AI-Accelerated Discovery of Ultra-High Strength-to-Weight Composite Materials

Principal Investigator: Dr. Sarah Chen
Co-Principal Investigator: Dr. Michael Rodriguez
Harvard University

Project Summary

Overview

This proposal seeks funding under the NSF Designing Materials to Revolutionize and Engineer our Future (DMREF) program to develop an integrated computational-experimental framework for discovering composite materials with order-of-magnitude improvements in strength-to-weight ratio. By combining machine learning, high-throughput computational materials modeling, and autonomous experimental validation, we will establish a closed-loop materials discovery pipeline targeting aerospace and transportation applications.

Intellectual Merit

The proposed research will advance fundamental understanding of structure-property relationships in multi-phase composite materials through:

- Development of physics-informed neural networks (PINNs) for predicting mechanical properties across length scales

- High-throughput density functional theory (DFT) calculations to populate training databases

- Automated synthesis and characterization protocols for rapid experimental validation

- Integration of manufacturing constraints into the materials design framework

Broader Impacts

This project will train graduate students and postdocs in the interdisciplinary skills required for modern materials research. We will partner with local community colleges to provide summer research opportunities for underrepresented minorities. Open-source software and curated datasets will be made publicly available through Materials Project and GitHub repositories. The discovered materials have potential applications in fuel-efficient aircraft, electric vehicles, and sustainable infrastructure.

1 Project Description

1.1 Introduction and Motivation

The search for materials with exceptional mechanical properties has been a cornerstone of materials science for over a century. Despite significant advances, the discovery of new high-performance materials remains largely empirical, requiring years of trial-and-error experimentation. This is particularly true for composite materials, where the complex interplay between constituent phases, interfaces, and microstructure creates an enormous design space that is impractical to explore through traditional methods.

Recent advances in artificial intelligence and machine learning offer transformative opportunities to accelerate materials discovery. However, most ML approaches in materials science have focused on single-phase materials or simple property prediction, neglecting the manufacturing constraints and multi-scale physics that govern real-world material performance.

This proposal addresses these limitations by developing an integrated AI-driven framework for discovering composite materials with strength-to-weight ratios exceeding current state-of-the-art by at least 10×. Our approach uniquely combines:

- **Multi-scale modeling**: From electronic structure to continuum mechanics

- **Physics-informed ML**: Incorporating conservation laws and symmetries into neural network architectures

- **Autonomous experimentation**: Robotic synthesis and characterization with real-time feedback

- **Manufacturing-aware design**: Optimizing not just for properties but for manufacturability

1.2 Background and Prior Work

1.2.1 Composite Materials

Composite materials—combining two or more distinct phases to achieve properties unattainable in single-phase materials—have revolutionized aerospace, automotive, and sporting goods industries. Carbon fiber reinforced polymers (CFRPs) currently represent the state-of-the-art, with specific strengths around 2-3 GPa/(g/cm³). However, theoretical calculations suggest that properly designed nanocomposites could achieve specific strengths exceeding 20 GPa/(g/cm³).

The key challenges in realizing such materials include:

1. **Phase selection**: Identifying optimal matrix and reinforcement materials

2. **Interface engineering**: Designing interfaces that effectively transfer load

3. **Microstructure control**: Achieving desired reinforcement distribution and orientation

4. **Processing limitations**: Developing scalable manufacturing routes

1.2.2 Machine Learning for Materials

Recent years have witnessed explosive growth in ML applications to materials science. Graph neural networks have shown remarkable accuracy in predicting crystal properties from structure alone [1]. Transfer learning has enabled property prediction with limited training data [2]. Active learning strategies have optimized experimental campaigns [3].

However, most existing ML materials work suffers from key limitations:

- Focus on equilibrium properties of crystalline materials

- Limited integration with synthesis and processing

- Black-box models lacking physical interpretability

- Insufficient attention to manufacturing constraints

Our preliminary work has demonstrated that physics-informed neural networks (PINNs) can overcome these limitations for composite materials [4]. By encoding conservation laws and material symmetries directly into network architecture, we achieved 3× better prediction accuracy with 10× less training data compared to standard approaches. Our open-source implementation is available at `https://github.com/chenlab/composite-ml`.

1.3 Research Plan

1.3.1 Aim 1: Computational Materials Screening (Years 1-2)

Objective: Develop and validate high-throughput computational pipeline for predicting mechanical properties of candidate composite systems.

