# MATERIALS ACCELERATION PLATFORM

Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods with Artificial Intelligence

Mission Innovation

January 2018

SENER
SECRETARÍA DE ENERGÍA

U.S. DEPARTMENT OF ENERGY

CIFAR
CANADIAN INSTITUTE FOR ADVANCED RESEARCH

"At its essence, the generation of the electricity that powers our progress is inextricably tied to chemistry and physics. The creation of the electron is fundamentally about the materials we deploy in converting of one form of 'raw' energy into finished electrons. How efficiently can we burn fossil hydrocarbons, or convert both wind currents and photons into electrons? And how efficiently can we store and transport them? Much of this ultimately boils down to the science of the materials we use, and whether we can improve the basic desired properties of those materials."

— "How Supercomputers Will Help To Accelerate The Emerging Energy Economy", Forbes Magazine, December 13, 2017

Materials Acceleration Platform

Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods with Artificial Intelligence

Report of the Clean Energy Materials Innovation Challenge Expert Workshop

Co-chairs:

Alán Aspuru-Guzik, Harvard University
Kristin Persson, Lawrence Berkeley National Laboratory and University of California – Berkeley

Executive Chair:

Hermann Tribukait-Vasconcelos, Mexico Energy R&D Innovation Funds
and Mission Innovation (Mexico)

This workshop, held in Mexico City, was co-led by the Mexican Ministry of Energy (SENER)
and the U.S. Department of Energy (US DOE) in partnership with
the Canadian Institute for Advanced Research (CIFAR).

Mission Innovation

January 2018

SENER
SECRETARÍA DE ENERGÍA

ESTADOS UNIDOS MEXICANOS

U.S. DEPARTMENT OF
ENERGY

CIFAR
CANADIAN
INSTITUTE
FOR
ADVANCED
RESEARCH

Contents

Executive Summary

Materials are an essential element of advanced energy technologies. Accelerating the discovery of new materials, and the associated research required for maturing these technologies into deployment, will require a multidisciplinary and international effort that brings together a wide variety of individuals working effectively across their specialties, as well as across sector and political boundaries. It will also require a radical departure from traditional forms of discovery.

The materials discovery process involves several stages summarized as conception, synthesis, and testing or characterization. Characterization encompasses the measurement of key properties of the material, followed by its incorporation into active devices to evaluate interactions with other device components and assess overall performance metrics. These steps have usually been carried out sequentially, and therefore, only a few materials can be tested at a time. Furthermore, sophisticated tools such as artificial intelligence (AI), large computational resources, and automated robotic systems have not been widely employed yet.

Recognizing the challenges and opportunities associated with materials discovery, Mission Innovation established the Clean Energy Materials Innovation Challenge and hosted its first international expert deep-dive workshop in Mexico City on Sept 11–14, 2017.[1] Leading scientists from throughout the world gathered to define the challenges, opportunities, and fundamental research needs related to materials discovery.

The main recommendation coming from the workshop participants is the need to develop the materials acceleration platform(s) (MAPs), which integrate automated robotic machinery with rapid characterization and AI to accelerate the pace of discovery. The deployment of the proposed

---

[1] Mission Innovation is a global initiative comprising 22 countries and the European Union that share the goal of accelerating clean energy innovation.

acceleration platforms will unleash a "Moore's law for scientific discovery" that will 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 an Edisonian 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. This, in turn, will benefit all seven Innovation Challenges of Mission Innovation.

Workshop participants identified six key priority research areas that comprise the MAP elements. These scientific and engineering challenges require the expertise of multidisciplinary, international teams for their successful implementation. These research areas include the development of the following tools, research activities and infrastructure:

1. "Self-driving laboratories" that design, perform and interpret experiments in an automated way;

2. The development of specific forms of AI for materials discovery;

3. Modular materials robotics platforms that can be assemblies of modular building blocks for synthesis and characterization;

4. Further research into computational methods for inverse design;

5. New methodologies for bridging the length and timescales associated with materials simulation; and

6. Sophisticated data infrastructure and interchange platforms.

These six MAP areas can be incorporated in facilities and implemented by collaborative teams. The resulting "autonomous materials discovery factories" could serve multiple and diverse international research groups in academia and industry in their materials discovery pipeline.

The workshop participants emphasized the need to develop multidisciplinary international teams of scientists and engineers that collectively have expertise in chemistry, materials science, advanced computing, robotics and AI, amongst other disciplines. Deep international collaborations and long-term support are also necessary to make MAPs a reality. The output of these platforms promises to accelerate the development of novel disruptive technologies that will in turn further the global transition to clean energy that can be deployed broadly in diverse settings.

1 Introduction

Materials discovery and development crosscut the entire energy technology portfolio, from energy generation and storage to delivery and end use. Materials are the foundation of every clean energy innovation: advanced batteries, solar cells, low-energy semiconductors, thermal storage, coatings, and catalysts for the conversion, capture, and use of CO_2. In short, new materials constitute one of the cornerstones for the global transition to a low-carbon future.

"There are many paths to cheap zero CO_2 free energy, what we need to do is fund the wild scientists who are looking at the early stages of these problems...I think we will get increased investments from government and private sector, I know that the scientific possibilities are pretty incredible."

