COVER PAGE

Project Title: Harnessing Nonnegative Matrix Factorization for Advanced Computational Materials Modeling	
Federal Award Identification Number: DE-SC0025801	
Agency Code: 8900	**Organization:** Office of Advanced Scientific Computing Research
Recipient Award Identification Number: DE-SC0025801	**Project Period:** 12/01/2024 - 11/30/2027
Reporting Period: 12/01/2024 - 11/30/2025	**Budget Period:** 12/01/2024 - 11/30/2025
Report Term: Once per Budget Period	**Submission Date and Time:** 09/17/2025 10:37 AM ET
Principal Investigator Information: Prof. Chinedu E Ekuma	**Recipient Organization:** Lehigh University 27 Memorial Drive West Bethlehem, PA 18015-3093 Country: USA **UEI:** E13MDBKHLDB5

ACCOMPLISHMENTS

1. Major Goals and Objectives

The overarching goal of this project is to **develop novel scientific machine learning (SciML) algorithms** that enhance data reduction, increase interpretability, and improve predictive performance in materials modeling. Specifically, the project aims to:

1. **Design physics-aware NMF models** that incorporate structure-aware regularization and domain knowledge (e.g., crystal symmetries).
2. **Develop scalable ML models** by integrating NMF into deep learning pipelines for material property prediction and inverse design.
3. **Integrate NMF with diffusion models (DMs)** to address data scarcity and advance high-throughput/inverse materials discovery.
4. **Build an open-source distributed platform**, with Docker-ready deployments, enabling broad community access to NMF-based algorithms.

These aims align with the DOE mission of advancing **energy efficiency, environmental sustainability, and scientific discovery**.

2. Accomplishments Toward These Goals

Significant progress has been achieved across all four aims:

- **Publications and Presentations**
  - Published in *J. Appl. Phys.* (2025): *Defect-induced bipolar magnetism in atomically thin GeS*.
  - Published in *Appl. Phys. Lett.* (2025): *Strain-tunable magnetism in hydrogen-functionalized WSe2 monolayer*.
  - Published at *ICML 2025*: *Improving Flow Matching by Aligning Flow Divergence*.
  - Published at *ICLR 2025*: *A Theoretically-Principled Sparse, Connected, and Rigid Graph Representation of Molecules*.
  - Presentations delivered at the upcoming NSBP/NSHP meetings (2025), and Carnegie Mellon Diaspora Fellowship (Nigeria).
- **Algorithmic Development**
  - Designed hierarchical NMF integrated with graph neural networks (GNNs), applied to both materials and brain network analysis.
  - Developed divergence-matching framework for flow-based generative models, offering superior fidelity and computational efficiency.
  - Built pipelines for graph-based learning on perovskite-derived and heterostructure datasets, incorporating uncertainty quantification and active design.
- **Collaborative Sub-Projects**
  - *Feature Identification*: Improved interpretability in large-scale datasets using NMF.
  - *Hierarchical NMF + GNNs*: Initial prototypes benchmarked with collaborators in Belgium.
  - *Flow-based Models*: Established cross-institutional collaborations for multimodal generative modeling.
- **Workforce Development**
  - Supported one REU (Arthur Lian), one PhD student (Anthony Iloanya), and an undergraduate (Thaksheel Alleck, DOE-funded, Summer 2025).
  - Additional personnel: Songlin Zhao (PhD), Jiaxin Chen (MS rotation), Joel Keller (Utah, supported 8 months).

3. Technical Approach

The project employs a **synergistic, multi-pronged approach**:

- **Physics-Aware NMF**: Developed with time-frequency domain regularization, informed by physical principles, ensuring interpretable and identifiable outcomes.
- **Hybrid Models**: Integration of NMF with deep learning and DMs to leverage scalability, handle data scarcity, and provide robust predictive capabilities.
- **Uncertainty Quantification**: Bayesian inference frameworks applied to NMF posteriors, supporting reproducible and reliable scientific modeling.
- **Graph-Based Learning**: Use of GNNs to encode periodic crystal structures, benchmarked against traditional ML baselines.
- **Distributed Infrastructure**: Development of Dockerized environments for broad deployment across platforms.

4. Progress Toward Major Activities

Key outcomes and achievements:

- **Algorithmic Innovations**:
  - Randomized deep NMF algorithms for efficient data representation.
  - SCHull graph representation ensuring sparsity, connectivity, and rigidity in molecular modeling.
  - Divergence-matching generative modeling framework improving fidelity across scientific tasks.
- **Materials Science Results**:
  - Demonstrated tunable magnetism in 2D materials (GeS, WSe_2) via hydrogen functionalization and strain engineering.
  - Identified pathways for defect- and strain-controlled spintronic properties in atomically thin materials.
- **Software & Datasets**:
  - Constructed pipelines for GNN-based property prediction across perovskite and heterostructure datasets.
  - Established reproducible computing configurations on NSF-MRI funded GPU infrastructure.

5. Plans for Next Reporting Period

Looking ahead, the project will:

- Finalize experiments on feature identification and submit a manuscript (Fall 2025).
- Optimize hierarchical NMF + GNN algorithms and apply them to brain network and materials datasets.
- Extend flow-based generative models, integrating probabilistic methods with NMF frameworks.
- Scale uncertainty quantification techniques (conformal prediction, ensembles).
- Release cleaned datasets, baselines, and software scripts for community use.
- Continue mentoring undergraduate and graduate students, with emphasis on diversity in computational science.