Approach:

1. **Database construction** (Months 1-6): Compile existing experimental and computational data on composite materials. Augment with new DFT calculations on 5,000+ matrix-reinforcement combinations, focusing on lightweight elements (C, B, Al, Mg, Ti) and compounds (carbides, borides, oxides). Calculate interface energies, elastic constants, and fracture properties.

2. **Multi-scale modeling framework** (Months 3-12): Develop hierarchical modeling approach linking:

   - DFT for interface properties and bonding
   - Molecular dynamics (MD) for nanoscale mechanics
   - Micromechanics models (Mori-Tanaka, finite element) for effective properties
   - Continuum models for macroscopic failure

3. **Physics-informed neural networks** (Months 6-18): Train PINNs to predict:

   - Elastic moduli and strength as functions of composition and microstructure
   - Failure modes (interfacial debonding vs. matrix cracking vs. fiber breakage)
   - Processing-microstructure relationships
   - Manufacturing feasibility scores

   Key innovation: Incorporate material symmetries, thermodynamic bounds, and Hall-Petch relationships as hard constraints in the loss function.

4. **Validation and uncertainty quantification** (Months 12-24): Validate predictions against held-out test data and new calculations. Implement Bayesian deep learning to quantify prediction uncertainties. Establish confidence thresholds for experimental validation.

Expected outcomes: Validated computational pipeline capable of screening 10^6 candidate compositions and predicting mechanical properties with <15% error. Identification of 100+ promising candidates for experimental synthesis.

1.3.2 Aim 2: Autonomous Experimental Validation (Years 1-3)

Objective: Establish robotic synthesis and characterization platform with closed-loop feedback to computational models.
   Approach:

1. **Synthesis automation** (Months 6-18): Develop automated protocols for composite fabrication via:

   - Spark plasma sintering for ceramic matrix composites
   - Infiltration casting for metal matrix composites
   - Automated fiber placement for polymer matrix composites

   Integrate with robotic liquid handling and powder dispensing systems. Target throughput: 5-10 samples per day.

2. **Characterization pipeline** (Months 6-18): Automate key measurements:

   - Density (Archimedes method)
   - X-ray diffraction (phase identification)
   - Scanning electron microscopy (microstructure)
   - Nanoindentation (local mechanical properties)
   - Tensile/flexural testing (bulk properties)

   Implement computer vision for automated image analysis and defect detection.

3. **Closed-loop integration** (Months 12-30): Establish feedback loops:

   - Experimental results automatically update computational model training data
   - Bayesian optimization selects next candidates to maximize expected information gain
   - Synthesis parameters adapted based on observed microstructure-property relationships
   - Failure analysis guides model refinement

4. **Scale-up studies** (Months 24-36): For top 5 candidates, investigate:

   - Larger specimen sizes (up to 10 cm)
   - Alternative synthesis routes for manufacturability
   - Long-term stability and environmental resistance
   - Preliminary cost analysis

   Expected outcomes: Synthesis and characterization of 200+ candidate materials. Validation of at least 3 compositions achieving >10× improvement in strength-to-weight ratio compared to conventional composites. Demonstrated closed-loop optimization reducing discovery time by 5× compared to traditional approaches.

1.3.3 Aim 3: Manufacturing-Aware Design Framework (Years 2-4)

Objective: Incorporate manufacturing constraints into materials design to ensure practical feasibility of discovered materials.
   Approach:

1. **Processing parameter models** (Months 12-30): Develop predictive models for:

   - Sintering kinetics and densification
   - Reinforcement distribution and orientation
   - Interfacial reaction and diffusion
   - Residual stress development

   Integrate with phase-field simulations and thermodynamic databases.

2. **Manufacturability metrics** (Months 18-36): Define and implement quantitative manufacturability scores based on:

   - Raw material cost and availability
   - Processing temperature and time
   - Equipment requirements
   - Yield and reproducibility
   - Health and environmental considerations

3. **Multi-objective optimization** (Months 24-42): Implement Pareto optimization to simultaneously maximize:

   - Mechanical performance (strength-to-weight ratio)
   - Manufacturability score
   - Material sustainability metrics

4. **Industry collaboration** (Months 24-48): Partner with aerospace and automotive companies to:

   - Validate manufacturability assessments
   - Identify application-specific requirements
   - Explore pathways for technology transfer

   Expected outcomes: Manufacturing-integrated design framework that identifies not just high-performance materials but *practically realizable* high-performance materials. Demonstrated pathway to commercial production for top candidates.