– Bill Gates [1]

The process of discovering and developing new materials currently entails considerable time, effort, and expense. Each newly discovered
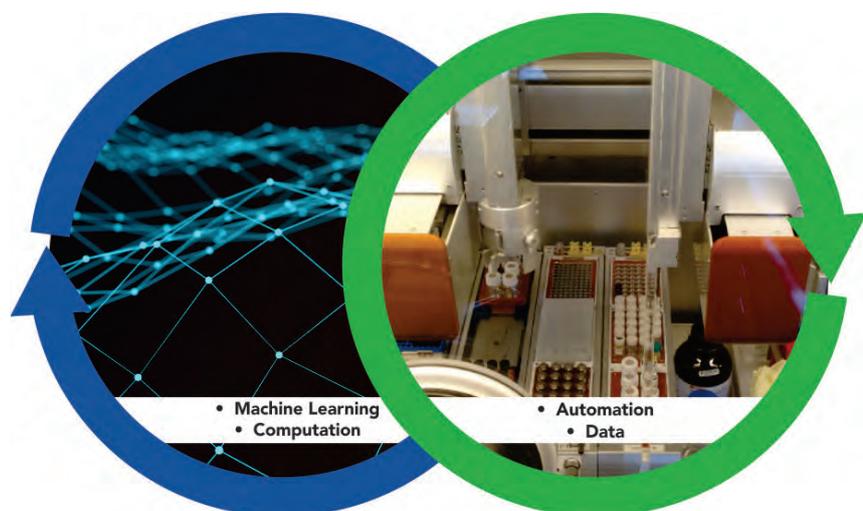

Figure 1.1 Virtuous cycle: Laboratory automation and data collection enable machine learning algorithms and computational methods to be more effective at planning experiments.

molecule is run through simulation, synthesis, and characterization, with synthetic procedures taking from 10 to 20 years at a very high cost. Materials discovery and development, however, are at the cusp of a transformational change that could reduce the time to design, optimize, and discover new materials by at least 10 times, cutting it down to one or two years. This enormous opportunity — and the associated technical challenges of realizing it — inspired the members of the global Mission Innovation (MI) initiative to launch the Clean Energy Materials Innovation Challenge. As a first step, international experts attended a technical workshop in Mexico City on September 11-14, 2017 to identify and explore the research and development (R&D) challenges, and the most promising breakthrough opportunities for accelerating the materials discovery process, with a long-term view towards 2030 and beyond.

The integrated materials innovation approach developed at this experts workshop, the Materials Acceleration Platform, envisages a Moore's law for research, where the rate of research doubles every two years [2]. This acceleration would result from leveraging emerging capabilities in next-generation computing, artificial intelligence (AI) and machine learning, and robotics. The reduced timeline would dramatically accelerate advances across the entire spectrum of clean energy technologies and generate widespread social and economic benefits, transforming industries beyond the energy sector. Major changes are already underway. The paradigm shift in materials design through the automation of computations over the last two decades has led to accelerated understanding and design of novel, high-performance materials [3,4].

Workshop Participants

Workshop participants shared a common interest in accelerated materials innovation. They represented a broad mix of expertise spanning advanced theory, applied physical chemistry and materials sciences, advanced computing, machine learning, and robotics. The workshop drew 133 attendees:

- 55 professors and scientists from top universities and research institutions;
- 6 keynote speakers and panellists, including Nobel Laureate Dr. Mario Molina and CIFAR President and CEO Alan Bernstein;
- 16 MI member governments represented: Australia, Canada, Denmark, Finland, France, Germany, European Union, India, Italy, Korea, Mexico, Netherlands, Norway, Saudi Arabia, United Kingdom, and United States;
- affiliates of Mexico- and U.S.-based universities, groups, labs, and companies;
- graduate students and postdoctoral researchers; and
- observers from the public and private sectors.

Attendees came from 17 countries: Australia, Belgium, Canada, Denmark, Finland, France, Germany, India, Italy, Korea, Mexico, Netherlands, Norway, Saudi Arabia, Switzerland, United Kingdom, and United States.

Computation and design, however, are only the first step in bringing novel materials to market. Materials synthesis and characterization have yet to benefit from automation and accelerated learning on a large scale. Integrating synthesis and characterization with advanced computing, machine learning, and robotics would automate the entire materials discovery process. Finally, closing the loop to create a virtuous cycle by using AI to direct experimentation and simulation for optimal learning would result in an accelerated, comprehensive, end-to-end materials innovation platform (Figure 1.1).

1.1 ABOUT MISSION INNOVATION AND THE INNOVATION CHALLENGES

"Although each [Mission Innovation] member will decide on priorities for its own expenditure, each challenge will be promoted through international conferences and workshops, and it is anticipated that many cross-country projects will be funded. The creativity and ingenuity that will flow from these projects will produce market-facing solutions through private finance and the engagement of private companies. It is critically important that the appropriate members of the research community are fully involved in these developments over the coming years."

– Sir David King [9]

Transforming traditional materials discovery pipelines into an integrated platform requires commitments from governments, academic research institutions, large industries, and capital providers [5]. This ambitious effort fits well in the framework of MI Innovation Challenges.

MI is a global initiative of 22 countries and the European Union with the goal of accelerating clean energy innovation. Participating countries have committed to seek to double their governments' investments in clean energy R&D over five years. They have also agreed to encourage greater levels of private-sector investment in transformative clean energy technologies and participate in an information sharing system to promote transparency, engage stakeholders, spur identification of collaborative opportunities, and provide the private sector with actionable information. These efforts will increase the availability of the advanced technologies that will accelerate the transition to a global energy mix that is clean, affordable, and reliable.