6. Training and Professional Development Opportunities

The project has provided:

- **Undergraduate Research Experiences (REU)**: Hands-on training for students in machine learning, materials modeling, and computational physics.
- **Graduate Training**: Advanced algorithm development in NMF, GNNs, and flow models.
- **Collaborative Mentorship**: Regular cross-team syncs and supervision, fostering collaboration across institutions (Lehigh, Utah, Mons, Auburn, Harvey Mudd).
- **Conference Exposure**: Opportunities for students and early-career researchers to present at international conferences (APS, NSBP/NSHP, SIAM CSE).

7. Dissemination of Results

Results have been shared across multiple communities:

- **Scholarly Publications**: Articles in leading journals (*J. Appl. Phys.*, *Appl. Phys. Lett.*) and top-tier conferences (*ICML 2025*, *ICLR 2025*).
- **Presentations**: Delivered at APS, NSBP/NSHP, ICLR conference, CMU Diaspora Fellowship, and internal seminars at partner institutions.
- **Workshops & Community Engagement**: Planned dissemination at SIAM CSE 2025 and DOE Computational Materials Workshops.
- **Open-Source Infrastructure**: Forthcoming releases of datasets, Docker images, and reproducible baselines for wide adoption.

PRODUCTS

PUBLICATIONS

Conference Paper/Presentation:
Yuhao Huang, Taos Transue, Shih-Hsin Wang, William M Feldman, Hong Zhang, Bao Wang, "Improving Flow Matching by Aligning Flow Divergence", *Forty-second International Conference on Machine Learning (ICML 2025)*, Published, 2025.

Conference Paper/Presentation:
Shih-Hsin Wang, Yuhao Huang, Justin M. Baker, Yuan-En Sun, Qi Tang, Bao Wang, "A Theoretically-Principled Sparse, Connected, and Rigid Graph Representation of Molecules", *Forty-second International Conference on Machine Learning (ICML 2025)*, Published, 2025.

Journal Article:
Tianhao Lian,Anthony C. Iloanya, and Chinedu E. Ekuma, "Strain-tunable magnetism in hydrogen functionalized WSe2 monolayer", *Appl. Phys. Lett*, Volume 126, 161901, Published, 2025, DOI: https://doi.org/10.1063/5.0269136.

Journal Article:
Tianhao Lian, Anthony C. Iloanya, Srihari M. Kastuar, Gour Jana, and Chinedu E. Ekuma, "Defect-induced bipolar magnetism in atomically thin GeS", *J. Appl. Phys.*, Volume 137, 233901, Published, DOI: https://doi.org/10.1063/5.0268105.

INTELLECTUAL PROPERTIES

There are no intellectual properties to report.

PARTICIPANTS AND OTHER COLLABORATING ORGANIZATIONS

The table below only contains participants who have identified an affiliation with the Awardee Institution. Participants from any associated subawards may not be included in this count.

PARTICIPANTS DETAIL

Project Role	Number of People	Total Person Months Worked
Co-Investigator	3	3
Graduate Student (Research Assistant)	4	25
Principal Investigator/Project Director	1	1
Undergraduate Student	2	11
Total Responses	**10**	**40**

PARTNERS DETAIL

There are no partners to report.

IMPACT

- This project has advanced the field of **computational materials science** by developing new machine learning algorithms that are more interpretable, efficient, and scalable than existing methods. These tools help scientists understand the relationship between material structures and their properties, which is essential for designing next-generation materials for **energy efficiency, clean technologies, and advanced manufacturing**.
  Impact on the principal discipline: The project has introduced physics-informed nonnegative matrix factorization (NMF) and new approaches combining NMF with generative modeling. These innovations provide powerful methods for analyzing large and complex datasets in materials science, making scientific models more reliable and interpretable.
  Impact on other disciplines: The algorithms and modeling frameworks developed are broadly applicable beyond materials science. For example, they can be used in areas such as **climate modeling, healthcare analytics, and molecular biology**, where large-scale data analysis and interpretable models are critical.
  Impact on infrastructure: The team has developed **reproducible computing pipelines** and is preparing open-source software, datasets, and containerized tools (Docker). These resources will strengthen U.S. research infrastructure by making advanced methods broadly available to universities, national labs, and industry.
  Impact on technology transfer: While the project is still in the research stage, its results lay the foundation for future technology transfer in **energy materials discovery, electronic devices, and advanced manufacturing**. The open-source release of methods and data will make it easier for government, industry, and academic groups to adopt these approaches in real-world applications.
  Impact on society: The research supports **clean energy innovation, sustainable materials design, and national competitiveness in advanced technologies**, directly benefiting the public by enabling more efficient use of resources and accelerating discovery of new materials for energy, electronics, and environmental applications.
  Impact on human resources: The project has trained multiple undergraduate and graduate students in computational science, artificial intelligence, and materials modeling, thereby strengthening the nation's workforce in high-demand areas. These students have gained valuable professional development through research training, conference presentations, and publications.
  International spending: Less than 1**% of the award budget** was spent outside the United States, primarily to support conference presentations in Canada. The vast majority of the investment directly supported U.S.-based research and workforce training.