OAK RIDGE National Laboratory

AI Initiative at ORNL

David Womble, Pradeep Ramuhalli,
Frank Liu, Dan Lu

June 23, 2022

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

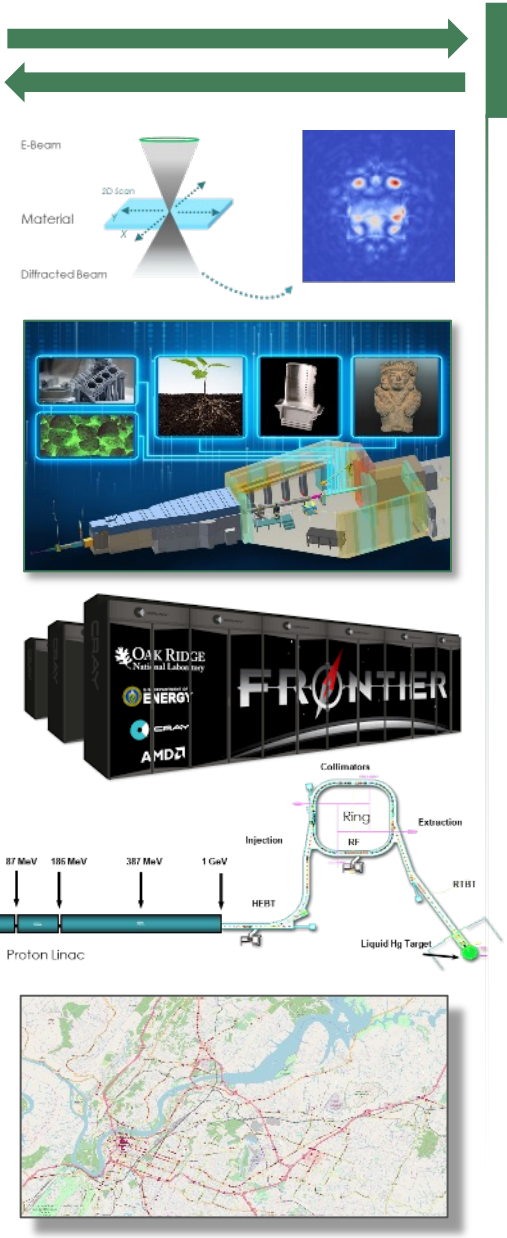
AI and the need for "foundational research"

- Best analogy is "traditional" modeling and simulation and HPC
  - Computational science, informed by math and C.S. developed as a discipline in response to the complexity of the programming model and environment, the application dependence of the algorithms and the need for robustness
- Foundational AI/ML is hard, arguably harder than traditional mod-sim
  - Driven by abstract form of the model and non-deterministic nature of algorithms
  - Difficulties with data
  - Lots of libraries and open-source software but very little focus on correct/robust use
- A complete capability includes applications, foundations and a hardware/software "ecosystem."

AI is Pervasive in ORNL Mission Applications

Notable Use Cases

- Science
  - Materials analytics and design
  - Neutron science
  - Bioscience and medicine
- Smart instruments and facilities
  - Manufacturing
  - Smart laboratories
- Scientific facilities operations
  - SNS, OLCF, HFIR
- Energy systems
  - Nuclear engineering and other energy generation systems
  - Energy distribution
- Engineering operations
  - Transportation
  - Building energy
- National security
  - Cybersecurity
  - Nonproliferation
  - Geospatial analysis
- Climate science / earth systems



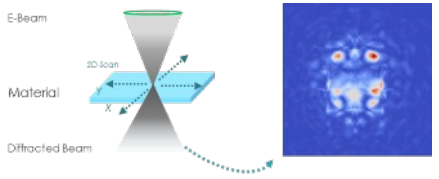
Translated to "foundational" R&D

- Physics-informed ML
- Multiscale surrogates
- Design, e.g., of
  - experiments
  - new materials, enzymes, catalysts, etc.
  - processes
- Data reduction
- Predictions
  - Time series, functional
- Data collection
  - Sensor placement
- Model form (e.g., GNNs)
- Learning
  - Continuous with aging data
  - Distributed/federated
- Ethics and privacy
- Validation
  - Data
  - Model
- Reproducible/replicable
- Interpretable/explainable
- Robustness and resilience
- Causal analysis
- Uncertainty quantification
- Anomaly detection
- Control
  - Distributed
  - Robust and resilient
- Co-design w/HW
- Scalable training/inference

Artificial Intelligence Initiative

Challenge

- Develop foundational AI capabilities supporting the DOE and ORNL science mission and laboratory of the future
  - Enable scientific discovery through the application of AI/ML
  - Research foundational algorithms for data reduction, design of experiment, data analysis and next-generation hardware
  - Develop algorithms and approaches using exascale computing for AI/ML
  - Enable autonomous scientific laboratories
  - Crosscut ORNL applications



Approach

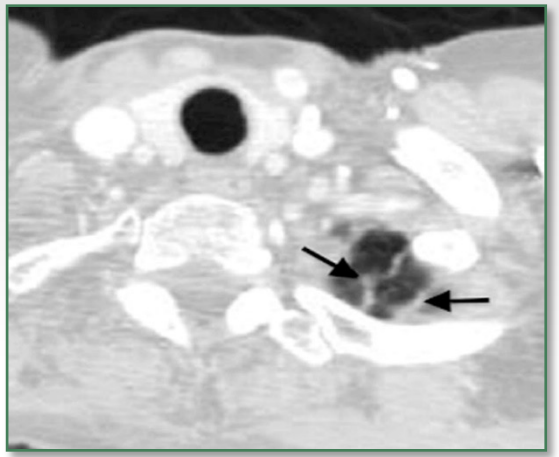
- Invest in two thrusts and two crosscuts
  - Robust engineering and science (AIRES) including digital twins and smart facilities
  - Science Discovery (AISD) to develop new materials and processes
  - Assurance for AI models and use
  - Exploratory projects
- Themes: "science-informed," "exascale," and "edge"