Figure 1.2 Workshop participants and observers.

MI was announced on November 30, 2015, as world leaders came together at the 21st Conference of the Parties (COP21) in Paris to undertake ambitious efforts to combat climate change. Clean energy innovations and disruptive technological breakthroughs are an essential element of broader efforts to significantly reduce greenhouse gas (GHG) emissions to limit the increase of the global average temperature to less than 2°C [6,7].

A year later, in November 2016 at COP22 in Marrakech, MI members launched seven Innovation Challenges, which are global calls to action aimed at accelerating research, development, and demonstration in specific technology areas where increased international attention would make a significant contribution to meeting the goals of the challenges. The following Innovation Challenges were selected and developed through a collaborative process in which policy and technical experts from all MI members presented proposals and exchanged information about national needs and priorities:

1. Smart Grids: To enable future grids that are powered by affordable, reliable, decentralized renewable electricity systems.

2. Off-Grid Access to Electricity: To develop systems that enable off-grid households and communities to access affordable and reliable renewable electricity.

3. Carbon Capture: To enable near-zero CO_2 emissions from power plants and carbon intensive industries.

4. Sustainable Biofuels: To develop ways to produce, at scale, widely affordable, advanced biofuels for transportation and industrial applications.

5. Converting Sunlight: To discover affordable ways to convert sunlight into storable solar fuels.

Figure 1.3 (Top) Workshop attendees at the plenary session on the first day. (Bottom) (From left to right) Dr. Paul Durrant, Dr. Horst Simon, Dr. Alan Bernstein, and Dr. Mario Molina participate in a panel moderated by Dr. Hermann Tribukait.

Figure 1.4 Through interactive workshop sessions, participants identify and discuss critical R&D priorities and gaps.

Figure 1.5 The writers' team discusses the findings and structure of the workshop report during a writing session at UNAM.

6. Clean Energy Materials: To accelerate the exploration, discovery, and use of new high-performance, low-cost clean energy materials.

7. Affordable Heating and Cooling of Buildings: To make low-carbon heating and cooling affordable for everyone.

A voluntary coalition of participating MI members, under the co-leadership of two to four countries, advances each Innovation Challenge through detailed work programs.

Mexico proposed and now leads the Clean Energy Materials Innovation Challenge, with the United States as co-lead. Participating MI members include Australia, Canada, Denmark, European Commission, Finland, France, Germany, India, Italy, Republic of Korea, Netherlands, Norway, Saudi Arabia, Sweden, United Arab Emirates, and United Kingdom [8].

"Perhaps of greatest interest to the theoretical physics, physical chemistry, and materials science communities that are working alongside the machine learning, robotics, and next-generation computing communities is the challenge of developing clean energy materials. The goal is to provide an integrated end-to-end materials innovation approach, or platform, to deliver the mix of solutions...Mission Innovation is intended to spur the interest of the creative community in what is now the most urgent series of demands facing humanity. The opportunities are, simply, immense."

– Sir David King [9]

1.2 CLEAN ENERGY MATERIALS INNOVATION CHALLENGE: WORKSHOP SUMMARY

The Mexican Ministry of Energy (SENER) and U.S. Department of Energy (DOE), in partnership with the Canadian Institute for Advanced Research (CIFAR), sponsored the Clean Energy Materials Innovation Challenge workshop. Professor Alán Aspuru-Guzik of Harvard University and Professor Kristin Persson of University of California-Berkeley co-chaired the workshop. Professor Carlos Amador-Bedolla of Universidad Nacional Autónoma de México (UNAM) was the local chair, and Dr. Hermann Tribukait, Mexico's Energy Innovation Funds Ambassador, the workshop Executive Chair and MI lead (Mexico).

The meeting followed a workshop methodology based on the proven model of U.S. DOE Basic Research Needs workshops. Adapted for an international setting, MI expert workshops help countries build robust domestic energy research portfolios by identifying and discussing high-priority basic research needs or other technical areas that are ripe for further investigation and

Figure 1.6 Workshop report writers' team meeting at UNAM. The team was joined by (front row) SENER Deputy Minister Leonardo Beltran, UNAM Board Member Dr. Eduardo Barzana, UNAM Chief Research Officer Dr. William Lee, and SENER Director General for Research and Technology Development Carlos Ortiz.

investments. The workshops also identify prospects for bilateral or multilateral R&D collaborations that would benefit from cooperation on high-impact scientific research. The workshops are multi-day, carefully planned, and structured working meetings.

Deputy Minister Leonardo Beltran of SENER, Deputy Office Director Maureen Clapper of DOE, and Assistant Deputy Minister Frank Des Rosiers of Natural Resources Canada opened the workshop. They were followed by a distinguished panel: Dr. Mario Molina, Nobel Laureate; Dr. Alan Bernstein, CIFAR President and CEO; Dr. Horst Simon, Lawrence Berkeley National Laboratory Deputy Director and Chief Research Officer; and Dr. Paul Durrant, Head of Strategy and Engagement at the U.K. Department for Business, Energy & Industrial Strategy (BEIS) and Head-Designate of the MI Secretariat, Dr. Hermann Tribukait moderated the

Figure 1.7 Workshop Co-Chairs Professor Kristin Persson of University of California-Berkeley and Professor Alán Aspuru-Guzik of Harvard University.

panel in its discussions of the importance of scientific R&D collaborations to the acceleration of energy innovation. Scientific thought leaders then delivered a series of plenary talks and provided the context for follow-on panel discussions.