1.4 Timeline and Milestones

1.5 Team Expertise and Roles

PI Chen (Harvard, Materials Science) brings expertise in computational materials science, machine learning, and high-throughput DFT. She will lead the computational screening efforts (Aim 1) and oversee overall project management. Her group has published 35+ papers on ML for materials and developed the open-source `MatML` software package.

Year	Milestones
1	Database compilation; Initial PINN development; Synthesis automation setup; Characterization pipeline development
2	PINN validation; First round experimental screening (50+ materials); Closed-loop optimization implemented; Manufacturing models development
3	Refinement of predictions based on experimental data; Second round optimization (100+ materials); Manufacturability framework complete; Scale-up studies initiated
4	Final validation; Industry partnerships; Scale-up completion; Documentation and data/software release; Technology transfer activities

Co-PI Rodriguez (Harvard, Mechanical Engineering) is an expert in composite materials processing and mechanical characterization. He will lead the experimental validation (Aim 2) and scale-up studies. His laboratory houses state-of-the-art synthesis and testing equipment including spark plasma sintering, automated fiber placement, and a universal testing machine with environmental chamber.

The team will be supported by:

- 2 graduate students (one computational, one experimental)

- 1 postdoctoral researcher with expertise in autonomous experimentation

- Undergraduate research assistants from our NSF REU program

1.6 Data Management and Dissemination

All computational and experimental data will be curated according to FAIR principles and deposited in public repositories:

- Crystal structures and DFT calculations $\rightarrow$ Materials Project

- ML models and training data $\rightarrow$ GitHub with DOI via Zenodo

- Experimental synthesis/characterization data $\rightarrow$ Figshare

- Processed datasets $\rightarrow$ Materials Data Facility

Software will be released under permissive open-source licenses (MIT/Apache 2.0). We will publish in high-impact open-access journals and present at major conferences (MRS, TMS, ICCM). Annual workshops will disseminate methods to the broader community.

Data will be retained for minimum 10 years post-project and made available within 6 months of generation or publication, whichever is sooner.

1.7 Broader Impacts

1.7.1 Education and Workforce Development

We will train the next generation of materials researchers in the integrated computational-experimental approach championed by MGI:

- Graduate students will gain experience across theory, simulation, ML, and experiment

- Annual summer program for community college students, targeting underrepresented minorities

- Development of course modules on AI for materials (to be shared via nanoHUB)

- Mentorship through established programs: Harvard PRISE, URIECA

1.7.2 Societal Impact

Ultra-high strength-to-weight materials enable transformative applications:

- **Transportation**: 50% weight reduction in aircraft and vehicles $\rightarrow$ major fuel savings and emissions reduction

- **Infrastructure**: Stronger, lighter construction materials $\rightarrow$ more resilient buildings and bridges

- **Renewable energy**: Lighter wind turbine blades $\rightarrow$ improved efficiency and reduced costs

1.7.3 Diversity and Inclusion

PI Chen is committed to broadening participation in STEM. As an Asian-American woman in a traditionally male-dominated field, she actively mentors women and minorities. Co-PI Rodriguez works extensively with the Harvard College Hispanic Association. The project will:

- Reserve 50% of summer research positions for students from underrepresented groups

- Partner with HBCUs for student exchange and collaborative research

- Develop and disseminate culturally responsive pedagogy materials

1.8 Prior NSF Support

PI Chen: DMR-2054321, "Machine Learning for Crystal Structure Prediction," $450,000 (2021-2024). *Intellectual Merit*: Developed graph neural network approaches achieving state-of-the-art accuracy in predicting formation energies and bandgaps. Published 8 papers, released open-source software with 500+ GitHub stars. *Broader Impacts*: Trained 3 PhD students and 8 undergraduates; 2 students from underrepresented minorities.

Co-PI Rodriguez: CMMI-1925476, "Autonomous Synthesis of Ceramic Matrix Composites," $380,000 (2019-2023). *Intellectual Merit*: Developed robotic platform for automated composite fabrication, demonstrating 10× throughput improvement. Published 6 papers, filed 2 patent applications. *Broader Impacts*: Trained 2 PhD students, developed hands-on lab module adopted by 5 universities.

2 References Cited

References

[1] Chen, C. et al. "Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals." *Chemistry of Materials* **31**, 3564-3572 (2019).

[2] Jha, D. et al. "Enabling deeper learning on big data for materials informatics applications." *Scientific Reports* **9**, 8174 (2019).

[3] Lookman, T., Alexander, F.J., Rajan, K. *Information Science for Materials Discovery and Design.* Springer (2019).

[4] Chen, S. et al. "Physics-Informed Neural Networks for Composite Materials Design." *Advanced Materials* **36**, 2301234 (2024).

3 Budget Justification

3.1 Personnel (Years 1-4)

- PI Chen (1 summer month/year): $15,000 \times 4 = \$60,000$
- Co-PI Rodriguez (1 summer month/year): $15,000 \times 4 = \$60,000$
- Graduate Student 1 (computational): $45,000/\text{year} \times 4 = \$180,000$
- Graduate Student 2 (experimental): $45,000/\text{year} \times 4 = \$180,000$
- Postdoctoral Researcher: $65,000/\text{year} \times 3 = \$195,000$
- Undergraduate researchers: $5,000/\text{year} \times 4 = \$20,000$

Total Personnel: $695,000

3.2 Materials and Supplies

- Precursor materials (carbides, borides, metals): $30,000/\text{year} \times 4 = \$120,000$
- Characterization consumables: $15,000/\text{year} \times 4 = \$60,000$
- Lab supplies: $10,000/\text{year} \times 4 = \$40,000$

Total Materials: $220,000

3.3 Equipment

- Robotic liquid handling system: $150,000 (Year 1)
- Automated testing apparatus: $80,000 (Year 1)

Total Equipment: $230,000

3.4 Computational Resources

- Cloud computing (via CloudBank): $25,000/\text{year} \times 4 = \$100,000$. This will provide access to GPU instances for training deep learning models (estimated 10,000 GPU-hours per year), high-memory nodes for large-scale DFT calculations, and storage for approximately 50TB of simulation data and experimental results.
- Software licenses: $5,000/\text{year} \times 4 = \$20,000$. Includes VASP (DFT software), Materials Studio (molecular dynamics), MATLAB (data analysis), and various Python ML libraries.

Total Computational: $120,000

3.5 Travel

- Conference attendance: $10,000/\text{year} \times 4 = \$40,000$
- Collaboration visits: $5,000/\text{year} \times 4 = \$20,000$

Total Travel: $60,000

3.6 Other

- Publication costs (open access): $8,000/year × 4 = $32,000

- Workshop organization: $15,000 (Year 3)

Total Other: $47,000

3.7 Indirect Costs

Harvard negotiated rate: 69% of MTDC
Base: $1,372,000 × 0.69 = $947,000
  Total Direct Costs: $1,372,000
Total Indirect Costs: $947,000
TOTAL PROJECT COST: $2,319,000

4 Data Management Plan

This project will generate three primary data types: (1) computational predictions from DFT and machine learning models, (2) experimental synthesis and characterization results, and (3) software for materials discovery workflows.

4.1 Data Types and Formats

- **Computational data**: Crystal structures (CIF format), DFT results (VASP OUTCAR), ML training data (JSON), model checkpoints (HDF5)

- **Experimental data**: Micrographs (TIFF), diffraction patterns (xy format), mechanical test curves (CSV), processing logs (JSON)

- **Software**: Python packages, Jupyter notebooks, workflow scripts

4.2 Standards and Metadata

All data will include comprehensive metadata following Materials Science and Engineering Data Interest Group (MSEDGI) standards. We will use established ontologies (Materials Science Ontology, MatPortal) to ensure interoperability.

4.3 Access and Sharing

Data will be made publicly available through established repositories:

- Computational results → Materials Project API

- Experimental data → Figshare with CC-BY license

- Software → GitHub under MIT license, archived in Zenodo

- Large datasets → Materials Data Facility

Timeline: Data released within 6 months of collection or upon publication, whichever comes first. All data released by project end.

4.4 Long-term Preservation

Harvard's research data management services will ensure minimum 10-year retention. Critical datasets will be submitted to domain-specific repositories (NIST materials databases) for permanent archival.

4.5 Roles and Responsibilities

PI Chen will oversee computational data management. Co-PI Rodriguez will manage experimental data. The postdoctoral researcher will serve as data curator, ensuring consistent formatting, metadata, and timely deposition.