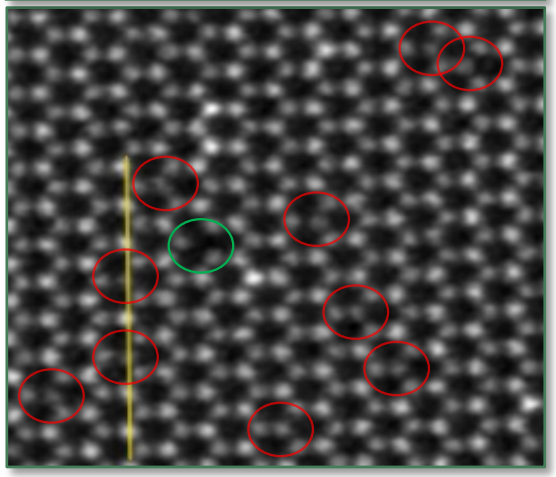
Expected Outcomes

- Scalable algorithms for physics-informed AI/ML
- New algorithms for scientific discovery, including design of experiment
- Efficient ML-based data analysis and reinforcement learning for smart facilities
- Assured AI, including UQ and validation

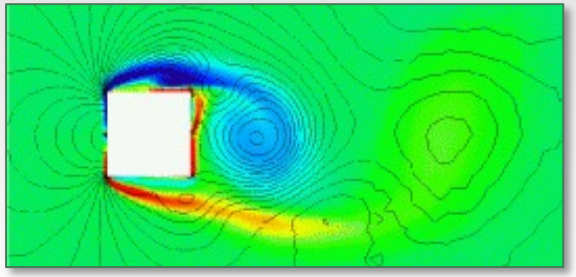
A Taxonomy of AI Use Cases

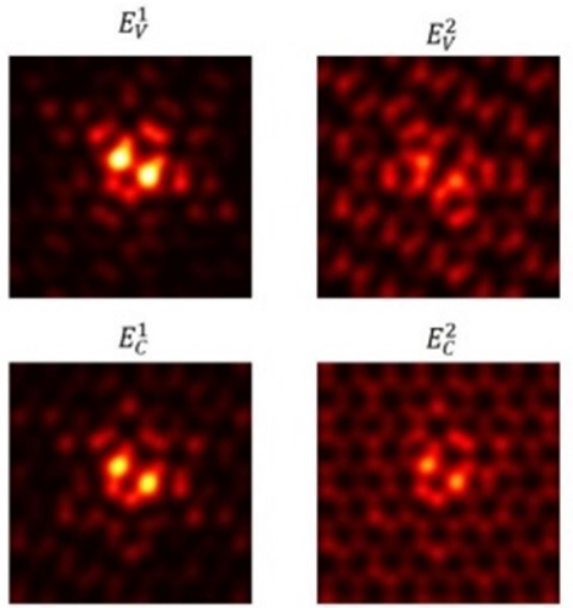
Classification and regression



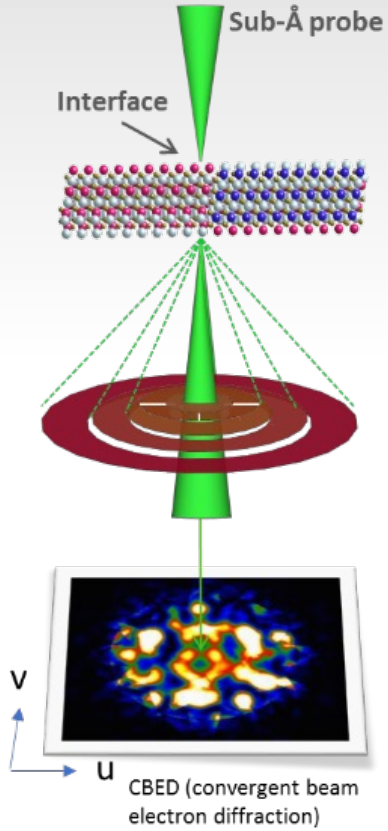


Surrogates

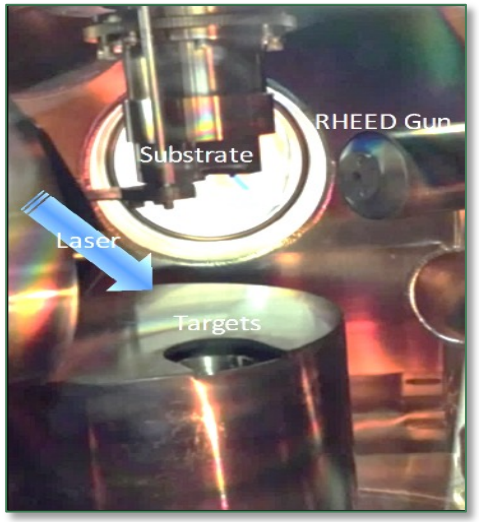




Inverse problems, design and optimization



Control systems

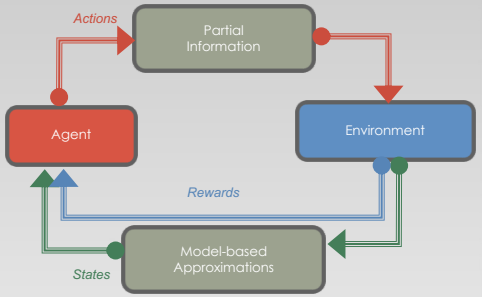


ORNL Strategic Directions and in AI/ML

Data

- Facilities operation and control
- Experimental design
- Data curation and validation
- Compressed sensing

Learning

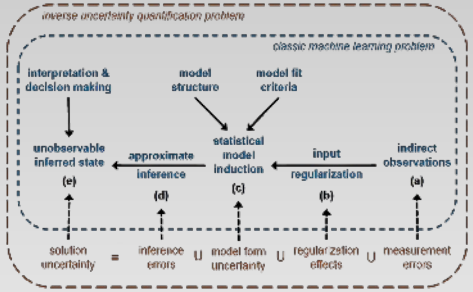


- Physics informed
- Accelerating learning
- Stability and robustness
- Foundations of ML formulations - RL, GANs, GNNs, BNNs
- Dimension reduction and encoding

Scalability

- Algorithms, complexity and convergence
- Levels of parallelization
- Mixed precision arithmetic
- Communication
- Implementations on accelerated-node hardware