On the afternoon of the first day, attendees divided into three panels on:

• inorganic functional materials;

• organic functional materials; and

• nanomaterials and composites.

The panels began by identifying the critical R&D priorities and gaps in each area of the energy materials innovation chain. They then explored opportunities for combining individual research pathways, high-throughput synthesis, high-throughput calculation, and high-throughput characterization into an integrated materials innovation approach or "platform". Participants engaged in technical discussions on combining advanced theoretical and applied physical chemistry with next-generation computing, machine learning, and robotics. Discussions continued through the second day of the workshop.

On the third day, the panels presented their recommendations on opportunities to be pursued. Panel co-chairs and workshop co-chairs then identified six priority

research areas — the Six Grand Goals for a materials innovation revolution — for further discussion. Attendees reconvened in work groups according to their materials focus areas to discuss how to leverage the cross-cutting themes of theory, synthesis, and characterization.

1.3 STRUCTURE OF THE REPORT

This workshop report draws from three main activities or sources: the pre-workshop report and reading materials prepared for the participating experts, workshop plenary presentations and discussions, and the large body of documentation and notes generated during the workshop panel sessions. A core writing team from the workshop panels and the workshop co-chairs curated all documents and information.

The report is organized as follows. Chapter 2 covers the state of the art in materials discovery. Chapter 3 presents the urgent need for a materials revolution and introduces the Six Grand Goals of the Materials Acceleration Platform. Chapters 4 through 9 discuss each of the goals in turn, identifying promising opportunities, the benefits and impact of achieving the goal, its relationship to the other five goals, and materials-specific challenges and approaches for organic materials, inorganic materials, and nanomaterials and composites. The conclusion and some final reflections are provided in Chapter 10.

References for Chapter 1

1. Bill Gates, "Beating Nature at Its Own Game," Gates Notes, March 14, 2017, www.gatesnotes. com/Energy/Beating-Nature.

2. George I. Seffers, "Scientists Pick AI for Lab Partner: Robotic research systems may foster a boom in scientific knowledge," *Signal*, September 1, 2017, www.afcea.org/content/scientists-pick-ai-lab-partner.

3. P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, "Machine-learning-assisted materials discovery using failed experiments," *Nature* 533 (2016): 73–77.

4. S. V. Kalinin, B. G. Sumpter and R. K. Archibald, "Big–deep–smart data in imaging for guiding materials design," *Nature Materials* 14 (2015), 973–980.

5. Jonathan D. Linton and Steven T. Walsh, "From bench to business," *Nature Materials* 2 (2003): 287–289, DOI: 10.1038/nmat882.

6. United Nations Framework Convention on Climate Change (UNFCCC), Adoption of the Paris Agreement, Report No. FCCC/CP/2015/L.9/Rev.1, 2015, unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf.

7. Eliza Northrop, Hana Biru, Sylvia Lima, Mathilde Bouyé and Ranping Song, "Examining the Alignment between the Intended Nationally Determined Contributions and Sustainable Development Goals," *World Resources Institute*, September 2016, www.wri.org/publication/examining-alignment-between-intended-nationally-determined-contributions-and-sustainable.

8. Mission Innovation, "Innovation Challenges," *Mission Innovation* (website), accessed 2017, mission-innovation.net/our-work/innovation-challenges/.

9. David King, "Global clean energy in 2017," *Science* 355 (6321), 111, DOI: 10.1126/science.aam7088.

2 State of the Art in Materials Discovery

This chapter briefly describes the successful implementation of components of an autonomous platform in the biological and medical sciences before identifying the key gaps preventing the full-scale deployment of autonomous discovery to materials science. It also summarizes the main materials-related approaches and opportunities of a range of clean energy technologies, including photovoltaics, batteries, and solar fuels.

2.1 SCALING UP SCIENCE PAYS OFF

From the dawn of the industrial revolution to the modern digital revolution, automation has driven the expansion of the global economy and improvements to average quality of life. In scientific research, robotic technologies have spurred gains in speed and efficiency, particularly in biology and medicine. In these fields, state-of-the-art robotic platforms have evolved to offload repetitive tasks from human researchers and execute experiments with greater precision, speed, and accuracy than ever before.

Such advances in the biological sciences are well documented. For example, the U.S. National Human Genome Research Institute (NHGRI) has tracked the costs of DNA sequencing at its many sequencing centres. Trends in these costs are an important metric for assessing the impact of improvements in DNA sequencing technologies and for benchmarking the capacity of the NHGRI Genome Sequencing Program (GSP). Figure 2.1 shows the actual costs per genome over time along with costs predicted by Moore's law, which describes a long-term trend in the computer hardware industry that involves the doubling of "compute power" every two years. Technology improvements that "keep up" with Moore's law are widely regarded to be doing exceedingly well, making it useful for comparison.

