Research Statement

Hanfeng Zhai

"As we go down in size, there are a number of interesting problems that arise. All things do not simply scale down in proportion." —————— *There's Plenty of Room at the Bottom*, Richard P. Feynman, 1959

A time-honored objective in engineering sciences is the theoretical and computational understanding of multiscale problems. As the world moves towards sustainability and biocompatibility, multiscale modeling and optimization can greatly accelerate scientific research and industrial productions, e.g., faster drug discovery, efficient battery design, novel energy sources exploration, and general materials design. However, there are two main issues in deploying computation techniques for real-world applications: (1) There exist theoretical gaps between the simulations at different scales, specifically on the **mesoscale**, i.e., people either add stochastic terms to continuum models (top-down) or coarse-grain full atomistic models (bottom-up) to describe the meso-world, with limited understandings of the relations across scales. (2) Seamlessly tailoring & optimizing materials properties across different scales based on computer simulations are difficult considering the algorithms **mismatch**, e.g., maximizing electronic conductivity at the atomic scale and maximizing toughness at the macroscale for battery design simultaneously is difficult, considering one would usually use molecular dynamics simulations to model the first scenario yet using finite element models for the latter.

To address these limitations, my research goal is to **develop computational methods** for better modeling, simulations, and optimizations of **multiscale engineering phenomena** that can bring good for mechanical, materials, and systems design with applications in **energy**, **bioengineering**, and **manufacturing**. I believe working in the following three aspects could roadmap such goals: ($\mathfrak{A}$) Multiscale Computational Modeling: This involves developing mathematical models, i.e., based on Langevin dynamics, Hamiltonian mechanics, etc., and computational models, i.e., Monte Carlo simulations methods, Coarse-Graining potential fields, etc. Related works could bring in better understanding regarding "(1), and potentially provide solutions for "(2). ($\mathfrak{B}$) Scientific Machine Learning: Most works in such areas are based on the idea of constructing surrogate models. Two applications interest me, (I) Machine learning (ML) potential for scale-bridging, i.e., training a regression model based on first principle calculations and applying the obtained energy metasurfaces to molecular simulations; and (II) Metaheuristics for optimizations, i.e., training a surrogate for the variable space for design optimization, which elicits our final research theme: ($\mathfrak{C}$) Materials Design Optimization: This is a specific tool to tackle "(2), involve designing new materials and structures by identifying the objectives according to engineering applications. Examples include designing new antibiofilm nanosurfaces with Bayesian optimization and designing thermo-magnetic generators with genetic algorithms.
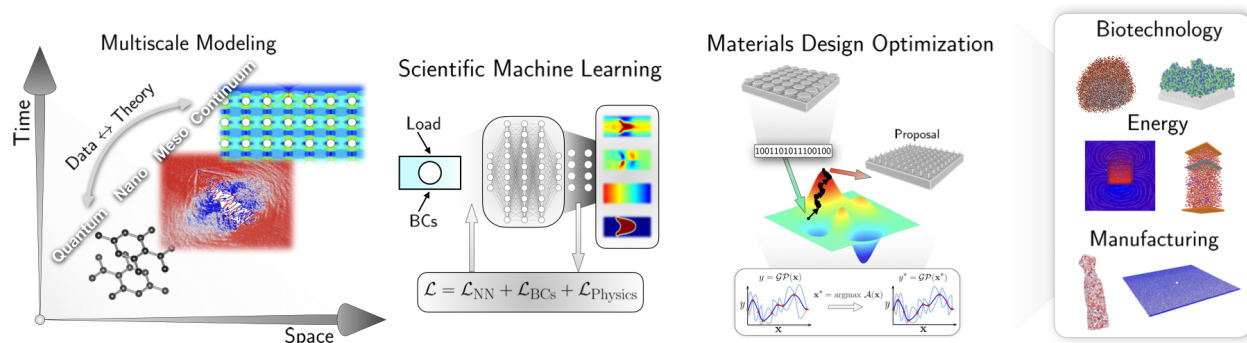

Figure 1: The schematic of my research themes & knowledge backgrounds, from theories to applications.