Assurance



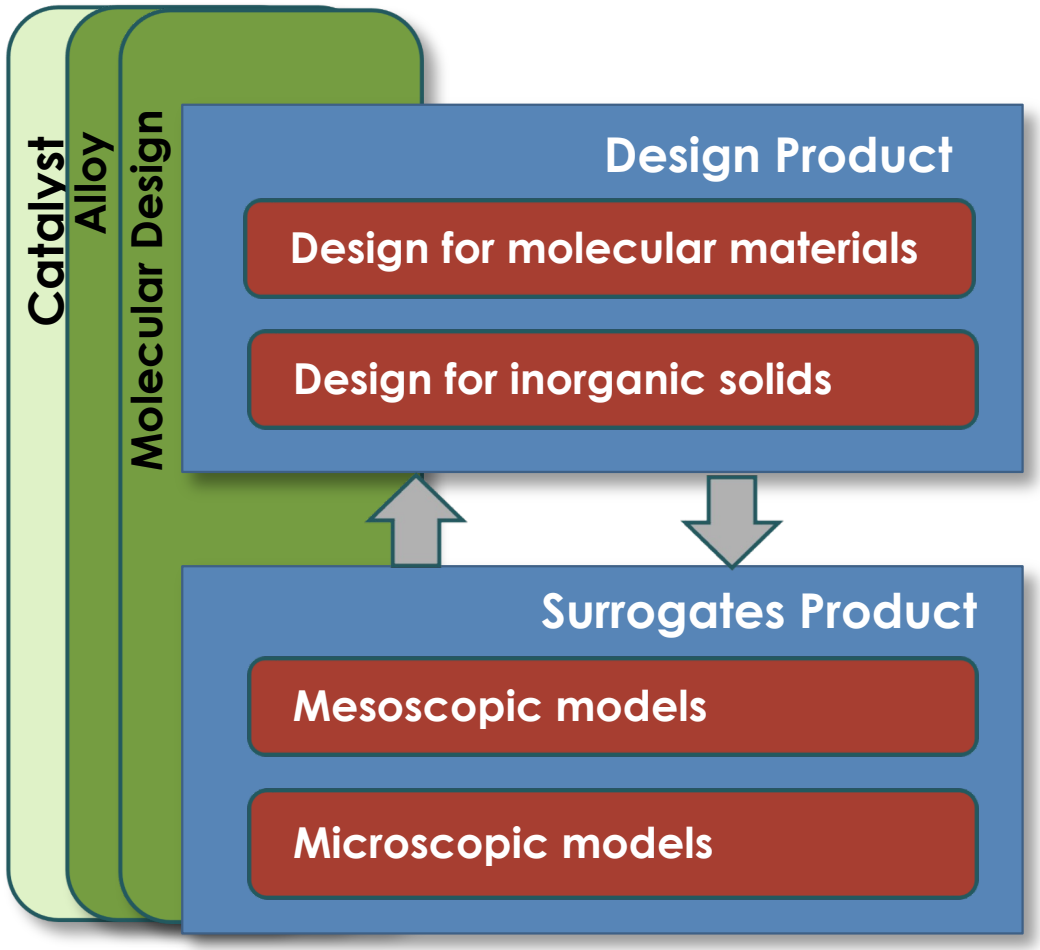
- Uncertainty quantification
- Robustness
- Explainability and interpretability
- Validation and verification
- Causal inference and hypothesis generation

Workflow

- Edge AI
- Compression
- Online learning
- Federated learning
- Infrastructure
- Augmented intelligence and HCI

AISD: Goal and Organization

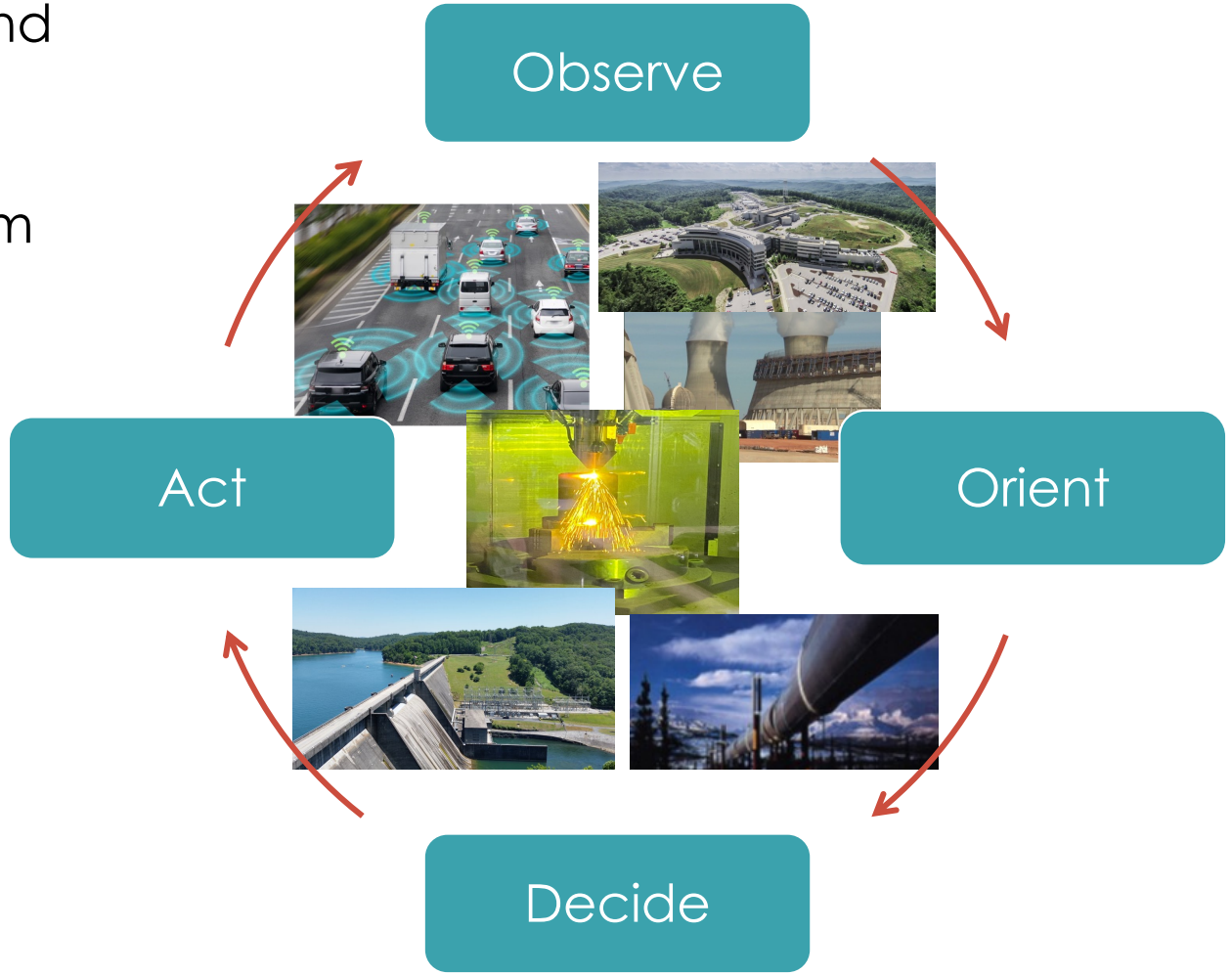
- The overarching objective of the AISD thrust is to **develop AI-based design capabilities for new materials and processes** with targeted applications in chemistry, biology and materials science