Driven by the mandate for an accelerated drug discovery process, the pharmaceutical industry has similarly pioneered automation in medicinal chemistry. Companies such as Eli Lilly, Merck & Co., and Aventis now
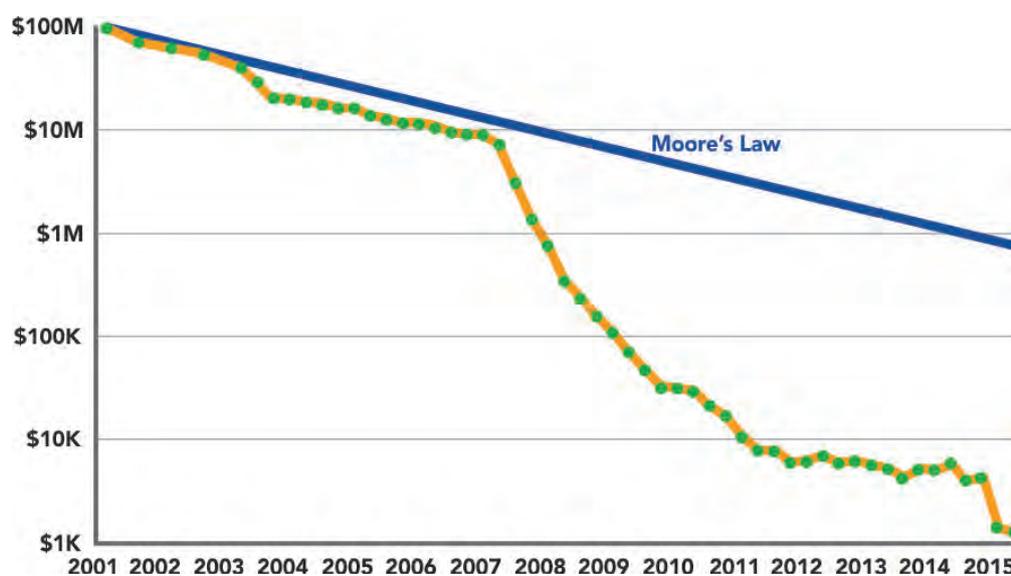
Figure 2.1 Annual trend in the "Cost per genome" determined by NHGRI. Data for cost estimates are presented alongside those predicted by Moore's law, which benchmarks performance based on a yearly doubling of computational power. [Source: U.S. National Human Genome Research Institute]

use automated synthesis to exhaustively search for compounds suitable for a given medicinal application [1,2,3]. These methods have yielded many successes in both medicine and materials science [4,5].

In addition to automated synthesis procedures, some aspects of automated characterization have been significantly advanced due to progress in flow chemistry [6]. For example, high-throughput characterization methods using nuclear magnetic resonance (NMR) spectroscopy and other radiation methods can now be used to monitor chemical reactions as they happen [7]. Determining a material's mechanical properties has also become increasingly automated. In recent years, Dow has begun using robotic systems to perform such characterizations of coatings [8]. Structural characterization is also increasingly being performed in a high-throughput and automated manner. There have been important advances in quality control for organic solar cells, for example, using these methods to detect defects and improve device performance [9,10].

Despite these advances, the automation of synthesis and characterization still has substantial room for improvement [11] and most of automated chemistry is based on a "trial and error" approach. We are now entering an age where integrated systems using increasing computer power, machine learning, and AI could not only enable exhaustive searching for new materials, but could also provide intermediate decision-making capabilities to accelerate the discovery and development of materials with specific, targeted properties. In principle, such AI-based algorithms can be adaptive, using feedback from experimental characterization procedures and enabling a more rational route to research objectives—either optimizing properties or even directly verifying hypotheses. If realized, such an approach would markedly increase research productivity, reduce its cost, and enable entirely new experimental paradigms that are critical for innovation [12].

In the broader context of pharmaceuticals and biotechnology, many start-up business models that reflect the paradigm of integrating AI and robotics into the discovery process have begun to appear [13, 14, 15]. For example, companies such as Zymergen, Transcriptic, Gingko Bioworks, Emerald Cloud Lab, Citrine, and BenevolentAI follow data-driven approaches and seek to maximally exploit existing knowledge bases. These efforts often use sophisticated human language processing methods to data mine existing literature. For example, ARES (Autonomous Research System) was the first AI-driven experimentation system to demonstrate closed-loop, iterative experimentation with in situ characterization for materials development [16]. Despite this emerging paradigm, a generalized end-to-end platform has yet to be developed. Moreover, decision science, the basic tool of discovery-enabling AI, and data-driven approaches specifically for materials science have not been sufficiently developed to support such an integrated platform.

2.2 CURRENT GAPS IN MATERIALS DISCOVERY

Several shortcomings in the current *modus operandi* of the synthesis, characterization, and theory communities must be addressed to develop the data-driven approaches necessary to achieve a successful materials discovery platform. This section briefly outlines key gaps in characterization, databases, and machine learning.

