U.S. DEPARTMENT OF ENERGY

Office of ENERGY EFFICIENCY & RENEWABLE ENERGY

Technical Workshop to Inform a

# Multi-Agency, Multi-Year Program Plan in Advanced Energy Materials Discovery, Development, and Process Design

## Utilizing High-Throughput Experimental Methods, Artificial Intelligence, Autonomous Systems, and a Collaboratory Network

Co-sponsored by the U.S. Department of Energy and the National Institute of Standards and Technology

**Workshop Summary Report**

July 12, 2018

Gaithersburg, MD

This document was prepared for the U.S. Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE) Advanced Manufacturing Office (AMO) and the National Institute of Standards and Technology (NIST) Material Measurement Laboratory (MML) as a collaborative effort by AMO, NIST, Allegheny Science & Technology, and Energetics.

## Disclaimer

# List of Acronyms

| | |
|---|---|
| AI | Artificial Intelligence |
| Al | Aluminum |
| AM | Additive Manufacturing |
| AMO | Advanced Manufacturing Office |
| API | Application Program Interface |
| AS | Autonomous Systems |
| Btu | British Thermal Unit |
| DBA | Database Administrators |
| DFT | Density Functional Theory |
| DOD | U.S. Department of Defense |
| DOE | U.S. Department of Energy |
| EERE | Office of Energy Efficiency and Renewable Energy |
| FAIR | Findable, Accessible, Interoperable, Reusable |
| FOA | Funding Opportunity Announcement |
| GDP | Gross Domestic Product |
| HPC4Mfg | High Performance Computing for Manufacturing |
| HTE | High-Throughput Experimentation |
| HTE-MC | High-Throughput Experimental Materials Collaboratory |
| IoT | Internet-of-Things |
| IP | Intellectual Property |
| Li | Lithium |
| MAP | Materials Acceleration Platform |
| Mg | Magnesium |
| MGI | Materials Genome Initiative |
| ML | Machine Learning |
| MML | Material Measurement Laboratory |
| MYPP | Multi-Year Program Plan |
| NIST | National Institute of Standards and Technology |
| NREL | National Renewable Energy Laboratory |
| PR | Public Relations |
| RAPID | Rapid Advancement in Process Intensification Deployment |
| R&D | Research and Development |
| UI/UX | User Interface/User Experience |

# Executive Summary

The Department of Energy's Advanced Manufacturing Office (AMO) and the National Institute of Standards and Technology (NIST) held a joint Technical Workshop on July 12, 2018 to inform *A Multi-agency, Multi-year Program Plan in Advanced Energy Materials Discovery, Development, and Process Design Utilizing High-Throughput Experimental Methods, Artificial Intelligence, Autonomous Systems, and a Collaboratory Network*. Scientists, researchers, and program managers from industry, academia, government, national laboratories, and non-governmental organizations gathered in Gaithersburg, MD, to identify key priorities moving forward and challenges in applying artificial intelligence (AI), autonomous systems (AS), and related techniques to the discovery, development, and process design of advanced energy materials. The results of this workshop provide important input for program plans and research directions that can advance the state of the art in materials discovery, development, and process design in energy and related materials.[1]

AS and AI tools such as machine learning (ML) and various optimization techniques present major opportunities to accelerate the materials discovery and development process and increase U.S. manufacturing competitiveness. Using advanced computing techniques and tools, researchers can better determine relevant materials properties of potential new materials in advance—much more quickly and at much lower cost than what is possible today. This has the potential to revolutionize the materials industry.

This workshop aimed to better understand the research and development (R&D) priority areas that are needed to overcome key challenges, and the federal role in this area. The workshop covered four main topic areas: 1) Priorities in Energy Materials R&D; 2) Database Infrastructure Needs in AI and Energy Materials R&D; 3) Expansion of the Collaboratory Network for Energy Materials Discovery and Process Design; and 4) Integration of AI, ML, and Experimentation for Energy Materials Design and Processing.

Overarching themes, several of which crosscut multiple breakout groups, emerged from the workshop. They include the following:

- Relevant, large, high-quality datasets that are broadly available are essential input for AI, ML, and related analyses. The techniques rely on ample and germane data to generate meaningful findings, but datasets for many material classes are too sparse. This is compounded by some corporate policies that restrict sharing of their process data, since it is considered proprietary.
- Data quality and availability could be improved with the implementation of data performance standards or common agreed practices. Without such standards, it is difficult to determine the quality of the data, and filtering it can be difficult and expensive. "Failures" (perceived non-successes with valid data) should be recorded and included in databases of experimental data, where researchers can access them and derive insights.
- Digitizing data from scientific literature, so that data are published in formats that can be curated and machine-readable, would also improve the quantity of data.
- Materials used in energy storage applications represent an important area where advancements made by AI techniques could enable new progress. Using AI methods to handle the complexity and enormous number of different configurations in chemical batteries, for example, presents an opportunity to discover new materials or optimize material combinations. Innovations in additive manufacturing (AM) are particularly conducive to materials advancements enabled by new high-throughput and AI techniques, since AM is a quickly evolving area with increasing amounts of data being generated. Alloys, for which there is ample performance data in both process and real

---

[1] This report summarizes the presentations and breakout group discussions that took place at this workshop. The results presented here provide a synopsis of the viewpoints expressed by the experts who attended the workshop and do not necessarily reflect those of the broader materials development community.

environments, also present a class of materials that have wide applicability to energy and manufacturing, and therefore could be excellent candidates for application of AI and related techniques.

- In-situ data processing technologies that enable in-process corrections in real time would be a major breakthrough in many industrial processes. AI and ML can enable autonomous synthesis and characterization, an approach that has the potential to quickly screen and optimize materials for composition and processing parameters in real time. In addition, ML algorithms could suggest the parameters for experimental design, and experimental equipment could then interact directly with database infrastructure, exchanging information automatically as needed. R&D and the development of autonomous synthesis and characterization tools that enable real-time analysis should be prioritized.
- Materials scientists and others trained in the materials community are typically not experts in advanced computer science or AI, and vice versa. Developing educational or training programs for AI and related applications would help bridge this gap between computer scientists and material scientists (both experimental and computational). Development of online training modules should be prioritized.
- The High-Throughput Experimental Materials Collaboratory (HTE-MC, the "Collaboratory")[2] presents an opportunity to build a sustainable, interactive ecosystem of resources across disciplines and affiliations. It has the potential to enable technological breakthroughs and foster standards and best practices. The HTE-MC could be a workable data hub for researchers and used to demonstrate value of AI, ML, and AS in the industry, facilitate model development for design and processing, and coordinate development of data standards, among other uses. Intellectual property is an important issue that will need to be addressed, and the HTE-MC will need a sustainable source of funding.

---

[2] The High-Throughput Experimental Materials Collaboratory is a concept for a new network of member institutions and a federated network of data infrastructure providers. The goal of Collaboratory is to accelerate materials innovation by building a sustainable ecosystem of resources to empower innovators, fostering standards and best practices, enabling game-changing scientific and technological breakthroughs, and recognizing intellectual property. The vision for the Collaboratory is to harness students and funding to tackle critical materials problems, which will create new knowledge, materials solutions, AI-ready public data, high-throughput experimental libraries, AI-materials data infrastructure, a new materials discovery commerce, and a next generation workforce. When a new measurement is required, a library would be synthesized and measured at the optimal member institution. Seamless data discovery and access will be possible via a dedicated registry for HTE resources and data. Data interoperability will be possible via creation of community data standards. See https://mgi.nist.gov/htemc for more information.

# Table of Contents

# 1 Background and Motivation

Materials are essential for innovations in nearly every energy technology: solar cells, catalysts, magnets, membranes, electrochemical batteries, thermal storage, structural materials, coatings, combustion, and so on. However, the materials design and development process can be expensive and extremely slow, taking 10 to 20 years from discovery to commercialization of a new material. More rapid and lower cost discovery of novel materials and their processing parameters can result in breakthroughs for cutting-edge clean energy applications in industry that would generate economic growth, advance environmental progress, and further improve our standard of living.

Traditionally, when trying to discover new materials, researchers and engineers combine materials in a lab, guided by intuition in what is essentially an advanced Edisonian approach. With recent improvements in computing power—along with advances in scientific theory, modeling, simulation, software, and experimental techniques—researchers and engineers now have the potential to design and develop materials much more rapidly and at lower cost than traditional approaches. Using these advanced computing techniques and tools, researchers can better determine relevant properties (e.g., strength, durability, reactivity, conductivity) of potential new materials which can inform appropriate experimentation. This has the potential to revolutionize the materials industry.

Among the approaches to realize this opportunity are autonomous systems (AS) and artificial intelligence (AI) tools such as machine learning (ML) and various optimization techniques to efficiently evaluate materials properties. While these approaches are being applied to other fields, such as pharmaceutical discoveries, web-search algorithms, and autonomous vehicles, the application to the advanced energy materials industry is still nascent.

Progress is being made, however, with scientists developing and sharing databases so that AI techniques can use the data for new materials discovery. Some notable examples include the Materials Resource Registry (https://materials.registry.nist.gov/) at NIST, the Open Quantum Materials Database (http://oqmd.org/) at Northwestern University, the Materials Project (https://materialsproject.org/) at Lawrence Berkeley National Laboratory, and the experiment-based High-Throughput Experimental Materials database (https://htem.nrel.gov) at the National Renewable Energy Laboratory. Other researchers (for example, the Dark Reactions Project (https://darkreactions.haverford.edu/) at Haverford College) are using AI machine learning algorithms to mine archived laboratory notebooks for data to predict new materials.

However, significant R&D challenges remain. AI techniques and high throughput computing models are not able to predict everything, and they are often limited by lack of available data, among other factors.

To accelerate the process—from newly discovered material through simulation, synthesis and testing, and processing—the U.S. Department of Energy (DOE) and National Institute of Standards and Technology (NIST) held a joint workshop to identify R&D priorities and collaborative opportunities for advancing energy materials discovery, development, and process design by utilizing techniques such as high-throughput experimental methods, AI, ML, and AS.

## Workshop Overview

To better understand the needs to overcome key challenges and the federal role in this area, DOE and NIST held a Technical Workshop on July 12, 2018 to inform *A Multi-agency, Multi-year Program Plan in Advanced Energy Materials Discovery, Development, and Process Design Utilizing High-Throughput Experimental Methods, Artificial Intelligence, Autonomous Systems, and a Collaboratory Network*. Representatives from industry, academia, government, the DOE national laboratories, and non-governmental organizations gathered in Gaithersburg, MD, to hear presentations from industry, academia, and DOE and NIST program managers, and participate in topical breakout sessions. Discussion topics focused on identifying key priorities moving forward and challenges in applying AI, AS, and related techniques to the discovery, development, and process design of advanced energy materials.

Manufacturing remains the essential core of U.S. innovation infrastructure and is critical to economic growth and national defense. As global competition to manufacture advanced products intensifies, America's innovation ecosystem must raise its performance. By sharing priorities and ideas, industry, academia, and government partners can more effectively leverage existing resources, collaborate, and co-invest to nurture manufacturing innovation.

## About the Workshop Coordinating Offices

The Advanced Manufacturing Office (AMO) within DOE's Office of Energy Efficiency & Renewable Energy (EERE) partners with private and public stakeholders to improve U.S. competitiveness, save energy, create high-quality domestic manufacturing jobs, and ensure global leadership in advanced manufacturing and clean energy technologies. AMO invests in cost-shared R&D of innovative, next-generation manufacturing processes and production technologies that will improve efficiency and reduce emissions, industrial waste, and the life-cycle energy consumption of manufactured products. These investments help the nation harness energy efficiency in manufacturing as a competitive advantage and competitively manufacture cutting-edge clean energy technologies in the U.S. AMO is particularly interested in pre-competitive, early-stage research collaborations that might overcome some of the critical challenges associated with advanced manufacturing technology.

The Materials for Energy and Sustainable Development Group within NIST's Material Measurement Laboratory (MML) develops and disseminates measurement science, measurement standards, and measurement technology pertaining to the measurement of all functional properties of advanced materials and devices, including chemical, electrical, thermal, optical, and magnetic properties. The Group also determines and disseminates key data needed to establish the relationships between structure, functional properties, and performance of inorganic and hybrid materials and devices. The research and measurement services provided by MML support innovation in both mature and emerging industrial sectors, and its programs are based in part on industry, academia, and government input, such as the information gathered at this workshop.

This workshop report summarizes the presentations and breakout group discussions that took place at this event. Note that the results presented here are a snapshot of the viewpoints of the experts who attended the workshop; they do not necessarily reflect the views of the broader materials development community.

## Workshop Process and Breakout Sessions

The one-day workshop consisted of plenary presentations in the morning, followed by breakout sessions and a plenary wrap-up session in the afternoon. Plenary presentations featured invited experts from relevant government programs at DOE and NIST, academia, and industry. The presentations helped to identify priorities with respect to energy materials research and the HTE-MC and raised specific challenges in applying advanced AI techniques to industrial material design, development, and processing. At the afternoon breakout sessions, experts discussed four topics:

1) **Priorities in Energy Materials R&D**

   Technical Chair: **Yifei Mo**, University of Maryland

   The main goal of this breakout group was to identify the high priority innovative or emerging energy materials (or material classes, categories, etc.) that are likely to experience major breakthroughs if AI, ML, AS, and/or inverse design were applied. Experts discussed which types of materials lend themselves particularly well to these methods. The group also briefly explored why these materials should be considered "high priority," considering in particular the potential benefits in terms of energy supply, use, emissions/environment, and security. For the highest priority materials area, experts suggested performance or technical targets in the short (< 2 years), medium (3-5 years), and long term (5+ years).

The group also identified key barriers and challenges to applying AI, ML, AS, and inverse design techniques to the materials identified as "high priority." The discussion focused on technical issues, limitations, problems, and gaps, rather than non-technical barriers such as limited program funding.

**2) Database Infrastructure Needs in AI and Energy Materials R&D**

Technical Chair: **Lei Cheng**, Argonne National Laboratory

The main goal of this breakout group was to identify the high priority needs for database infrastructure to support AI, ML, and AS. The group discussed needs in software development, development of shared services (e.g., repositories, registries), community engagement activities (e.g., hackathons), data demonstration projects, and others.

Experts briefly explored why these database infrastructure needs should be considered "high priority" in the context of energy materials, considering in particular the potential benefits to energy supply, storage, distribution, and use. The group also identified key technical barriers and challenges to addressing the high priority database infrastructure needs. The breakout group suggested and discussed overarching goals and technical targets in the short (< 2 years), medium (3-5 years), and long term (5+ years).