- Main AI/ML challenges:
  - Physics-informed, multiscale surrogates
    - **Training of transferable models**
    - **Interpretability**
    - **Data reduction, including feature selection and clustering**
  - Design of experiment and scientific discovery
    - **Optimization and design of experiment**
    - **Robust and stable generative models**
    - Efficient, physics-informed RL algorithms
    - Pareto-optimal decision in high dimensional design space
  - Crosscutting
    - **Uncertainty quantification**
    - Causality and validation
    - **Computational scalability (OLCF, CADES)**

AIRES: Develop and Demonstrate AI Methods for Robust Decisions in Complex Dynamic Systems

- High fidelity data-driven, knowledge-informed representations that accurately represent and predict system behavior at relevant spatiotemporal scales
- AI algorithms for detecting changes in system behavior or configuration
- Algorithms for continual learning
- Control/decision approaches that adapt to changing system conditions and achieve autonomy
- Algorithms that are edge-deployable and scalable



Assurance crosscut

Goal: enable an optimal, robust, transparent, and safe application of AI in lab-wise scientific and engineering problems.

Four components/products

- **Uncertainty Quantification (UQ)**
  - Assess model prediction's credibility;
  - Identify data/domain shift;
  - Guide data collection & inform decision.
- **Verification & Validation**
  - Ensure data has information to build model;
  - Ensure the model acts as expected.
- **Explainability & Interpretability**
  - Assure the model's physical consistency;
  - Explain feature importance.
- **Causal Analysis**
  - Explore knowledge and interactions;
  - Advance scientific understanding.

Three research foci

- Develop fundamental methods;
- Support AISD and AIRES projects in the AI initiative;
- Apply the methods to scientific problems.

Assurance	FY22	FY23	FY24	FY25
UQ				
Validation				
Interpretability				
Causal Analysis				

OAK RIDGE National Laboratory

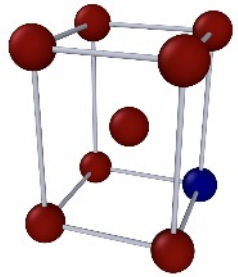
Example Foundational AI Projects

HydraGNN: scalable multi-task graph convolutional neural network for prediction of multi-scale material properties from atomic information

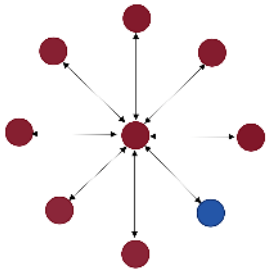
Challenges:

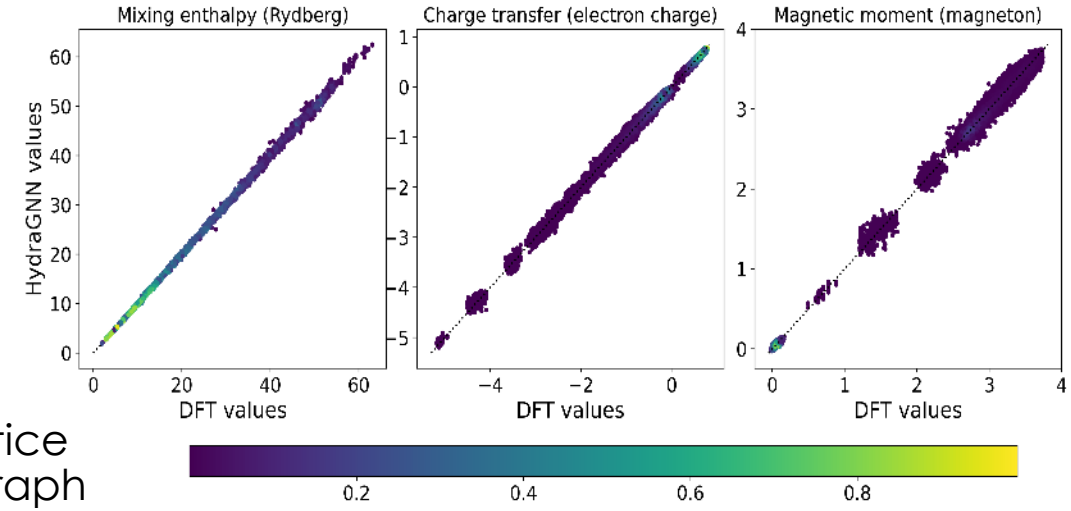
- High quality atomic modeling data is expensive to collect, and thus is accessible only in small volumes
- graph convolutional neural networks take advantage of data structure, but are prone to overfit on low volume of data