Characterization techniques, in particular, are often applied on a case-by-case basis with specially designed instruments located in geographically dispersed facilities. Coordination between facilities is limited by traditional methods of data dissemination (e.g., journal articles, conferences), which are slow and highly selective about what data are shared. In particular, negative results, crucial for machine learning, are rarely shared or reported [17]. As samples have become more complex, the time it takes to replicate them at, or transport them to, technique-specialized labs has become a significant bottleneck. The inefficiency of this process is illustrated in Figure 2.2 (left). In comparison, a sample-centric approach (Figure 2.2, right)
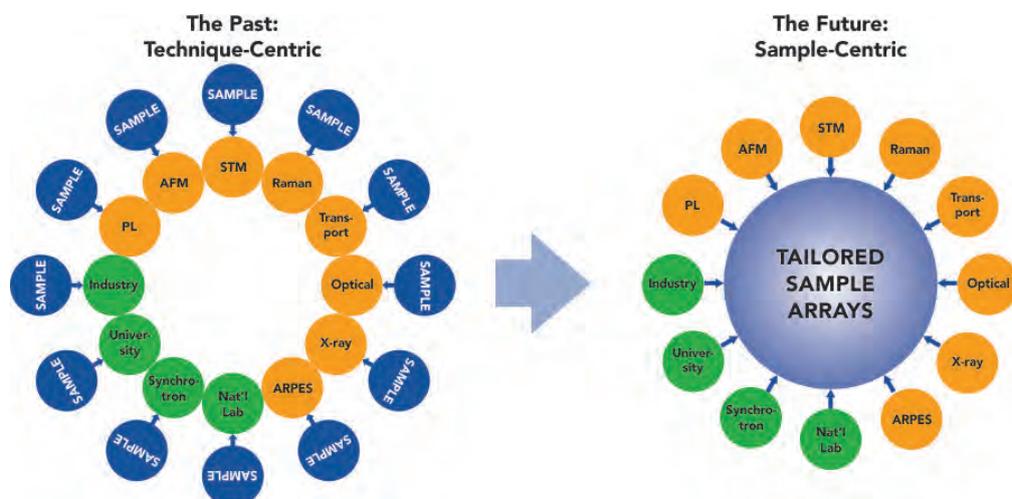
Figure 2.2 Comparison of (left) technique-centred and (right) sample-centred approaches. The technique-centred approach involves replicating growth procedures or shipping samples among geographically separated institutes. The sample-centred approach brings the people and techniques to a central growth/characterization facility, saving time and allowing techniques to be applied to the identical sample arrays. [Source: Dr. Eli Rotenberg, Lawrence Berkeley National Laboratory]

illustrates a central facility that concentrates efforts on preparing the best samples (or arrays of samples) to be shared among many co-located techniques, such as synchrotron beamlines or high-resolution electron microscopy, that cannot be replicated easily in small lab settings.

Databases underlie all machine learning and data-driven approaches. In the course of the Materials Genome Initiative, databases such as NOMAD (Novel Materials Discovery), OQMD (Open Quantum Materials Database), AiiDA (Automated Interactive Infrastructure and Database for Computational Science), and AFLOW, have been developed [17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27]. However, as with characterization techniques, many of these databases are specialized to address case-by-case issues; an international data ecosystem does not yet exist.

One significant gap in data-driven approaches to inorganic materials science is the lack of meaningful descriptors and knowledge of structure-property relationships to derive insights from abundant data. A number of efforts using motifs from both crystal and electronic structures have already yielded some insights. However, further investment and progress may be required to support an AI-integrated platform for materials discovery. Databases also exist for organic materials. For example, the Harvard Clean Energy Project has calculated more than 3 million organic compounds in more than 300 million density functional theory calculations [28, 29, 30], although not yet at the scale of inorganic materials. The Reaxys database contains the most complete collection of known organic reactions, setting an important foundation for data-driven approaches for synthesis planning [31, 32]. It enables chemists to specify strategies for synthesizing small molecules selected as potential candidates with specific desired properties.

A final gap is the lack of understanding of how to most effectively apply machine learning to materials discovery. Most of the machine learning methods applied to date, both theoretical and experimental, are relatively straightforward adaptations of methods originally developed for other problems, such as image recognition, text generation, and translation [33, 34, 35, 36]. These methods include deep neural networks and Bayesian optimization. However, the development of new AI methods tailored for materials could provide a breakthrough that greatly increases the effectiveness of these techniques, thus accelerating the discovery process. Currently, materials discovery is still a "trial and error" process; however, developing new machine learning algorithms that generate candidate materials, rather than only predicting properties, could lead to an environment where the inverse design of materials is possible. Therefore, investing in this area would help derive new machine learning algorithms and better tune existing ones specifically for materials discovery [37].

2.3 MATERIALS OPPORTUNITIES FOR CLEAN ENERGY TECHNOLOGIES

Table 2.1 summarizes the current status of seven clean energy technologies and the key opportunities associated with each from a materials perspective. Many of the technologies need new materials to continue to grow because they are fundamentally limited by their current materials. The table also lists important targets published by governments and consortia.

Table 2.1: Current status and opportunities for major energy technologies

Energy Technology	Current Status	Materials Opportunities
Photovoltaics	• PV module production is dominated by a small number of materials: polycrystalline and single crystalline Si and CdTe. • Reducing capital and manufacturing costs could lower full levelized cost of solar electricity to 0.03 US$/kWh (i.e., US DOE 2030 target).	• Further reductions in cost of solar electricity require improved efficiency, longer module/ electronics lifetimes, and reduced efficiency degradation. • New materials and new mechanisms for solar energy harvesting are required. • Hybrid perovskites are one of the most promising classes of new materials, with power conversion efficiencies >22%, but suffer from low long-term stability and presence of toxic elements. • New mechanisms for boosting efficiency (e.g., photon up-conversion, carrier multiplication, luminescence concentrators) require new materials, but performance of current materials is lacking.

