, Introduction to the Ethical Challenges of Research series, and the EPIC Bootcamp, which helps postdocs understand their role and succeed during their training at UCSD and beyond. Additionally, UCSD Extension offers certificate programs, such as the grAdvantage Certificate in Leadership and Teamwork, Project Management Certificate to equip postdocs with skills in leadership, teamwork, project management, and more. Furthermore, the UCSD Postdoctoral Association offers training and volunteer opportunities, including The Postdoc Survival Guide, events calendar, and volunteer opportunities with local non-profits. These resources aim to support postdocs in their career development and provide them with the skills and knowledge necessary to succeed in their chosen field.

Recruitment of Project Personnel

Given the multi-disciplinary nature of the project, postdocs will be recruited from several different areas.

To recruit physics-oriented postdocs, the most effective job posting is on the INSPIRE job ad board. Large experimental collaborations regular send out job announcements to their membership, and we will likewise utilize that mechanism.

This approach to posting job ads is effective in recruiting candidates who are already working in an NP sub-area. However, the project also includes a strong data-science component, and for some postdoc positions a different approach to recruitment is needed. We will broadcast such positions on a variety of data science job boards, including MathJobs, the IMS Job Board, the ASA JobWeb, and the ISBA Job Board.

To ensure that our recruitment effort has reached under-represented populations, we will likewise post job ads to Under-Represented Minority (URM) serving fora connected to physics and data science. These include the National Society of Hispanic Physicists, the National Society of Black Physicists, the Society of Advancement of Chicanos and Native Americans in Science, and the National Society for Black Engineers. The Nuclear Science Division at LBNL will pay for the expenses for the advertisements in these URM publications.

As part of the postdoc recruitment process, we will also contact other colleagues to aid in identifying potential candidates from under-represented groups, and the encourage such candidates to apply.

For graduate student recruitment, UC Berkeley, UC San Diego, and Duke University are university members of the GEM Consortium with the access to an impressive pool of under-represented STEM candidates. LBNL hosts GEM fellows in performing their research, and its Workforce Development & Education department will provide the required administrative support, while the research programs at LBNL will provide the financial support.

To build a diverse workforce pipeline in Nuclear Physics, the Nuclear Science Division at LBNL leads the GREAT-NS program, which was funded by DOE NP to provide undergraduate students at Minority Serving Institutions (MSIs) with traineeship opportunities in nuclear physics. PI Poon of this proposal is also co-PI in the GREAT-NS program, mentoring a veteran from a local MSI. We intend to work with the GREAT-NS program and the WD&E department to benefit more undergraduate students for this proposal.

The Physics Department at the University of California, Berkeley has an active program for “transfer” students — undergraduates who were transferred from community colleges — including research opportunities for these students. A similar program supported by the Patricia & Christopher Weil Family Foundation also exist at UC San Diego. We intend to recruit suitable candidates from this pool of undergraduate students, which consists of mostly first-generation and/or minority students.

Programs supported by the Department of Physics, such as the Berkeley Physics Undergraduate Research Scholarship and the Physics Innovators Initiative, provide financial support for students working with faculty members during the academic year and the summer months. In addition, the MPS Scholars program in the College of Letters and Sciences provides support to the students doing research in STEM fields with the focus on under-represented populations. PIs on this proposal will continue to participate in these programs; postdocs will be encourage to mentor undergraduates on the specific research projects.

The Co-PI at Wayne State (Shen) commits to actively recruit and support under-represented minority and first-generation graduate and undergraduate students. Wayne State University has a large base of minority students. About 35%/28% of its undergraduates/graduates are minorities, and 16% of the undergraduate and graduate students are African-American, according to the 21-22 Wayne State Fact book. The Co-PIs will formulate summer research projects on Machine Learning to engage more undergraduate students, especially those from the programs at Wayne State targeted at improving retention, namely the Wayne State Warrior Vision and Impact Program (VIP) and Academic Pathways to EXcellence (APEX) program.

The Co-PI at Duke University (Mak) has a strong track record of mentoring students from diverse backgrounds and experiences, and will continue to actively do so for this project. Two of Mak's seven current PhD students are women, and five of Mak's ten undergraduate thesis advisees are women. Mak is also on the leadership board of the Institute of Mathematical Statistics (IMS) New Researchers Group, which aims to provide career development opportunities for early-career statisticians, particularly those from underrepresented backgrounds and institutions. Mak will leverage these connections to further promote diversity, equity and inclusion, providing students from underrepresented backgrounds and institutions with ample research training and opportunities. The students mentored in this project will also learn and benefit from thinking from these diverse perspectives.

The co-PI at UC San Diego (Li), who holds a joint appointment in the Halıcıoğlu Data Science Institute and Department of Physics, has demonstrated excellence in mentoring students at the intersection of AI/ML research and physics analysis, with particular success supporting students from underrepresented groups. Notable achievements include:

- Mentoring three PhD students (including one LGBTQI+ woman) and 10 undergraduate researchers from diverse backgrounds
- Facilitating the public release of AI-ready particle physics data and integrated it into data science education, engaging over 160 data science students in group projects and in-class particle identification challenges. Remarkably, these students mastered complex physics analysis despite having no prior physics background.
- establishing the Germanium Machine Learning (GeM) group within the LEGEND collaboration. The GeM group features a comprehensive training pipeline that enables physics students with no prior AI/ML experience to develop expertise in the field. To date, the group has successfully launched 15 AI/ML projects within LEGEND by training international students and postdocs.