**3) Expansion of the Collaboratory Network for Energy Materials Discovery and Process Design**

Technical Chair: **John Perkins**, National Renewable Energy Laboratory

The main goal of this breakout group was to describe key aspects of the Collaboratory (HTE-MC). Key aspects included the mission of the Collaboratory, physical attributes (e.g., infrastructure, size, number of members, centralized/decentralized), technical capabilities, and governance and participation (e.g., partner commitments, information sharing requirements). The breakout group also explored ways for the Collaboratory to become self-sustaining, including models for financial sustainability.

Considering the key aspects brainstormed by the group about the Collaboratory, the group suggested specific next steps, actions, and activities that should be undertaken to formalize and expand the Collaboratory. The group also considered which organizations would be well-suited to participate or lead these activities.

**4) Integration of AI, ML, and Experimentation for Energy Materials Design and Processing**

Technical Chair: **Joshua Schrier**, Fordham University[3]

The main goal of this breakout group was to identify the high priority areas for using integrated AI, ML, and experimentation in high throughput energy materials design and processing. Experts in this group focused on the "integrated" aspect in particular, discussing breakthrough opportunity areas when considering these methods working in concert. The group discussed which materials (or types/classes of materials) would have the largest impact (for example, benefits in terms of energy

[3] At the time of the workshop Dr. Schrier was an Associate Professor of Chemistry at Haverford College. He is now the Bepler Professor of Chemistry at Fordham University.

supply, use, emissions and environment, and security) if additional R&D emphasis on integrated AI, ML, and other high throughput methods were applied.

The group also identified the key technical barriers and challenges to applying integrated AI, ML, and experimental techniques to the areas identified as “high priority.” The group suggested overarching goals or technical targets in the short (<2 years), medium (3-5 years), and long term (5+ years).

# 2 Overview and Perspectives

The workshop began with a series of presentations from invited experts from academia and industry, as well as context-setting overviews from government officials. Summaries of the perspectives offered by the speakers are provided below.

### Summary of Findings: 2017 DOE Workshop on Artificial Intelligence Applied to Materials Discovery and Design

**Brian Valentine**, DOE Advanced Manufacturing Office, Technology Manager

DOE held a related workshop on August 9-10, 2017 in Pittsburgh, PA, on AI applied to materials discovery and design. The purpose of the workshop was to learn about the R&D needs for accelerating AI applied to materials design and gather expert opinions on the appropriate role by the federal government to advance the approach. Several experts from industry, national laboratories, and government delivered presentations at the workshop with valuable insights and unique perspectives from both the private and public sectors. Breakout sessions at the workshop covered three topics: data quantity and quality; platforms and infrastructure; and collaboration, partnerships, and education/training. Each breakout group responded to topical questions related to future capabilities and targets, technical and scientific challenges, and education and training.

Discussions at the workshop primarily explored issues that are common across diverse material types, such as data formatting, algorithms, models, and tools. In addition, a few industry participants shared their experiences in applying AI to alloy design for additive manufacturing and to the development of glasses and catalysts. Based on the common themes identified during the workshop, four high priority areas emerged: common data formatting and quality, integrating multi-scale models, public-private partnerships, and cross-discipline education. Detailed results are available in the workshop report.

### Summary of Findings: 2018 NIST Workshop on High-Throughput Experimental Materials Collaboratory (HTE-MC)

**Martin Green,** NIST Materials for Energy and Sustainable Development Group, Leader

NIST hosted a workshop on the concept of a High-Throughput Experimental Materials Collaboratory (HTE-MC) on February 28-March 2, 2018. The goals of the workshop, attended by participants from industry, government, and academia, were to socialize the HTE-MC concept among stakeholders, strategize expansion of HTE-MC membership, and define technical, operational, and business models for the HTE-MC. Presentations included speakers from industry, government, and academia.

The Materials Genome Initiative (MGI), formed in 2011, is a combination of digital data, computational tools, and experimental tools. At the intersection of experimental tools and digital data, there is a need for high-throughput experimental data. The Collaboratory is a solution for this need. The Collaboratory would consist of an integrated, delocalized network of high-throughput experiment synthesis and characterization tools and a best-in-class materials data management platform. Conclusions of the HTE-MC workshop determined that the Collaboratory will accelerate materials innovation by 1) building a sustainable ecosystem of resources to empower innovators, 2) fostering standards and best practices, 3) enabling game-changing scientific and technological breakthroughs, and 4) recognizing intellectual property.

In the Collaboratory concept, government agencies would provide the structural funding and members (academia, national labs, industry, and small business) receive the funding to do the experimental work. Members provide their staff, students, and infrastructure. Users pay access fees and create new data for the consuming public. Contributors publish the data and receive benefits (such as credits) for published work. The Collaboratory model would generate new knowledge and solutions for the public.

HTE-MC members would be required to provide experimental and/or data infrastructure, adopt HTE-MC data infrastructure, actively participate in applicable consortia and working groups, and champion development of new capabilities. NIST is testing and improving the standards for exchange of data and samples through a proof of the Collaboratory principle with National Renewable Energy Laboratory (NREL). Next steps include determining the type of entity and rules for the HTE-MC, obtaining funding, and encouraging participation.

### Overview: DOE AMO

**Rob Ivester,** DOE Advanced Manufacturing Office, Director

Manufacturing represents $2 trillion in U.S. GDP, 12.4 million direct employment jobs, and 25% of U.S. energy consumption. Of the 98 quadrillion Btu of energy consumed across all sectors in the United States, nearly 62 quadrillion Btu is estimated to be wasted during the end-use stage, where consumers use the manufactured goods. The percentage of energy lost in the industrial manufacturing sector is smaller, in part because the energy intensive industries have a direct incentive to become more energy efficient. The DOE AMO Multi-Year Program Plan (MYPP) describes the Office mission, vision, and goals, and identifies the technology, outreach, and crosscutting activities the Office plans to focus on over the next five years to enable increased industrial energy efficiency.

AMO's Technical Assistance programs address market barriers that limit the adoption of available technologies. The Better Plants Program & Challenge works with industry to commit to, take action, and report results from energy saving measures. Regional Clean Energy Application Centers provide market assessment, education and outreach, and technical assistance to increase the adoption of energy-savings Combined Heat and Power technologies. Industrial Assessment Centers train the energy engineers of tomorrow while simultaneously providing free energy assessments to small and medium sized manufacturers who may not be able to allocate overhead to energy management programs as easily as large firms. The HPC4Mfg Program provides modest funding to the national labs to utilize resources to solve advanced manufacturing problems, such as the drying of paper towels. The Lab-Embedded Entrepreneurship Program brings technologies into the lab and accelerates these technologies. Participating labs include Lawrence Berkeley National Laboratory, Argonne National Laboratory, and Oak Ridge National Laboratory.

The AMO consortia model brings together diverse perspectives and opinions from academia, government, and industry, to aggressively work to get input from the right stakeholders. There are fourteen Manufacturing USA Institutes, including RAPID[4] – Modular Chemical Process Intensification, which enables development of breakthrough technologies to boost energy productivity and energy efficiency in domestic chemical manufacturing. A successful consortia model will assist with creating the most efficient technologies domestically and selling them around the world.

[4] Rapid Advancement in Process Intensification Deployment

### AMO Multi-Year Program Plan: Moving Applications of AI and ML from Materials Design and Discovery Through Process Design and Development

**Brian Valentine,** DOE Advanced Manufacturing Office, Technology Manager

AMO is interested in bridging a gap between scientific discovery and technology application, with an interest in manufacturing issues. The AMO R&D Projects Program supports cross-cutting R&D in technology coming from DOE Basic Energy Sciences. The Program supports competitive solicitations for new technology development and manages the projects selected from the solicitations. Program goals include discovering and designing processes for new, advanced materials required for next generation energy applications and reducing the time and costs needed to bring new energy materials to market.

Advanced energy technologies require increasingly complex, advanced materials for systems and processes. A new advanced material typically requires about two decades to reach deployment, while the product development and manufacturing cycle has been reduced by computer-aided design to as little as three years, so this timing mismatch can negatively affect performance and competiveness. AMO envisions a solution with tools based on AI/ML methods can expedite materials discovery by integrating synthesis, characterization, and modeling of complex materials and chemical processes and reduce design and development costs.

The Materials Genome Initiative (MGI) is a government led effort to discover and optimize new materials for higher performance and advanced products. AMO aims to bridge MGI with AI to expand materials discovery methodology to include process design. There is a need for a unified framework of approaches that would be useful across broad areas and useful to investigators not specialized in particular techniques. This need can be addressed through inter-agency partnerships. AI and ML have an important role in materials research to accelerate product design and processing.

### Summary of Findings: 2018 Mission Innovation Workshop on Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods with Artificial Intelligence

**Joshua Schrier,** Fordham University, Professor of Chemistry

Mission Innovation is a global initiative of 23 countries and the European Union, collectively representing approximately 80% of the global public sector clean energy R&D budget. The goal of the initiative is to accelerate clean energy innovation and make clean energy widely affordable. As part of the initiative, countries seek to double their government/state-directed clean energy R&D over five years (2016 – 2021). Among the Mission Innovation activities are "Innovation Challenges," which are areas of common R&D interest among participating countries. One of those Challenges is the Clean Energy Materials Innovation Challenge, which held an international experts' workshop on materials discovery in September 2017 with over 130 attendees from 17 countries. This Innovation Challenge aims to accelerate the innovation process for high-performance, low-cost clean energy materials and automate the processes needed to integrate these materials into new technologies.

The 2017 workshop resulted in an influential report that outlines integrated Materials Acceleration Platforms (MAPs), which consists of six main elements: Closing the Loop, AI for Materials, Modular Materials Robotics, Inverse Design, Bridging Length and Timescales, and Data Infrastructure and Exchange. Closing the Loop requires digital experiment plans across diverse lab environments and programming language for experiments. AI for Materials will emphasize the need for material-specific ML methods and the importance of interpretability (machine rhetoric); ideas should be presented in a way that human decision makers can process. Modular Materials Robotics represents techniques and materials as modular "building blocks," fosters human-machine communication, and simplifies materials exploration. Inverse Design enables automated generation of candidate materials designed to meet the performance, cost, and compatibility requirements of a given clean energy technology. Bridging Length and Timescales would require using ML as the "glue" between existing experimental and computational methods for multiscale modeling and characterization. Data

Infrastructure and Exchange involves extracting and distributing data from structured instrumental sources and unstructured natural language sources.

The main recommendation coming from the workshop is the need to develop MAPs to integrate automated robotic machinery with rapid characterization and AI to accelerate the pace of discovery. The deployment of the acceleration platforms has the potential to speed up the discovery of materials at least by a factor of ten—from 20 years to 1 to 2 years. This will catalyze a transition from a trial and error approach to scientific discovery to an era of inverse design, where the desired property drives the rapid exploration, with the aid of advanced computing and AI, of materials space and the synthesis of targeted materials. The inverse design of materials allows for their accelerated scale-up into installed technologies, accelerating energy technology innovation.

### Materials Discovery at Solvay: Applying AI Tools to 150 Years of Historical Data

**Jean-Yves Delannoy,** Solvay Chemicals, Manager of Statistical Modeling and Artificial Intelligence Team

Solvay is a European company that emphasizes sustainable chemistry to meet the challenges of society. Solvay's 24,500 employees work at locations in Asia, Europe, and North and South America, using chemistry as a way to create materials. The company has a goal that, in 5 years, every chemist will have a chatbot on their bench with the ability to suggest the next experiment.

Several examples illustrate the work that Solvay is doing to develop new materials with more efficient properties. The company's approach is a three step process: document classification, information extraction, and cognitive solution. The need is to be able to systematically structure unstructured data.

An example Solvay project can be found in gas creation in batteries and examining how much gas batteries can emit while being used. Formulations are examined that minimize gas generation. Solvay is also studying green/biodegradable materials and how to predict if a new molecule is going to be biodegradable, through data analysis of similarities and clustering. Solvay has successfully tested for this property for small molecules.

Solvay has assessed current needs in the application of AI tools. For example, there is a need for a digital platform to share ways to communicate and more uniform data sources. Data should be available in the same format or with a translator. There is also a need for more iterative learning and improvement on robotization and automation. Data transformation is a key enabler.

# 3 Summary of Results

This section provides a synopsis of key findings from the four workshop breakout sessions. It is organized by focus discussion topic, which varied between sessions. Additional results and details from the breakout groups are provided in Appendix A.

## 1. Priorities in Energy Materials R&D

The main goal for this breakout session was to identify the high priority innovative or emerging energy material classes or categories that are likely to experience major breakthroughs if AI, ML, AS, and/or inverse design were applied. Participants in this breakout group discussed what types of materials lend particularly well to these methods and briefly explored why these materials should be considered high priority. Participants then identified key challenges and barriers to applying these methods (i.e., AI, ML, AS, and/or inverse design) to the materials that the group determined to be high priority for further R&D, which were decided by vote. In addition, the breakout group briefly considered specific performance targets for the highest priority materials.

### Promising Materials Areas

**FOCUS QUESTION 1:** *What energy materials are likely to experience major breakthroughs if AI, ML, AS, and/or inverse design techniques are applied?*

During the discussion on promising materials areas, the following themes were determined by participant voting to be the highest priorities. Additional important ideas raised during the session are listed in the Appendix A.

**Materials for additive manufacturing:** Innovations in additive manufacturing (AM) are particularly conducive to materials advancements enabled by new high-throughput and AI techniques such as machine learning. As a quickly evolving area and with increasing amounts of data being generated, energy materials for AM presents opportunities for major breakthroughs that could have significant impacts across the spectrum of energy supply, distribution, and use.

Specifically, thin film alloys represent a class of materials where high volumes of data can be collected through in-situ monitoring and analyzed using ML algorithms to quickly assess 3D-printable materials. AI techniques can be used to rapidly filter data to extract the useful data from the noise. AM alloys in the transportation industry that are durable and lightweight (e.g., Al and Mg alloys), which reduce fuel use, provide an example of how significant energy savings could be achieved. In addition, the use of AI to optimize discovery and development of AM materials for safety and structural components can result in major breakthroughs. Currently there are no known AM components qualified for primary structural or safety applications, and advanced high-throughput AI methods could enable this use.

**Catalysts:** AI techniques could stimulate significant discoveries in catalytic materials. In particular, metal oxide frameworks for catalysts represent a major opportunity to find and develop novel materials. A breakthrough in this area that results in stable and effective materials for $CO_2$ conversion to valuable chemicals, for example, would have enormous energy, economic, and environmental benefits.

**Process optimization:** AI and ML approaches have the potential to accelerate the determination of appropriate processing techniques (e.g., whether the material is additively manufacturable) and quickly screen and evaluate materials to detect and correct deviations from desired properties instantly. Advanced informatics techniques can also improve functionally-graded AM parts, whereby AI methods enhance process optimization and have the potential to link the process to microstructures.