Batteries (for electric vehicles and grid storage)	• Technology is dominated by lithium-ion batteries. • While single electrode materials have demonstrated high-power capabilities, the current Li-ion cell suffers from power limitation due to interfacial reactions. • Current materials have about half the specific energy (200–250 Wh kg–1) of the goal proposed by the Battery500 Consortium: >10-year life and total mileage of ~150,000 miles. • Current materials have problems of chemical and structural stability. • Grid storage applications require performance stability, i.e., >6,000 cycles and >20-year lifetimes.	• New materials architectures, interfacial treatments, and cell design are needed to improve power density. • New materials with higher energy and power densities and improved chemical and structural stability are required. • Lower-cost, low supply-risk element replacements for current battery components are required to scale up battery storage. • There have been rapid developments in solid-state Li batteries, but interfacial reactions and cell design need further R&D. • Redox flow batteries show promise for grid storage, but solubility and stability of redox molecules at any state of charge must be improved. • The cost of ion-exchange membranes must decrease.
Solar fuels	• Key reactions to produce fuels (by light-induced or dark catalysis) are limited in efficiency and long-term cycling. • Use of solar fuels (in fuel cells) requires high temperatures, which increase operating costs and cause corrosion of cell materials. Or, if run at low temperatures, they rely on precious metal catalysts, e.g., platinum.	• More efficient catalysts based on cheaper, earth-abundant elements are needed.
Wind Power	• Wind power is one of the fastest growing renewable energy sources, currently providing 2.8% of the world's electricity generation. • If growth continues, wind could provide 20-30% of global power generation, displacing 3 billion tons of CO_2 per year. • To generate electricity, wind relies on electric motors, which today require permanent magnets. • In cold and icy or sandy desert climates, wind blades erode and foul, which can cause up to 25-50% degradation in power production.	The following new materials are needed: • erosion-resistant coatings and materials; • hydrophobic, light-weight materials and coatings; • impact-resistant materials for the leading edge of wind blades; and • high-performance permanent magnet materials that do not contain elements with supply risk (critical materials). • Offshore installations (which provide much more power than onshore) need improved salt-water corrosion-resistant materials.

Thermal energy conversion	• Harnessing waste heat would help mitigate global energy needs, but most heat sources are not harnessed.	• More efficient thermoelectric materials are required to convert waste heat to electricity. Promising strategies include nanostructures, heterostructures, and alloying. • More thermally stable thermoelectric materials that do not experience degradation under severe thermal cycling conditions are needed. • New materials for smart-windows using electro-, photo-, and thermochromic mechanisms are needed to control infrared radiation. • Materials are needed to improve building energy efficiency.
Gas separation and storage	• The petrochemical industry currently fractionally separates hydrocarbons based on molecular weight. • Recent developments in metal-organic frameworks have resulted in excellent storage capabilities for CH_4 and CO_2.	• Clearer structure-property relationships are needed to optimize materials. • Finding new materials and processes that generate useful products from captured CO_2 would provide a financial incentive for further development.
Power transmission	• High-voltage alternating current (AC) transmission is widely implemented due to its ability to transform the voltage. However, as lines get longer and the voltage higher, the losses due to power dissipation increase dramatically. • Efficient power transmission across long distances and underwater would improve offshore, high-power wind energy.	• High-voltage direct current (DC) is more efficient to transport large amounts of power over a long distance. However, improved materials for high-power electronics are needed for widespread implementation. • Discovery of superconducting materials at ambient temperatures would revolutionize electric power transmission by effectively eliminating losses.

Refs for table: [38, 39].

References for Chapter 2

1. Alexander G. Godfrey, Thierry Masquelin and Horst Hemmerle, "A remote-controlled adaptive medchem lab: an innovative approach to enable drug discovery in the 21st Century," *Drug Discovery Today* 18, Issues 17–18 (2013): 795–802, DOI: 10.1016/j.drudis.2013.03.001.

2. Michael Shevlin, "Practical High-Throughput Experimentation for Chemists," Merck & Co., Inc., *Medicinal Chemistry Letters* 8, Issue 6 (2017): 601–607, DOI: 10.1021/acsmedchemlett.7b00165.

3. Angelika Weber, Erich von Roedern and Hans Ulrich Stilz, "SynCar: An Approach to Automated Synthesis," Aventis Pharma Deutschland GmbH, *Journal of Combinatorial Chemistry* 7, Issue 2 (2005): 178–184, DOI: 10.1021/cc049838z.

4. Shevlin, op. cit.

5. Gurpur Rakesh D. Prabhu and Pawel L. Urban, "The Dawn of Unmanned Analytical Laboratories," *TrAC Trends in Analytical Chemistry* 88 (2017): 41–52, DOI: 10.1016/j.trac.2016.12.011.

6. Prof. Steven V. Ley, Daniel E. Fitzpatrick, Dr. Richard. J. Ingham and Dr. Rebecca M. Myers, "Organic Synthesis: March of the Machines," *Angewandte Chemie International Edition* 54, Issue 11 (2015): 3449–3464, DOI: 10.1002/anie.201410744.

7. Michael Maiwald, Holger H. Fischer, Young-Kyu Kim, Klaus Albert and Hans Hasse, "Quantitative high-resolution on-line NMR spectroscopy in reaction and process monitoring," Journal of Magnetic Resonance 166, Issue 2 (2004): 135–146, DOI: 10.1016/j.jmr.2003.09.003.

