Plan Overview

A Data Management Plan created using DMPTool

DMP ID: https://doi.org/10.48321/D1HH0S

Title: Graph-Based Stochastic Learning to Build Physics-Informed AI for Discovering Defect Mechanisms in Powder-based Fabrication

Creator: Hui Wang - **ORCID:** 0000-0002-9412-3516

Affiliation: Florida State University (fsu.edu)

Contributor: Jun Zeng, Jingjing Li, Rebekah Sweat, Tarik Dickens

Funder: United States Department of Defense (DOD) (defense.gov)

Funding opportunity number: FOA-AFRL-AFOSR-2023-0009

Template: Department of Defense (DOD)

Project abstract:

Powder-based fabrication, such as binder jetting technologies using metal and ceramic powders, has emerged as a promising technology in aerospace and defense industries, enabling the creation of high-performance components with complex geometries at high productivity levels. However, a major challenge in this field is understanding the porosity formation mechanisms due to powder spread and powder-droplet interactions, which are difficult to observe directly. Traditional methods, including state-of-the-art multi-physics simulations, struggle to address this challenge effectively due to a limited understanding of fundamental material science, process physics, intrinsic uncertainty, and changes in measurements.

To overcome these limitations, the proposed research aims to integrate computational material sciences, multi-physics simulations, statistical modeling, in-situ sensing, and AI algorithms (e.g., Graphnets) to create physics-informed AI software that accelerates the quantitative discovery of porosity formation root causes. A theoretic framework will be established in this project to study combinatorial generalization problems (CG) that can empower the AI to make inferences on new scenarios given piecewise knowledge learned from multiple data sources, such as multi-physics simulation and experimental data. The research will develop theories and algorithms based on physics-informed Graphnets and generative AI to solve CG problems in three thrust areas, thereby helping discover unmeasurable defect mechanisms in powder-based fabrication. In Thrust 1, the graph-based architecture, along with multi-scale multi-physics simulation, can effectively fuse the results from multi-fidelity modeling to uncover the powder-droplet interactions with limited measurements available. Thrusts 2-3 provide essential modeling tools as building blocks and data to support CG algorithm development in Thrust 1 by conducting experiments and simulations. Through close collaboration among Florida A&M University (a historically black college/university-HBCU), HP Inc., and Pennsylvania State University, four use cases will be developed to explore the concept of CG and its applications in accelerating discoveries of defect mechanisms.

The proposed research has the potential to greatly benefit a wide range of powder-based fabrication processes (e.g., powder spray) and contribute to the development of live digital twins that can interact with physical processes in near real-time to

progressively guide experimentation and simulation, leading to new discoveries in defect formation for manufacturing engineering and material sciences. This project also features a strong educational aspect involving the interaction of all teams with graduate students from HBCU in their research efforts. HBCU graduate students will have opportunities to conduct experimental and simulation research in industrial facilities during their internships at HP.

Start date: 01-08-2024

End date: 01-07-2027

Last modified: 06-09-2023

Copyright information:

Graph-Based Stochastic Learning to Build Physics-Informed AI for Discovering Defect Mechanisms in Powder-based Fabrication

Types of data produced

The types of data, software, curriculum materials, and other materials to be produced in the course of the project that are publicly releasable.

The types of data, software, curriculum materials, and other materials to be produced in the course of the project that are publicly releasable.

The data in this project includes experimental data, simulation data, synthetic data from data models and machine learning algorithms, and associated codes and models.

Experimental data that will be collected include in-situ sensing and ex-situ characterization of microstructures from different binder jetting processes. In-situ sensing generates the image and stream video data from high-speed stereo optical cameras, high-resolution infrared cameras, and acoustic sensors. Ex-situ characterization includes images of microstructures performed by different microscopy and spatial data from micro-CT.

Simulation data that will be collected include powder spread dynamics and binder-powder interaction from FLOW-3D, LAMMPS, Siemens simulation software, HP proprietary simulation software, and other simulation software. The data from these software packages will be exported from the software and organized to be saved as spatiotemporal data in .csv or mat format or other text formats for further processing. Data sets generated using HP proprietary simulation software will be made publicly available.

Synthetic data include those from predicted outcomes or generated images from trained machine learning models. These data will be in .csv or .mat

All codes will be developed and implemented in Python, Matlab, C++, or other programming languages. Simulation models will be saved in the format of the corresponding software package. Jupyter Notebooks will be created to develop tutorials for student training and workshops. Funding will not be used to modify the HP proprietary simulation software, and the HP proprietary simulation software will not be made publicly available. These do not affect the public sharing of the data and newly developed tools in this project.

The existing data, including those in the prior publications of the project team, are insufficient for this project. Most data from this project will be newly generated. The project team will explore methods of integrating these data to build physics-informed AI based on Graphnets and the data generator.

Data and metadata standards

The standards to be used for data and metadata format and content.

The standards to be used for data and metadata format and content.

Metadata covers the date, time, methods, testing conditions, and personnel to conduct experiments, simulations, and training/running of machine learning algorithms and data models. The data will be organized under csv and txt formats or other formats, for example, JPG.. For different case studies, various metadata files will be created, including ReadMe, to explain the data structure. These data formats are chosen since they are widely accepted in engineering and data science communities for code/algorithm sharing on the cloud (e.g., GitHub, Zenodo).

Conditions for access and sharing

Conditions for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

Conditions for access and sharing include provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

All PI and co-PI will have equal access to the generated data and existing data from prior publications. University researchers (FAMU and PSU) do not have restrictions on sharing data. HP will review the codes and data to ensure that no proprietary information from HP is included in the data or codes of this project for public disclosure and access.

Conditions and provisions for reuse, redistribution, and derivatives

Conditions and provisions for reuse, redistribution, and the creation of derivative works.

Conditions and provisions for reuse, redistribution, and the creation of derivative works:

The PI/co-PIs and their affiliated institutions will own the copyright and intellectual properties of the corresponding tools and data generated. The university researchers have no particular restrictions on sharing the tools, code, and data generated by them. The data can be made available by journal/conference paper publication or upon request. HP will review the codes and data to ensure that no proprietary information from HP is included in the data or codes of this project for public disclosure and access.

Plans for archiving and preservation

Plans for archiving datasets, or data samples, and other digitally formatted scientific data, and for preservation of access thereto. Explicitly describe how the data that underlies scientific publications will be available for discovery, retrieval, and analysis. In accordance with OSTP Memorandum, digitally formatted scientific data resulting from unclassified, publicly releasable research supported wholly or in part by DoD funding should be stored and publicly accessible to search, retrieve, and analyze to the extent feasible and consistent with applicable law and policy; agency mission; resource constraints; and U.S. national, homeland, and economic security.

During project periods, all data will first be stored in each PI's storage space. The PI has RAID5 hard disk array to store these data in the lab. These data will then be selected to upload onto a cloud drive with accessibility to all team members, such as the PI's OneDrive with 6TB storage space or Dropbox enterprise version. Meanwhile, data will also be stored on portable SSD drives for data transfer and exchange.

For long-term archival, the cleaned data and codes will be selected to be included in conference/journal publications, GitHub and Zenodo. The files will also be uploaded to DigiNole for storage and distribution. DigiNole, Florida State University’s institutional repository, will be used to provide public access to the final de-identified dataset(s) necessary to reproduce the results presented in journal articles. DigiNole is a digital dissemination and archiving platform for all forms of scholarly research outputs. Researchers utilize DigiNole for public/open access compliance and to increase the access and impact of various products of original research by, including publications and data. Built on the open-source software Islandora, DigiNole is a highly flexible platform for storing, accessing, and managing digital assets, and includes a binary module that facilitates the long-term storage and download of various file formats. In addition, DigiNole is supported by a team of librarians with expertise in digital asset management and preservation. This team will be available to consult on data curation and descriptive metadata throughout the project to ensure optimal accessibility, discoverability, and reusability.

Justification for the restriction of data

If, for legitimate reasons, the data cannot be preserved and made available for public access, the plan will include a justification citing such reasons.

Planned Research Outputs

Dataset - "Multi-scale multi-physics data from binder jetting "

The dataset will include experimental and simulation data collected from binder jetting for metal and ceramics powders. The data will be organized to correlate properties, microstructures, and defect distributions to corresponding process conditions. The dataset will be valuable for machine learning and generalization in various applications.

Software - "Graphnet-based learning tool for combinatorial generalization"

This software package will include codes based on the algorithms/models developed in this project. It has the function of creating Graphnet for prediction of defects under new conditions or scenarios by leveraging piecewise knoweldge/models obtained from multiple data sources.

Data paper - "Journal/conference papers on solving combinatorial generalization problems by developing physics-informed AI"

Multiple conference and journal papers in the field of Industrial Engineering, Material Science, Computer and Information Sciences, and Manufacturing Engineering will be generated and submitted for publication upon approval by HP and AFOSR.

Model representation - "Multi-scale multi-physics simulation models for powder-droplet interaction in binder jetting"

Through application programming interface (API), python-based user interfaces will be available to coordinate the simulaiton runs and exchange of results among different simulation packages to create a pipeline of simulation workflow.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Multi-scale multi-physics data from binder jetting ...	Dataset	2027-06-29	Open	None specified		Custom Data Use Agreements/Terms of Use	None specified	No	No
Graphnet-based learning tool for combinatorial gen ...	Software	2027-12-30	Open	None specified		Custom Data Use Agreements/Terms of Use	None specified	No	No
Journal/conference papers on solving combinatorial ...	Data paper	Unspecified	Open	None specified		None specified	None specified	No	No
Multi-scale multi-physics simulation models for po ...	Model representation	2027-12-30	Open	None specified		None specified	None specified	No	No