**Materials for energy storage:** Materials are fundamental to the performance, cost, and durability of batteries. Materials that are candidates for use in energy storage applications represent an area where advancements made by AI techniques could enable new innovations. For example, for solid-state batteries, there are many parameters, chemical conditions, and process variables, and using AI methods to handle the complexity and enormous number of different configurations and options presents an opportunity to discover new materials and to optimize the battery manufacturing. Electric vehicles in particular would benefit from advancements in this area. In addition, substitutes for lithium in batteries would be beneficial. There are several candidates for Li replacements that, with additional emphasis on applying AI, could lead to development of commercial alternatives. Cobalt is another material for which the discovery of substitutes through use of computational AI techniques would have major benefits for energy storage.

**Materials for use in harsh conditions:** "Harsh conditions," for the purposes of the breakout discussion, was defined as environments where the current best-in-class materials struggle to survive. This includes high temperature and/or corrosive environments, such as those present in molten salt applications, thermal power generation units, and recuperators. Harsh conditions also include situations where the material must withstand high load and wear, and chemical/electrochemical environments such as for high temperature superconductivity transmission lines. AI and high-throughput computational methods present an opportunity to design and discover novel materials suitable for these conditions in just a few years where using traditional methods could take decades or longer. AI and ML can be particularly useful in analyzing datasets, and then optimizing material designs based on the data, when testing under harsh conditions parameters and various exposure conditions. Heat transfer materials for waste heat recovery is another example of materials use in harsh conditions where large gains could be realized when AI is applied.

**Alloys:** Alloys are a class of materials that have wide applicability, are extensively used, and have ample data available regarding their performance in both process (e.g., manufacturing, laboratory) and real (in-situ) environments, and therefore could be ideal for application of AI techniques that would make a major impact for energy technologies. Largely because of their broad use and data availability, alloys may be considered a low hanging fruit for further emphasis on R&D using AI. Examples include anti-gunk materials for coatings on nuclear fuel rods. New materials design and development is already underway in this area, but AI could help fine-tune the most promising materials options to help further advance the R&D. Self-healing and metamaterials (artificial materials that cannot be found in nature) for energy applications also present opportunities for leapfrog technological advancements using AI approaches.

## Barriers and Challenges

**FOCUS QUESTION 2**: *What are the most important barriers and challenges to using AI, ML, AS, and inverse design for the priority materials areas?*

Participants identified the following barriers and challenges as the most critical areas that inhibit application of AI techniques to the priority material areas.

**Lack of relevant, large datasets:** AI and related techniques rely on ample and germane data. Currently, most relevant datasets for many material classes are too sparse to yield meaningful findings via application of AI. To address this challenge may require greater participation and buy-in from stakeholders to add data, high-throughput experimentation to generate data, and automated aggregation and organization of data from the literature. As a further complication, data are often saved in different propriety formats, with no standard format that is conducive to sharing in a database.

**Lack of data sharing by industry:** Many companies maintain close hold of their process data, since it is considered confidential and proprietary. This is usually the initial instinct for industry when asked to share, although often companies may be able to share a subset of their data or anonymize entire datasets without compromising company secrets. Nonetheless, this is a major hurdle to data availability and use of AI tools. Both historic data from processes no longer in use, and access to process data in real manufacturing settings are commonly kept within company networks only.

**Inconsistent data quality:** Limited or non-existent data performance standards and common agreed-upon practices present a major challenge to the use of the data that is available. Without such standards, it is difficult to determine “good” data from “bad” data, and filtering the data can be difficult and expensive.

**Lack of real-time, in-situ data processing technologies:** In-process corrections in real time would be a major breakthrough in many industrial processes, but this is hampered by a lack of technologies that enable quick data processing and real-time analysis. Some materials developments require long processing times, which contributes to the complexity and difficultly of rapid data collection and analysis. The lack of effective, fast, and reliable materials synthesis and experimentation analysis techniques is also complicated by the fact that many processes have inadequate or non-existent in-situ sensors.

**Disconnect between materials community and computer scientists:** Materials scientists and others trained in the materials community are typically not experts in advanced computer science or AI, and vice versa, computer scientists typically are not experts in materials science. This presents a challenge when marrying these two fields. The technical languages are unique to each domain, and skillsets do not fully overlap. Thus, applying AI to materials is impossible for either discipline to do alone. Combining these unique areas and increasing collaboration will require removal of institutional or traditional work silos. This challenge can be partly overcome by a concerted education effort to get buy-in from both communities – materials and computer science – that joining forces presents opportunities for developing exciting technological breakthroughs and solutions to some of the world’s most difficult energy and environmental problems.

## Goals and Targets

**FOCUS QUESTION 3**: *Given the priority materials areas identified and considering the key barriers and challenges, what specific performance or technical targets would be reasonable in the near, mid, and long term?*

During this session, breakout group participants suggested performance targets for the priority materials area deemed most important (in terms of votes received). Due to time constraints at the breakout session, the discussed focused solely on the top priority area: materials for additive manufacturing.

In the near term (within about two years), reasonable performance targets for researchers incorporating AI techniques with materials for additive manufacturing include incremental process improvements on existing materials, enhanced size control, and yield improvements for powder. In the medium term (about three to five years), targets in this area could include in-situ corrections in closed loop systems, and improvements in more challenging materials such as ceramics, oxides, and carbides. In the long term (beyond five years), research targets for AI techniques in AM could include materials qualification certification and reproducibility, and development of primary structural and safety components.

## 2. Database infrastructure needs in AI and Energy Materials R&D

During this breakout session, participants discussed the future of AI, ML, and AS for materials discovery as it relates to the development of database infrastructure. Participants discussed goals and targets that database infrastructure performance should achieve in the future; high priority needs related to the development of database infrastructure; and barriers and challenges that could hinder the development of the needed database infrastructure.

### Overarching Goals and Targets

**FOCUS QUESTION 1:** *During the time periods specified, what performance targets for database infrastructure will need to be met to enable innovation in AI/ML/AS for Energy Materials R&D? The time periods of interest are specified as short term (<2 years), mid term (3–5 years), and long term (>5 years).*

During the discussion on overarching goals and targets for database performance, the following themes were most prominently discussed by workshop participants:

Short Term Goals and Targets

**Improved User Experience:** In the near future, material scientists will be interacting with databases, repositories, and tools through developed software. Material scientists are not, inherently, computer scientists. Software tools must be usable by researchers who are not software developers. A near term goal is for improved user interfaces and experiences through which researchers interact with database infrastructure.

**Improved Data Speed:** To facilitate working with large volumes of data, database infrastructure must operate at improved speeds, exhibiting at least a 5X improvement over current baselines. Near term goals include improved compression algorithms to facilitate sharing large datasets, improved search engines to allow for faster and more relevant results from queries in large and complex datasets, and topically-focused datasets and databases (as opposed to covering a broad range of applications or industries). Achieving these goals will facilitate faster discovery using AI techniques.

**Improved Data Standards:** Data sharing will be greatly enhanced by the creation of clear and widely understood standards for data organization, submission, and access. Near term goals include clearly defined data quality and trustworthiness standards that specify the metadata and are included along with experimental measurements. Lack of agreement concerning data standards can greatly hinder developments and innovations in materials discovery.

Mid Term Goals and Targets

**Metadata Inclusion Mandates:** A mid-term goal is to publish standards for scientific journal papers that mandate the inclusion of metadata in publications that include experimental data. This rule should be strictly enforced. Publishers have an opportunity to support the development of the industry by taking a strong stance. In a similar manner, publicly funded research should enforce mandates concerning data sharing. At the least, it is recommended that federally funded research efforts mandate the use of FAIR (Findable, Accessible, Interoperable, Reusable) Guiding Principles[5] for supported research.

**Full Data Threads:** Materials are impacted by the processes they undergo and conditions that they are exposed to, from material sourcing, through production, to end use. However, data related to these processes and conditions are often unavailable, and when materials fail during use, it can be difficult to determine the

[5] For information about FAIR Guiding Principles for scientific data management and stewardship, see https://www.nature.com/articles/sdata201618.

underlying cause. In the mid-term time range, data threads should follow a material throughout its lifetime, providing full traceability.

**Curation of Historical Data:** While much of the conversation around AI for materials discovery has focused on automated collection of experimental data, not all data used for materials discovery will derive from new experiments. A goal for the mid-term is to curate a tremendous amount of historical data from text books, manuals, laboratory notes, and other publications.

Long Term Goals and Targets

**Automated experimental data collection:** In the long term, researchers anticipate that ML algorithms will be able to suggest the parameters for experiments that should be performed. Experimental equipment will be able to interact directly with database infrastructure, receiving instructions and exchanging information as needed. In this way, experimentation can be optimized, allowing for more rapid identification of successful means for manufacturing novel materials for the energy industry. As a prerequisite to this functionality, researchers must first be able to collect experimental data automatically in real time. Material information, process information, machine parameters, and other relevant variables must be automatically monitored and recorded while experiments are being conducted. This information will eventually form the basis for the use of AI/ML for optimizing experimentation.

**Automated data quality:** Another long term target is that researchers, when utilizing AI algorithms, will have the option to only consider data that meets their quality requirements. Alternatively, experimental equipment will have the ability to auto-calibrate to the needed quality levels specified by the investigator.

**APIs (Application Program Interfaces) Everywhere:** In the long term, useful scientific data are produced from a variety of sources, including instruments, computer systems, analog measurement devices, and laboratory notebooks. At present, the collection, digitization, and unification of these disparate data sources is nearly impossible. In the future, this process should be automatic. This will necessitate the development of APIs to facilitate software development and interaction at every stage in the research process.

## High Priority Needs

**FOCUS QUESTION 2**: *What are the high priority needs related to database infrastructure to support AI, ML, AS in the following areas: software development, development of shared services, community engagement activities, and data demonstration projects?*

During this session, participants discussed high priority needs related to database development to support AI for materials discovery. The following discussion summarizes the high priority needs discussed for each area:

**Software Development:** The highest priority need is software development involving digitizing data from scientific literature. Data are routinely published in formats that are not machine readable. Furthermore, many sources of relevant data are unstructured, including the lab notebooks kept by researchers. Though these primary sources of research data contain troves of useful information, they are not represented in a fashion compatible with database infrastructure.

**Development of Shared Services:** Accessibility to the development of shared services is another high priority need. Shared services are only helpful to the extent that researchers, vendors, contributors, etc. are able to easily access, interact with, and derive value from the services. Shared services that have been developed in the past have neglected to consider user accessibility and experience, and this has resulted in low utilization.

**Community Engagement Activities:** The most pressing need with regard to community engagement involves establishing buy-in from stakeholders. To engage, stakeholders need to be properly incentivized, and incentives will vary for different types of stakeholders (e.g., academic vs. corporate researcher, private institute vs. federal government). Unless stakeholders are properly incentivized, they will continue to withhold data or use ineffective means to share and communicate experimental and research results.

**Data Demonstration Projects:** Demonstration projects are needed to promote best practices associated with data capture, management, dissemination, and access. Projects should be referenceable, allowing researchers to gain a full understanding of expectations around the use of database infrastructure for materials discovery.

**Intellectual Resources:** Additional experts and volunteer resources are needed to improve likelihood of success. Citizen scientists, retirees, Emeritus professors, undergraduate and community college students, and other volunteers, may be able to contribute needed time and expertise.

## Barriers and Challenges

**FOCUS QUESTION 3**: *What are the barriers/challenges to developing the needed database infrastructure to support AI/ML/AS applications for energy materials development?*

During this session, participants discussed barriers and challenges impacting the future development of AI/ML/AS for energy materials discovery, as it relates to database infrastructure. The following topics were highlighted as most important by the breakout group participants:

**Scope and Complexity:** Materials discovery for energy applications represents a broad field, involving a great diversity of applications, and an enormous variety of research questions to be investigated. This complexity presents a challenge to the use of AI/ML/AS algorithms, and to the development of the supporting database infrastructure on which they rely. Algorithms converge more rapidly and produce more useful insights when the underlying data is well organized and structured. The volume, diversity, and complexity of data available for consideration can complicate the process of specifying a relevant dataset, specific to a research inquiry. Database infrastructure should be designed to facilitate useful down-selection of available data. However, down-selection may result in a loss of insight, as unexpected relationships may exist within larger sets of data. The design of database infrastructure will greatly influence the extent to which relationships and insights can be derived from large and complex data sources.

**Inclusion of "Failures":** The present and historical culture surrounding scientific publication is one that rewards successes and novelty but excludes experiments that are considered "failures." In some instances, errors occur, and experiments fail to produce useful information. These instances are of no use in developing AI and ML algorithms. However, some experiments fail to validate an initial hypothesis and are considered non-successes. These type of "failures" or "uninteresting findings" are often deemed as unfit for publication, although they still retain inherit value and can be useful for training algorithms and informing future experimental efforts. There is much to be learned from experiments that produced results that were unintended, undesirable, or uninteresting. Results from non-successes should be recorded and included in databases of experimental data, where researchers can access them and derive insights. In science, there are many more non-successes than there are successes, and each contains a lesson.

**Unfriendly Software:** The majority of the experimental data produced is created by people that are not computer scientists or software engineers. The software used to extract data from research, publish data to databases, or utilize data in analyses must be user friendly. Software must be designed with an intentional focus on providing simple, intuitive interfaces and functionality.

## 3. Expansion of the Collaboratory Network for Energy Materials Discovery and Process Design

Participants in this breakout group discussed key aspects of the High-Throughput Experimental Materials Collaboratory (HTE-MC or Collaboratory for short), including the mission, physical attributes (e.g., infrastructure, size, number of members, centralized/decentralized), technical capabilities, and governance and participation (e.g., partner commitments, information sharing requirements). The breakout group also explored ways for the Collaboratory to become sustainable and suggested specific next steps that should be undertaken to formalize and expand the Collaboratory.

**FOCUS QUESTION 1:** *Define the key elements of the Collaboratory's mission, desired physical attributes, required/ideal technical capabilities, and governance and participation.*

### Defining the Collaboratory

As described in a previous workshop, the High-Throughput Experimental Materials Collaboratory (HTE-MC or Collaboratory) will accelerate materials innovation by:

- Building a sustainable ecosystem of high-throughput experiment and data resources to empower material innovators
- Fostering standards and best practices
- Enabling game-changing scientific and technological breakthroughs
- Recognizing intellectual property.

This "center without walls" will serve a range of stakeholders in the research community from government agencies, dues paying members, visitors or public users, and other contributors. However, to collaborate with these stakeholders, the Collaboratory first needs to develop a more robust vision for how it will operate.

**Mission:** The mission of Collaboratory should be to build a sustainable ecosystem of resources that empower innovators, whether they hail from industry, government, or academia. Ultimately, the primary goals of the Collaboratory should be to enable game-changing scientific and technological breakthroughs and to foster standards and best practices. However, the organization will likely need a well-defined technical focus that aligns with the goals of the sponsors (i.e., the funding sources) to sustain itself. Demonstrating a positive impact on the development of AI and ML for materials in the form of usable results will help to secure the future of the Collaboratory and support the mission.