Here, "($\mathfrak{A}$) and "($\mathfrak{B}$) enable targeted applications in "($\mathfrak{C}$), e.g., one can build up an optimization workflow enabled by molecular simulations and construct a design space surrogate via Gaussian process machine learning for the design of desalination graphene to maximize the desalination rate and the mechanical strength simultaneously. "($\mathfrak{A}$), "($\mathfrak{B}$) & "($\mathfrak{C}$) together provides potential solutions for "(1) and "(2). We will highlight some of our previous works and provide potential directions for future plans based on the above.

Multiscale Computational Modeling: From Macro to Micro

In a report by Oden *et al.* on virtual design, it is pointed out that *"In order to make virtual design a reality in the next decade, radically new computational tools with the ability to handle **multiscale phenomena**, ..., must be available."* Current state-of-the-art multiscale methods include finite element method (FEM), extended FEM (XFEM), discrete elements method (DEM), and molecular dynamics (MD), which are usually implemented in software like ANSYS APDL & workbench, Abaqus CAE, COMSOL Multiphysics, LAMMPS, etc.

Finite element enabled structural analysis & design. Basic structural design proffers promising guidelines in civil engineering, and product design like phones, chairs, etc. I lead a project team tailoring *ad hoc* compressive structures for an optimal structural design using ANSYS APDL & workbench with 3D printing and compression tests [1]. The *in silico* experiments verify our proposal from FEM simulations and largely reduce experimental costs, leading to an *Outstanding Undergraduate Student Course Project*. Interested in the thermal effect on microelectronics devices, we proposed analytical solutions coupling the thermal and linear elastic model of plate theory and used ANSYS workbench structural & thermal modules to verify [2].

Continuum modeling of mechanical properties of bio- and bio-inspired materials. The intriguing nature of biomaterials sparked interest in the mechanics' communities in recent decades, where continuum modeling for fast accurate predictions of various properties offers potential applications like biomedical devices and orthopedic surgery. We implemented the perfect plasticity model (Abaqus CAE) to study the effect of the initial defect on the constitutive response of fiber-reinforced composite, using an RVE approach [3]. Inspired by the microstructure of enamel and nacre, through running FEM simulations, we proposed an optimized structure combining the two that achieve higher stiffness and toughness in both directions, using a plain-strain approach [4]. Inspired by the enamel microstructures, we compared the perfect plastic and linear elastic model of inter-tubular dentin (ITD) bonding with high stiffness elastic peri-tubular dentin implemented in Abaqus CAE using XFEM and proposed the plasticity of ITD resists crack propagation [5]. This series of projects may inspire the design of next-generation bio-inspired structures and materials.

Molecular dynamics study of thermal gradient-coupled graphene fracture. The fracture behaviors of graphene under thermal gradients are essentially non-equilibrium processes that possess rich industrial applications as potential materials for desalination, gas separation, etc. We apply optimized Tersoff, REBO, AIREBO, and AIREBO-M potentials for non-equilibrium MD (NEMD) to model the crack propagation of single-layer defected graphene implemented in LAMMPS [6]. We also compare the empirical potentials with state-of-the-art ML potentials. Our simulations indicate that the thermal gradients do not show an evident trend regarding the fracture directions. The NEMD data fitted well with quantized fracture mechanics theories and literature experimental results. The ML potentials underestimate the fracture stresses.

Scientific Machine Learning: Data-Enhanced Scientific Computing

In a review by Karniadakis *et al.*, they state: *"..., there is a pressing need for integrating fundamental physical laws and domain knowledge by 'teaching' ML models about governing physical rules, ..."*, indicating the importance of integrating domain knowledge and AI for industrial productions and scientific discoveries.

Physics-informed deep learning for bubble dynamics. ML surrogate modeling is a heated topic in scientific computing recently. Among, encoding physical laws via auto-differentiation as constraints to accelerate the NN training for less required labeled data is becoming a promising direction, termed physics-informed neural network (PINN). However, an essential question is how much physical information is sufficient to "inform" the NN? We implement PINN with continuity equation, and Poisson equation (symbolized as $\mathcal{P}$) comparing with the normal NN [8] trained with FEM simulation data of microbubbles constrained in channels. We found that the continuity equation with a proposed time-discretized normalizer can greatly improve the NN predictions, whereas the inclusion of the $\mathcal{P}$ does not offer evident improvements. Our works shed light on PINN surrogates' construction strategies for bubbly flows with energy and biomedical applications.