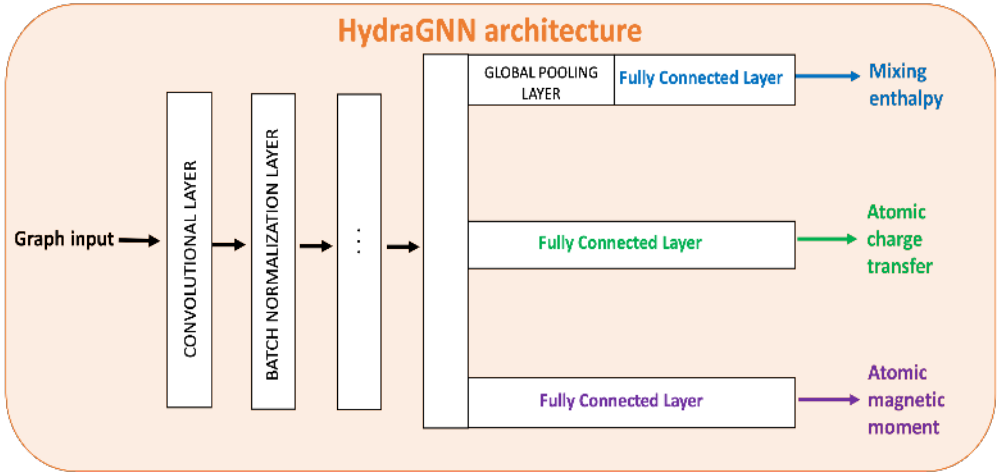
Solution: exploit physics correlation between material properties to stabilize the training via multi-task learning



Conversion of lattice structure into a graph







https://www.osti.gov/doecode/biblio/65891

OAK RIDGE National Laboratory

PI: Max Lupo Pasini

Scan Path Planning in Additive Manufacturing using Deep RL

Challenges: Cannot be dealt with using current methodologies

1. Incredibly large, inconsistently-sized action space
2. Policy is complex (human performance, including algorithms, remains low)

Solution to #1: Novel methodology for tiled MDPs with inherent locality.

1. Decomposes global MDP into local MDPs
2. Learns Q-value function for local MDPs
3. Composes global policy from conglomeration of local MDPs

Solution to #2: Novel model-free policy algorithm called *FlowSAC*

1. Leveraged soft-actor critic (SAC) for better policy exploration
2. Leveraged normal flows to allow a multimodal, nonGaussian policy
3. Combines these to algorithms for efficient exploration of an arbitrarily complex policy

Next phase of work is to combine these algorithms and apply them to the AM problem

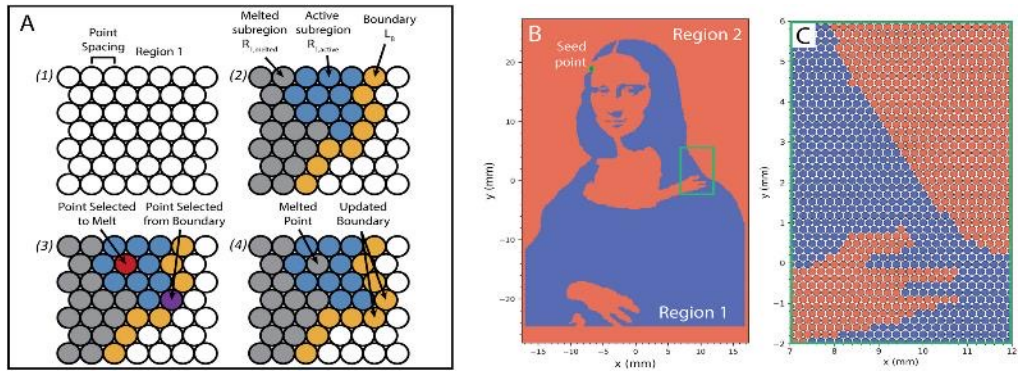


Figure 1: Tide Fill Scan Strategy and Resulting Microstructure
[Plotkowski, A., et al. "A Stochastic Scan Strategy for Grain Structure Control in Complex Geometries using Electron Beam Powder Bed Fusion." Additive Manufacturing (2021): 102092.]

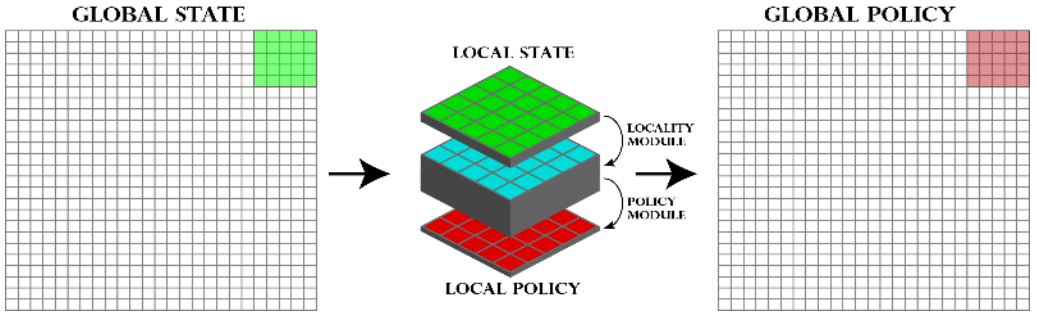


Figure 2: Workflow for Constructing an Arbitrarily Large Global Policy from Decomposed Local States

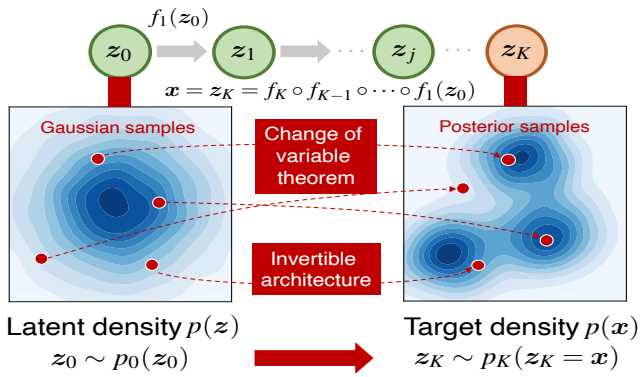


Figure 3: Normal Flows Result in Richer Flexibility for Policy Capture

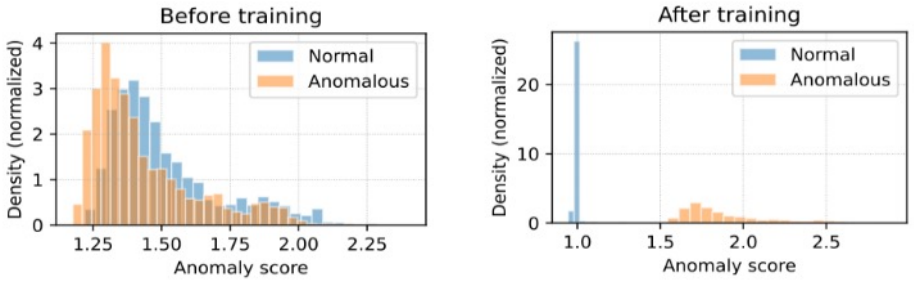
OAK RIDGE National Laboratory

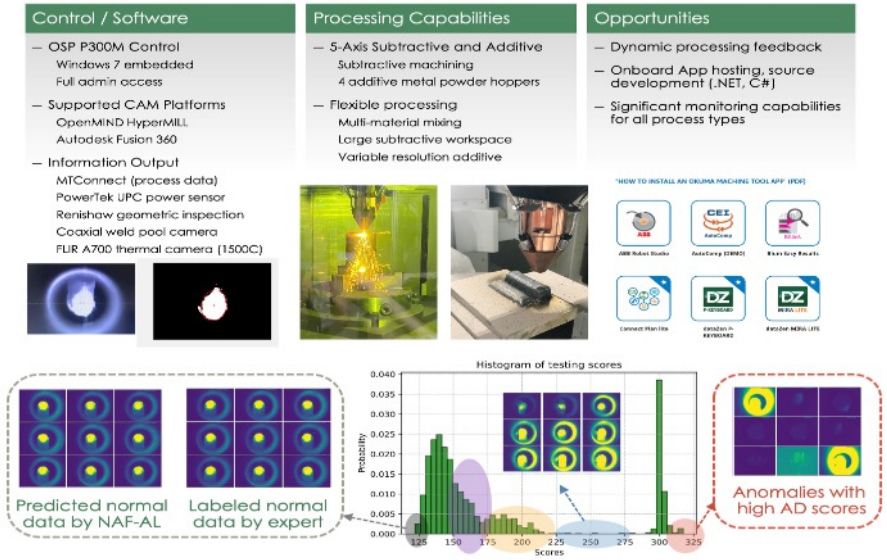
PI: Jamie Stump

Self-Supervised Anomaly Detection via Normalizing Flows

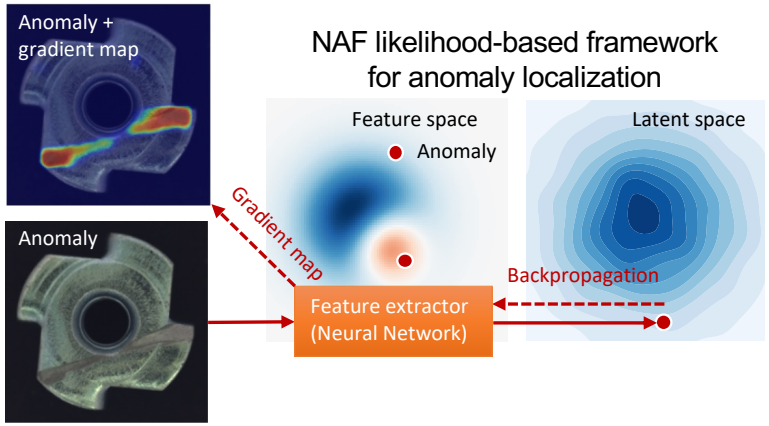
We developed a novel self-supervised anomaly detection approach based on normalizing flows for determining processing parameters and defects region in additive manufacturing.

Applied to detect anomalies of the meltpool images in AM, and allows automatically labeling data, which improve the working efficiency.

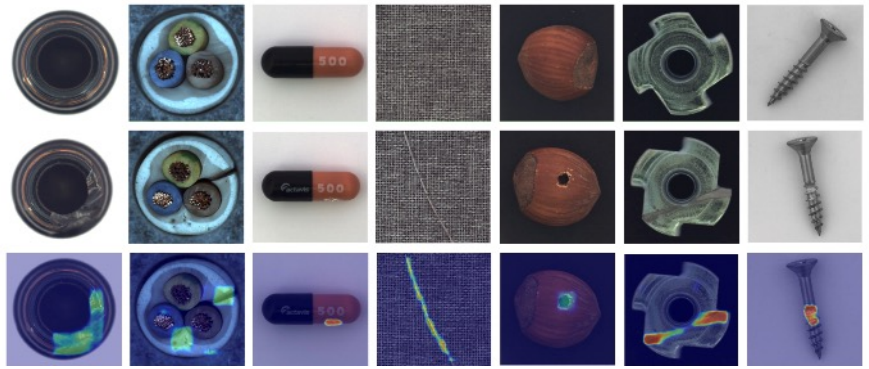




The feature extracted NN identifies the anomalous regions by backpropagating the likelihood loss up to the input image which yields a gradient map. This allows detailed analysis of defect/anomaly positions and shape – anomaly localization



MVTec AD dataset (www.mvtec.com, CVPR 2019, 2021)



OAK RIDGE National Laboratory

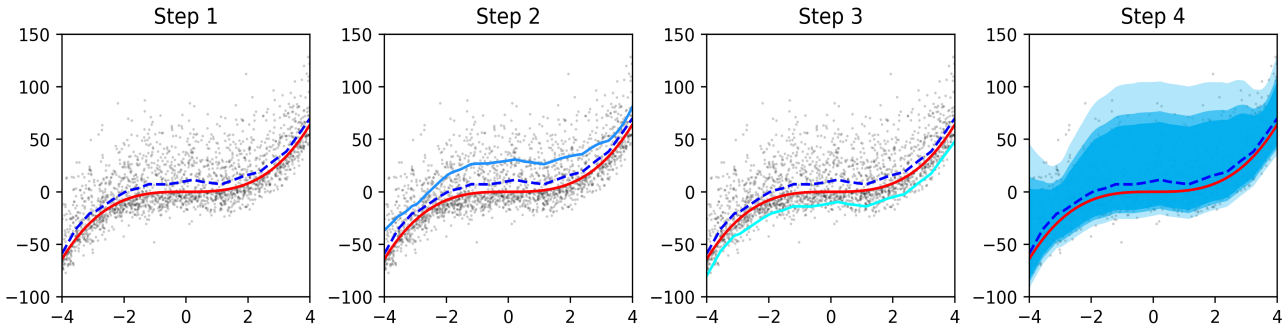
J. Zhang, K. Saleeby, T. Feldhausen, S. Bi, A. Plotkowski, and D. Womble. "Self-Supervised Anomaly Detection via Neural Autoregressive Flows with Active Learning." *NeurIPS 2021 Deep Generative Models and Downstream Applications.* [PDF file]

PI: Jiaxin Zhang

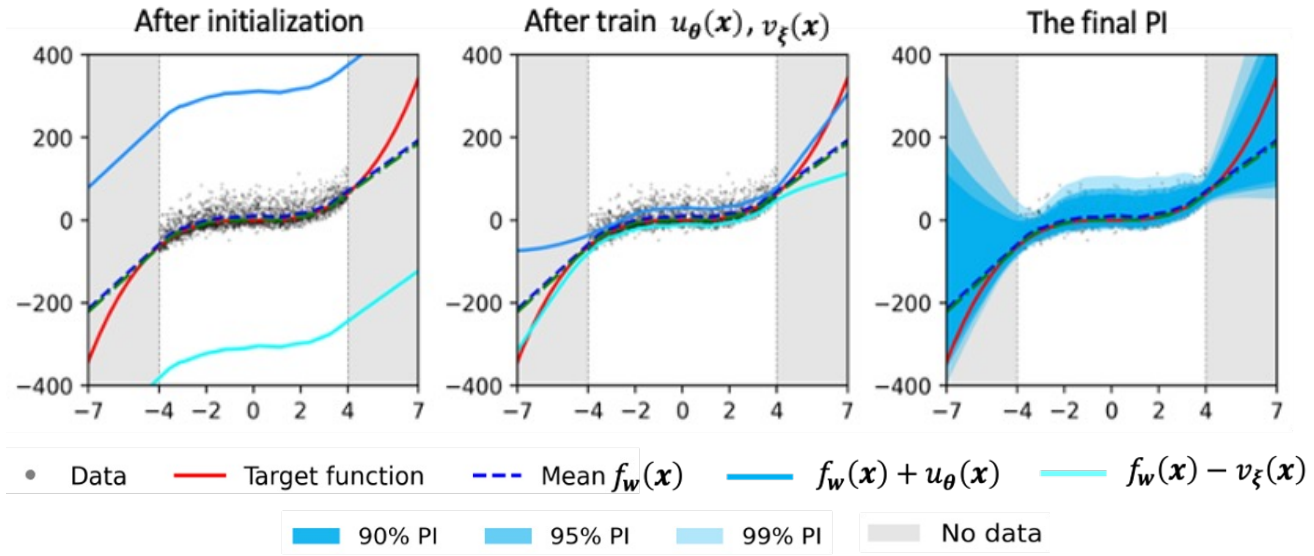
PI3NN produces well-calibrated UQ and identifies OOD data

OOD: out of distribution

Illustrate the main PI3NN algorithm



Illustrate the novel OOD identification feature



- For in-distribution data, PI3NN produces a well-calibrated UQ from 3 NNs training.
- PI3NN is computationally efficient, reliable and can be combined with different network structures to produce credible ML model predictions.
- For OOD data identification, we proposed a novel initialization scheme to enable PI3NN producing an increasing uncertainty bound as the data is further away from training set.
- OOD identification can be used for data collection and model improvement.

OAK RIDGE National Laboratory

PI: Dan Lu

Ultra Low-Latency ML Solutions

Background

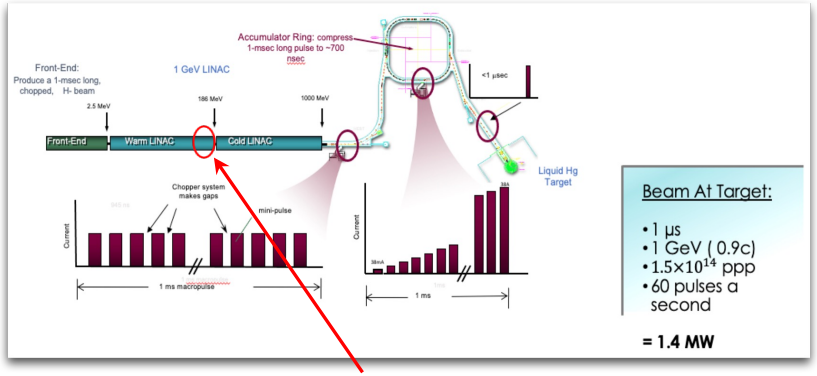
- Fast machine learning inference (~1 μs) is highly desirable for edge deployment
- ML implementations on CPU/GPU usually have latencies of milliseconds due to communication time

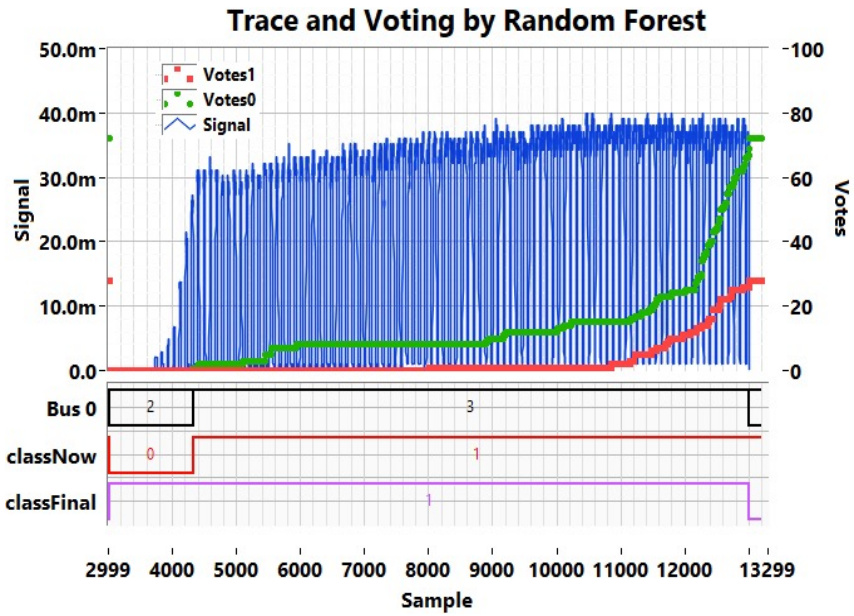
Highlights

- Development of a novel out-of-order inference method for complex machine learning models (Random Forest, 8,000 nodes with 10,000 input features)
- Development of a machine learning inference kernel on FPGA for errant beam prediction
- Fully integrated at SNS edge infrastructure. Demonstrated 60 ns inference latency (after the last signal is ingested).

Reference

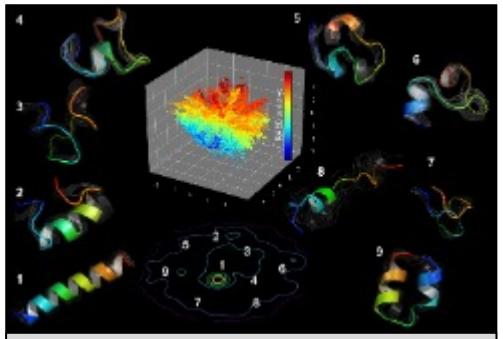
- Narasinga Miniskar, Aaron Young, Frank Liu, Willem Blokland, Anthony Cabrera and Jeff Vetter, "Ultra low latency machine learning for scientific edge applications", International Conference on Field Programmable Logic and Applications, 2022



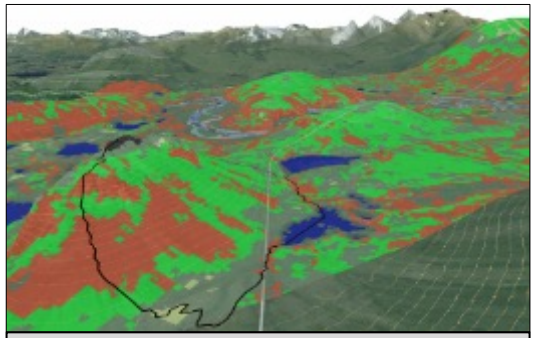


benchmark	# feature	# tree(max depth)	# node (internal/terminal)	hls4ml(Conifer) LUT	hls4ml(Conifer) FF	ours LUT	ours FF
Hastie	10	20 (3)	140 / 160	38,151	1,242	1,043	12
IRIS	4	20 (7)	165 / 185	**failed**	**failed**	1,402	37
BreastCancer	30	100 (20)	1994 / 2094	**failed**	**failed**	6,504	271
Olivetti	4096	100 (20)	6074 / 6174	**failed**	**failed**	52,656	8,513
SNS	10000	100 (20)	3978 / 4078	**failed**	**failed**	35,558	4,559

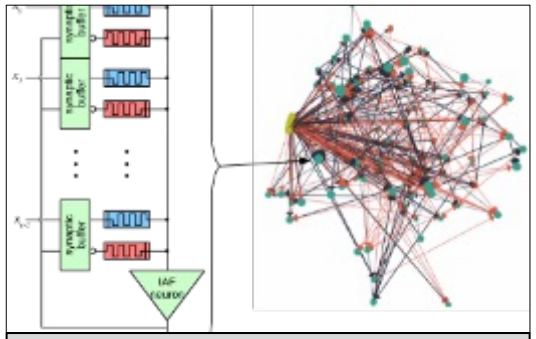
A few more examples of AI at ORNL



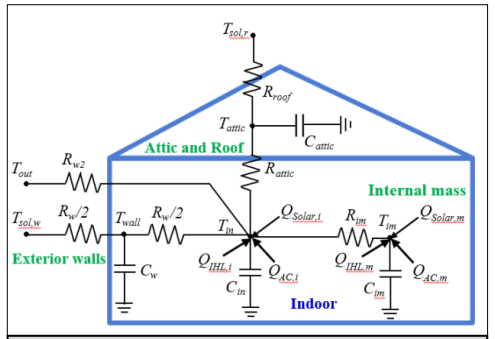
Atomistic and molecular-scale modeling



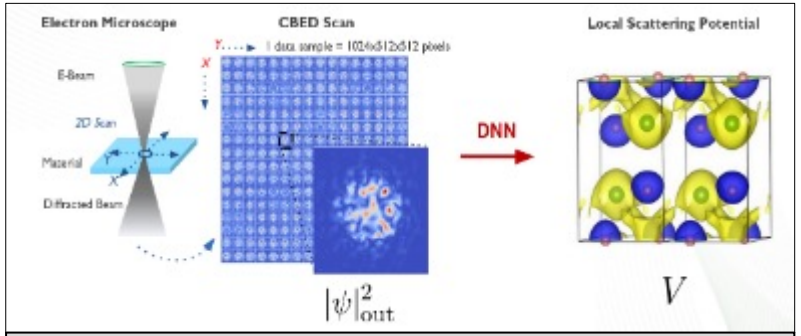
Arctic vegetation mapping



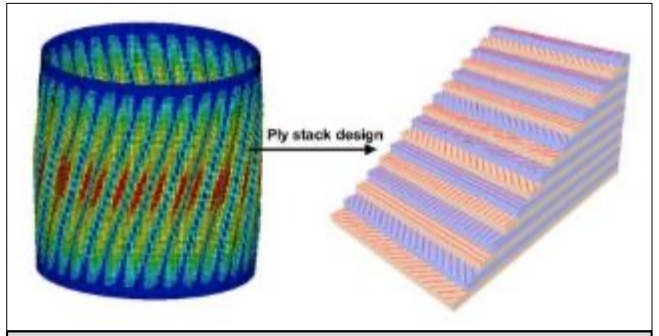
Quantum computing



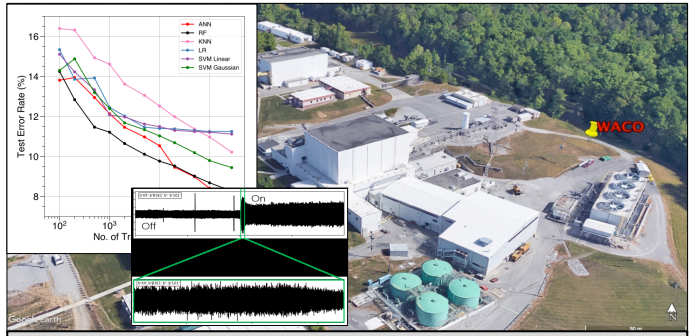
Building energy efficiency



Inverse problems in materials



Design of composite structures



Seismic monitoring