Intellectual property (IP) developed in pursuit of the mission of the Collaboratory will likely be a major point of contention that the Collaboratory will have to address from the outset. There is inherent tension in a collaborative organization like this one between the desire of industry participants to develop IP by investing in research using shared infrastructure and expertise of the organization and the desire of government, national laboratory, and academic participants to revolutionize dissemination of knowledge and make all results "open source" for anyone who would build on them.

**Desired Physical and Cyber Attributes:** The Collaboratory will likely bring together disparate researchers, so a well-developed online platform will be necessary to provide access to members and promote interaction. An online platform could also serve as a central point for public relations to cultivate a consistent message. Leveraging internet resources requires high-quality security, especially due to the IP that would likely reside online. Businesses would need confidence in the security of the site before they would share confidential business information through any online platform.

One approach to the physical research facilities that compose the Collaboratory would be clustering of synthesis and analysis techniques for different classes of materials. That is, the facilities that specialize in certain synthesis and analysis techniques specific to a class of materials would be co-located in order to encourage shared expertise and collaboration. While co-location of synergistic resources often happens naturally, a key component of the Collaboratory would remain the ability for any given project to easily draw upon the most relevant resources available across the entire Collaboratory.

**Required/Ideal Technical Capabilities:** Ideally, the Collaboratory would use (or develop) standardized, scalable lab capabilities that support commercialization as a key end goal. Subject matter experts in fields like nanomaterials and bulk materials, fatigue and fracture, corrosion resistance, and certification would enhance lab capabilities.

Hardware and software used in the lab must be compatible with the mission to perform HTE. Modular instrumentation and data interfaces, both open source designs and commercial implementations, could provide the flexibility to scale capabilities to fit opportunities in HTE. Automated capture of the metadata describing the materials synthesis, processing and characterization is critical. Furthermore, more robust data handling capabilities could feed optimization algorithms (e.g., using AI and ML) to design subsequent generations of experiments.

Data security, management, organization, and administration will be key capabilities for the Collaboratory. These aspects must be incorporated into a workable data hub that allows researchers to aggregate and access quality data. One possible filter to control the quality of data managed by the organization would be to maintain only data that results in publications. However, in keeping with the above stated goal of including "failures" as well as successes, the Collabotory will have to develop internal standards other than just publication for accessing data quality.

**Governance and Participation:** The governance structure will incentivize or discourage a range of potential participants in the Collaboratory. To reach a broad stakeholder group, the governance should be designed to be flexible and opportunistic enough to not only serve large and small institutional members, but also to draw in new partners and groups as well as to allow visiting researchers.

At the same time, the participation might be served best by having a focused problem set that is relevant and practical, but ranges across different length scales from nanomaterials to bulk materials. Users could also have the option to post their specific needs to define the most relevant problems. The governance structure could also look to other entities in the value chain for materials for input, possibly through advisory board members.

Finally, the governance structure should provide mechanisms for data quality control, such as having an authorized entity on data quality evaluation grading the quality of data, allowing comments from users, and providing capabilities for data conditioning and synthesis.

**Steady-State:** In the steady state condition, the Collaboratory should consistently generate new publications and data for the research community’s benefit and maintain an ecosystem for shareable code and data (e.g., in Jupyter notebooks). Researchers can benefit not just from raw data and results in the shared ecosystem, but also from the tools and analysis developed along the way. Finally, workshops and conferences organized or sponsored by the Collaboratory would bring together researchers to discuss past and future work.

The Collaboratory will likely also support education and training to expand the number of knowledgeable researchers. Initial training courses could introduce researchers to the institution and familiarize them with the operations and capabilities. Continuing education opportunities would ensure sustained growth in skills among the HTE research community.

One open question about the Collaboratory is whether the institution could exist without a benefactor like the U.S. government, and instead subsist only on revenues from members and partners.

## Strategies for a Self-Sustaining Collaboratory

**FOCUS QUESTION 2:** *How can the Collaboratory become self-sustaining, and what is the financial sustainability model?*

There are many strategies the Collaboratory could follow to become sustainable. Primarily, the Collaboratory will need a reliable source of funding, whether that comes from the U.S. government or from industry through subscriptions or fees for use of facilities or IP. While not truly "self-sustaining," government funding could provide the seed needed to transition to primarily industry funded in the longer term, assuming that industry could support the Collaboratory completely. While subscribing members would support the sustainability of the institution, they would likely need flexibility in how they pay for their memberships. For example, in addition to paying cash for memberships, organizations could contribute in-kind with data, shorter data embargos, or sponsoring capabilities. Moreover, a fee structure will likely need to account for the financial resources of members, charging different subscription rates to large corporations and research universities than to small businesses and interested colleges.

As the Collaboratory is being launched and seeks to become more self-sustaining, it should focus on making a name for itself. One inexpensive possibility would be to initiate HTE-MC "branded" projects and proposals based on existing funding. By pursuing branded projects, the Collaboratory could build trust in the community as a leader and authority. In addition to funding projects, the Collaboratory could also encourage researchers to identify themselves as affiliated with the HTE-MC as well as with their home university, lab, or business in any research they publish.

Another key to become self-sustaining will be data management. Making data submission as easy as possible will clearly be a requirement to boost participation. Data from the national labs and academia, where publishing is imperative, could seed a library, which curates the collection. To deliver greater value to the researchers and to encourage collaboration, the data should retain the association with the researcher(s) that developed it.

## Steps Forward

**FOCUS QUESTION 3:** *What steps, actions items and activities should be undertaken to formalize and expand the Collaboratory, and potentially what organizations should participate/lead these activities?*

The Collaboratory only exists as long as people are willing to make it a functioning operation that keeps the community engaged and motivated. A core group, such as the people providing input in this workshop session, is needed to advance the Collaboratory's mission, develop an organizational structure, delineate collaborators' capabilities and interests, and estimate the required funding level for the Collaborative. Because funding is critical, the core group will need to create the business case—including the value to taxpayers—and messaging needed to explain the Collaboratory to decision makers and lobby for funding from DOE, the Department of Defense, directly from Congress, a private sector sponsor, or a combination of all these sources. With funding, the Collaboratory should focus on research valuable to the community. In addition, the Collaboratory could seek funding through existing funding opportunity announcements (FOAs) from a range of federal agencies. The Collaboratory could also demonstrate its value through projects that lead to faster material discovery.

# 4. Integration of AI, ML, and Experimentation for Energy Materials Design and Processing

The breakout group identified several priorities and needs that, if addressed, will advance AI, ML, and experimentation for energy materials design and processing. Workshop participants discussed capabilities for meeting these future needs, including accessible high speed computing, advanced AI and ML that is relevant to industry, and next generation robotics.

## Goals and Targets

**FOCUS QUESTION 1:** *What goals and targets (in the near term: 1-2 years, medium term: 3-5 years, and long term >5 years), would be appropriate for integrating AI and ML to improve materials design and processing for high throughput systems?*

This session identified overarching goals and targets for integration of AI and ML to improve materials design and processing for high throughput systems for important industrial applications. A summary of findings is below, with additional details provided in Appendix A.

### Experimentation and Data Management

Incorporating theory into physics-based models create challenges for integrating data. Material science needs to define foundational principles as well as broad translatable principles. Reproducibility and the treatment of uncertainty (in both data and predictions) are also significant challenges. Work is needed to develop tools for consistent data reporting in a way that facilitates subsequent analysis. There is value in providing “data lakes” (centralized repositories of raw, structured, semi-structured, and unstructured data, stored in their natural format) to reduce barriers to depositing information and facilitate creative reuse. Rather than requiring deposited data to have a predefined schema, a “schema on read” is applied upon data extraction. There is also a need for datasets that can be used as benchmarks of existing and future ML methods, and which could be used by the broader community of computer scientists looking for data upon which to test their methods. The process for data and retrieval must be consistent.

***Near Term (0 – 2 years)***
In the next two years, the research community should aim to assemble curated datasets for real-world problems, e.g., by developing open collections of labeled and annotated materials characterization images. Publicly available datasets, with consistent and well-documented data representations representing a wide range of materials science problems and hosted in a central location, will facilitate widespread engagement.

***Medium Term (3 – 5 years)***
In the next three to five years, autonomous inorganic materials synthesis and processing experiments should be directed using AI in concert with the integration of experimental and computational databases. ML-assisted characterization methods should also be developed to reduce characterization time and cost. Examples include combining the results of multiple low-cost characterization techniques to approximate the results of higher cost, low-throughput techniques, and applications of compressed sensing and active learning methods to reduce the acquisition time or number of data points needed to measure properties with a desired level of error. Internet-of-things (IoT) data collection directly from manufacturing processes can be used to construct models (“online learning”), and in turn used to identify bottlenecks and variations in operational manufacturing processes that can be targeted for experimental design. Techniques for integrating sparse and/or noisy legacy data with new high-throughput experimental and simulation results should be developed in this timeframe. This would allow a determination of how data can be used to bridge time and length scales starting at the atomistic level. Technical journals should require machine-readable experimental plans and outcomes for

publication (e.g., "GitHub for experiments"). An ambitious medium term goal is to have a DOE user facility manage experiments by means of an API for user experiment specification and data access.

***Long Term (>5 years)***

Beyond five years, routine, single-property material property optimizations are expected to be largely driven by AI techniques. Given a property, AI will be able to identify the best system and growth parameters by drawing upon a wide range of datasets, simulations, and models. Tools that predict chemical or physical performance of novel compounds—and provide routes to their synthesis—should be generally available. Furthermore, these tools will provide reliable estimates of the extent to which properties can be increased based on known physical understanding. New multiscale methods and software libraries for incremental autonomous decision making are needed to achieve this aim. The predictions will then be tested using autonomous high-throughput experimental facilities. This will require the development of new types of high-throughput experiments and hierarchical methods, facilitated by standardization of autonomous experimental platforms. This combination of software and hardware advances may reduce the time from conception to demonstration by 90%.

Incorporating new material into a final product currently requires a high level of human expertise and lengthy, iterative process development. AI/ML tools providing decision support for process-manufacturing of materials will be essential for reducing the time from demonstration to market by as much as tenfold. While these tools will not replace scientists and engineers, making these types of AI decision support tools widely available will help humans empower innovation, thus improving our nation's global economic competitiveness.

An ambitious goal is to use these methods to discover new multifunctional materials (e.g., simultaneously better mechanical and electrical properties). Making progress in this field will require investments in foundational AI techniques (e.g., multi-objective optimization methods), simulation (e.g., multi-scale, multi-physics modeling), and experimental methods (hierarchical synthesis, self-assembly, and high-throughput methods for conducting them).

**Model Development**

In contrast to the wealth of representations and models available for machine learning on discrete organic molecules and proteins, machine representations of materials are relatively undeveloped. One challenge is representing the material diversity (atomic types, extended bonding, disorder, alloying, etc.). A second challenge is the diverse range of physical properties to predict, which often depend on processes occurring over many orders of magnitude of length- and time-scales. The wide diversity of possible structures implies that property goals can be achieved with many different material choices. For instance, many possible material compositions have qualities needed for high-temperature magnetic applications, but economic, environmental, synthesizability, and other practical considerations must be incorporated to constrain the search space. These external constraints must be optimized together with the physical properties of interest.

***Near Term (0 – 2 years)***

In many cases, existing physics-based simulation methods (e.g., density functional theory, reactive molecular dynamics, etc.) provide reasonable property estimates, but often the computational cost of using these methods to search the entire design space is prohibitive. Methods that help limit the number of candidate materials are valuable. There are many ways to achieve this end in the near term. One approach is to train ML-based proxy models on previous physics-based calculations, and then use these models to focus subsequent calculations on the best performing materials. In addition, machine learning could assist with bridging length- and time- scales in multiscale simulation methods. Existing ML models for extracting constitutive laws for "new" materials from experimental data—suitable for use in existing finite element analysis packages—often violate conservation laws, and near term opportunities exist to develop models that correct this flaw. A second approach build models and representations directly on experiments to predict the cost and difficulty of the synthesis and processing stages. A third approach combines experimental and computational results. This can provide uncertainty estimates on theoretical and ML model predictions, as well as provide systematic corrections to these predictions.

There is a great need for better decision-making optimization tools at each stage in the R&D process, e.g., selecting the most informative simulations and laboratory experiments to perform and helping optimize manufacturing processes. Many of these techniques have been developed in industry but are not yet widely used by material scientists and engineers.

In the near term, most researchers will continue use of physics-based simulation models for forward prediction of material properties and function. Applications of ML techniques in this area will include development of computationally low-cost proxy models trained to reproduce these simulation results. In parallel, a growing community will advance the agenda of analyzing experimental data with ML and using AI to plan experiments based on the results. One open question is how to combine physics-based and ML-based models to provide insight into foundational principles and broad translatable principles for materials science.

There is significant debate about the role of physics-based theory to constrain ML models. On the one hand, many recent advances in language translation, speech recognition, and image recognition were achieved by dispensing with preexisting theoretical models and learning directly from raw data. Could similar performance gains be made by dispensing with limited (often linear or low-order polynomial based) theories in materials science? A potential downside to this approach is that the models may not improve human insight into the underlying processes and reasoning. Furthermore, these types of data-only approaches require vast quantities of training data, a scale that is unrealistic for materials datasets. Methods that use physical theories to constrain the parameter space might be able to learn better with less data. However, it is unclear how to construct models that incorporate these constraints in a disciplined way. Both “theory-free” and “theory-full” strategies should be pursued to determine which is more successful.

***Medium Term (3-5 years)***

In the medium term, one main focus should be making ML tools accessible to the widest range of practicing materials scientists and engineers. General methods for fast and accurate approximations of physics-based models—ideally with uncertainty quantification—is an area of high interest and requires significant research effort. ML models should ideally be human interpretable, unless there are significant performance benefits to “black box” models. Reducing the technical obstacles to dispatching experiments and collecting results from autonomous robotic experimentation platforms is crucial. Furthermore, distributed and secure computational resources for data management and rapid data processing with low latency times are needed, so that results can be easily used, reused, shared (rather than lost) and processed in real time (rather than days). This will be facilitated by efforts within the materials research community to standardize representations and data formats.

The tremendous diversity of material systems and characterization techniques means that there will probably be many independently created representations and standards developed to suit the needs of particular communities of researchers. Interoperability should be encouraged, but without imposing undue burden on users. Not all experiments are created equal—and so methods that identify appropriate representations and the domain of applicability of models are needed to prevent egregious misuse of these techniques by inexperienced users.

***Long Term (>5 years)***

In the long term, constitutive laws from atomistic or other fine-scale simulations leading to the bridging of time and length scales needs to be developed. A user-friendly graphical interface for comparing multiple systems would be beneficial. A moonshot program would be the development of a “one platform for all” model for materials discovery within this timeframe.

**Machine Learning, Robotics and Automation**

***Near Term (0 – 2 years)***

A near term goal for both basic research and process modeling is to demonstrate a combination of ML, AI, and laboratory robotics for semi-autonomous basic research and process development. This type of hybrid system would involve "learning by doing"—acquiring new information by performing experiments, using the data to train ML models, and using AI to select the most informative next experiment. This type of system is needed both to obtain data of sufficient quantity and quality for general ML model construction, and also to put AI predictions into practice at scale for a physical system. This type of automation should be made readily accessible to the research community. Moderate cost, tabletop-scale laboratory robots are already commercially available, but it is important to make them user friendly. A sub-goal within two years is that these systems must be collaborative, intelligent and increasingly autonomous, so that detailed programming is not needed for every step. A second sub-goal is that these systems should facilitate automatic and comprehensive data collection.

***Medium Term (3-5 years)***

The following is needed for the medium term: two major research directions, two enabling technologies to develop, and two key demonstrations.

The first major research direction is extending these approaches to "hard to measure" properties. This includes efforts at developing autonomous high-throughput characterization hardware for processes that are currently difficult to measure or difficult to reproduce, as well as ways of using data from these measurements to build models that allow for inferring "hard to measure" properties from collections of easy measurements, drawing upon simulations and past data and not just data mining. A second major research direction is increasing the level of autonomy. This includes efforts at hardware (e.g., high throughput synthesis and characterization systems that can run without significant human involvement for extended periods), and software (e.g., processing background data with minimal human involvement to generate new knowledge in real time). This will require collaborations between vendors and users to set standards for interaction and data transfer.

Although many enabling technologies need to be developed to support these efforts, there is a critical need for better probabilistic active-learning algorithms for ML and for better multiple-feature/multiple-objective learning and optimization algorithms, particularly for cases with conflicting objectives.

In the medium term, there should be several demonstrations of the value of this approach. Proof-of-concept demonstrations that "close the loop" between ML and automation will need to be accepted by the broader materials research community. There should be some demonstration of the ML tools and data-driven approaches in a practical production setting. Some possible examples include: (i) combining catalyst development with process conditions to accelerate process optimizations by as much as 50%; (ii) rapid identification of the material properties from experimental and modeling outputs, and the ability to tune process parameters for desired microstructure for manufactured material; (iii) use of AI for process control (e.g., improved proportional-integral-derivative controllers). There are a number of possible applications for this approach, in which product performance enhancements seem feasible by changing materials processing. Examples included: high-temperature alloys, solar catalysis, and structural alloys for reducing vehicle weight.

***Long Term (> 5 years)***

The aspirational target is dramatic (10x or more) reduction in time and cost needed for materials discovery, product design, and process development. Possible applications include reducing the use of critical materials by 50% (which are often sourced from countries with strained U.S. relations), modifying processes to improve energy efficiency (e.g., to save one quadrillion Btu of energy savings), and using ML to perform product characterization and quality control. The common need is reduced experiment time and cost, increased scale of data collected from experiments, and generation of higher fidelity models from the data in less time.

There are several simultaneous paths towards this target. One realization would be at research-level facilities capable of performing bulk synthesis/characterization with real-time feedback and performing experiments that generate enormous amounts of data, resulting in high-fidelity models; these may be laboratory or national scale facilities. Another realization would be at production plants, in which data from process measurements and robotic product characterizations process measurements are used to actively learn models and perform real-time adaptations of processing conditions during production, especially where robotics are used. A third realization would be the creation of an ecosystem of suppliers that would allow "fabless" materials companies to purchase laboratory experiments and production capabilities on an incremental basis. An analogy was made to the way cloud computing has reduced the initial investment for internet startups; similar reduction of initial barriers would enable economic growth by reducing the cost needed for materials product startups as demonstrated in the electronics industry.

### Work Flow and Interfaces

A distinction was made between high-level workflows (the overall investigative pipeline of experimentation, data acquisition, processing, ML, and decision making) and interfaces (the specific "glue" that connects ML models, AI experiment planning, and instrumentation). Because progress in all of the areas discussed above relies upon having interfaces for communicating between systems and workflows that combine systems, the discussion focused on near term needs. At present, humans direct workflows. Experimentation and ML are not directly connected to one another. Bottlenecks include lack of acceptance by the R&D community, lack of experimental hardware, lack of software interfaces, and lack of general integration tools.

***Near Term (0 – 2 years)***

In the near term, cultural resistance should be overcome by more case studies showing performance exceeding the state-of-the-art R&D processes. A significant experimental bottleneck is the lack of high-throughput *bulk* synthesis methods capable of real-time measurement and feedback; most current HTE systems consider only non-bulk sized samples. Better software tools are needed to provide the "glue" between experiment and ML/AI. There are many different tools and data processing requirements that can be used to program a robot, and systems that facilitate constructing "recipes" are needed that link together these pieces. There is a value in having an ecosystem of different types of modeling, inference, and planning tools. One possible form is the development of high level programming languages for describing lab experiments and manufacturing processes, data acquisition and processing. These would provide vendor and device independent descriptions that could be compiled into specific directives for experiments by humans at a laboratory bench, high throughput robotic systems from different vendors, specific industrial process controllers, etc.

Progress in developing and standardizing interfaces is also a near term target for progress. Adopting data interchange practices/requirements and continued development of standards for describing experiments and their results would facilitate the ML process. Standardized APIs can be developed within two years that would allow software to integrate with remote experiments performed in different parts of the world simultaneously. Enforcing standards on data output through these APIs will also facilitate data collection and sharing efforts, and in turn provide more, higher-quality data for training ML models.

## High Priority Areas

**FOCUS QUESTION 2:** *What are the high priority areas for expanding the use of integrated AI, ML and experimentation for energy materials design and processing?*

### Experimentation and Data Management

A high priority area should be experimentation that can be done quickly to generate copious amounts of data for multiple material properties across the widest possible range of qualitatively different property regimes. Existing datasets focus primarily on single properties, characterized over a narrow range of values, using a single characterization method. A strength (unrealized) of ML approaches is interpolation in high-dimensional spaces that allow a functional form for engineering use. Current experimental methods and datasets often do not provide appropriate data for this, as they tend to focus on single variable changes. New types of experimental design plans that hierarchically sample the material design space are needed. Measuring and reporting uncertainties are important for diagnosing failures of ML approaches due to data versus shortcomings of the models. The experimental uncertainties need to be quantified but need not be small; a potential strength of ML models is the ability to provide meaningful interpolation using noisy datasets.

### Model Development

Four high priority areas were identified. Models should better handle experimental uncertainty and propagate these through model predictions. In many cases, predicting observed properties requires combining several types of model outputs, further adding to the difficulty. A second high priority need is for phased iterative predictive models for learning from experiments and simulations. The goal here is to generate low cost proxy ML models that approximate a high-dimensional response function based on a small number of examples. These models could be used to accelerate simulations (by providing a shortcut for lengthy calculations) or to guide autonomous decision-making algorithms (e.g., identifying regions with the highest probability of success or informativeness, and then guiding actions). A phased iterative approach would try to maintain a user-specified level of accuracy statistically by monitoring and reassessing the model performance throughout the simulation/experiment process. The model would adaptively switch between exploration (adding new points to improve the model) and exploitation (delivering successful predictions) to optimally describe the region of interest with minimal computational/experimental cost. A third priority is the need for models that describe the synthesizability of materials. Most demonstrations to date have focused on predicting material properties, but there is a need for models that can predict whether a material can be made and provide a recipe for how to make it. So far, most work on the applications to ML to synthesizability have focused almost exclusively on organic small molecule synthesis, rather than materials. The final high priority need is for a better understanding of the domains within which ML approaches are successful. Most models are constructed from ad hoc, human selected representations and models. A more systematic approach would have AI select the appropriate representations or have models learn the appropriate representation based on some type of raw, based representation of the data. The software tools developed for these four areas should be open source.

### Materials

Most work to date has focused on single property optimizations of bulk materials and composites at the research scale. High priority areas for future research-scale work include: strongly correlated oxides, materials for energy storage and sustainability, magnets (especially reducing the use of critical rare earth elements, as noted above), thin films for coatings and energy applications that are *not* bulk or composites. High priority areas for manufacturing/development-scale work include: improving metal additive manufacturing material quality and performance, and applications to inorganic materials manufacturing.

**Education and Training**

Applying AI/ML methods to materials discovery and design is a multi-disciplinary effort requiring experts across many fields, including, but not limited to, material scientists, chemists, mathematicians, and computer scientists. This is a relatively young discipline (arguably since the early 2000s), so there are not many practitioners. All participants agreed that it was difficult to find adequately trained research staff and postdoctoral fellows. The field also requires advanced tools and infrastructure that is costly and difficult to maintain without the required expertise. These factors necessitate robust collaborations between the public and private sectors and identification of gaps in education and training needed to ensure that the United States can remain competitive in this area.

There is an overall need for developing an educational/training framework for AI/ML tools and prioritizing education to train more computer scientists and material scientists (both experimental and computational) that work together for developers and users. Material scientists need to be trained in knowledge extraction and programming. It may be easier for material scientists to learn a baseline level of programming and data science than it is for computer scientists to learn an adequate level of chemistry and physics. More graduate students and undergraduates are needed who understand materials informatics, ML, AI, and robotics. This is not the case in most universities today; most students learn ML/AI methods in an ad hoc way. Many materials science programs include numerical simulation and programming coursework, but more work for developing a curriculum that uses ML tools for materials sciences is needed. This might include knowledge extraction, image analysis for characterization, data mining for cluster analysis, among other topics. Training modules need to be developed and put online that will provide the fundamentals at all educational levels. Some computer science programs offer a minor degree in fields like data analytics, but this is rare in materials science (notable exceptions are programs at University of Buffalo and Georgia Tech). A process-specific AI educational program leading to a domain expert degree or certification should be pursued. This could be minor degrees or certificate programs in AI/ML/HTE integration. Research and training centers could develop the principles, tools, and training curricula to integrate AI, ML and HTE.

**Machine Learning for Process Design Applications**

High priority areas include: context-aware data representations, collaborative and intelligent robots for automated data collection, AI (using ML and reasoning) frameworks capable of complex process modeling, strategies for capturing processing conditions, and process representations. It was noted that even if initial ML predictions are poor, they can still be can be useful as long as feedback is used for multiple iterations to learn and self-correct. A better understanding is needed of problem domains where ML is more applicable versus a conventional approach.

## Barriers and Challenges

**FOCUS QUESTION 3:** *What are the barriers/challenges to developing the needed science or technologies to meet the high priorities identified above?*

**Experimentation and Data Management**

Experimentation barriers include: difficulty of using HTE techniques especially when the overall process/experiment is slow and expensive, lack of a standardized approach or case studies for integrating ML and AI with HTE, and existence of only a limited number of experiments that generate a large volume of sharable data.

Data Management barriers include: lack of machine-readable open data with reproducibility for sharing, the dearth of data from “failed” experiments, lack of procedures to share intellectual property and proprietary control (closed) software, limited scientific publications with access to standardized data, lack of standardized experiment descriptions that allow ease of data extraction, lack of data standardization across research organizations (or using ML to learn how to standardize data), limited incentives from funding agencies to execute data management plans, and limited incentives from journals to demand complete dataset and reproducible modeling details.

Expanding on these points, experiments are costly and slow, and as a result, many of the applications of ML to material science thus far have used density functional theory (DFT) calculations to create data sets. Future emphasis should be placed on HTE to collect an abundance of data in a reasonable time. Applying what has been learned from ML modeling of DFT datasets to experimental data will not be straightforward. There is a tremendous disconnect between what is predicted at a molecular level and what goes into a (chemical) reactor. Experimental data is often limited, and there needs to be knowledge of when ML methods become feasible and when traditional physics-based heuristics should be used instead. A promising area is the use of hybrid approach that utilize simulation to interpolate between sparse experimental data.

A cultural shift is needed in the norms by which researchers share data, software, and methods, and accept ML as a tool. Information management should promote data sharing. Synthesis and characterization devices lack vendor-independent APIs, limiting the ability to specify and record comprehensive descriptions of experimental procedures and results. Proprietary instrument control software control is a barrier for developing autonomous experimentation. Standardizing the data sets or code that can be applied to data sets is desirable, but there are concerns about intellectual property. Funding agencies should require more rigorous data management plans. A requirement such as depositing data in centralized repositories should be a prerequisite for publication and continued funding, analogous to the norms in fields such as crystallography and biological sequencing. The data stored in these centralized repositories would also be useful for benchmarking ML methods more generally. Because energy-related materials is a large field, norms instituted by this program would propagate through the broader materials research community.

Current resources for experimental data are insufficient to inform AI and ML models. Academic journals do not always present data with sufficient detail to reproduce the experiment or computation, much less the data needed to train ML models. Technical journals should require reproducibility in reporting results of experiments, simulations, and ML/AI model construction, ideally in machine-readable formats. Open-software journals and dataset journals are one step towards this goal, but the software and data must also be in public repositories. All journals should handle data in the same manner. Publishing information from “failed” and “marginally successful” experiments represents a huge but necessary cultural shift. Ideally all data, independent of success, would be published alongside the manuscript. Technical societies and publishers should be encouraged to demand increasingly detailed datasets as a prerequisite for publication.

**Model Development**

Barriers include lack of real-time data analysis, complexity of model transfer from ideal systems to real systems, and difficulty in developing accurate models. Models often describe ideal systems at the experimental level that are not useful or accurate enough to describe real systems.

**Machine Learning for Process Design Applications**

Large scale manufacturing is much more complicated than experimental research. Maintaining the same material quality for large manufacturing processes is very complicated relative to a small scale or even a pilot scale (e.g., roll-to-roll processing). One should not force ML/AI methods into a problem domain in which it is not suitable. Current demonstrations have used ML as a tool to solve particular scientific problems, and a barrier is finding the niches in process design and manufacturing for which ML/AI are appropriate.

**Education and Training**

The barriers in this area are lack of sufficient expertise and communication barriers between adjacent fields.

As noted in the previous section, the typical material scientist is insufficiently trained in data science techniques, and the typical computer scientist or statistician is insufficiently trained in the language of material science and manufacturing processes. There are not enough cross-disciplinary experts with expertise in both the underlying science and engineering and AI/ML. There is difficulty in finding graduate students and postdoctoral research fellows with baseline expertise for funded projects in AI and ML. Large information technology companies attract this expertise to the detriment of the rest of the economy and the research community. Educational initiatives to increase both the general quality (of non-specialists) and quantity (of specialists) should be a high priority focus.

A second barrier is communication between material experimentalists, material theorists, statisticians, data scientists, and computer scientists. Incompatible technical languages prevent communication. Information or data may be lost which prevents models from being accurate. This “silos” developers and users according to their discipline or domain, and sometimes even *within* their discipline or domain.

# Appendix A

Appendix A provides a compilation of the input and ideas raised by workshop participants during the breakout sessions. The appendix is organized by breakout group and by focus question. For some focus questions, experts were asked to vote on the inputs/ideas they perceived as most important. For those questions, the number of votes received (indicating participants' highest priorities) are shown in the tables below with dots; the vote totals are listed in parentheses.

## Breakout Group 1: Priorities in Energy Materials R&D

### Promising Materials Areas

**FOCUS QUESTION 1:** *What energy materials are likely to experience major breakthroughs if AI, ML, AS, and/or inverse design techniques are applied?*

Table A-1. Priorities in Energy Materials R&D – Promising Materials Areas

**Alloys ●●●●●●● (7)**

- Fe-based alloys. These are widely used and therefore have a broad impact. This class of materials has wide applicability and a relatively simple electronic structure. Data are available for these materials. This could be considered low hanging fruit.
- Anti-gunk materials on nuclear fuel rods. The cost of fouling on nuclear fuel rods is billions of dollars. These are based on carbon materials, similar to crud resistant materials.
- Graphene/graphite materials for batteries, through the use of more modern designs.
    - Currently used in all forms and are required for all future forms: fuel cells, ultra-capacitors, flow batteries, lithium-ion batteries
- Densification of composites to capture complex mechanisms with slow processing time to accelerate the processing time and capabilities of the materials.
- Metamaterials (artificial materials/materials that cannot be found in nature) in energy applications; self-healing materials.
- Materials that are high strength, corrosion resistant, and light weight. Composites. AI might help in design and development since there are several materials close but need fine tuning

**Materials Operating in Harsh Conditions ●●●●●●● (7)**

(Definition of harsh conditions: any conditions where the current material of choice will struggle to survive)

- Materials in high temperature/corrosive environments (power generation/recuperators) e.g., high load/wear/fatigue, chemical/electrochemical environment.
- Extreme temperature materials
    - Materials that operate at higher temperatures than currently possible, while having longer life, cheaper, and lightweight, e.g., reactors, engines.
    - Target temperature: 2000ºC
    - Value proposition: Higher efficiency operations and cost savings

**Table A-1. Priorities in Energy Materials R&D – Promising Materials Areas**

- Relationships in testing for harsh environment materials between tests and various exposure conditions.
- Corrosion resistant alloys for molten salt applications.

**Materials for Waste Heat Recovery● (1)**

- Low manufacturing cost, high efficiency alloy thermoelectrics.
- Heat transport materials. Heat capture materials.
  - Energy/heat conservation to forestall heat wasted. Currently 67% of generated heat is wasted. Huge savings will be realized if engineers can better capture heat and transport it.
  - Uni-directional heat transportation. Moves heat from production point to point of use
  - Substantial gains will be realized when AI is applied to heat capture. For heat transfer, heat is typically generated to create electricity. But heat is also needed for industry. The class of materials needed to do this are not currently available.
  - One application is nuclear imaging and heat recovery.

**Materials for Additive Manufacturing ●●●●●●●●●●●● (12)**

- Thin film alloys where, using AI, data can be collected and analyzed in high volumes to quickly assess materials that can be additively manufactured.
  - The most important use are tools to filter out data that are useful and not useful, resulting in saved money and time.
- Manufacturable materials using AM techniques and process.
- Additively manufactured materials for safety and structural components.
  - There are currently no additively manufactured components qualified for primary structural or safety application. Potential of AM will not be realized until this can be achieved. AI could optimize on this development.
- Lightweight AM alloys (e.g., Al, Mg alloys) for transportation technologies.
- Functionally-graded additively manufactured parts: machine learning can enable process optimization and link process to microstructure.

**Electronics ●●● (3)**

- Organic semiconductors, perovskites.
  - Processable, low-cost photoelectrical devices (solar cells, lighting).
  - Experimentally driven; processing can significantly affect properties.
- Degradable/recyclable plastics for consumer electronics (or catalysts to enable this).
- Photovoltaics.
  - Primarily driven by "bulk" properties that are simple for ML.
  - Need new materials with range of energy gaps (band gaps).

Table A-1. Priorities in Energy Materials R&D – Promising Materials Areas

**Storage ●●●●●●●●● (9)**

- Solid-state battery.
  - Electrolyte, interface engineering processing and manufacturing.
  - Many parameters, chemical composition, process conditions.
  - AI could lead to storage materials with better performance, lower costs, and improved safety.
  - Application in electric vehicles, for example, where high energy density is required.
- Next generation charge storage, e.g., Na-ion, Li-S.
  - These are high priority materials because lithium supply chain is somewhat unstable, and there is a need to diversify for growing demand. Compounds like Li-S have application or chemistry issues to overcome to be realistic. There are several candidates for storage materials that AI could push to fruition.
- Lithium and cobalt. Substitutes for these elements in battery materials need to be identified to overcome supply-chain issues. Computational techniques may create advancements in this class of materials.
- Thermal storage conductors.

**Catalysts, Process Optimization ●●●●●●●●● (9)**

- Metal oxide frameworks for catalysts.
  - AI techniques may have large impact on these materials because performance is very difficult to scale.
- Stable and active materials for $CO_2$ conversion to valuable chemicals. There is a need to find novel and earth-abundant materials. For example, looking for oxide materials.
- Optimizing and discovering catalysts. For example, to discover new, stable catalysts for energy applications.
- Superconducting wire/tape (multi-materials) to optimize higher temperature capability and bring down costs.

**Functional Materials**

- Magneto/Elasto-responsive systems.
- Superconductor materials with strongly correlated or “entangled” electron systems – hard to solve theoretically.

**Novel Approaches**

- Designing critical material classes for turbines. Use data and learnings from wood fibers.
- Use Nature (biomimetics) as inspiration for novel design.
- Classify parameters and explaining/predicting material performance from data (across life cycle).

Table A-1. Priorities in Energy Materials R&D – Promising Materials Areas

| Materials Processing Optimization |
|---|
| • In situ monitoring/diagnosis to enable "fail-fast" screening experiments (e.g., evaluating new AM alloys) to detect and correct materials defects identified during processing. This would evaluate current materials and use information to determine your next step immediately.<br>• Techniques to make the process faster to determine whether something is additively manufacturable. |

## Barriers and Challenges

**FOCUS QUESTION 2:** *What are the most important barriers and challenges to using AI, ML, AS, and inverse design for the priority materials areas?*

Table A-2. Priorities in Energy Materials R&D – Barriers and Challenges

**Data Availability**

- Industry's desire to maintain control over proprietary process data - historic and mostly new innovations ●●●●●● (6)
    - Accessing processing data in real manufacturing setting.
- Generating and sharing reliable, quality data ●●● (3)
    - Mature topic, as it has been a barrier for a long time.
- Lack of relevant and large datasets ●●●●●●● (7)
    - To overcome it may require buy-in from community to add data, high throughput experimental tools to generate data, and automated aggregation and organization of data from the literature.
- Data are saved in different proprietary formats – no standard data format conducive to sharing

**Data Quality**

- **Scarcity of quality data (not always data quantity) can make it difficult to develop a well-performing AI, ML ● (1)**
- **Lack of data consistency (quality, repeatability) in AM models ●● (2)**
- **Lack of high quality and quantity data.**
- **Limited or non-existent data performance standards and ethics ●●●●●● (6)**
- **Lack of assessments of data quality, including from unknown/other sources ● (1)**

**Data Generation**

- **Lack of surrogate (proxy) property data to help overcome time or length scale issues ●●●● (4)**

Table A-2. Priorities in Energy Materials R&D – Barriers and Challenges

**Data Processing and Visualization**

- Lack of flexible frameworks for data exploration and visualization ●●● (3)
- There is a need to visualize the different parameters if a sufficient database exists. However, there is a lack of easily accessible ways to visualize the data in the database to extract meaningful and timely information.
- Rather than a programming challenge, the barrier pertains to designing tools that will assist with the discovery and help determine what scientists should be looking for.
- Limited suitable data analytics software options.
- Complexity in determining how to extract relationships from massive amounts of data.
- Lack of easy-to-use toolsets and solutions for data use and visualization ●●●●● (5)
- Lack of techniques or protocols for processing information quickly to show data correlations

**Real-time, In-situ data processing**

- Lack of technologies/tools that enable in-process corrections in real time ●● (2)
- Long processing times, preventing availability of rapid analytic data ●●●●● (5)
  - Lack of ability to quickly process real time data and incorporate analysis.
- Lack of effective, efficient, fast, and reliable material synthesis and experimental analysis techniques ●●●●●● (6)
- Inadequate or non-existent in-situ sensors ●●● (3)

**Physical Constraints**

- Disparate location of resources, making it difficult to share information and implement AI, ML at facilities.

**Hardware**

- Lack of universal high-volume test apparatus, which requires physical equipment where hardware must be engineered ● (1)

**Talent/Human Factors**

- Difficulty in interpreting the data/findings and models.
- Materials community ill-equipped to communicate materials problems to computer scientists, and vice versa ●●●●●●● (7)
- Determining and tracking potential covariates.
- Relatively new broad recognition/appreciation of importance of AI and ML in science and technology. Not many people recognize the potential impacts of these tools. ● (1)
- Lack of community buy-in/education. Most people are not capable of determining machine learning models or evaluating these models critically ●●●● (4)
- Competition for top talent from other fields. Challenge to convince best new talent in AI and ML to come into field for energy materials ● (1)

## Goals and Targets

**FOCUS QUESTION 3:** *Given the priority materials areas identified and considering the key barriers and challenges, what specific performance or technical targets would be reasonable in the near, mid, and long term?*

Table A-3. Priorities in Energy Materials R&D – Goals and Targets

| Promising Materials Area | Near term (<2 years) | Medium term (3-5 years) | Long term (>5 years) |
|---|---|---|---|
| **Additive manufacturing** | ○ Incremental process optimization improvements on existing materials – there could be a push for some improvement.<br>○ Size control – optimizing powder production.<br>○ Yield improvement (powder) | ○ In-situ corrections (closed loop)<br>○ More challenging materials (e.g., ceramics, oxides, carbides, titanium, 3D dental implants) and new materials discovery | ○ Material qualification, certification<br>○ Reproducibility<br>○ Primary structural and safety components |
| **Catalysts, Process Optimization** | ○ Existing reactor geometries | | |

## Breakout Group 2: Database infrastructure needs in AI and Energy Materials R&D

**FOCUS QUESTION 1:** *What performance targets for database infrastructure will need to be met to enable innovation in AI/ML/AS for Energy Materials R&D?*

### Overarching Goals and Targets

| Near term (<2 years) | Medium term (3-5 years) | Long term (>5 years) |
|---|---|---|
| ○ Better compression algorithms – factor of 5x improvement<br>○ Data transfer speed greater than 5x the data generation speed of experimental equipment, (e.g., on-the-fly analytics, additive manufacturing)<br>○ Improved capabilities of search engines to query data<br>○ Daily calibration information for experimental equipment captured in metadata file<br>○ Incentivize the development of database structure (monetary may be the best way)<br>○ Researchers receive credit for uploading data (e.g., citations, klicks, etc.)<br>○ Build databases with more focused content, for better convergence in AI algorithms<br>○ Data that are system/application focused/relevant enough that scientists can reach a solution with AI faster<br>○ Build alternative databases with more broad content, to identify non-obvious connections<br>○ Diverse data, e.g., cross application, synthesis protocols, process information | ○ Entity relationships, including formulations and devices, captured along with data<br>○ Complete meta-data included for greater than 75% of all records in a materials database (e.g., how was the data produced?)<br>○ Data quality/trustworthiness, i.e. how to partition data into low, medium, and high quality.<br>○ Experimentation with alloys-need to specify in papers should contain meta-data on contamination and verification of chemistry<br>○ Enforcement of meta-data accessibility in journals; publishers should enforce this standard<br>○ Data thread: connect data from materials sourcing, through production, through end use<br>○ Ensure data from publicly funded experiments conform to FAIR standards<br>○ Public / private cooperation on data services<br>○ Full data thread: provenance, attribution, linkages<br>○ Curation of historical data | ○ Wet lab automated data collection from instruments (describing experiment procedures, equipment, materials, and results)<br>○ Facilitation of automated, fully integrated, collection of data- as experiments are run<br>○ Map of data locations and connections;<br>Similar to graph database (like a tree)<br>○ Automatize quality of data which is formatted to your specs<br>○ Metrology tools that take instruction from database<br>○ API everywhere; everything interfaces easily and allows secure data exchange<br>○ Analog of science for recapture – can scientist look at messy data and improve the data?<br>○ Automated curation (to specified specs) |

- Reinforce experimental methodology to report the true chemistry of the alloy/materials produced
- Guidelines and standards for data organization, submission, and receiving; data quality, trustworthiness
- User customizable/ease of use

## High Priority Needs

**FOCUS QUESTION 2:** *What are the high priority needs related to database infrastructure to support AI, ML, AS in the following areas: software development, development of shared services, community engagement activities, and data demonstration projects?*

Table A-5. Database infrastructure needs in AI and Energy Materials R&D – High Priority Needs

| Software Development |
|---|
| • Digitalization of Scientific Literature Data/ Unstructured to Structured ●●●●●●●● (8)<br>• Robust software/APIs for data preprocessing & ingestion; keeping useful data and discarding redundant data ●●●●●●● (7)<br>• idata App: easy application that anyone can use to access or contribute data ●●●● (4)<br>• Software development to help in selecting appropriate physical models for a dataset, or to help in analyzing data quality ● (1)<br>• Incentivize adoption/use of software products |
| **Development of shared services** |
| • Accessibility for user, researcher, vendor, keeper ●● (2) |
| **Community Engagement Activity** |
| • Buy-in from stakeholders ●●● (3)<br>• Engaging industry to share their data ● (1) |
| **Data demonstration projects** |
| • Pilot Projects to identity best practices for large scale data deployment which ensures data quality ●●●● (4) |
| **Other** |
| • Money and volunteers: utilize not-for-profits, students, and retirees ●●●●● (5)<br>• Prioritize input of existing data ●● (2) |

**FOCUS QUESTION 2:** *What are the high priority needs related to database infrastructure to support AI, ML, AS in the following areas: software development, development of shared services, community engagement activities, and data demonstration projects?*

**Table A-5. Database infrastructure needs in AI and Energy Materials R&D – High Priority Needs**

- Researchers need training on data management ●● (2)
- University curriculum: Integration of ML/AI into physical sciences educational curriculum ●● (2)
- Funding for a sustained amount of time ● (1)

## Barriers and Challenges

**FOCUS QUESTION 3:** *What are the barriers/challenges to developing the needed database infrastructure to support AI/ML/AS applications for energy materials development*

**Table A-6. Database infrastructure needs in AI and Energy Materials R&D – Barriers and Challenges**

- Enormous diversity of material sets/applications and research questions in this field ●●●●●●●● (8)
- Include 'failures' in databases ●●●●●●● (7)
- Unfriendly software, complicated, poor user interfaces ●●●●●●● (7)
- IP Ownership:  who owns it at the end? ●●●● (4)
- Cultural change related to data use and management– urgency to implement changes ●●●● (4)
- Standardization of data collection and maintaining pedigree ●●● (3)
- Platform general enough to store complex data and in easily accessible format ●● (2)
- Long term funding to maintain data infrastructure ●● (2)
- Coordination and agreement on standards on data submission and sharing ●● (2)
- Physics in AI algorithms ● (1)
- Experimental variability and distribution of outcomes ● (1)
- Access to journal articles ● (1)
- Instrument issues; data in closed/dark format; old operating systems ● (1)
- Cross-discipline terminology
- Enough funding for host institutes
- Enormous diversity of ML topics, approaches, software solutions and implementations
- Identifying right set of tools, software, infrastructure that meet requirements, e.g., scalability
- To some extent data will be incomplete and unstructured
- Incorporating good coding / design practices and knowledge of science for robust software development
- Balancing cost and robustness: broad or focused databases? All materials or one class of materials? Varying scopes require varied resources to develop and sustain.

Table A-6. Database infrastructure needs in AI and Energy Materials R&D – Barriers and Challenges

- Volunteers aren't enough- need UI/UX developers and dba's and ITD
- Software teams - researchers want to research, not develop software
- Getting people to move on from their individual systems and use new data management tools, e.g., their excel spreadsheets
- Mixed incentives (collaboration vs competition)
- Demonstration of usefulness of data, especially for AI
- Some people are only interested in data; others are interested in physics and data
- Money is required to enable jobs for database development

## Breakout Group 3: Expansion of the Collaboratory Network for Energy Materials Discovery and Process Design

### Defining the Collaboratory

**FOCUS QUESTION 1:** *Define the key elements of the Collaboratory's mission, desired physical attributes, required/ideal technical capabilities, and governance and participation.*

Table A-7. Expansion of the Collaboratory Network for Energy Materials Discovery and Process Design – Characteristics

**Mission**

- Building a sustainable ecosystem of resources to empower innovators
- Fostering standards and best practices
- Enabling game-changing scientific and technological breakthroughs
- Recognizing intellectual property
- Revolutionize dissemination of knowledge; "open source" science"
- Well-defined technical focus, aligned with sponsors' (funding sources) interest. "Quick wins," e.g., impacts/usable results in 2-3 years.

**Desired Physical Attributes**

- Central website and PR product for dissemination
- IP issues: input "ideas" securely
- Co-located clusters of synthesis and analytical techniques (for a given class)

**Required/Ideal Technical Capabilities**

- Standardized scalable lab capabilities (with "industrialization" as a key end goal)
- Cyber security data management organization and administration
- Must have a workable data hub
- Secure data
- Quality control (e.g., only data that resulted in a publication)
- Subject matter expertise
  - Nanomaterials/bulk
  - Fatigue and fracture
  - Corrosion resistance
  - Certification
- "HTE" compatible hardware/software interface
- Modular instrumentation and data interfaces
  - Public domain designs
  - Commercial implementation
- Use optimization (data analytics/AI/ machine learning) to design experiments (save money)

**Table A-7. Expansion of the Collaboratory Network for Energy Materials Discovery and Process Design – Characteristics**

## Governance and Participation

- Focused problem
    - Relevant/practical
    - Nanomaterials/bulk
    - Length scales
- Opportunities/flexibility to bring in new partners or groups
- Powerful index (improving)
- Data quality control
    - Authorized entity on data quality evaluation and grade
    - Comments from users
    - Synthesis methods and conditioning
- Can users post a need?
- Connection to different value chain entities for inputs (not just competing organizations) → Advisory board
- Allow visiting researchers

## Steady State

- Journal ecosystem (ML)
- Sharable code/paper/data (Jupyter notebook hub)
- Library ecosystem
- Resources for
    - Initial training
    - Sustainable training
- Workforce training:
    - Existing employees
    - New employees
- Organize conferences/workshops
- Should have educational component
- Can it exist without any funding sponsor?

## Strategies for a Self-Sustaining Collaboratory

**FOCUS QUESTION 2:** *How can the Collaboratory become self-sustaining, and what is the financial sustainability model?*

Table A-8. Expansion of the Collaboratory Network for Energy Materials Discovery and Process Design – Strategies for Self-Sustainment

| Strategies for Becoming Self-Sustaining |
|---|

- Initiate HTE-MC "branded" projects and proposals based on existing funding
- Makeup: GL/universities/industry → engagement benefactors
- Build trust
- EAG, Intermolecular (companies) and multi-user facilities
- Subscription: Pay money, loan (free trial), pay data
- Name association with the data → encourage participation and collaboration
- Easy data submission (all data formats including video and figure)
- Understand the value proposition for all members
    - Why would researchers contribute to the collaboratory in the long run?
- Money
    - No money, no work
    - Volunteering: ok
- Government funding to get launched; industrial utilization for sustaining funds
- Lab/academia data input to create library
- "Pay fees" through:
    - Ideas of what to measure
    - Shorter data embargo
    - Sponsor of capabilities
- Funding: Basic research or applied research side of government?
- Should generate new knowledge
    - Results speak for themselves. Drivers: cost/benefits

## Steps Forward

**FOCUS QUESTION 3:** *What steps, actions items and activities should be undertaken to formalize and expand the Collaboratory, and potentially what organizations should participate/lead these activities?*

Table A-9. Expansion of the Collaboratory Network for Energy Materials Discovery and Process Design – Steps Forward

| Steps Forward |
|---|

- Have a list of already established collaborators
- Map who is doing what, who is interested in what

**Table A-9. Expansion of the Collaboratory Network for Energy Materials Discovery and Process Design – Steps Forward**

- Form the “core” group
  - Invite comments
  - Finalize an organizational chart
  - Keep us engaged/motivated
- Lobby (DOD, DOE, Congress) to publish a FOA to start this collaborative
- Specific projects to define, specify, and create prototype components
- Show tax payer value
  - Comes from industry
- Demonstration of ML and DOE user facility → faster material discovery
- Share capability statement among participants
- Identify and write proposals to federal FOAs
- 1) Expand agreements/contracts within national lab system; 2) involve big corporations; 3) university and small businesses (smaller fee)
- 1) Publish outcome of this meeting; 2) hold a public meeting to solicit broader inputs
- Determine the order of magnitude that is needed in terms of funding; explain to decision makers
- How to support collaboration through funding agencies before formal funding/seed funding (first seed funding)
- Identify a relevant problem a sponsor would fund or find an example of something already done
- Beyond the workshop report, form a core group to develop a communication piece, in layman language

# Breakout Group 4: Integration of AI, ML, and Experimentation for Energy Materials Design and Processing

## Goals and Targets

**FOCUS QUESTION 1:** *What goals and targets (in the near term: 1-2 years, medium term: 3-5 years, and long term >5 years), would be appropriate for integrating AI and ML to improve materials design and processing for high throughput systems?*

Table A-10. Integration of AI, ML, and Experimentation for Energy Materials Design and Processing – Goals and Targets

| Overarching Goals and Targets |
|---|
| • Consistent experimentation and data management for accessibility<br>• Model development using uncertainty principles and prediction for design and processing<br>• Improved ML using AI frameworks for use with robotics and process automation<br>• Better definitions in workflow and interfaces from experimentation to ML<br>• New requirements for education and training of material scientists in AI/ML<br>• Advanced design of specific materials using AI/ML |

Table A-11. Integration of AI, ML, and Experimentation for Energy Materials Design and Processing – Goals and Targets Near Term (<2 years)

| Experimentation and Data Management |
|---|
| • Determining the optimal strategy for integrating data with a physics-based model<br>  ○ A physics-based model that is even partially useful would be good for forward prediction.<br>  ○ Inputting data into a theoretical approach requires an inverse design.<br>• Data representation for ML and ability to capture the content of all data<br>• For fields such as natural language and computer vision, there are fundamental representations that implicitly code these areas. Develop fundamental representation for materials science.<br>• Systematic (or optimal) algorithms using AI for length-time scale bridging of models<br>  ○ Examine straightforward processes to incorporate scale bridging in additive manufacturing for modeling at the macroscopic level<br>• Autonomous high-throughput characterization of “hard to measure” properties<br>• Benchmark datasets (real world problems)<br>  ○ A community of ML usable images is needed. If the data sets and storage are open, then the (materials science) community will work with the data.<br>  ○ Datasets need to be defined and a determination made of what can be reproduced.<br>  ○ Proof-of-integration at the subscale to be generated to improve confidence in the data.<br>  ○ Technically, at a high level, there are different tools and data processing to be used with robotics. |

  - Connected data length. There are a lot of tools from the community. Data lengths: put it in and parse it out, otherwise it fails.
- Data lakes (schema on read)
- Develop more proof of concepts for integration for closing the loop
- Along the path from experiments to measurements to data to information to labeled data (knowledge) to ML, research goes from humans (experiments, measurements) to without humans (ML)
  - Machine learning does not come from experiments; they are not connected. Knowledge of available data is learned. For integration to happen, background data needs to be converted into available information without humans.
  - The first step is to generate new knowledge on the fly.
- Bulk synthesis (high throughput)
  - More scientists are needed to work on the HTE so real-time feedback can be achieved.
- Constitutive laws for “new” materials using ML from near-term experimental data
  - The major ingredient in any finite element law is the conservation law and the correct conservative laws should be developed and used.
- Develop API standard for machine / experiment integration
  - Standardized API for a machine to be able to talk to experimental data, especially if it is dealing with experiments in different parts of the world simultaneously. This could be done in 0 – 2 years.
- Adoption of data interchange practices / requirements and standards in experiments and beyond
  - There are standards that already exist for data exchange developed over past 5 years that are not being used, and their use should be enforced. In this way, there will be enough data to train AI with machine learning. This should be done until there is enough data in the public domain. There is not enough now.

**Model Development**

- Model development for design and processing
  - There are few models that handle frames, such as a magnet, with specific qualities for high-temperature applications. A frame could be selected, and then new materials forecasted with their intrinsic properties.
  - Determine whether there are suitable models for greater predictive capabilities that can translate into helping experimentalists to design processes more easily. Modeling lags in this area and should be a target to develop in a few years.
- Effective means of incorporating context into models and representations
  - Not all predictions are true under experimental conditions. The types of results obtained, for instance, temperature and pressure, can be difficult to incorporate into ML.
- Conduct initial training by models that translate to experimental data
  - Initial training is needed in AI with simulation tools. Determine how to build AI in a simulated environment.
- Develop programming languages for lab experiments and manufacturing that go from compiler to bench, robots, etc.
  - Better programming languages and compilers

  - An ecosystem of different types of languages is needed and a method to resolve them. The algorithms in a program could be described as a recipe that can be linked to a set of instructions that a person or machine could perform. Once the language description is obtained, it can compile.
  - There are a lot of existing tools. Add domain-specific components and use an interface that, in certain applications, could be developed in six months.
  - Develop a new paradigm and build into fundamental methods for machine language. Write the algorithm that is simple enough to connect.
- Develop domain-specific components and interfaces

**ML, Robotics and Automation**

- ML has been evolving and has had recent success in image processing and speech recognition.
- Hybrid AI framework, ML, and automated reasoning: 1. feasible data; 2. complex knowledge
  - Process modeling should be a hybrid that is structured whereby ML goes into components that are connected into another part of AI reasoning. A hybrid could have ML as a modeling part with automated reasoning to give a physical system for ML.
- Robotics for flexible automation shop for data collection: 1. collaborative; 2. intelligent; 3. autonomous
- Generate data for ML, which is very data intensive.
  - People have been doing ML to discover catalysts for decades. Scientists need to learn what is out there and be inclusive in whatever is available. This is not just data mining but includes methodologies that have been developed in different fields that can be leveraged.
  - For a simple ML program to recognize a single room with a chair and desk requires 12,000 images, which is too much data.

**Work Flow and Interfaces**

- Develop high-level workflow / pipeline system / language for experiments, data, processing, ML
- Articulate case studies with performance exceeding state-of-the-art. Success examples are needed.
- Test cases demonstrating proof-of-concept that machine learning and AI can improve
  - For catalysis, there is an entire building for demonstrating proof-of-concept, but it is not widely known.
  - Algorithms based upon decision theory have been used in industry for many years including in automobile design.
  - Decision-making optimization models have been developed to precisely guide the next calculation.
  - Researchers are borrowing and importing ideas from cancer genomics.

**Table A-12. Integration of AI, ML, and Experimentation for Energy Materials Design and Processing – Goals and Targets Medium Term (3-5 years)**

### Experimentation and Data Management

- Journals require machine-readable experimental plans for publication (GitHub for experiments)
- Re-envision a DOE facility as an API / cloud service
- Integration of computational and experimental databases for inorganic materials
  - Autonomous synthesis / processing experiments driven by AI for inorganic materials (much exists for organic materials)
- Adaptive characterization techniques (for material synthesis)
- Identify bottlenecks to target efficiency improvement
  - Determine specific cases where adaptive scanning can predict where in-depth scans are needed and generalize to see where there are bottlenecks
- Multi-objective and multi-fidelity experimentation
  - Determine how the design can be developed given conflicting objectives
- Integration of sparse / noisy legacy data with high-throughput data and high-throughput computational results
  - How should data be used to bridge scales? In the near-term, development of conservative laws but at the longer-term, development from the atomistic level.
- Autonomous synthesis and processing experiments incorporating physics-based models with AI
- Fully autonomous platforms for rapid data processing

### Model Development

- Fast and accurate approximations of physics-based models
- Identify domain of applicability of models and representations
  - Identify the domain of applicability; "democratization"
  - Not all experiments are created equal, so doing something on one set of equipment and something else on another set, these differences need to be standardized.
- Rapid identification of material properties from experimental and modeling outputs
- Probabilistic active-learning algorithms for ML
- Enhance the performance of existing models
- Probabilistic ML for active learning
- ML tools for production setting (e.g., working on big data)
- Interpretable ML models
  - In a fully autonomous platform, it should be rapid (minutes or hours, not days).

### ML for Process Design Applications

- AI for process control
- Ability to tune process parameters for desired microstructure for any manufactured material

- Determine the process parameters to make sure that what is printed is what is desired. It is a collection of experimental and modeling outputs that can be used to tune any material to what is required for a given application.
- Synthesizability – combine catalyst development with process conditions to predict 50% faster
- Reduction of criticality of materials by 50%
  - Aids with natural resource security (less reliance on imports for critical materials)
  - In 5 years, develop method for materials processing in a plant, i.e. materials design, discovery, and processing with the help of ML to reduce characterization costs by 50 to 90%.
- Active AI/ML projects in diverse applications (e.g., high-temperature alloys, solar catalysis, vehicle light-weighting, etc.)

**Education and Training**

- Process specific AI program leading to a domain expert
- Manpower development for integration theory
- Funded research and training centers for HTE/AI/ML integration (education)
  - Research and training centers for career development focusing on principles, tools and coursework that integrate AI, ML and HTE to train people
  - Have 3 to 5 centers for integrative research and training.
- Lower entry barrier by developing training standards and languages
- Conduct two workshops to identify pedagogy and principles of HTE and integration with AI
  - Integrate academic fields. Students can get a minor in fields such as data analytics, but there is nothing corresponding to that in materials science.
  - Materials science needs to define foundational principles based on broad translatable principles.
  - Education modules need to be developed that can be made available online and teach the fundamentals to everyone from the community college level to Ph.D. level.
- Five minor or certificate programs in HTE/AI/ML integration nationally
  - Universities with minors or certificate programs (or possibly online training)

**Table A-13. Integration of AI, ML, and Experimentation for Energy Materials Design and Processing – Goals and Targets Long Term (>5 years)**

**Experimentation and Data Management**

- AI-driven discovery of new multifunctional materials (e.g., mechanical and electrical properties)
- More combinatoric experiments and new methods (hierarchical)
- Standardization of autonomous experimental platforms
- Develop constitutive laws from atomistic or other fine-scale simulations; bridging of time and length scales. This could be either a medium-term or long-term goal.
- Reduction of experimental time by 90% in 5+ years

- Property enhancement (factor increase depends on property); some cases might be small, some be orders of magnitude
  - Given a property, AI identifies best system and growth parameters
- Autonomous decision framework (incremental) (library)
  - Scientists can synthesize thousands of compounds, but there are no tools to predict their catalytic performance. Such tools are needed that can at least predict this performance.
  - It is a very process-intensive effort to integrate into the final product. This would reduce development time significantly.

**Model Development**

- Tools to predict chemical and physical performance of new multi-functional materials compared to market products
- User-friendly graphical interface for comparing multiple systems
- Development time for new models are less than 3 years
- One common platform for all materials

**ML for Process Design Applications**

- Active learning during the discovery-manufacturing cycle to drive process measurements robotically
- ML process design that could lead to energy efficiency savings on the order of 1 quadrillion Btu
- AI/ML to optimize HTE during materials discovery phase for intensive, process manufactured materials integration; increase discovery and development time by 10x
- AI/ML to “predict” process performance during material discovery stage; 10x to 100x reduction in discovery time
- Provide case study examples of manufacturing processes improved by HTE/AI method.
- Real-time feedback with bulk synthesis /characterization
  - Data itself is not very meaningful whereas knowledge is. If an experiment can be done cheaply and generate a lot of data, then a high-fidelity model could be developed.
  - Using uneven quality legacy data, can researchers add to it and not get biased results?
- “Fabless” materials company ecosystem with less than $20,000 for startup
  - Create materials catalyst company that would enable buying lab time for less than $20,000 and selling products.
  - An example is hardware electronics, which allows technologies to be fabricated at low cost. This needs to be done for materials as well.
- Close the cycle – active learning to drive next measurements robotically

## High Priority Areas

**FOCUS QUESTION 2:** *What are the high priority areas for expanding the use of integrated AI, ML and experimentation for energy materials design and processing?*

Table A-14. Integration of AI, ML, and Experimentation for Energy Materials Design and Processing – High Priority Areas

**Experimentation and Data Management**

- Intelligent integration applied to theory and experiment ●●●●●●● (7)
- Length & time scale bridging ●●● (3)
- Standardize data ●● (2)
- Interpretability ●● (2)
- Predictive models ●● (2)
- Identifying strengths and weaknesses of AI in materials development and assessment ●● (2)
- “Easy” experiment ● (1)
- Use of AI for design of experiments for hard-to-measure materials properties ● (1)
- Predicting materials performance in integrated product during discovery phase ● (1)
- Strategies for extracting labeled knowledge from experimental observations ● (1)
- Rapid data reduction and analysis for true operando characterization ● (1)
- Predicting material composition to meet process performance requirements
- Ecosystem of experimental services on demanded and easy data sources
- Utilization of ML for truly predictive models and not just validation
- Standard datasets

**Model development**

- Addressing uncertainty (modeling, data) ●●●●●●●●●●● (11)
- Develop open tools ●●●●●●● (7)
- Combine tools to predict synthesizability ●●●● (4)
- Develop specialized, key tools for enhancing materials discovery and design

**Materials**

- Strongly correlated oxides ●●●● (4)
- Improvement of metal additive material quality and performance through ML ●●● (3)
- Materials for energy storage and sustainability ●●● (3)
- Thin films and coatings for energy applications (not bulk; not composites) ●●● (3)
- Magnets ●● (2)
- Application of HTE/AI to inorganic materials manufacturing (not molecules, catalysts, polymers, steel as alloys) ● (1)

Table A-14. Integration of AI, ML, and Experimentation for Energy Materials Design and Processing – High Priority Areas

**Education and Training**

- Education / training ●●●●●●●●●●● (11)
  - Develop education framework for tools
  - Priority education – human talent for developers and users
- Computer scientists and materials scientists (experimental and computational) working together ●●●● (4)

**Machine Learning for Process Design Applications**

- Context-aware data representations ●● (2)
- Collaborative and intelligent robots for automated data collection ●● (2)
- AI (ML and reasoning) frameworks capable of complex process modeling ● (1)
- Strategies for capturing processing conditions ● (1)
- Process representations ● (1)

## Barriers and Challenges

**FOCUS QUESTION 3:** *What are the barriers and challenges to developing the needed science or technologies to meet the high priorities identified above?*

Table A-15. Integration of AI, ML, and Experimentation for Energy Materials Design and Processing – Barriers and Challenges

**Experimentation and Data Management**

- Machine readable open data reproducibility sharing ●●●●●● (6)
- Including "failed" data in datasets used for ML ●●●●●● (6)
- Obtaining HTE when overall process / experiment is slow and expensive ●●●●● (5)
- Addressing intellectual property (IP) issues ●●●●● (5)
- When addressing a scientific problem, focusing on science and not the tool (ML is a tool) ●●●● (4)
- Developing more useful scientific publications to allow data extraction ●●● (3)
- Proprietary control software (closed) ●●● (3)
- Standardizing data across many teams (or using ML to learn how to standardize it for us) ●● (2)
- Integration with experiments ●● (2)
- Cultural shifts ●● (2)
- High-throughput measurement techniques ● (1)
- Standardized data and standardized experiment access and competition; unified benchmark ● (1)

**Table A-15. Integration of AI, ML, and Experimentation for Energy Materials Design and Processing – Barriers and Challenges**

- Difficulty of getting quantity of data required for AI ● (1)
- Lack of policy incentives from funding agencies to execute data management plans ● (1)
- Complex information management
- Theory versus high-throughput experiment (in catalysis)
- Lack of reproducible experimental procedures in literature

**Education and Training**

- Integrating diverse expertise ●●●●●●●●●●● (11)
- Lack of human power (postdocs) to solve the complex problems ●●●●●●●●● (9)
- Sociological – integration of experimentalists with theorists with statisticians (data scientists) ●● (2)
- Developers and users are siloed in specific domains or applications ● (1)
- Cross disciplinary skill sets (lack of)

**Machine Learning for Process Design Applications**

- Integrating ML and AI into high-throughput experiments ●●● (3)
- Making optimal AI frameworks combining ML and other methods for feasible AI ●● (2)
- Force-fitting of ML/AI to problems for which it is not suitable ●●● (3)
  - Focus on one or two areas from many possible energy materials ● (1)

**Model Development**

- Real time data analysis does not exist. Move models from ideal systems to real systems ●●●● (4)
  - Lack of accurate models ● (1)

# Appendix B. Agenda

## A Multi-Agency, Multi-Year Program Plan in Advanced Energy Materials Discovery, Development, and Process Design

*Utilizing High-Throughput Experimental Methods, Artificial Intelligence, Autonomous Systems, and a Collaboratory Network*

National Institute of Standards and Technology (NIST) Campus
100 Bureau Drive
Gaithersburg, MD 20899

Portrait Room
Administration Building 101

### Agenda

| Thursday, July 12 | |
|---|---|
| 7:30 – 8:30 am | **Registration and Continental Breakfast** |
| 8:30 – 8:45 am | **Opening Remarks**<br>Rob Ivester, AMO Director & Eric Lin, Director, Material Measurement Laboratory |
| 8:45 -9:15 am | **Summary of Findings: 2017 DOE Workshop on Artificial Intelligence Applied to Materials Discovery and Design**<br>Brian Valentine, AMO Technology Manager |
| 9:15 – 9:45 am | **Summary of Findings: 2018 NIST Workshop on High-Throughput Experimental Materials Collaboratory (HTE-MC)**<br>Marty Green, NIST |
| 9:45 – 10:00 am | **BREAK** |
| 10:00 – 10:30 am | **Overview: DOE AMO Multi-Year Program Plan (MYPP) Development**<br>Rob Ivester, AMO Director |
| 10:30 – 10:45 am | **AMO MYPP: Moving Applications of AI and ML from Materials Design and Discovery through Process Design and Development**<br>Brian Valentine, AMO Technology Manager |
| 10:45 – 11:15 am | **Summary of Findings: 2018 Mission Innovation Workshop on Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods with Artificial Intelligence**<br>Professor Joshua Schrier, Associate Professor of Chemistry, Haverford College |
| 11:15 – 11:50 am | **Invited Presentation: Materials Discovery at Solvay : Applying AI Tools to 150 Years of Historical Data**<br>Dr. Jean-Yves Delannoy, Manager of Statistical Modeling and Artificial Intelligence Team, Solvay Chemicals, Inc. |
| 11:50 – noon | **Breakout Summary: Assignments and Objectives of Each Session** |
| 12:00 – 1:00 pm | **LUNCH** (provided) |
| 1:00 – 3:30 pm | **Facilitated Breakout Sessions:**<br>1. Priorities in Energy Materials R&D: Barriers, Timeline, and Metrics<br>2. Database infrastructure needs in AI and Energy Materials R&D: Moving Materials Discovery through Materials Processes<br>3. Expansion of the Collaboratory Network for Energy Materials Discovery and Process Design<br>4. Integration of AI, ML, and Experimentation for Energy Materials Design and Processing |
| 3:30 – 3:45 pm | **BREAK** (and networking) |
| 3:45 – 4:30 pm | **Summary of Breakout Sessions**<br>Representatives from Each Group |
| 4:30 – 5:00 pm | **Summary of Next Steps**<br>Brian Valentine, AMO<br>Martin Green, NIST |
| 5:00 pm | **ADJOURN** |

# Appendix C. Workshop Participants

| Name | Organization |
|---|---|
| Alexander Abboud | Idaho National Laboratory |
| Assaf Anderson | MaterialsZone |
| Matt Antes | Energetics |
| Donald Anton | Savannah River National Laboratory |
| Mark Bailey | Wildcat Discovery Technologies |
| Prasanna Balachandran | University of Virginia |
| Robert Bartolo | Transformational Liaisons (TRL), LLC |
| Ram Bhagat | AECOM |
| Satyaveda Bharath | PHMSA/Department of Transportation |
| Isaac Chan | DOE Advanced Manufacturing Office |
| Lei Cheng | Argonne National Laboratory |
| Fred Crowson | Energetics |
| Jean Yves Delannoy | Solvay |
| Laura Fabeny | Allegheny Science & Technology |
| Michael Gao | National Energy Technology Laboratory |
| Andrew Gellman | Carnegie Mellon University |
| Michael Greenwood | Natural Resources Canada |
| James Greer | PVD Products |
| Varun Gupta | Pacific Northwest National Laboratory |
| Vipul Gupta | GE Global Research |
| Anna Hiszpanski | Lawrence Livermore National Laboratory |
| Mikel Holcomb | West Virginia University |
| Huilong Hou | University of Maryland, College Park |
| Gabriel Ilevbare | Idaho National Laboratory |
| Robert Ivester | DOE Advanced Manufacturing Office |
| Theodore Krause | Argonne National Laboratory |
| Christoph Kreisbeck | Harvard Universtity |
| Gregory Krumdick | Argonne National Laboratory |
| Oh-Hun Kwon | Saint-Gobain |
| Anton Lauterbach | University of South Carolina |
| Xiaodong Li | University of Virginia |
| Kenneth Lipkowitz | Office of Naval Research |
| Turab Lookman | Los Alamos National Laboratory |
| Tommi Makila | Energetics |
| Robert Maxwell | Lawrence Livermore National Laboratory |
| Apurva Mehta | SLAC National Accelerator Laboratory |

| Name | Organization |
|---|---|
| Carson Meredith | Georgia Tech |
| Scott Morgan | Energetics |
| George Muntean | Pacific Northwest National Laboratory |
| Yellapu Murty | University of Virginia |
| Ryan Ott | Ames Laboratory |
| Stefanos Papanikolaou | West Virginia University |
| Jake Parduhn | McAllister & Quinn |
| Young Soo Park | Argonne National Laboratory |
| Durga Paudyal | Ames Laboratory |
| John Perkins | National Renewable Energy Laboratory |
| Ellen Piccioli | Worcester Polytechnic Institute |
| Sudarsan Rachuri | DOE Advanced Manufacturing Office |
| Eli Rotenberg | Lawrence Berkeley National Laboratory |
| Erdem Sasmaz | University of South Carolina |
| Joshua Schrier | Fordham University / Haverford College |
| Lawrence Scipioni | PVD Products |
| Barak Sela | MaterialsZone |
| Somnath Sengupta | Powerhouse Consulting Group |
| Dennis Sheberla | Harvard University |
| Kenneth Smith | United Technologies Research Center |
| Joshuah Stolaroff | Lawrence Livermore National Laboratory |
| Paul Syers | DOE Advanced Manufacturing Office |
| Anjana Talapatra | Texas A&M University |
| Emmanuel Taylor | Energetics |
| Richard Todaro | Allegheny Science & Technology Corporation |
| Zachary Trautt | National Institute of Standards and Technology |
| Dallas Trinkle | University of Illinois, Urbana-Champaign |
| Alexander Umantsev | Fayetteville State University |
| Brian Valentine | DOE Advanced Manufacturing Office |
| Anthony van Buuren | Lawrence Livermore National Laboratory |
| Bogdan Vernescu | Worcester Polytechnic Institute |
| Jeff Vervlied | Superior Graphite |
| kaisheng wu | Thermo-Calc Software Inc. |
| Baoxing Xu | University of Virginia |
| Andriy Zakutayev | National Renewable Energy Laboratory |
| Yuepeng Zhang | Argonne National Laboratory |
| Stacey Young | The Building People |

U.S. DEPARTMENT OF
ENERGY

Office of
ENERGY EFFICIENCY &
RENEWABLE ENERGY

For more information, visit:
energy.gov/eere/amo
DOE/EE-1867 ▪ October 2018