Machine learning for dynamical systems. Nonlinear dynamics are omnipresent in engineering sciences. Specifically, the van der Pol dynamics have wide applications in robotics, biology, chemical engineering, etc. We encode stabilized system signals as circular trajectories in phase portraits into a PINN to answer the question: can a NN be trained on nonlinear dynamic systems but infer different desired output through encoding signals as constraints but still possessing characteristics of the training data? This would contain

huge potential in control sciences. Answering this, we achieve good signal implementation quality for stabilizing the van der Pol systems of lower nonlinearities for outputting different desired circular trajectories [9]. However, a comparison with traditional control methods including nonlinear feed-forward, linearized feedback, and combined controls indicates machine learning methods took a significantly longer time on training as for total computational burden and nonlinear idealized feed-forward achieve overall better signal implementation accuracy [10]. We also apply Gaussian process regression (GPR), autoregressive modeling (AR), and autoregressive moving average (ARMA) to train and infer real-time ocean turbulence dynamical data from drifters as part of a DARPA FFT Challenge [7], ending up with a global ranking of #27.

Materials Design Optimization: Algorithm to Applications

Osanov and Guest state that *"..., materials architectures must be optimally designed according to the target application, base material used, ..."* in a review on topology optimization. The "*optimally designed*" as the optimization algorithms and "*target application*" as the engineering objectives are bridged by digital twins.

Thermo-magnetic generator design using multi-objective genetic algorithm. Considering the urgency of global warming and energy transformation, efficient development methods of clean energy power generators are of urgent need. Specifically, thermo-magnetic generators that can convert waste heat into electricity roadmap the path to a clean and carbon-neutron future. Using FEM simulations in COMSOL Multiphysics, coupling with multi-objective genetic algorithm (MOGA) implemented in MATLAB, we provide engineering design strategies of thermo-magnetic generator through running 100 generations and select the optimal designs based-upon the objective values [11]. Considering both the efficiency, power output, and cost as objectives, the designed geometry provides actual guidelines according to real-world scenarios.

Antimicrobial active surfaces by design via automated Bayesian optimization. Bacteria biofilms create huge problems in diverse engineering disciplines from marine science, and biomedicine to general global health. We perform the discrete element method (DEM) for simulating multiscale & multiphysical properties of biofilms. Inspired by the topology of hydrophobic nanosurfaces, a constrained optimization for such topological designs is formulated. By implementing DEM in LAMMPS coupling with Bayesian optimization (BO) algorithms implemented in Python, we generated metamodels of the design domain through GPR and extract the optimal antimicrobial active surfaces considering different biofilm removal methods [12]. The proposed autonomous simulation-based BO workflow can be applied to more general materials design.

Future Research

My previous skills and research built me a foundation in numerical modeling, statistical learning, and optimization. My potential future work centers on bridging these for materials modeling & design. Striving to answer our fundamental questions in "(1) and "(2), we propose the following potential directions.
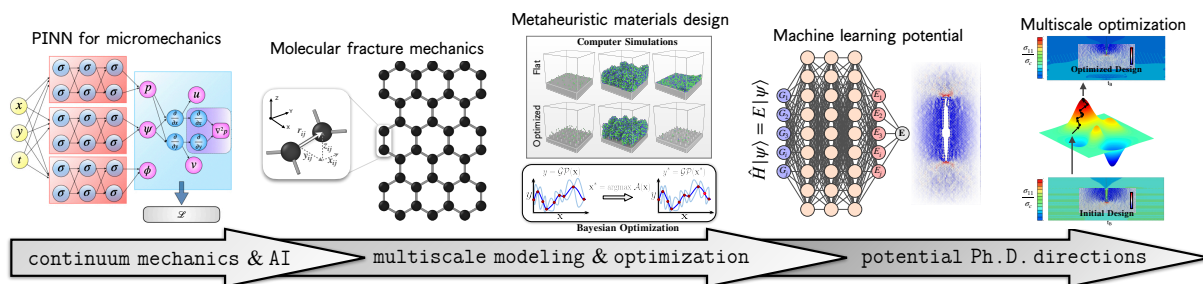

Figure 2: From my undergraduate and master's research to potential Ph.D. directions.

Multiscale materials characterization under extreme environments. Materials' behaviors under extreme environments, i.e., high temperature, shock, high pressure, etc., have been an active research area for decades. Recent advancements in high-performance computing and computational methods, i.e., coarse-graining, model reduction, etc., could provide significant insights into this field. Many sponsors and institutions funded related ideas and projects recently. For example, PNNL's CM4 project proffers the theoretical foundation for materials mesoscopic modeling, A recent IDA report suggests that ML techniques could greatly advance materials characterizations under extreme environments, An Air Force project implies that

reduced-order modeling could potentially advance alloy damage modeling. Our future plan for this research theme is: ($\mathbb{A}$) Develop and benchmark different mesoscale computational methods, e.g., compare dissipative particle dynamics, coarse-grained molecular dynamics, and Monte Carlo simulations on fracture modeling of polymers, alloys, etc. This could help us understand how different computational methods perform for extreme conditions modeling, and further provide guidelines to specific applications like which methods to adopt when modeling hydraulic fractures, gas separations, etc. ($\mathbb{B}$) Develop and benchmark different model reduction techniques for mesoscopic simulations of extreme conditions, e.g., compare linear and quadratic reduced order modeling on crystal plasticity simulations. This provides answers for which modeling methods to use in simulating extreme materials behaviors to reduce computational burden while simultaneously preserving model quality. Our previous works in "($\mathfrak{A}$) & "($\mathfrak{B}$) prepared me with skills in using FEM, MD, DEM implemented in LAMMPS and other software to conduct simulations of different materials and training ML surrogate models based on simulation data. These skills ensure this research theme can be carried out.

Machine learning potentials for scale-bridging and extreme conditions. The development of quantum-molecular concurrent modeling was awarded the 2013 Nobel Prize in Chemistry. The pioneering work of Behler and Parrinello of using neural networks for learning molecular potentials via training data from first principle calculations inspired further data-driven scale-bridging efforts for molecular simulations. Recently, both Argonne and Sandia National Labs are making efforts in this area. Many publications and works have been developed on new machine learning potentials for different materials like graphene, copper, aluminium, etc. However, two main points these works lack direct our future research: ($\mathbb{A}$) Most state-of-the-art ML potentials are only developed for single element systems or simple alloys, creating a vacancy for complex materials like metallic glass, graphene oxide polymer composites, etc. We hope to develop new ML potentials for these complex materials. Related works may potentially benefit both academia and industry by adopting new targeted materials in advanced research & developments. ($\mathbb{B}$) Most ML potentials are only able to model atomic stable configurations under equilibrium, lacking abilities to model non-equilibrium behaviors like fractures. We hope to develop *ad hoc* ML potentials or ML-based surrogates for those extreme conditions following our previous direction, e.g., an ML potential for describing graphene oxide fracture. Related works will potentially scale up high fidelity simulations, and further, potentially unveil hidden mechanisms within materials' extreme behaviors. Previous works in "($\mathfrak{A}$) & "($\mathfrak{B}$) proffer me solid skills in training ML models, performing *ab initio* & molecular simulations, ensuring related works can be carried out.

Scientific machine learning for materials design optimization. Design and optimization are identified as key factors and further introduced as a prototype study in Materials Genome Initiative Strategic Plan. The recent boost in metaheuristics such as grey wolf optimization and whale optimization provides new insights for materials design compared with many traditionally used heuristics like GA or BO. Based on these, we identify two potential directions: ($\mathbb{A}$) Apply and benchmark novel optimization methods in materials design, e.g., compare grey wolf optimization with GA on the design of hydrogel with both high toughness and high thermal conductivity. This could help us understand how different algorithms perform on new materials design, e.g., BO is maybe more suitable for Li-ion battery design considering the more complicated system whereas gradient-based optimization is maybe more suitable for high-strength architectured structures design due to the convenience of gradients computations in existed FEM models. ($\mathbb{B}$) Combining state-of-the-art ML models as surrogates for materials optimization, e.g., apply and compare DeepONet with traditional models like GPR or NN as surrogates in BO for the design of block copolymers with high strength and ion-conductivity. This could help us understand the role of surrogate models in metaheuristic optimization specifically in materials design. Our previous works in "($\mathfrak{B}$) & "($\mathfrak{C}$) roadmaps this theme: the experiences in research theme "($\mathfrak{B}$) offers me to fast adopt and train novel ML models; and researches in "($\mathfrak{C}$) provide me skills of optimizing materials with predefined design variables, constraints, and their application-based objectives through running optimization loops between the heuristic algorithms and multiscale simulations.

All those potential directions combined may help us answer our fundamental questions "(1) and "(2): the first two directions may help us better understand the theoretical relations between scales in "(1), and the latter two directions may provide better materials optimization strategies for "(2). Importantly, the three potential directions are not simple follow-ups to our previous works in "($\mathfrak{A}$), "($\mathfrak{B}$), "($\mathfrak{C}$) – these novel ideas are challenging for researchers in the field, and possess huge potential impacts for both industry and academia. To be more specific, there are various technical challenges corresponding to each of our prementioned potential directions. For both the first and second directions, many novel computational

toolboxes and methods are developed based on specific versions of software with numerous limitations, e.g., the original PINNs depend on TensorFlow v.1, OpenKIM requires targeted C++ compiler or `sudo` access to install, and the package installer and they do not favor specific versions of `wget`, which are not user-friendly for researchers lack experiences in high-performance computing. For the second and the third directions, the automation of the design optimization requires building up interfaces between the simulation programs with tunable variables and optimization algorithms to variate these variables, e.g., our nanosurface design optimization requires compiling the LAMMPS-Python interface with their corresponding C++ interpreter, simultaneously implementing NUFEB in LAMMPS. This would require the programmer to be familiar with both simulation and optimization codes and theories implemented in the corresponding software (e.g. LAMMPS, FEniCS) and languages (e.g. Python, Julia), and understand fundamental debugging techniques to compile their interfaces. We have shown a successful history of solving these tough problems. My unique skills & knowledge background, and perspectives ensure our niché in the field to drive future innovations.

References

[1] **H. Zhai**, S. Weng, B. Liang, C. Tang, Z. Wang. Structural Design of Lightweight Compressive Interlayer. *Group Project Poster for Finite Element and Its Applications* (2019). [`Poster`] (*Selected as Outstanding Undergraduate Course Project*)

[2] **H. Zhai**. Thermal Estimation of Smartphone Chipset: Mechanical Distribution of Chipset in Multiphysics Field. *Course Paper for Elastic Mechanics* (2019). [`PDF`]

[3] **H. Zhai**. Formulation and application of rate-independent stress update algorithm of hydrostatic pressure: elastoplastic yielding in composite. *Course Paper for Plasticity Theory* (2020). [`PDF`]

[4] **H. Zhai**. Structural design of composite materials with superior mechanical behaviors: lesson from the microstructure of nacre and enamel. *Course Paper for CAD Application in Structural Mechanics* (2020). [`PDF`]

[5] **H. Zhai**. An investigation of the elastoplastic nature of ITD on the toughness of the dentin microstructure. *Shanghai University Summer Research Project Report* (2020). [`PDF`]

[6] **H. Zhai**, J. Yeo*. Multiscale Characterization of Graphene Fracture using Machine Learning & Empirical Interatomic Potentials and Mechanics. *Molecular Machine Learning Conference, MIT, Cambridge, MA*. (2022). DOI: `10.1063/5.0079602`.

[7] **H. Zhai**, T. Sands*. Autoregressive modeling for turbulence predictions. *DARPA FFT Challenge*, (2021). [`PDF`]

[8] **H. Zhai**, Q. Zhou, G. Hu*. Predicting micro-bubble dynamics with semi-physics-informed deep learning. *AIP Advances* **12**, 035153 (2022). DOI: `10.1063/5.0079602`. (Featured Article)

[9] **H. Zhai**, T. Sands*. Controlling Chaos in Van Der Pol Dynamics Using Signal-Encoded Deep Learning. *Mathematics*. 2022; 10(3):453. DOI: `10.3390/math10030453`. (Highly Accessed Article)

[10] **H. Zhai**, and T. Sands*. "Comparison of Deep Learning and Deterministic Algorithms for Control Modeling". *Sensors*. 22, 6362; (2022). DOI: `10.3390/s22176362`

[11] W. Hintlian, **H. Zhai**, M. Berg. Engineering Design of Thermo-Magnetic Generator using Multi-objective Genetic Algorithm. *Course Paper for Multidisciplinary Design Optimization*. [`PDF`]

[12] **H. Zhai**, and J. Yeo*. "Computational design of antimicrobial active surfaces via automated Bayesian optimization". *arXiv preprint*. (2022). DOI: `10.48550/arXiv.2209.00055`