8. Tzu-Chi Kuo, Niranjan A. Malvadkar, Ray Drumright, Richard Cesaretti and Matthew T. Bishop, "High-Throughput Industrial Coatings Research at The Dow Chemical Company," The Dow Chemical Company, *ACS Combinatorial Science* 18, Issue 9 (2016): 507–526, DOI: 10.1021/acscombsci.6b00056.

9. Johannes Hepp, Florian Machui, Hans-J. Egelhaaf, Christoph J. Brabec and Andreas Vetter, "Automatized analysis of IR-images of photovoltaic modules and its use for quality control of solar cells," *Energy Science & Engineering* 4, Issue 6 (2016) 363–371, DOI: 10.1002/ese3.140.

10. G. Zhou, A. Moment, S. Yaung, A. Cote and T.-E. Hu, "Evolution and Application of an Automated Platform for the Development of Crystallization Processes," *Organic Process Research and Development* 17, Issue 10 (2013): 1320–1329, DOI: 10.1021/op400187h.

11. S. Szymkuć, E. P. Gajewska andT. Klucznik. "Computer-Assisted Synthetic Planning: the End of the Beginning," *Angewandte Chemie International Edition* 55 (2016), DOI: 10.1002/anie.201506101.

12. Claudia Houben and Alexei A, Lapkin, "Automatic discovery and optimization of chemical processes," *Current Opinion in Chemical Engineering* 9 (2015) 1–7, DOI: 10.1016/j.coche.2015.07.001.

13. E. C. Hayden, "The Automated Lab," *Nature News* 516 (2014): 131–132.

14. S. V. Ley, D. E. Fitzpatrick, R. J. Ingham and R. M. Myers, "Organic Synthesis: March of the Machines," *Angewandte Chemie International Edition* 54, Issue 11 (2015): 3449–3464.

15. J. Bohannon, "A New Breed of Scientist, with Brains of Silicon," *Science* (2017), DOI: 10.1126/science.aan7046.

16. Pavel Nikolaev, Daylond Hooper, Frederick Webber, Rahul Rao, Kevin Decker, Michael Krein, Jason Poleski, Rick Barto and Benji Maruyama, "Autonomy in materials research: a case study in carbon nanotube growth," *npj Computational Materials* 2 (2016), DOI: 10.1038/npjcompumats.2016.31.

17. P. Raccuglia, K. C. Elbert, P D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, "Machine-learning-assisted materials discovery using failed experiments," *Nature* 533 (2016): 73–77, DOI: 10.1038/nature17439.

18. Scott Kirklin, James E Saal, Bryce Meredig, Alex Thompson, Jeff W Doak, Muratahan Aykol, Stephan Rühl & Chris Wolverton, "The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies," *npj Computational Materials* 1, Number 15010 (2015), DOI: 10.1038/npjcompumats.2015.10.

19. Pizzi et al, op. cit.

20. Dr. Jens S. Hummelshøj, Dr. Frank Abild-Pedersen, Dr. Felix Studt, Dr. Thomas Bligaard and Prof. Jens K. Nørskov, "CatApp: A Web Application for Surface Chemistry and Heterogeneous Catalysis," *Angewandte Chemie International Edition* 51, Issue 1 (2012): 272–274, DOI: 10.1002/ange.201107947.

21. Ivano E. Castelli, Thomas Olsen, Soumendu Datta, David D. Landis, Søren Dahl, Kristian S. Thygesena and Karsten W. Jacobsen, "Computational screening of perovskite metal oxides for optimal solar light capture," *Energy & Environmental Science* 5, Issue 2 (2012): 5814–5819, DOI: 10.1039/c1ee02717d.

22. S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo and O. Levy, "AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations," *Computational Materials Science* 58 (2012): 227–235, DOI: 10.1016/j.commatsci.2012.02.002.

23. A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Du□ak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, "The atomic simulation environment–a Python library for working with atoms," *Journal of Physics: Condensed Matter,* 29, Number 27 (2017), DOI: 10.1088/1361-648X/aa680e.

24. J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway and A. Aspuru-Guzik, "The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid," *The Journal of Physical Chemistry Letters* 2 (2011): 2241–2251, DOI: 10.1021/jz200866s.

25. G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari and B. Kozinsky, "AiiDA: Automated Interactive Infrastructure and Database for Computational Science," *Computational Materials Science* 111 (2016): 218–230, DOI: 10.1016/j.commatsci.2015.09.013.

26. "NOMAD Repository," The NOMAD Laboratory (website), accessed 2017, nomad-repository.eu/.

27. A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Materials* 1, 11002 (2013), DOI: 10.1063/1.4812323.

28. Johannes Hachmann, Roberto Olivares-Amaya, Adrian Jinich, Anthony L. Appleton, Martin A. Blood-Forsythe, László R. Seress, Carolina Román-Salgado, Kai Trepte, Sule Atahan-Evrenk, Süleyman Er, Supriya Shrestha, Rajib Mondal, Anatoliy Sokolov, Zhenan Baod and Alán Aspuru-Guzik, "Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry – the Harvard Clean Energy Project," Energy & Environmental Science 7, Issue 2 (2014): 698, DOI: 10.1039/C3EE42756K.

29. Steven A. Lopez, Edward O. Pyzer-Knapp, Gregor N. Simm, Trevor Lutzow, Kewei Li, