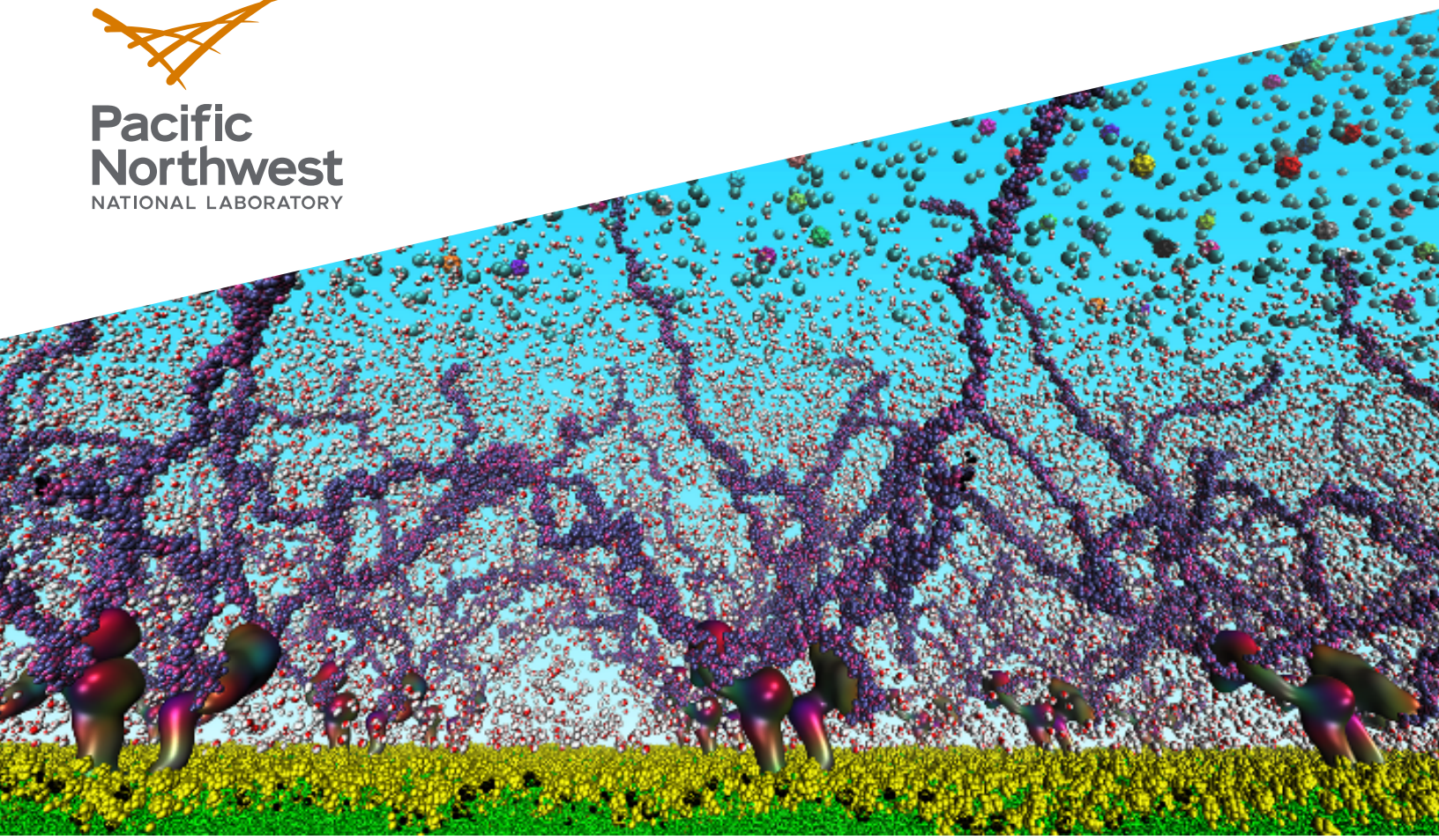
Pacific Northwest
NATIONAL LABORATORY



PhILMS: Collaboratory on Mathematics and Physics-Informed Learning Machines for Multiscale and Multiphysics Problems

October 2019

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <https://www.ntis.gov/about>
Online ordering: http://www.ntis.gov

PhILMS: Collaboratory on Mathematics and Physics-Informed Learning Machines for Multiscale and Multiphysics Problems

ANNUAL REPORT

October 2019

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Executive Summary

Motivation

Complex systems are governed by the hidden physics of interfaces and inhomogeneous cascades of scales, e.g., multifunctional materials, subsurface transport, reactive transport, hence requiring new data driven multiscale modeling that is accurate, efficient, and easy to implement. In particular, inhomogeneous cascades of scales involve long-range spatio-temporal interactions and often lack proper closure relations to form complete and mathematically well-posed systems of governing equations. Existing multiscale and multiphysics approaches and classical integer-order partial differential equations (PDEs) have been ineffective in addressing nonlocal interactions, inhomogeneous cascades of scales, or propagation of uncertainty and stochasticity across scales. They are inadequate for solving *inverse* stochastic multiscale problems, especially in the context of noisy *multifidelity* data.

Research Plan

We propose to develop stochastic multiscale modeling in conjunction with emerging deep learning techniques to seamlessly fuse physical laws and multifidelity data both for forward and inverse multiscale problems. Hence, we propose a synthesis of physics-based and data-driven tools and approaches, including nonlocal operators, multifidelity data and information fusion, deep neural networks (DNNs), meshless methods, uncertainty propagation, and stochasticity to simulate complex multiscale systems. Our high-level objective is to establish the Collaboratory on Mathematics and Physics-Informed Learning Machines for Multiscale and Multiphysics Problems (**PhILMs**) as a new Department of Energy (DOE) center at the interface of mathematics, physics, data science, and deep learning. With emphasis on predictability *and* uncertainty quantification, the research plan of PhILMs includes:

- *Developing physics-informed learning machines by encoding conservation laws and prior physical knowledge into deep learning networks and analyzing their mathematical properties.*
- *Demonstrating the effectiveness of PhILMs in designing functional materials with tunable properties, and extending PhILMs to other DOE-relevant multiscale problems, e.g., combustion, subsurface, and earth systems, all exhibiting inhomogeneous scaling cascades.*
- *Establishing scientific machine learning (ML) as a new meta-discipline at the interface of computational mathematics, data science, information fusion, and deep learning.*

Our integrated mathematical and computational activities can be classified into four research areas: (*RA-I*) PDE-based modeling of macroscales; (*RA-II*) stochastic modeling of mesoscales; (*RA-III*) bridging methods to connect the scales; and (*RA-IV*) deep learning approximations and algorithms to support RA-I to RA-III.

Coordination, Integration, and Evaluation Plan

Integrating and assimilating the four PhILMs research areas will lead to the development of PhILMs. Its usefulness in modeling multiscale phenomena and discovering new governing equations from multifidelity data will be demonstrated in the diverse driver applications motivated by DOE problems. In **RA-I**, we will develop flexible high-order meshless methods, a unified theory of nonlocal operators with corresponding high-order numerical algorithms, and PDE-based PhILMs for forward and inverse problems. We will emphasize the mathematical analysis of the new approximations induced by DNNs and their variants constrained by PDEs. In **RA-II**, we will develop stochastic PDEs and DNN-based algorithms to capture fluctuations in coarse-grained (CG) systems. We will emphasize new mathematical foundations based on statistical mechanics and derive rigorously stochastic reduced models based on the Mori-Zwanzig formalism. In **RA-III**, we will consider concurrent coupling of heterogeneous domains, including nonlocal to local PDEs (e.g., fractional/local), and propagation of stochastic fluctuations seamlessly across domains. In **RA-IV**, we will combine deep learning with Bayesian techniques for uncertainty quantification and integration of domain knowledge. We will develop nonlinear multifidelity statistical models for fusing information from multiscale simulations and available experimental (noisy) data. We will propose novel (physics-informed) neural network topologies and develop mathematics to explain the success of deep learning constrained with physical laws. Ultimately, these advances will be combined with the knowledge and inherent structure of physical laws to develop PhILMs for material design, enhancing efficiency in combustion, and making predictions for subsurface systems and ice sheets, as we explain next.

For the evaluation plan, we will define benchmarks of increasing complexity from one year to the next to measure progress in each research area, we will plan annual workshops and mini-symposia as well as mutual

visits to each Principle Investigator's (PI's) laboratories and student and postdoc exchanges. In addition, we have formed a dissemination panel so that our work is presented across all the DOE labs and beyond so that the DOE scientists will be the first who can benefit from our findings. During the first year we held almost weekly webinars on Monday afternoons attended by all researchers associated with PhILMs at Pacific Northwest National Laboratory (PNNL), Sandia National Laboratories (SNL), Stanford University, Massachusetts Institute of Technology (MIT), University of California, Santa Barbara (UCSB), and Brown University.

Highlights of Accomplishments and Outcomes

In the first year, we established substantial and meaningful collaborations between the six groups by mutual visits, weekly webinars, and a workshop at SNL in California (May 27, 2019). Many of the PIs participated previously in the Collaboratory on Mathematics for Mesoscopic Modeling of Materials (CM4) so there was particular emphasis in establishing interactions and eventually specific collaborations with the two new members of PhILMs, Valiant at Stanford and Daskalakis at MIT, who work on theoretical computer science. In particular, both of these team members presented their work to the group on three different occasions, highlighting different aspects that could benefit physics-informed learning. In terms of awards, Karmiadakis was named American Association for the Advancement of Science fellow (2018) and Daskalakis won the prestigious Rolf Nevanlinna prize (2018). Also, our work on coupling across scales for flows over soft multifunctional surfaces was on the cover of Soft Matter (2019) while our work on self-cleaning of hydrophobic rough surfaces by coalescence-induced wetting transition was on the cover of Langmuir (2019). In terms of software, we have several open-source software releases thus far, e.g., the library ADCME.jl with powerful metaprogramming features in Julia, GMLS-Nets allowing for general meshfree operator regression, and DeepXDE, a deep learning tool to solve PDEs of any type.

Research Area I

PDE-based Modeling of Macroscales: Generalized Moving Least Squares (GMLS) is a new meshless method developed during CM4 by Brown and SNL teams for simulating multiphysics problems in arbitrary geometries. Recently, the SNL and UCSB teams collaborated to develop solvers for PDEs on manifolds and GMLS-Nets, a meshfree regression operator as an alternative to convolution neural networks for physics-informed learning. The capability of GMLS-Nets was demonstrated in discovering data-driven PDE models from molecular data and to regress engineering quantities of interest from datasets characteristic of computational fluid dynamics (CFD) simulations. SNL has also been investigating data-driven network models based on GMLS and DNNs with application to design and simulation of electrical circuits. PNNL and Brown developed physics-informed neural networks (PINNs) for transport in porous media, and in particular for learning unknown parameter fields in PDE models and applied for assimilating measurements of hydraulic conductivity and hydraulic head and tracer concentration. PINNs can also be used to learn unknown physics, e.g., estimating an unknown term as in unsuaturated porous media. We extended PINNs for learning unknown physics by modeling the hydraulic conductivity and hydraulic head with two DNNs and using a third DNN to represent the governing PDE residual. These DNNs share weights and are trained jointly. This allows us to accurately learn the DNN approximations of the two fields without any measurements of the hydraulic head, showing superior accuracy to other methods such as Bayesian inference and maximum *a posteriori* probability (MAP) estimation. The Stanford team used a variant of PINNs to discover the constitutive law of a complex material. Specifically, they use DNN to approximate the unknown constitutive law (stress/strain relation) and finte elements to discretize the momentum conservation equation. They also studied the difference between DNN and traditional approximation functional forms, and found that DNNs are superior in cases where highly non-uniform sample points are used (adaptivity), where the function to be approximated exhibits sharp gradients in some regions (concentrated jumps similar to shocks), and in high dimension situations. DNNs are also robust to noise and consistently generalize better in many of our test cases. Finally, SNL and Brown developed nonlocal PINNs that can obtain the parametrized nonlocal kernels from data, and they also developed a unified theoretical framework that includes all nonlocal and fractional operators as special instances of nonlocal truncated weighted operators.

Research Area-II

Stochastic Modeling of Mesoscales: Brown and PNNL have developed an efficient parallel multiscale method that bridges the atomistic and mesoscale regimes via concurrent coupling of atomistic and mesoscopic models by combining an all-atom molecular dynamics description for specific atomistic details in the vicinity of the functional surface with a dissipative particle dynamics (DPD) approach that captures mesoscopic hydrodynamics in the domain away from the functional surface. They also employed the many-body DPD method to study self-cleaning

of hydrophobic rough surfaces by coalescence-induced wetting transition. PNNL developed a probabilistic ML approach to estimate the effective CG potential parameters for water-hexane mixture, a typical immiscible binary liquid/liquid mixture. PNNL researchers also proposed several generative adversarial network models for modeling stochastic dynamical systems given their time series measurements by learning the flow map and enforcing dynamic constraints during training for all three major modes of learning, namely supervised, unsupervised, and reinforcement learning.

Research Area-III

Bridging of Methods to Connect the Scales: We have developed methods for the concurrent coupling of heterogeneous domains, including coupling across scales, to discover the hidden physics models at the interfaces and provide the glue functions/functionals to seamlessly connect the cascade of scales. Brown, PNNL, and SNL have contributed to this research area. We designed a new composite network that learns from multifidelity data by exploiting correlations across different data sets and applied it to high-dimensional function approximation and to inverse geophysics problems. We also demonstrated how PINNs work for nonlocal problems, using molecular dynamics data, and how PINNs can be used to learn physical parameters for use in physical simulations across multiple length scales, and in particular in learning a nonlocal surface tension model. Finally, for problems with heterogeneous properties we proposed a domain decomposition PINN (DD-PINN), where the spatial domain is decomposed in two (or more) non-overlapping subdomains such that the mean value of the field is (nearly) constant within each subdomain. These DD-PINNs, which are amenable to parallel GPU computing, yield superior results compared to single-domain PINNs.

Research Area–IV

Statistical Learning and Deep Learning Approximations and Algorithms: We have explored foundational algorithms and mathematics for deep learning. We have combined deep learning with statistical models integrating domain knowledge about spatial correlations and physical invariances, and developed parallel-in time algorithms for long-time simulations. Brown, Stanford, PNNL, and MIT have contributed to this research area. For long-time integration of PDEs PINNs become inefficient and not trainable and to this end we developed parareal neural networks that use a CG supervisor and parallel small PINNs to gain efficiency. In some of the applications of PINNs it is also advantageous to combine Karhunen-Loève expansions to represent correlated fields and add this with PINNs for better accuracy. Similar to PINNs, we introduced the notion of equivariant transformer networks, which consist of typical convolutional neural networks, with the addition of equivariant transformer layers, which are a lightweight and flexible class of functions that improve robustness towards arbitrary predefined groups of continuous transformations. We have designed optimization and learning approaches, which are robust with respect to: (i) noise in the training data due to sensing and communication imperfections or adversarial perturbations; (ii) static and dynamic model misspecifications; (iii) changes in perception modalities leading to missing or corrupted features in test data; (iv) noise injected in the algorithms for privacy or other algorithmic considerations; and (v) strategic interaction against one of multiple other learning agents. Training algorithms that incorporate these types of robustness solve a robust optimization problem, a.k.a. a min-max problem, where we seek to find model parameters that minimize a loss function against worst case perturbations within allowable constraints.

Challenges and Implemented Course Corrections

H. Gao and Zhen Li of Brown and Nathan Hodas of PNNL are no longer in the PhILMs due to new jobs but we have hired other experienced personnel and re-assigned tasks to cover this change.

Contents

PhILMs Progress Report 2018-2019

Most multiscale problems can be broadly classified into three categories: (i) data poor but complete knowledge of the governing physics; (ii) data rich but little knowledge of governing physics; and (iii) only moderate data and moderate knowledge of the governing physics. Existing multiscale and multiphysics approaches have been ineffective, especially for the latter two categories, in addressing key scientific challenges, such as inhomogeneous cascades of scales, propagation of uncertainty/stochasticity across scales, and nonlocal interactions, and are inadequate for solving inverse stochastic multiscale problems, especially in the context of noisy multifidelity data. In the Collaboratory on Mathematics and Physics-Informed Learning Machines for Multiscale and Multiphysics Problems (PhILMs), we address these shortcomings by seamlessly fusing physical laws and multifidelity data, for both forward and inverse multiscale problems, with the goal of revealing, modeling, and understanding the hidden physics of interfaces. Our integrated mathematical and computational activities can be classified into the following research areas (see Figure 1): (RA-I) partial differential equation (PDE)-based modeling of macroscales; (RA-II) stochastic modeling of mesoscales; (RA-III) bridging methods to connect the scales; and (RA-IV) statistical learning and deep learning approximations and algorithms. Physics-informed neural networks (PINNs) are a workhorse tool for the project. In the following, we summarize research accomplishments, highlights, and current directions from the first year of the project.

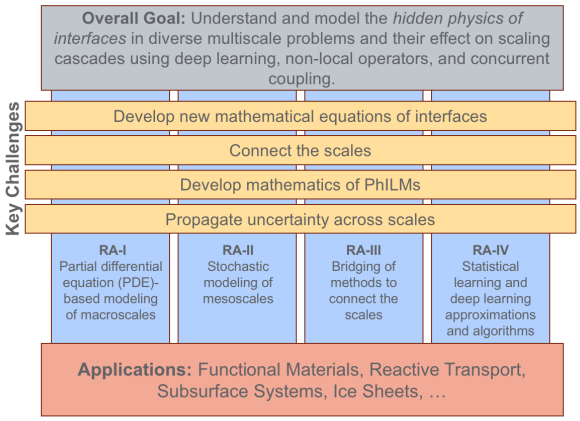


Figure 1: *Overview of the PhILMs proposal:* Relation between the overall goal of PhILMs, the thematic key challenges derived from this goal, and its four research areas motivated by exemplar Department of Energy (DOE) applications.

1 RA-I: PDE-based modeling of macroscales

We developed PINNs to seamlessly enforce PDE conservation laws in training neural networks for applications including data assimilation, inverse modeling, and learning unknown constitutive relationships. To deal with the complexity of grids and nonlocal multiphysics interactions, we developed flexible high-order meshless methods and unified theory of nonlocal operators with corresponding high-order numerical algorithms. Sandia National Laboratories (SNL), Stanford, University of California, Santa Barbara (UCSB) and Pacific Northwest National Laboratory (PNNL) teams contributed to this research area.

1.1 GMLS-Nets: meshfree operator regression

Convolutional networks have proven effective in machine vision, where data are characterized by a Cartesian arrangement of pixels and one may learn an operator in the form of a compactly supported finite difference stencil. Over the past year, we (UCSB (P. Atzberger) and SNL (N. Trask, R. Patel, M. Perego)) teams developed a machine learning (ML) method utilizing Generalized Moving Least Squares (GMLS). We have introduced GMLS-Nets as an extension of convolutional networks to SciML applications [1]. These frameworks are able to handle heterogeneous unstructured data associated with arbitrary differential forms characteristic of mixed finite element discretizations (i.e., face flux data in addition to point sample data). GMLS is a non-parametric functional regression technique to construct approximations of linear, bounded functionals, and was developed within the Collaboratory on Mathematics for Mesoscopic Modeling of Materials (CM4) project. On a Banach space $\mathbb{V}$ with dual space $\mathbb{V}^*$, we estimate target functional $\tau_{\tilde{\mathbf{x}}}[u] \in \mathbb{V}^*$ acting on $u = u(\mathbf{x}) \in \mathbb{V}$, where $\mathbf{x}, \tilde{\mathbf{x}}$ denote locations in compact domain $\Omega \subset \mathbb{R}^d$. We assume u is characterized by an unstructured collection of sampling functionals, $\Lambda(u) := \{\lambda_j(u)\}_{j=1}^N \subset \mathbb{V}^*$. We construct the estimate by considering $\mathbb{P} \subset \mathbb{V}$ and seek an element $p^* \in \mathbb{P}$ that provides an optimal reconstruction of the samples in weighted-ℓ_2 sense

$$p^* = \underset{p \in \mathbb{P}}{\text{argmin}} \sum_{j=1}^{N} \left(\lambda_j(u) - \lambda_j(p)\right)^2 \omega(\lambda_j, \tau_{\tilde{\mathbf{x}}}).$$

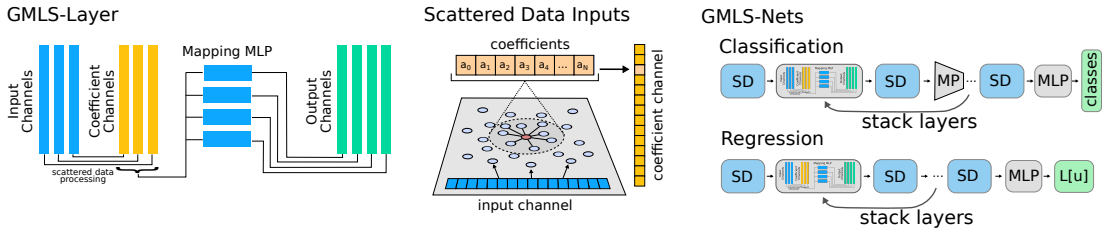


Figure 2: Scattered data inputs are processed by learnable operators $\tau[u]$ parameterized via GMLS estimators. At each location, point data are encoded as coefficient vectors. GMLS-Layers can be stacked to obtain deeper architectures for classification and regression tasks *(inset, SD: scattered data, MP: max-pool, MLP: multilayer perceptron). For more information see [1].*

Operators in $\tau \in V^*$ are then characterized by approximating $\tau(u) \approx \tau(p^*)$; that is one may estimate the action of an operator acting on data by applying the operator to the optimal reconstruction. This framework is supported by a rigorous approximation theory developed at SNL over the last year, and has been used extensively to develop *Compatible Meshfree Discretizations* which mimic the ability of compatible mesh-based discretization (e.g., mimetic finite differences or finite elements) to discretely preserve properties of continuum operators, such as conservation or exact sequence properties [2].

To develop a SciML framework, we assume the operator is unknown. The GMLS process provides an optimal local encoding of the data in terms of $\mathbb{P}$, and we seek operators $\mathcal{L}_\xi : \mathbf{P} \to \mathbb{R}$ to characterize V^*. A simple choice for this mapping is to introduce a multilayer perceptron. In this manner, we obtain stackable layers that may be used in an identical manner to convolutional networks (Figure 2). These architectures have been shown to be an effective platform to develop data-driven PDE models from molecular data and to regress engineering quantities of interest from datasets characteristic of computational fluid dynamics simulations (Figure 3).

1.2 PDE solvers on manifolds: meshless methods

We (UCSB and SNL teams) have developed meshless methods for solving PDEs on manifolds [3]. The approach is based on a collection of local regression problems (least-squares) used both to perform local reconstructions of the manifold geometry and for approximating the action of differential operators, see Figure 4. A central challenge in this work is the interplay between discretization errors arising from the geometric approximations with the errors associated with the operators, here from exterior calculus of differential geometry. Methods based on GMLS were developed to obtain accurate robust approximations. It was shown these techniques could be used for approximating the hydrodynamic flow responses of fluid interfaces taking into account the incompressibility constraints and contributions of the surface geometry.

1.3 Data-driven network models

Network models are an effective modeling and simulation tool for many complex systems for which first-principle descriptions are either non-existent or intractable. Examples include electrical circuits, biological systems [4], brain models [5], porous media, *de novo* design of synthetic gene circuits [6], and subsurface flows [7]. Network models replace detailed physics descriptions of subsystems by "compact models" performing specific functions. For example, the solution of a drift-diffusion PDE to simulate a semiconductor device such as a diode is replaced by an empirical model such as Shockley diode equation. Likewise, Darcy's law is inadequate for some types of porous media flows and can be replaced by a pore-level network model with empirical pore morphology descriptions [7]. However, traditional development of compact models relies on human experts and is an expensive, time-consuming effort. It requires highly skilled experts combining knowledge of physics and numerical analysis. Besides the long development times, compact models do not always generalize well and adding new physics may require starting from scratch.

A data-driven approach could help mitigate these issues by providing the means to automate the development of compact models directly from data. However, unlike traditional ML applications where "big" data is the default setting, in the context of compact models we are often dealing with the availability of only "small" or at best "medium" data sources. The reasons for this data scarcity can range from cost limitations to regulatory restrictions due to environmental hazards. In addition, data may not be available in all regimes of interest because of equipment limitations.

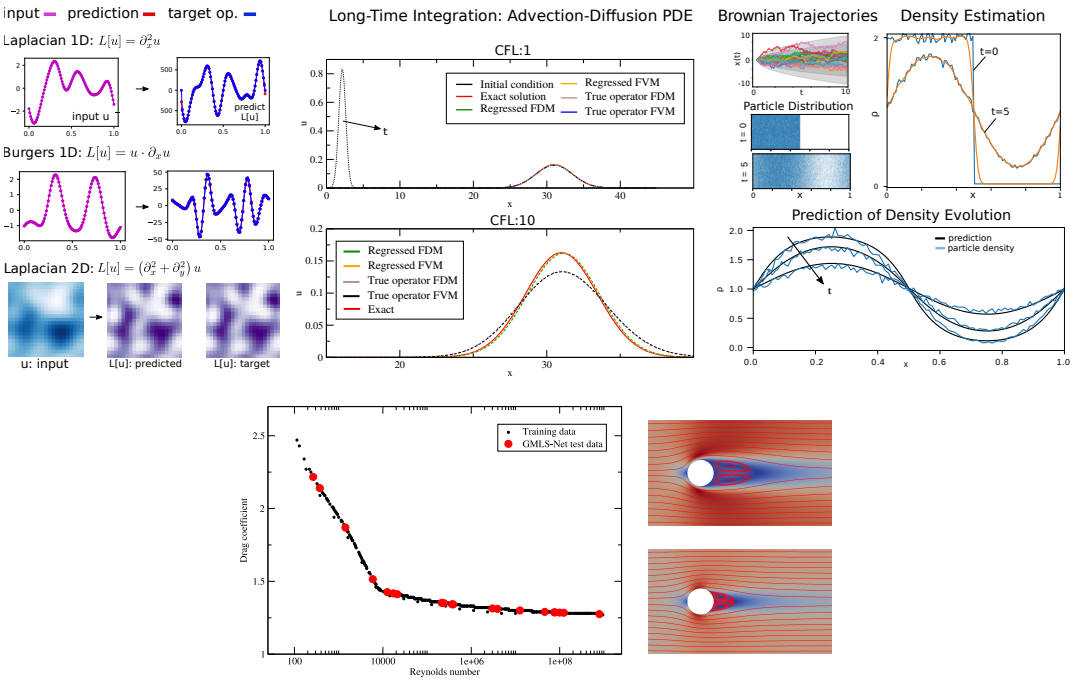


Figure 3: Top left: Learning differential operators using the GMLS-Net framework. **Top center:** From solutions to the exact solution of the advection-diffusion equation, GMLS-Nets are able to learn implicit time integrators that are stable for large timesteps but lack the excess numerical dissipation of implicit Euler integration. **Top right:** A short burst of molecular dynamics (MD) simulation (Brownian motion) over a short time period is used to learn an operator evolving the particle number density over longer time scales (unsteady heat equation). **Bottom:** Cell center velocities from a Reynolds-averaged Navier-Stokes prediction of flow past a cylinder are provided as features to learn the drag over a cylinder. The network is predictive using solely flow features, despite the fact that pressure and viscosity information are not provided.

We seek to address two scientific ML questions: [Q1] If an expensive but accurate physics model is available, can we leverage this model to automate generation of generalizable compact models? and [Q2] If a cheap but inaccurate physics model is available, can we combine this partial knowledge of the physics with the available "small" data to obtain a generalizable compact model?

We explore data-driven approaches ranging from GMLS non-parametric regression to deep neural networks (DNNs). As a driver application, we have selected the development of data-driven compact models for design and simulation of electrical circuits. This application both exposes most of the essential research challenges and is also relevant to multiple aspects of DOE and SNL mission problems.

We have focused our initial efforts on the development of data-driven models for a PN-diode and demonstrating these models using a simple rectifier circuit. Figure 5 illustrates the steps in the model development workflow. This choice is motivated by (i) the availability of an idealized physics model (the Schokley diode equation) that can be used to address research question Q2 and (ii) availability of experimental facilities at SNL that allow us to collect $I - V$ (voltage-current) measurements from actual devices. Using these measurements we have implemented two data-driven diode models. The first one uses the GMLS regression to generate approximations of the $I - V$ curve and its derivative. The second model used a DNN and was intended to evaluate how DNN will operate in the "small" data regime. Figure 5 shows initial tests of the two data-driven devices in the rectifier circuit. Results in the middle plot show that the DNN device using the available 175 measurements does not perform correctly. The zoom in the right plot clearly shows unphysical negative voltage. Upscaling the experiment data to 300 points improves the DNN model to a point where negative voltages disappear as they should in a rectifier circuit. In contrast, we found that the GMLS diode performs well with the original experimental data.

1.4 PINNs for learning unknown physics

Conservation laws governing most natural and engineered systems are, at best, "partially" known as conservation laws in general do not have a closed system of equations. Accurate theoretical models for closing the system

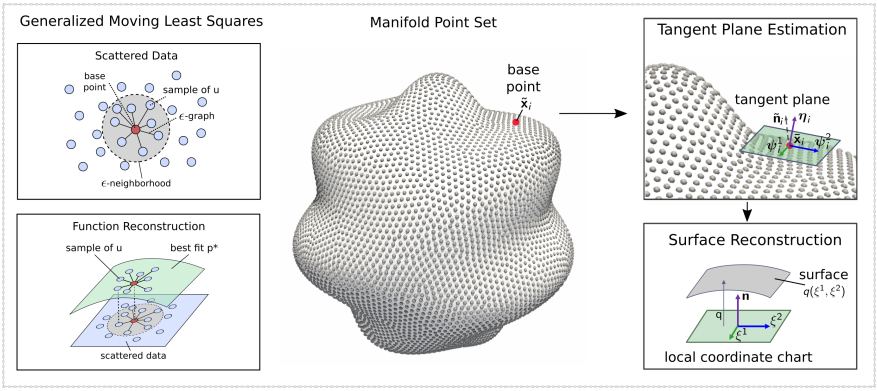


Figure 4: GMLS Approximation of Operators and Surface Reconstructions. **Top left:** A target functional $\tau_{\tilde{\mathbf{x}}}[u]$ is approximated using data within an ϵ-neighborhood around the base point $\tilde{\mathbf{x}}$. **Bottom left:** For values of u the best fitting function $p^* \in \mathbb{V}_h$ is identified using the sampling functionals $\{\lambda_j\}$ for computing $\tau^h_{\tilde{\mathbf{x}}}[u] = \tau_{\tilde{\mathbf{x}}}[p^*]$. **Center:** For geometric reconstructions, a principle component analysis is used to find local parameterization of the surface of form $(\xi^1, \xi^2, s(\xi^1, \xi^2))$ **(top right)**. **Bottom right:** The $s(\xi^1, \xi^2)$ and its derivatives are approximated by GMLS to obtain general geometric quantities of the manifold and approximate operators, see [3].

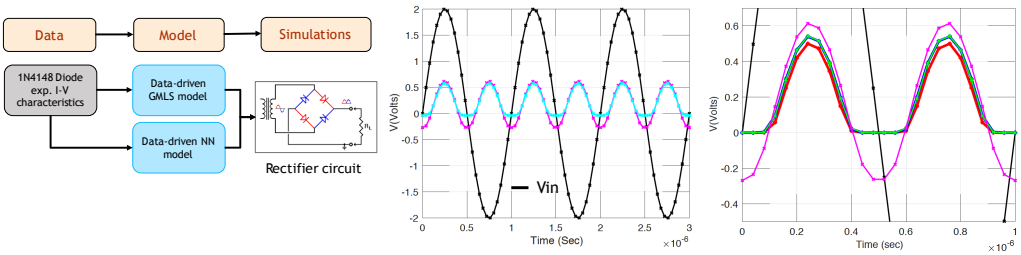


Figure 5: Left: Workflow for the development and testing of a data-driven device model (1N4148 Diode). **Center, Right:** Simulation of a rectifier circuit using data-driven diode models. Black: input AC voltage. Green: GMLS diode. Magenta: DNN model using experimental measurements (175 data points). Light blue: DNN model using experimental data upsampled to 300 points.

of conservation equations are available for homogeneous systems exhibiting time and length scale separation. Examples of accurate closures include Newtonian stress for homogeneous (Newtonian) fluids, Fick's law for mass flux in diffusion processes, and the Darcy law for fluid flux in porous media. For more complex systems, including non-homogeneous turbulence, non-Newtonian fluid flow, multiphase flow and transport in porous media, and granular materials, accurate theoretical closures are not available. Instead, phenomenological constitutive relationships are used, which are usually accurate for a narrow range of conditions.

While there are a number of established methods for parameter estimation, such as Bayesian inference and maximum *a posteriori* probability (MAP) estimation [8, 9], existing approaches for learning unknown physics in partially known models are few and not fully mature. The complicating factor here is that the unknown processes (e.g., stresses, fluxes) are very difficult to measure directly.

To be more specific, here we consider unsaturated flow in porous media described by the PDE $\nabla \cdot (K(u) \nabla u(x)) = 0$, where $K(u)$ is an unknown function of pressure $u(x)$, and assume that some measurements of $u(x)$ and *no measurements* of $K(u)$ are available. We extended PINNs for learning unknown physics by modeling $K(u)$ and $u(x)$ with two DNNs and using a third DNN to represent the governing PDE residual [10]. These DNNs share weights and are trained jointly. As shown in Figure 6, this allows us to accurately learn the DNN approximations of $u(x)$ and $K(u)$ without any measurements of u.

1.5 Hybrid DNN and classical approximation schemes

Traditional numerical analysis for engineering and science problems focuses on the development of discretization schemes, the study of convergence, and stability. In hybrid ML, we combine the best of two worlds: efficient, accurate numerical discretization schemes for PDEs and DNNs to approximate unknown physical relationships.

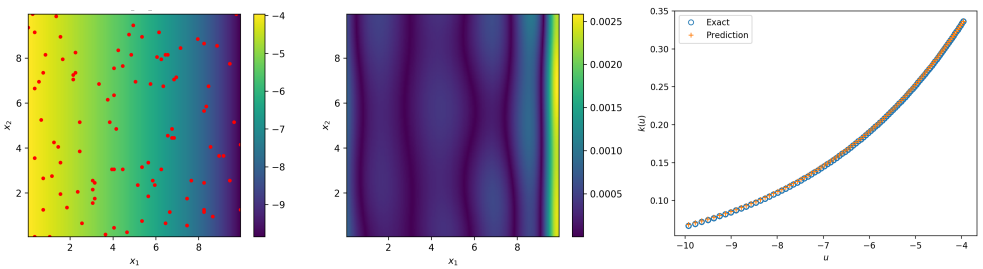


Figure 6: Left: Referenced $u(\mathbf{x})$ field generated using STOMP with the van Genuchten model for $K(u)$ and the locations of u observations. **Center:** Absolute error in the $u(\mathbf{x})$ field estimated with the physics informed DNN. **Right:** The comparison of the learned $K(u)$ and the ground truth $K(u)$ given by the van Genuchten model.

This is done by expressing the numerical physical modeling as a computational graph, where each node in the graph can be a finite element method (FEM) discretization scheme, a time marching step, a linear solver, a Newton–Raphson algorithm, etc. Each edge represents the flow of data.

In a data-driven modeling setting, we assume that some of the parameters (represented as edges in the graph) are unknown and need to be estimated based on observing parts of the solution. Using gradient-based methods and back-propagation algorithms, we estimate these unknown parameters by minimizing the discrepancy between the predictions from our model and the observations. The gradients are computed using automatic differentiation (AD) methods and specialized adjoint state methods (e.g., efficient checkpointing schemes).

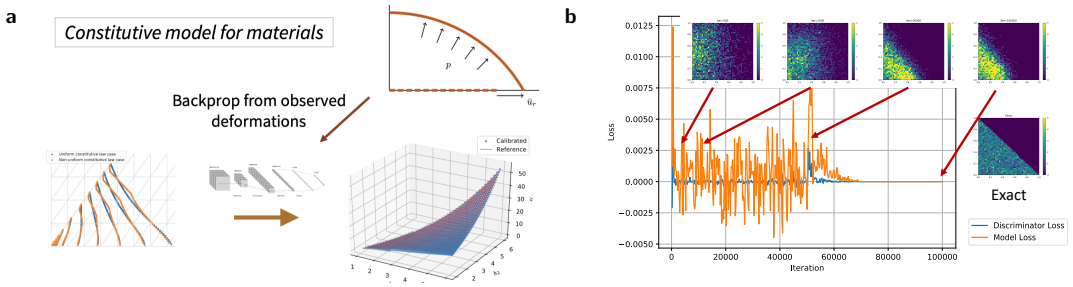


Figure 7: a: DNN for constitutive law. **b:** Statistical inverse modeling with adversarial numerical analysis. The estimated distribution converges to the true distribution.

We have automated adjoint-state methods, AD, and specialized numerical discretization schemes in the library `ADCME.jl`, an open source software capable of coupling general numerical schemes and ML algorithms and data-structures by leveraging powerful metaprogramming features in Julia. Note that this hybrid ML approach is in general different from PINNs, where both states and unknown constitutive relationships (or space-dependent parameters) are represented with neural networks and can leverage advances in numerical methods. On the other hand, PINNs do not require discretizing and solving equations in the entire domain. Our immediate plan is to compare the merits of both approaches.

We demonstrated the efficiency of the hybrid ML method for solid mechanics and geophysics applications. In Figure 7(a), we learn a nonlinear constitutive law from displacement and external applied force observations. We use DNN to approximate the unknown constitutive law (stress/strain relation) and FEM to discretize the momentum conservation equation.

Also, we proposed an adversarial numerical analysis for estimating unknown (non-Gaussian) distributions of parameters in PDE systems. Similarly to generative adversarial networks (GANs), we define two neural networks, generator and discriminator. The generator is used to parametrize the unknown distribution. Then we can leverage numerical schemes to solve the PDE system and compute a (stochastic) solution. The discriminator network is used to measure the discrepancy (or distance) between the predictions and the true observations. Then, the two neural networks are trained adversarially such that at equilibrium, the generator DNN converges to the unknown probability distribution. We applied this method to learn the distribution of an (unknown) space-dependent coefficient in the Poisson equation. Figure 7(b) shows the rapid convergence of the estimated distribution to the

true distribution.

We also studied the difference between DNN and traditional approximation functional forms, such as piecewise linear functions in FEM, radial basis functions, etc. We found that DNNs are superior in cases where highly non-uniform sample points are used (adaptivity), where the function to be approximated exhibits sharp gradients—or even discontinuity—in some regions (concentrated jumps similar to shocks), and in high dimension situations. DNNs are also robust to noise and consistently generalize better in many of our test cases. We created an interactive website for visualizing and comparing different basis functions[1], which can be used to develop intuition for how the different schemes perform in practice.

1.6 PINNs for data assimilation: transport in porous media

Mathematical models of most complex natural and engineered systems contain spatially-heterogeneous parameters, which can be measured only at very few locations. Computational approaches for parameter estimation cast this inherently ill-posed problem as an optimization or a statistical inference task, usually requiring repeated evaluation of expensive forward solvers until the parameters that minimize a given error metric are found or until space of parameter configurations satisfying available data is explored. Although significant progress has been made over the last two decades involving high-order schemes for PDEs, AD of computer code, PDE-constrained optimization, and optimization under uncertainty, Bayesian inference and MAP estimation [8, 9], parameter estimation in large-scale problems remains a significant challenge [11]. We (Brown and PNNL) developed a PINN-based approach for learning unknown parameter fields in PDE models and applied for assimilating measurements of hydraulic conductivity $K(x)$ (parameter, a propriety of the porous medium) and hydraulic head $h(x)$ and tracer concentration $c(x)$ (both are state variables) in the transport in porous media application. We assume that this problem can be described by the PDE model $\nabla \cdot u = 0$, $u(x) = -\gamma k(x) \nabla h(x)$, and $\nabla \cdot (u(x)c(x)) = \nabla \cdot \left[\left(\frac{D_w}{\tau} + \boldsymbol{\alpha} \|u(x)\|_2 \right) \nabla c(x) \right]$ ($u(x)$ is the average pore velocity), subject to known boundary conditions. We model $K(x)$, $h(x)$, and $c(x)$ with feed-forward DNNs and first attempt to train them using measurements only. Figure 8 shows both the reference fields and fields learned with the "data-based" DNNs, and reveals that for the given number of measurements, data-based DNNs are very inaccurate. Next, we apply the PINN framework (proposed by the Brown team in [12, 13]), where in additional to the "primary" K, h, and c DNNs we define to residual DNNs to represent the residuals of the governing PDEs: $R_1(x) = \nabla \cdot (K(x) \nabla h(x))$ and $R_2(x) = \nabla \cdot (\gamma k(x) \nabla h(x) c(x)) + \nabla \cdot \left[\left(\frac{D_w}{\tau} + \boldsymbol{\alpha} \|u(x)\|_2 \right) \nabla c(x) \right]$. The primary and residual DNNs share the same weights and we train them jointly by minimizing the mean square error of primary DNNs with respect to data and mean square of residuals at a selected number of points in the domain. Figure 8 demonstrates PINNs provide a much better estimation of K and c than the "data-based" DNNs.

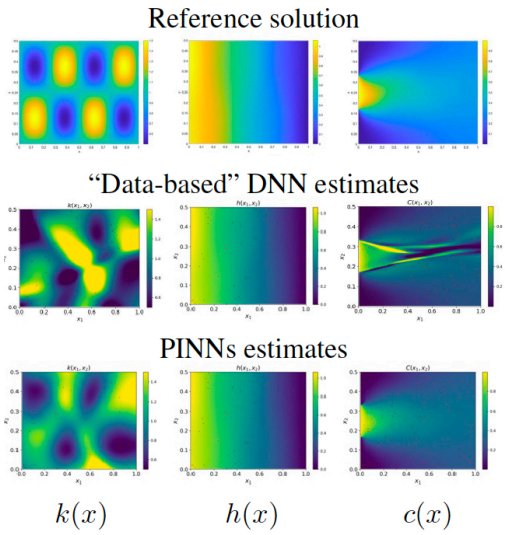


Figure 8: Reference conductivity $K(x)$, hydraulic head $h(x)$, and the tracer concentration field $c(x)$, same fields reconstructed from data only using feed-forward DNNs, and the fields learned with PINNs using the same data and Darcy and advection-dispersion equations.

1.7 nPINNs: nonlocal physics-informed neural networks

Nonlocal equations are model descriptions for which the state of a system at any point depends on the state in a neighborhood of points, i.e., every point in a domain interacts with a neighborhood of points. Nonlocal models provide an improved predictive capability for several scientific and engineering applications including fracture mechanics [14, 15, 16], anomalous subsurface transport [17, 18, 19], phase transitions [20, 21, 22], image processing [23, 24, 25, 26], multiscale and multiphysics systems [27, 28], MHD turbulence [29], and stochastic processes [30, 31, 32, 33].

[1] http://stanford.edu/~kailaix/Hy5d3sYsCFSFPPlhkq2i.html

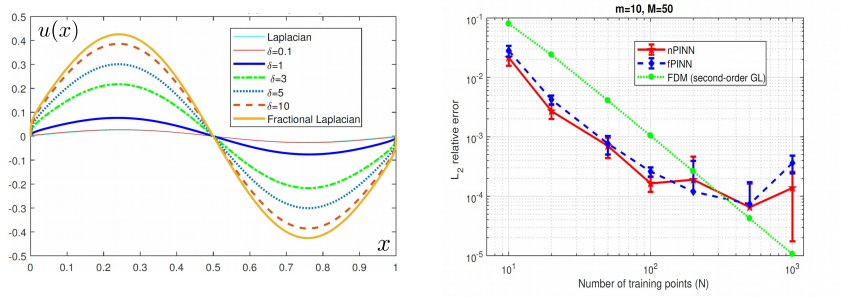


Figure 9: Left: Solutions, for different δ, of $-L_{\delta,0.5}u = \sin(2\pi x)$: as $\delta \to 0$ u approaches the solution of the classical Laplacian whereas as $\delta \to \infty$, u approaches the solution of the fractional Laplacian. **Right:** Convergence results for nPINNs, fractional PINNs (fPINNs), and standard finite difference discretization for fractional problems.

Independent definitions and formulations of nonlocal models exist, namely *unweighted truncated nonlocal*, *weighted truncated nonlocal*, and *fractional* models. Similarities are evident, but they have not been rigorously proven; furthermore, the corresponding communities barely interact with each other, making it impossible to benefit from each other's theoretical and computational findings.

In the first year of PhILMs we have developed a unified theoretical framework that includes all the operators mentioned above as *special instances* of nonlocal truncated weighted operators and derived a unified variational theory for their analysis [34]. This allows us to

- Connect the nonlocal and fractional communities that would benefit from each other's research;
- Include as special cases the well-known classical differential calculus at the limit of vanishing interactions and the fractional calculus at the limit of infinite interactions;
- Provide the groundwork for *new model discovery* thanks to the broad class of operators that it describes;
- Describe intrinsically nonlocal phenomena that have not been analyzed or used due to the lack of theory (see the nonlocal behavior in the cascade of scales in von Kármán flow that lacks proper closure relations).

The unified nonlocal vector calculus provides a universal definition of parametrized nonlocal operators that describe both well-known nonlocal phenomena and may describe new intrinsically nonlocal phenomena not yet analyzed and used due to lack of theory. We propose a new approach to model learning that is in stark contrast with previously developed uncertainty quantification and PDE-constrained-like optimization techniques. By combining *ML* with *physical principles* and *the new unified nonlocal calculus* with *versatile surrogates*, we arrive at a data-driven physics-informed tool for learning new complex nonlocal phenomena. *This is a center-wide effort that involves collaborations between SNL and (1) Brown University (G.E. Karniadakis and G. Pang) on the development of nonlocal PINNs (see Section 1.7); (2) Stanford University (E. Darve and K. Xu) on operator regression for nonlocal kernels; and (3) PNNL (A. Tartakovsky and A. Howard) on PINNs for operator learning in nonlocal surface tension models (see Section 3.2).*

We describe the identification of parameters for the generalized nonlocal operator

$$L_{\delta,s}u(x) = \int_{B_{\boldsymbol{\delta}}(x)} (u(x) - u(y))k(x,y;s)dy = C_{\delta,s}\int_{B_\delta(x)} \frac{u(x)-u(y)}{|x-y|^{n+2s}}\,dy,$$

where $C_{s,\delta}$ is such that the corresponding solutions span a broad range of nonlocal diffusion processes including local and fractional diffusion at the limit of vanishing and increasing nonlocality, see Figure 9(left). Specifically, $\lim_{\delta\to 0} L_{\delta,s}u = \Delta u$ and $\lim_{\delta\to\infty} -L_{\delta,s}u = (-\Delta)^s u$.

The nPINNs algorithm consists of three simple steps:

1 Collect measurements of solution and data in some training sets: $f_m(x_i)$, $x_i \in \mathcal{T}_f$, and $u_m(x_j)$, $x_j \in \mathcal{T}_u$;

2 Approximate the solution with a neural network: $u(x) = u_{NN}(x)$;

3 Minimize the loss function: $\min_{u;\delta,s} \mathcal{L}oss(u;\delta,s) = \frac{1}{2}\sum_{x_i\in\mathcal{T}_f}(L_{\delta,s}u_{NN}(x_i) - f_m(x_i))^2 + \frac{\beta}{2}\sum_{x_i\in\mathcal{T}_u}(u_{NN}(x_j) - u_m(x_j))^2$.

Note that $\mathcal{L}oss$ has both a physics-driven and a data-driven component: the first term controls the residual of the nonlocal equation, whereas the second term controls the mismatch between solution and data. The outcome of the optimization are the weights and biases of the neural network and the model parameters. The proposed strategy is as **accurate** as any other discretization method for the forward problem. As an example, in Figure 9(right) we report the convergence of the solution error with respect to the number of training points for nPINNs, fPINNs, and a standard finite difference discretization; here, nPINNs has the same convergence rate as finite difference (up to some number of training points, after which it reaches stagnation). However, due to the increased computation cost, nPINNs is not recommended for the solution of forward problems. As PINNs, nPINNs, and fPINNs are not tied to any discretization method, they easily handle *sparsity* and require *minimal implementation effort*, i.e., available solvers can be used as black boxes.

2 RA-II: Stochastic modeling of mesoscales

We developed DNN-based algorithms to capture statistical spatial correlations that cannot be readily incorporated into DNNs, and to bridge atomistic and mesoscale systems. The Brown and PNNL teams have contributed to this research area.

2.1 Concurrent coupling of atomistic simulation and mesoscopic hydrodynamics for flows over soft multi-functional surfaces

We developed an efficient parallel multiscale method that bridges the atomistic and mesoscale regimes, from nanometers to microns and beyond, via concurrent coupling of atomistic and mesoscopic models. In particular, we combined an all-atom MD description for specific atomistic details in the vicinity of the functional surface with a dissipative particle dynamics (DPD) approach that captures mesoscopic hydrodynamics in the domain away from the functional surface, as shown in Figure 10. In order to achieve a seamless transition in dynamic properties we

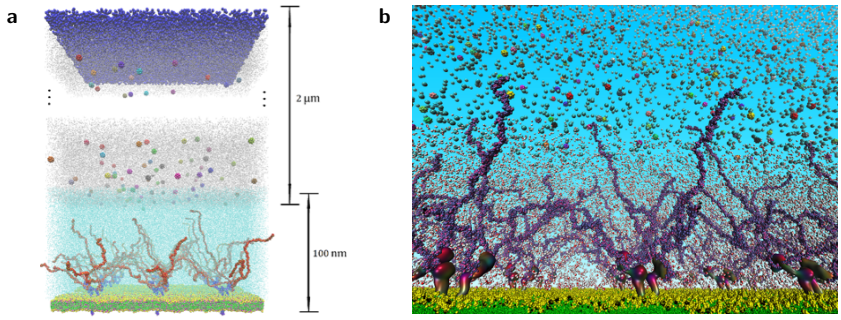


Figure 10: a: Sketch of a setup for simulating the transport of drug delivering nanoparticles to the glycocalyx. The size of the DPD domain is typically 1–2 μm to represent the cell-free layer in arterioles. The nanoparticles are transported in the DPD solvent (plasma) and as they cross into the MD domain, they are endowed with molecular functionality. **b:** Visualization of the MD-DPD coupling method for flows over soft multi-functional surfaces, from nanometer to micron and beyond, with atomic accuracy and ultra-high computational efficiency.

endowed the MD simulation with a DPD thermostat, which is validated against experimental results by modeling water at different temperatures. We then validated the MD-DPD coupling method for transient Couette and Poiseuille flows, demonstrating that the concurrent MD-DPD coupling can accurately resolve the continuum-based analytical solutions. Subsequently, we simulated shear flows over grafted polydimethylsiloxane (PDMS) surfaces (polymer brushes) for various grafting densities, and investigated the slip flow as a function of the shear stress.

2.2 Self-cleaning of hydrophobic rough surfaces by coalescence-induced wetting transition

The superhydrophobic leaves of a lotus plant and other natural surfaces with self-cleaning functions have been studied intensively for the development of artificial biomimetic surfaces. The surface roughness generated by hierarchical structures is a crucial property required for superhydrophobicity and self-cleaning. Here, we demonstrate a novel self-cleaning mechanism of textured surfaces attributed to a spontaneous coalescence-induced wetting transition. In particular, we perform many-body DPD simulations of liquid droplets (with a diameter of 89 μm) sitting on mechanically textured substrates. We quantitatively investigated the wetting behavior of an isolated droplet as well as coalescence of droplets for both Cassie–Baxter and Wenzel states. Our simulation results

reveal that droplets in the Cassie–Baxter state have much lower contact angle hysteresis and smaller hydrodynamic resistance than droplets in the Wenzel state. When small neighboring droplets coalesce into bigger ones on textured hydrophobic substrates, we observe a spontaneous wetting transition from the Wenzel state to the Cassie–Baxter state, which is powered by the surface energy released upon coalescence of the droplets, as shown in Figure 11(left).

For superhydrophobic surfaces, the released surface energy may be sufficient to cause a jumping motion of droplets off the surface, in which case adding one more droplet to the coalescence may increase the jumping velocity by one order of magnitude. When multiple droplets are involved, we found that the spatial distribution of liquid components in the coalesced droplet can be controlled by properly designing the overall arrangement of droplets and the distance between them, as shown in Figure 11(right). These findings offer new insights for designing effective biomimetic self-cleaning surfaces by enhancing spontaneous Wenzel-to-Cassie wetting transitions, and for developing new noncontact methods to manipulate liquids inside the small droplets via multiple-droplet coalescence.

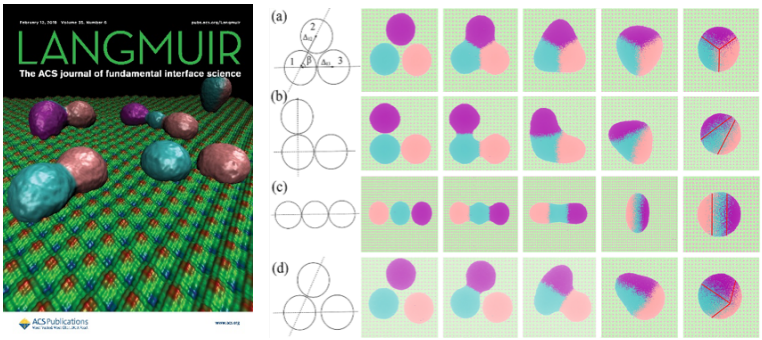


Figure 11: Left: Visualization of spontaneous coalescence-induced wetting transition and jumping motion of droplets on textured hydrophobic surfaces. **Right:** Snapshots of the initial droplet distribution configurations (top view), in which the top three cases (**a**)-(**c**) are in concentrated configurations ($\Delta_{12} = \Delta_{13}$) with $\beta = 60°, 90°$, and $180°$, whereas the bottom case (**d**) is in spaced configuration ($\Delta_{12} \neq \Delta_{13}$). The red lines in the last column show the spatial distribution of liquid components in the coalesced droplet for different configurations.

2.3 Application of learning coarse-grained potential to the liquid/liquid interface

Many physical, chemical, and biological processes occur in presence of hydrophobic-hydrophilic interfaces. However, the microscopic structures of the interface are still not revealed in full detail. Coarse-grained (CG) simulation can provide detailed information about the interface as the spatial distribution and interaction energy at mesoscale, which complements experimental approaches. In CG models, the results greatly depend on the accuracy of the CG force field (FF). In [35], we developed a probabilistic ML approach to estimate the CG FF parameters for water-hexane mixture, a typical immiscible binary liquid/liquid mixture. The first step in our approach is to build a polynomial-regression-based response surface relating the CG FF parameters to target properties. The standard approach is to estimate parameters by minimizing the loss function given by the square difference between the surrogate model and the target properties (in our case, the surface tension of flat and curved interfaces) observed in the experiments and/or first-principle calculations. Figure 12(a) demonstrates that this problem does not have a unique solution. To address this issue, we randomly perturb the training set and compute the probability density function of parameters minimizing the loss function. Figure 12(b) demonstrates that, unlike the loss function, the probability density function has a single peak corresponding to the unique set of optimal parameters. Figure 13 depicts that the obtained CG FF is significantly more accurate for predicting the structure (intrinsic density) at the water and hexane-interface than the commonly used MARTINI CG FF.

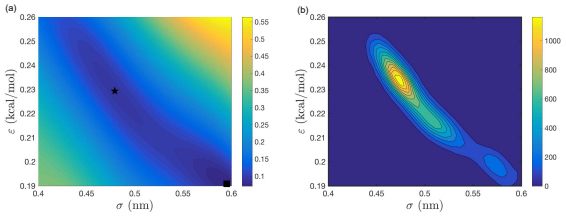


Figure 12: a: Energy surface in the parameter space for the CG potential between water and hexane. **b:** PDF function of the optimal parameter set with 4% random Gaussian noise on the training set.

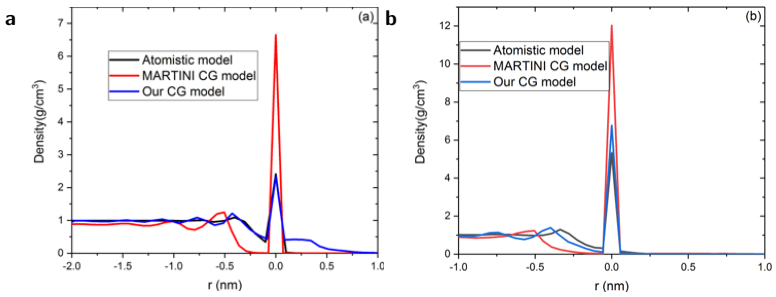


Figure 13: Intrinsic density of planar (**a**) and curved (**b**) water-hexane interface obtained from the atomistic model, the MARTINI CG model, and our CG model.

2.4 Enforcing constraints for time series prediction and continual learning

Complex stochastic mesoscale systems usually have memory terms, which could be derived using model reduction formalism [37]. We have proposed several methods for modeling dynamic systems given time series measurements by learning the flow map with NNs. We have developed ways to enforce dynamic constraints during neural network training for all three major modes of learning, namely supervised, unsupervised, and reinforcement learning. These constraints include terms analogous to memory terms and control terms in control theory [38]. Such terms act as a restoring force which corrects the errors committed by the learned flow map during prediction.

For supervised learning, the constraints are added to the objective function [36]. For the case of unsupervised learning, in particular GANs, the constraints are introduced by augmenting the input of the discriminator [39]. Finally, for the case of reinforcement learning and in particular actor-critic methods, the constraints are added to the reward function. In addition, for the reinforcement learning case, we have developed a novel approach based on homotopy of the action-value function in order to stabilize and accelerate training (see Figure 14) [36].

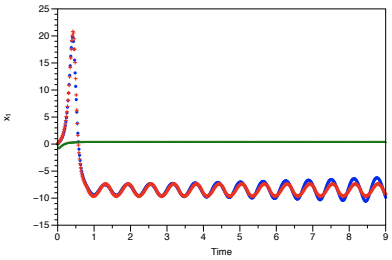


Figure 14: Reinforcement learning for the Lorenz system [36]. Comparison of ground truth for $x_1(t)$ (blue dots), the neural network flow map prediction *with* homotopy for the action-value function during training (red crosses) and the neural network flow map prediction *without* homotopy for the action-value function during training (green triangles).

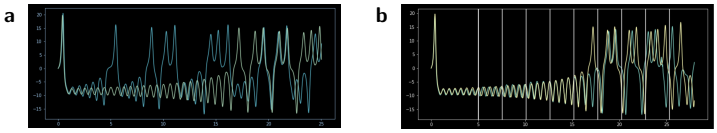


Figure 15: Supervised learning for the Lorenz system. Comparison of ground truth for $x_1(t)$ (yellow line) with trained neural network predictions (green line). **a:** One-time learning; **b:** Continual learning.

Most real-world dynamical systems are strongly chaotic. This makes training NNs to accurately represent a flow map for long time intervals a non-trivial task even if we have data for the whole interval (one-time learning). At the same time, we want to utilize new information that may be available e.g., online measurements. We have constructed "online" training approaches, which can use information from new measurements to keep training the parameters of the network (continual learning). This can allow the trained neural network to adapt to a temporally changing environment (see Figure 15) [40].

3 RA-III: Bridging of methods to connect the scales

We have developed methods for the concurrent coupling of heterogeneous domains, including coupling across scales, to discover the hidden physics models at the interfaces and provide the "glue" functions/functionals to seamlessly connect the cascade of scales. Brown, PNNL, and SNL have contributed to this research area.

3.1 A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems

Currently the training of NNs relies on data of comparable accuracy but in real applications only a very small set of high-fidelity data could be available while inexpensive lower fidelity data may be plentiful. We proposed a new composite NN, which can be trained based on multi-fidelity data. It is comprised of three NNs, with the first neural network trained using the low-fidelity data and coupled to two high-fidelity NNs, one with activation functions and another without, in order to discover and exploit nonlinear and linear correlations, respectively, between the low-fidelity and high-fidelity data. We first demonstrated the accuracy of the new multi-fidelity neural network for approximating some standard benchmark functions but also a 20-dimensional function that is not easy to approximate with other methods, e.g., Gaussian process regression. Subsequently, we extended the recently developed PINNs to be trained with multi-fidelity data sets (MPINNs), as shown in Figure 16.

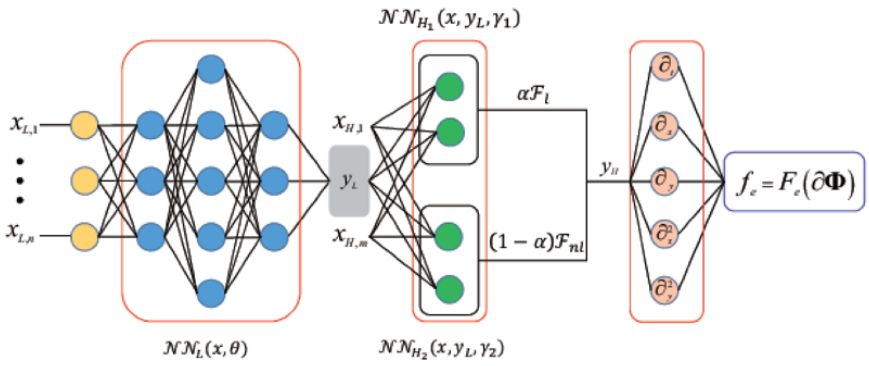


Figure 16: Schematic of the multi-fidelity DNN and MPINN. The left box (blue nodes) represents the low-fidelity DNN $\mathcal{NN}_L(x, \theta)$ connected to the box with green dots representing two high fidelity DNNs, $\mathcal{NN}_{H_i}(x, y_L, \gamma_i)(i = 1, 2)$. In the case of MPINN, the combined output of the two high-fidelity DNNs is input to an additional PDE-induced DNN.

MPINNs contain four fully connected NNs, where the first one approximates the low-fidelity data, while the second and third construct the correlation between the low- and high-fidelity data and produce the multi-fidelity approximation, which is then used in the last neural network that encodes the PDEs. Specifically, in the two high-fidelity NNs a relaxation parameter is introduced, which can be optimized to combine the linear and nonlinear subnetworks. By optimizing this parameter, the present model is capable of learning both the linear and complex nonlinear correlations between the low- and high-fidelity data adaptively. By training the MPINNs, we can: (1) obtain the correlation between the low- and high-fidelity data; (2) infer the quantities of interest based on a few scattered data; and (3) identify the unknown parameters in the PDEs. In particular, we employ the MPINNs to learn the hydraulic conductivity field for unsaturated flows as well as the reactive models for reactive transport. The results demonstrate that MPINNs can achieve relatively high accuracy based on a very small set of high-fidelity data. Despite the relatively low dimension and limited number of fidelities (two fidelity levels) for the benchmark problems in the present study, the proposed model can be readily extended to very high-dimensional regression and classification problems involving multi-fidelity data.

3.2 Learning the kernel function for the nonlocal surface tension model

The surface tension in multiphase fluid flows is typically modeled using the Young-Laplace law; however, the Young-Laplace law has been shown to break down for very small bubbles with radius less than 10nm [41]. This motivates the need for a multiscale surface tension model proposed in [42]. The surface tension is represented as the integral of a molecular-force-like function, with a length scale that corresponds to the relative size of the droplet. As shown in [43, 44], the nonlocal surface tension model has non-physical results for very small droplets for certain force shape functions. In this project, we demonstrate how PINNs work for nonlocal problems, and how PINNs can be used to learn physical parameters for use in physical simulations across multiple length scales. This research is done by the PNNL, SNL, and Brown teams in close collaboration with the nPINN task 1.7.

To use a PINN to learn the nonlocal force shape function from MD simulations, we first formulate the integral in terms of matrix operations. One key feature is that the force shape function can be shown to be radially

symmetric, and we enforce this symmetry. Using this, we can train a PINN to learn the force shape function that minimizes the error in the resulting pressure compared to MD simulations. This new shape function can then be used to inform future continuum-scale simulations with the nonlocal surface tension model.

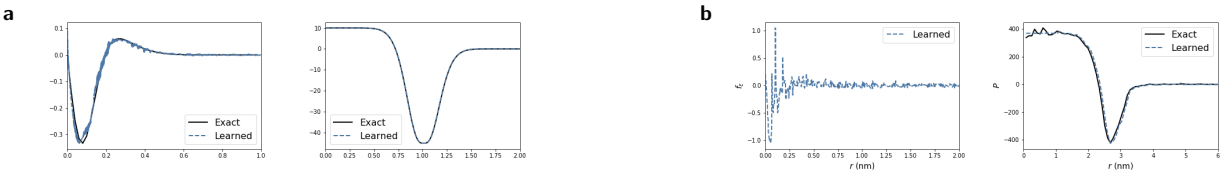


Figure 17: Force shape function and pressure profile for (**a**) an analytical form of the force shape function f_ϵ and (**b**) the pressure profile from an MD simulation for a 3nm droplet.

3.3 PINN with domain decomposition for estimating multiscale parameters

In many porous materials (including geological porous media), parameters (and their statistics) vary in space by orders of magnitude. Figure 18 displays a porous medium, where the mean of $Y(x) = \log K(x)$, the log hydraulic conductivity, in the left half of the domain is 10 times larger than in the right half of the domain. We assume that N measurements of $K(x)$ and $h(x)$ (the hydraulic head) are available. As evident from Figure 18, the standard PINN method (described in Section 1.6) results in relatively large errors in estimated Y and u and converge very slowly with respect to N, i.e., the errors decrease very slowly with increasing number of measurements N. The reason is that DNNs are not very efficient at representing fields that vary at different scales.

To address this challenge, we proposed a domain-decomposition-based PINN (DD-PINN), where the spacial domain Ω is decomposed in two non-overlapping subdomains Ω_i such that the mean of $K(x)$ is (near) constant within each subdomain. In Figure 18, we decompose Ω with two subdomains (left and right halves of Ω) and replace the governing equation $\nabla \cdot (K(x)\nabla u(x)) = 0$ $(x \in \Omega = \Omega_1 \cup \Omega_2)$ with $\nabla \cdot (K_i(\mathbf{x})\nabla u_i(\mathbf{x})) = 0$, $(\mathbf{x} \in \tilde{\Omega}_i$, $i = 1, 2)$, subject to the boundary conditions on $\Gamma = \Omega_a \cap \Omega_2$: $u_1(\mathbf{x}) = u_2(\mathbf{x})$ and $K_1(\mathbf{x})\nabla u_1(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = K_2(\mathbf{x})\nabla u_1(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x})$ $(\mathbf{x} \in \Gamma)$. Then, we define DNNs for $K_1(\mathbf{x})$, $u_1(\mathbf{x})$, $K_2(\mathbf{x})$, and $u_2(\mathbf{x})$ and the residual DNNs corresponding to the governing equations and boundary conditions and train them jointly. Figure 18 shows that the error in DD-PINN is much smaller and the convergence rate with respect to N is much faster than in the "single-domain" PINN, where a single DNN is used to represent the entire $K(x)$ field.

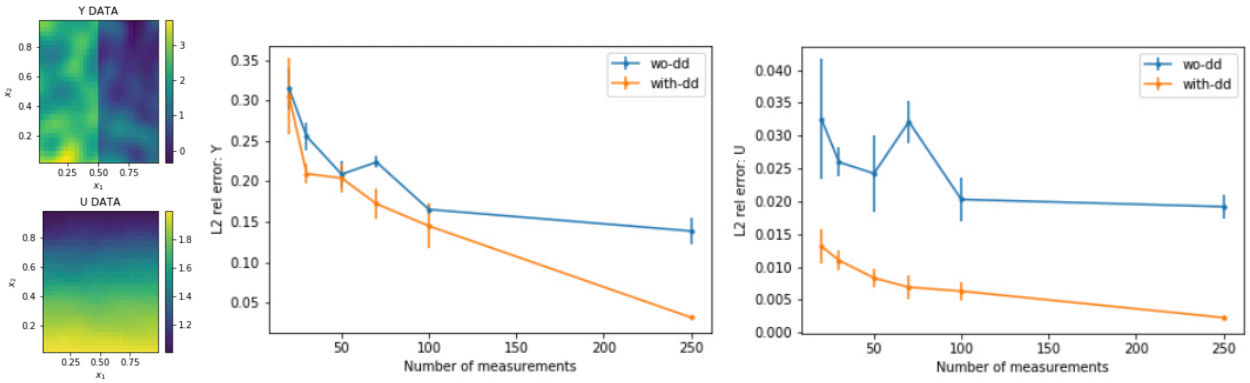


Figure 18: Left: reference $Y(x) = \log K(x)$ and $h(x)$ fields. **Center:** comparison of L_2 errors (and associated uncertainty) in Y (center) and u (right) predictions with PINN and DD-PINN methods.

4 RA-IV: Statistical learning and deep learning approximations and algorithms

We have explored foundational algorithms and mathematics for deep learning. We have combined deep learning with statistical models integrating domain knowledge about spatial correlations, physical invariances and developed parallel-in time algorithms for long-time simulations. Brown, Stanford, PNNL, and MIT have contributed to this research area.

4.1 Physics-informed learning with hybrid Karhunen-Loève DNN approach

In the PINN approach presented in Sections 1.6 and 3.3, we represent both the system's parameters and states with DNNs. Many spatial parameters exhibit spatial correlation that cannot be readily incorporated in DNNs. On the other hand, Karhunen-Loève (KL) representation of partially observed spatial fields explicitly incorporates the correlation structure and can be exactly conditioned on the fields measurements. When the fields statistics are available (e.g., can be accurately determined from measurements) the conditional KL expansion can provide a more accurate representation of spatial fields as compared to DNN, and with significantly less weights. We have developed a hybrid physics-informed KL-DNN approach for parameter and state estimation in PDE models. Consider the $\mathcal{L}(x, \kappa(x), u(x)) = 0$ PDE model with appropriate boundary conditions, where $\kappa(x)$ is the unknown space-dependent parameter to be estimated, $u(x)$ is the state/solution to the model, $\hat{\kappa}_i$ is the parameter observation at $x = \hat{x}_i$ for $i = 1, \ldots, N$ and $\tilde{u}_i$ is the state measurement at $x = \tilde{x}_i$ for $i = 1, \ldots, M$. We use the conditional KL expansion $\kappa(x, \boldsymbol{\xi})$) to represent the unknown parameter $\kappa(x)$ given its known mean and covariance functions and N measurements. By construction, $\kappa(x, \boldsymbol{\xi})$ exactly satisfies the given statistics and measurements [45]. As in PINNs, we use a DNN to model the state $u(x)$, $u(x) \approx u_{NN}(x, \boldsymbol{\theta})$. Then, we jointly train $\kappa(x, \boldsymbol{\xi})$ and $u_{NN}(x, \boldsymbol{\theta})$ by solving the optimization problem: $\arg\min_{\boldsymbol{\xi}, \boldsymbol{\theta}} \sum_{i=1}^{M} (|u_i - u_{NN}(x_i, \boldsymbol{\theta})|^2) + \delta \|\mathcal{L}(x, \kappa(x, \boldsymbol{\xi}), u_{NN}(x, \boldsymbol{\theta}))\|_{L_2}^2$. We applied this approach to one-dimensional Darcy equation describing flow in porous media with (unknown) conductivity $\kappa(x)$, $\mathcal{L} = \frac{\partial}{\partial x} \kappa(x) \frac{\partial u(x)}{\partial x}$, subject to Dirichlet boundary conditions. Figure 19 shows the comparison of $u(x, \boldsymbol{\theta})$ and $\kappa(x, \boldsymbol{\xi})$ with their reference values given 50 measurements of u and 50 measurements of κ. The relative mean square error (RMSE) of u is 1.94% and the RMSE of the conductivity κ is 2.87%.

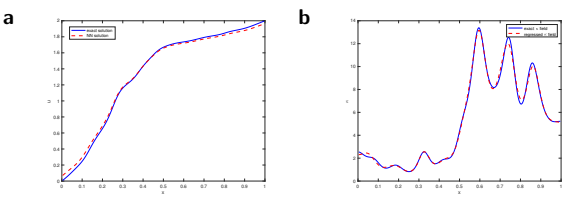


Figure 19: Estimated hydraulic head $u(x)$ (**a**) and conductivity field κ (**b**) using the hybrid KL-DNN approach and the corresponding reference (ground truth) fields.

4.2 Parallel-in-time algorithms

Parallel physics-informed neural networks (PPINN). PINNs encode physical conservation laws and prior physical knowledge into the NNs, ensuring the correct physics is represented accurately while alleviating the need for supervised learning to a great degree. While effective for relatively short-term time integration, when long time integration of the time-dependent PDEs is sought, the time-space domain may become arbitrarily large and hence training of the neural network may become prohibitively expensive. To this end, we develop a PPINN, hence decomposing a long-time problem into many independent short-time problems supervised by an inexpensive/fast CG solver. In particular, the serial CG solver is designed to provide approximate predictions of the solution at discrete times, while initiate many fine PINNs simultaneously to correct the solution iteratively. There is a two-fold benefit from training PINNs with small-data sets rather than working on a large-data set directly, i.e., training of individual PINNs with small-data is much faster, while training the fine PINNs can be readily parallelized. Consequently, compared to the original PINN approach, the proposed PPINN approach may achieve a significant speedup for long-time integration of PDEs, assuming that the CG solver is fast and can provide reasonable predictions of the solution, hence aiding the PPINN solution to converge in just a few iterations. To investigate the PPINN performance on solving time-dependent PDEs, we first apply the PPINN to solve the Burgers equation, and subsequently we apply the PPINN to solve a two-dimensional nonlinear diffusion-reaction equation. Our results demonstrate that PPINNs converge in a couple of iterations with significant speedups proportional to the number of time-subdomains employed.

Supervised parallel-in-time algorithm for long-time Lagrangian simulations of stochastic dynamics: Application to hydrodynamics. Lagrangian particle methods based on detailed atomic and molecular models are powerful computational tools for studying the dynamics of microscale and nanoscale systems. However, the maximum time step is limited by the smallest oscillation period of the fastest atomic motion, rendering long-time simulations very expensive. To resolve this bottleneck, we propose a supervised parallel-in-time algorithm for stochastic dynamics (SPASD) to accelerate long-time Lagrangian particle simulations. Our method is inspired by bottom-up CG projections that yield mean-field hydrodynamic behavior in the continuum limit. As an example, we use the DPD as the Lagrangian particle simulator that is supervised by its macroscopic counterpart, i.e., the Navier-Stokes simulator. The low-dimensional macroscopic system (here, the Navier-Stokes solver) serves as a predictor to supervise the high-dimensional Lagrangian simulator, in a predictor-corrector type algorithm. The

results of the Lagrangian simulation then correct the mean-field prediction and provide the proper microscopic details (e.g., consistent fluctuations and correlations).

The unique feature that sets SPASD apart from other multiscale methods is the use of a low-fidelity macroscopic model as a predictor. The macro-model can be approximate and even inconsistent with the microscale description, but SPASD anticipates the deviation and corrects it internally to recover the true dynamics. We first present the algorithm and analyze its theoretical speedup, and subsequently present the accuracy and convergence of the algorithm for the time-dependent plane Poiseuille flow, demonstrating that SPASD converges exponentially fast over iterations, irrespective of the accuracy of the predictor. Moreover, the fluctuating characteristics of the stochastic dynamics are identical to the unsupervised (serial-in-time) DPD simulation. We also simulate a two-dimensional cavity flow that requires more iterations to converge compared to the Poiseuille flow, and we observe that SPASD converges to the correct solution. Although a DPD solver is used to demonstrate the results, SPASD is a general framework and can be readily applied to other Lagrangian approaches including MD and Langevin dynamics.

4.3 Enforcing invariances in deep architectures

In many domains, including the evolution of physical systems and computer vision, we are equipped with prior knowledge regarding certain transformation invariances. For example, the dynamics of certain physical systems might be invariant to translation and rotation of the domain, or possess other symmetries and structures. Here, we are developing methods for enforcing invariances in DNNs, most specifically, in the convolutional neural network (CNN).

We introduced the notion of *equivariant transformer networks*, which consist of typical CNN networks, with the addition of *equivariant transformer* (ET) layers, which are a lightweight and flexible class of functions that improve robustness towards arbitrary predefined groups of continuous transformations. An ET layer for a transformation group G (such as rotations, scaling, or perspective transformations) is an image-to-image mapping that satisfies the following local invariance property: for any input image ϕ and transformation $T \in G$, the images ϕ and $T\phi$ are both mapped to the same output image. Crucially, the invariance property of ETs holds by construction, without any dependence on additional heuristics during training.

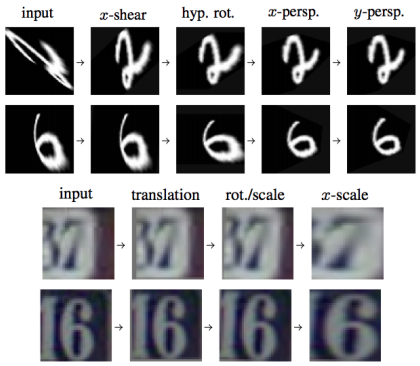


Figure 20: Example illustrating the effects of ET layers. **Top left**: On Projective MNIST, ETs reverse effect of distortions such as shear and perspective, despite being provided no direct supervision on pose parameters (final images remain rotated and scaled since classification CNN operates over their log-polar representation). **Bottom left:** On Street View House Numbers dataset, the final x-scale transformation has cropping effect that removes distractor digits.

We evaluate the performance of ETs with both synthetic and real-world image classification using the Street View House Numbers dataset, see Figure 20. We demonstrated that ET layers reduce the error of image classifiers by up to 15% relative to standard *Spatial Transformer* layers [46]. Moreover, we show that a ResNet-10 classifier augmented with ET layers is able to exceed the accuracy achieved by a larger, more complicated ResNet-34 classifier without ETs, thus reducing both memory usage and computational cost.

Our next step is to extend this approach to training DNNs for physical systems. One shortcoming of ET networks is that they require the transformation invariances to be known *a priori*. In addition to enforcing known invariances, we will explore methods for *learning* unknown invariances over the course of training, and incorporating the appropriate ET layers for these learned transformations.

4.4 Min-max optimization

We have designed optimization and learning approaches, which are robust with respect to: (i) noise in the training data due to sensing and communication imperfections or adversarial perturbations; (ii) static and dynamic model misspecifications; (iii) changes in perception modalities leading to missing or corrupted features in test data; (iv) noise injected in the algorithms for privacy or other algorithmic considerations; and (v) strategic interaction against one of multiple other learning agents. Training algorithms that incorporate these types of robustness

solve a robust optimization problem, a.k.a. a *min-max problem*, where we seek to find model parameters that minimize a loss function against worst case perturbations within allowable constraints. As an example, GANs are trained by means of defining a zero-sum game between two feed-forward NNs, the generator and the discriminator, parametrized by vectors of parameters, w and θ, respectively. The discriminator takes as input samples from a target distribution and samples generated by pushing random seeds through the generator. Its goal is to tune its weights θ so that it distinguishes between the two distributions as best as possible. The goal of the generator is to choose its weights w to fool the discriminator. For some function $f(w, \theta)$ that expresses how well the discriminator distinguishes between the target distribution and that output by the generator, the two NNs are competing, striving to solve a min-max problem of the following form: $\min_w \max_\theta f(w, \theta)$. We are developing (no-regret) variants of online gradient descent that result in fast, last iterate convergence to the min-max solution for arbitrary convex-concave objectives f, and when w and θ are constrained to lie in convex sets, as well as develop practical first-order methods that are guaranteed to converge for non convex-concave objectives. In [47], we have demonstrated this approach for f linear in both w and θ and when w and θ are constrained to lie in convex polytopes. We have also studied the limit points of first-order methods in min-max problems, understand the basins of attraction of locally min-max solution, and design algorithms (with higher order information or properly introduced randomness) converging to points with better optimality properties. In [48], we took a dynamical systems approach to study the limit points of gradient descent-ascent and optimistic gradient descent-ascent in non convex-concave settings. We showed that unstable limit points have measure-zero basins of attraction, that there may be stable limit points of these dynamics that are not locally min-max, that the stable limit points of gradient descent-ascent form a subset of those of optimistic gradient descent-ascent, and that this inclusion can be strict. Going forward, we aim to further our understanding of the basins of attraction of locally min-max solutions, and develop methods that make these basins broader.

5 Software dissemination

Although early in the project, we have three open-source software releases thus far (see Section A.3): (1) the library `ADCME.jl`, which has automated AD, adjoint-state methods, and specialized numerical discretization schemes and leverages powerful metaprogramming features in Julia, (2) GMLS-Nets, which allows for general meshfree operator regression, and (3) `DeepXDE`, a ML tool to solve PDEs.

6 Integration and outreach

We have a regular webinar presenting forum for external speakers (listed on the PhILMs website) and internal discussions about strategic research directions. We have been working with and training postdocs and Ph.D. students, who are fully engaged in all PhILMs activities, including PhILMs webinars, and relevant conferences and workshops. We emphasize person-to-person interaction, so several students from Stanford and UCSB have visited PNNL and SNL for extended periods to work with our co-PIs. All academic PIs have visited both PNNL and SNL, as well as each other's universities, for seminars or PhILMs project meetings. National lab PIs have been active in engaging with ML-related projects and application owners to locate additional opportunities for impact. Publications and presentations have served as a primary means of outreach for year one (see Appendix A). We plan to have a substantial presence at select future conferences, including the July 2020 Mathematical and Scientific ML (MSML) Conference at Princeton.

M. D'Elia, M. Perego, P. Stinis, A. Tartakovsky, M. Parks, and N. Trask visited Brown University over the course of the year. We held a one-day project workshop in SNL. In January, 2019 G. Karniadakis (PI) organized a three-day workshop on Scientific ML at ICERM in Providence, Rhode Island. The talks were live streamed and recorded for viewing. We are in the process of proposing a semester-long program on Scientific ML at ICERM and a minisymposium centered around PhILMs on physics informed machine learning for the upcoming SIAM Mathematics of Data Science meeting in May 2020.

References

1. Trask, N., Patel, R. G., Gross, B. J. & Atzberger, P. J. GMLS-Nets: A framework for learning from unstructured data. *arXiv preprint arXiv: 1909.05371* (2019).
2. Trask, N., Bochev, P. & Perego, M. A conservative, consistent, and scalable meshfree mimetic method. *arXiv preprint arXiv: 1903.04621* (2019).
3. Gross, B. J., Trask, N., Kuberry, P. & Atzberger, P. J. Meshfree Methods on Manifolds for Hydrodynamic Flows on Curved Surfaces: A Generalized Moving Least-Squares (GMLS) Approach. *arXiv preprint arXiv: 1905.10469* (2019).
4. Oishi, M. & May, E. *Addressing biological circuit simulation accuracy: Reachability for parameter identification and initial conditions* SAND2007-6817C (Sandia National Laboratories, 2007).
5. Schiek, R. L. *et al. Simulating neural systems with Xyce* SAND2012-10628 (Sandia National Laboratories, 2012).
6. Marchisio, M. A. & Stelling, J. Automatic Design of Digital Synthetic Gene Circuits. *PLoS Comput Biol 7(2) e1001083* (2011).
7. Thauvin, F. & Mohanty, K. K. Network Modeling of Non-Darcy Flow Through Porous Media. *Transport in Porous media* **31,** 19–37 (1998).
8. Stuart, A. M. Inverse problems: a Bayesian perspective. *Acta Numerica* **19,** 451–559 (2010).
9. Carrera, J., Alcolea, A., Medina, A., Hidalgo, J. & Slooten, L. J. Inverse problem in hydrogeology. *Hydrogeology Journal* **13,** 206–222 (2005).
10. Tartakovsky, A. M., Marrero, C. O., Tartakovsky, D. & Barajas-Solano, D. Learning Parameters and Constitutive Relationships with Physics Informed Deep Neural Networks. *arXiv preprint arXiv:1808.03398* (2018).
11. Lieberman, C., Willcox, K. & Ghattas, O. Parameter and state model reduction for large-scale statistical inverse problems. *SIAM Journal on Scientific Computing* **32,** 2523–2542 (2010).
12. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations. *arXiv preprint arXiv:1711.10561* (2017).
13. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations. *arXiv preprint arXiv:1711.10566* (2017).
14. Ha, Y. D. & Bobaru, F. Characteristics of dynamic brittle fracture captured with peridynamics. *Engineering Fracture Mechanics* **78,** 1156–1168 (2011).
15. Littlewood, D. *Simulation of Dynamic Fracture using Peridynamics, Finite Element Modeling, and Contact* in *Proceedings of the ASME 2010 International Mechanical Engineering Congress and Exposition, Vancouver, British Columbia, Canada* (2010).
16. Silling, S. A. Reformulation of elasticity theory for discontinuities and long-range forces. *Journal of the Mechanics and Physics of Solids* **48,** 175–209 (2000).
17. Benson, D., Wheatcraft, S. & Meerschaert, M. Application of a fractional advection-dispersion equation. *Water Resources Research* **36,** 1403–1412 (2000).
18. Schumer, R., Benson, D. A., Meerschaert, M. M. & Baeumer, B. Multiscaling fractional advection-dispersion equations and their solutions. *Water Resources Research* **39,** 1022–1032 (2003).
19. Schumer, R., Benson, D. A., Meerschaert, M. M. & Wheatcraft, S. W. Eulerian derivation of the fractional advection-dispersion equation. *Journal of Contaminant Hydrology* **48,** 69–88 (2001).
20. Bates, P. W. & Chmaj, A. An integrodifferential model for phase transitions: Stationary solutions in higher space dimensions. *J. Statist. Phys.* **95,** 1119–1139 (1999).
21. Delgoshaie, A. H., Meyer, D. W., Jenny, P. & Tchelepi, H. Non-local formulation for multiscale flow in porous media. *Journal of Hydrology* **531,** 649–654 (2015).
22. Fife, P. in, 153–191 (Springer-Verlag, New York, 2003).
23. Buades, A., Coll, B. & Morel, J. Image denoising methods. A new nonlocal principle. *SIAM Review* **52,** 113–147 (2010).

24. Gilboa, G. & Osher, S. Nonlocal linear image regularization and supervised segmentation. *Multiscale Model. Simul.* **6,** 595–630 (2007).

25. Gilboa, G. & Osher, S. Nonlocal operators with applications to image processing. *Multiscale Model. Simul.* **7,** 1005–1028 (2008).

26. Lou, Y., Zhang, X., Osher, S. & Bertozzi, A. Image recovery via nonlocal operators. *Journal of Scientific Computing* **42,** 185–197 (2010).

27. Alali, B. & Lipton, R. Multiscale dynamics of heterogeneous media in the peridynamic formulation. *Journal of Elasticity* **106,** 71–103 (2012).

28. Askari, E. Peridynamics for multiscale materials modeling. *Journal of Physics: Conference Series, IOP Publishing* **125,** 649–654 (2008).

29. Schekochihin, A. A., Cowley, S. C. & Yousef, T. A. *MHD turbulence: Nonlocal, anisotropic, nonuniversal?* in *In IUTAM Symposium on computational physics and new perspectives in turbulence* (2008), 347–354.

30. Burch, N., D'Elia, M. & Lehoucq, R. The exit-time problem for a Markov jump process. *The European Physical Journal Special Topics* **223,** 3257–3271 (2014).

31. D'Elia, M., Du, Q., Gunzburger, M. & Lehoucq, R. Nonlocal convection-diffusion problems on bounded domains and finite-range jump processes. *Computational Methods in Applied Mathematics* **29,** 71–103 (2017).

32. Meerschaert, M. M. & Sikorskii, A. *Stochastic models for fractional calculus* (Studies in mathematics, Gruyter, 2012).

33. Metzler, R. & Klafter, J. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports* **339,** 1–77 (2000).

34. D'Elia, M., Gulian, M., Olson, H. & Karniadakis, G. E. *A unified calculus for fractional, nonlocal, and weighted nonlocal models* tech. rep. *In progress* (Sandia National Laboratories, 2019).

35. Gao, P., Yang, X. & Tartakovsky, A. A New Approach For Learning Coarse-Grained Potentials with Application to Immiscible Fluids. *arXiv preprint arXiv:1907.06144* (2019).

36. Stinis, P. Enforcing constraints for time series prediction in supervised, unsupervised and reinforcement learning. *arXiv preprint arXiv:1905.07501* (2019).

37. Chorin, A. J. & Stinis, P. Problem reduction, renormalization, and memory. *Communications in Applied Mathematics and Computational Science* **1,** 1–27 (2007).

38. Isidori, A. *Nonlinear Control Systems* (Springer, London, 1995).

39. Stinis, P., Hagge, T., Tartakovsky, A. M. & Young, E. Enforcing constraints for interpolation and extrapolation in Generative Adversarial Networks. *Journal of Computational Physics* **397** (2019).

40. Masters, M. & Stinis, P. Continual learning for time series prediction. *(in preparation)* (2019).

41. Kashchiev, D. Determining the curvature dependence of surface tension. *J. Chem. Phys.* **118,** 9081–9083 (2003).

42. Tartakovsky, A. M. Continuum Model for Nanoscale Multiphase Flows. *arXiv preprint arXiv:1805.08319* (2018).

43. Howard, A. A., Zhou, Y. & Tartakovsky, A. M. Analytical steady-state solutions for pressure in multiscale non-local model for two-fluid systems. *arXiv preprint arXiv:1905.08052* (2019).

44. Howard, A. A. & Tartakovsky, A. M. Non-local model for surface tension in fluid-fluid simulations. *arXiv preprint arXiv:1906.10153* (2019).

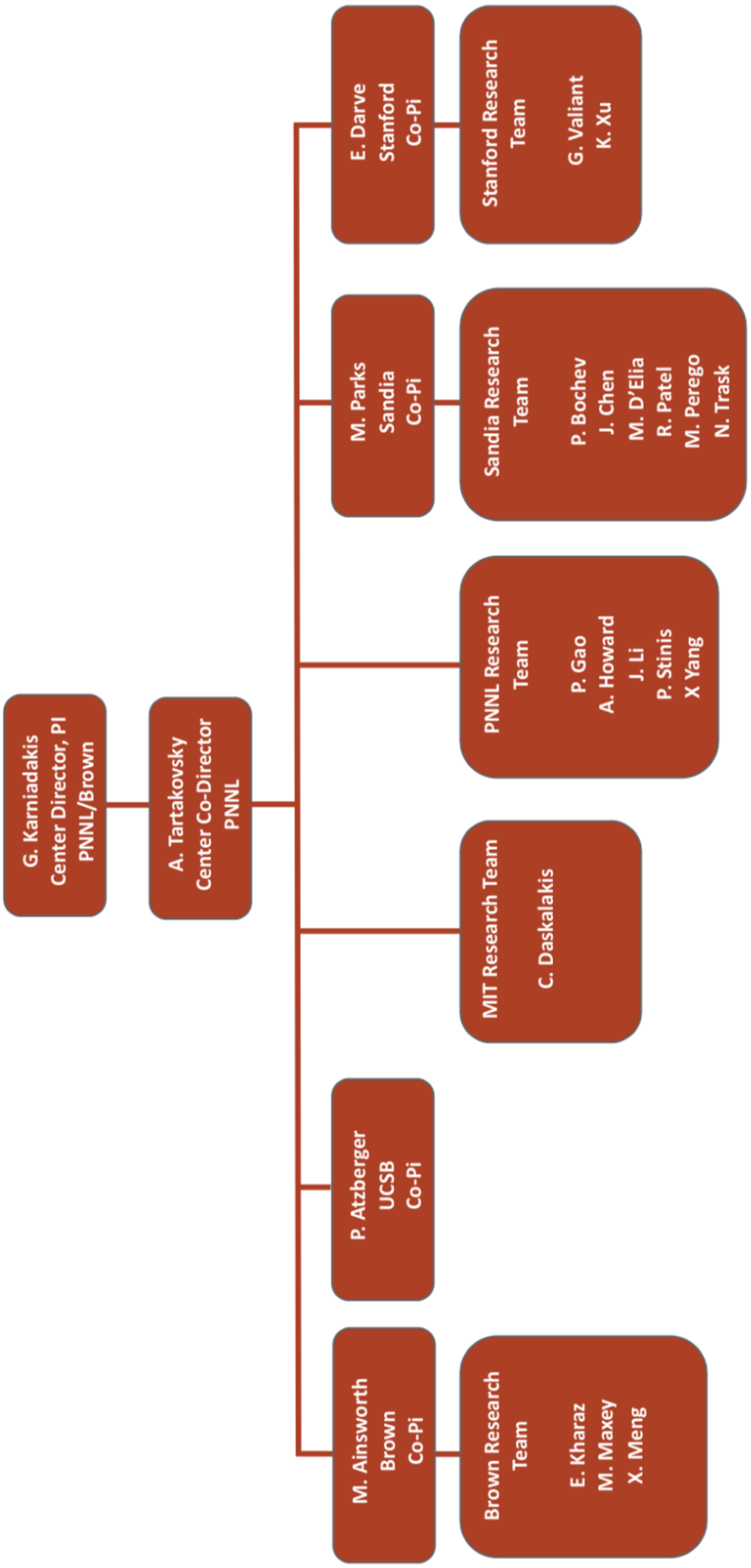
45. Li, J. & Tartakovsky, A. M. Gaussian Process Regression and Conditional Polynomial Chaos for Parameter Estimation. *arXiv preprint arXiv:1908.00424* (2019).

46. Jaderberg, M., Simonyan, K., Zisserman, A. & Kavukcuoglu, K. in *Advances in Neural Information Processing Systems 28* (eds Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R.) 2017–2025 (Curran Associates, Inc., 2015).

47. Daskalakis, C. & Panageas, I. *Last-Iterate Convergence: Zero-Sum Games and Constrained Min-Max Optimization* in *10th Innovations in Theoretical Computer Science Conference (ITCS)* (2019).

48. Daskalakis, C. & Panageas, I. *The Limit Points of (Optimistic) Gradient Descent in Min-Max Optimization* in *Advances in Neural Information Processing Systems (NeurIPS)* (2018).

B Org. chart


G. Karniadakis
Center Director, PI
PNNL/Brown
A. Tartakovsky
Center Co-Director
PNNL
E. Darve
Stanford
Co-Pi
Stanford Research
Team
G. Valiant
K. Xu
M. Parks
Sandia
Co-Pi
Sandia Research
Team
P. Bochev
J. Chen
M. D'Elia
R. Patel
M. Perego
N. Trask
PNNL Research
Team
P. Gao
A. Howard
J. Li
P. Stinis
X Yang
MIT Research Team
C. Daskalakis
P. Atzberger
UCSB
Co-Pi
M. Ainsworth
Brown
Co-Pi
Brown Research
Team
E. Kharaz
M. Maxey
X. Meng

C Work responsibilities and timelines

Research Area I: PDE-based modeling of macroscales (Lead: Ainsworth-Brown; co-PIs: Parks, D'Elia, Trask, Bochev-SNL; Karniadakis, Hodas, Stinis, Tartakovsky, J. Li-PNNL; Li, Maxey-Brown; Atzberger- UCSB; Darve-Stanford; MIT-Daskalakis)
Year 2: Continue developments and publish papers in the following areas: physics-informed neural networks (PINNs) that learn from multi-fidelity data for forward & inverse problems; unified nonlocal vector calculus, theory, and computation of nonlocal models: nonlocal PINNs (nPINNs); inference of constitutive laws of complex hard materials and of polymers from synthetic data using PINNs; learning the hidden fluid mechanics and hidden fractional dynamics in seismology using PINNs Machine learning methods / Regression on scattered data sets and manifolds (GMLS and neural networks); nonlinear functional and operator regression using GMLS parametrization and trunk-branch networks.
Years 3-4: Integrate the advances in RA-III and RA-IV. Develop PhILMs for multiscale and multiphysics problems in exemplar applications, specifically subsurface reactive transport and ice sheets. Combine nonlocal and fractional operators with PhILMs to discover (stochastic) closures in the exemplar applications. Compare the approximation properties of PhILMs and meshfree high-order methods for multiphysics problems at the macroscale. Study the ability of PhILMs to quantify uncertainty in long-term predictions in geophysical applications and validate it using historical data.

Research Area II: Stochastic modeling of mesoscales (Lead: Stinis-PNNL; co-PIs: Atzberger-UCSB; Tartakovsky, Yang, J. Li-PNNL; Trask, Parks-SNL; Kharazmi, Meng, Maxey-Brown; Darve-Stanford)
Year 2: Continue developments and publish papers in the following areas: learning the nucleation of nano-bubbles using many-body Dissipative Particle Dynamics (mDPD); physics-informed Gaussian process regression (GPR) Bayesian methods for incorporating physical information data sets on manifolds; connections between machine learning and model reduction with applications to unsupervised (GANS) and reinforcement learning; parallel-in-time PINNs for long-time integration of measoscale stochastic equations.
Years 3-4: Integrate the advances in RA-I and RA-IV; synthesize and scale up peptoids to achieve desired properties and functionality at the macroscale. Employ the MZ-derived mDPD to study nucleation in soft materials and obtain phase diagrams. Evaluate the ability of SRNNs to deal with multiple timescales in realistic soft material applications. In collaboration with RA-I, use fractional operators to represent nonlocal interactions (due to aggressive coarse-graining) in scaled-up functional materials and other systems.

Research Area III: Bridging methods to connect the scales (Lead: Bochev-SNL; co-PIs: Chen, Trask, D'Elia, Perego, Parks-SNL; Ainsworth, Kharazmi, Meng -Brown; Karniadakis, Tartakovsky, Yang-PNNL; Atzberger-UCSB; Daskalakis-MIT; Darve, Valiant-Stanford)
Year 2: Continue developments and publish papers in the following areas: active learning of constitutive laws via GPR for multiscale modeling of non-Newtonian fluids using DPD data; SPH-SPH interface for viscoelastic media using the Multiscale Universal Interface (MUI); domain decomposition for PINNs for porous media with largely disparate conductivities; learning coarse grained potentials for multiphase fluids (water-hexane system); learning the kernel in nonlocal models for surface tension based on MD data; learning surrogate models for turbulent mixing and ignition (DSMC, DPD, MD); learning (via GPR of fractional PDEs) of nonlocal flocking dynamics from particle trajectories (agent models).
Years 3-4: Integrate the advances in RA-I, RA-II, and RA-IV. Apply active learning and upscaling to peptoids and combustion and examine accuracy and cost; validation using data from PNNL and SNL, respectively. Apply active learning and domain decomposition to subsurface reactive transport and ice sheets and examine feasibility and scaling up from laboratory scales to field scales. Validation with existing (classical) solvers and partial data available at PNNL and SNL.

Research Area IV: Bayesian Deep Learning (Lead: Darve-Stanford; co-PIs: Valiant-Stanford; Daskalakis-MIT; Karniadakis, Yang, Stinis-PNNL; Parks, Trask, Bochev-SNL; Atzberger-UCSB; Ainsworth, Kharazmi, Meng - Brown)
Year 2: Continue developments and publish papers in the following areas: generalized existing techniques for estimating learnability of the best classifier in a specified class in the data regime in which there is insufficient data to learn even an approximation of such a classifier; explored approaches for integrating certain classes of invariances within a convolutional DNN architecture; developed Generative Adversarial Networks for the optimal transport problem and for solving high-dimensional stochastic PDEs (e.g., 10,000 dimensions in a porous media model for the Hanford site); designed stable optimization methods ('optimistic' gradient descent) for Generative Adversarial Networks.
Years 3-4: Develop the next generation of PhILMs for continuum- and molecular-based physical systems. Systematically study the findings of RA-I-III and incorporate lessons learned into the new deep learning architectures and SGD and HNN algorithms. Develop and finalize information-theoretic approaches for designing a priori PhILMs with a specific number of layers and neurons and document accuracy bounds.

D Abbreviations

- AD: Automatic differentiation
- CG: Coarse-grained
- CM4: Collaboratory on Mathematics for Mesoscopic Modeling of Materials
- CNN: Convolutional neural network
- DD-PINN: Domain decomposition physics informed neural network
- DNN: Deep neural network
- DOE: Department of Energy
- DPD: Dissipative particle dynamics
- ET: Equivariant transformer
- FEM: Finite element method
- FF: Force field
- GAN: Generative adversarial metworks
- GMLS: Generalized moving least squares
- KL: Karhunen-Loève
- MAP: Maximum *a posteriori* probability
- MD: Molecular dynamics
- MIT: Massachusetts Institute of Technology
- ML: Machine learning
- MPINN: PINN trained with multi-fidelity data sets
- nPINN: nonlocal physics informed neural network
- PDE: Partial differential equation
- PINN: Physics-informed neural network
- PNNL: Pacific Northwest National Laboratory
- PPINN: Parallel physics informed neural network
- SNL: Sandia National Laboratories
- SPASD: Supervised parallel-in-time algorithm for stochastic dynamics
- UCSB: University of California, Santa Barbara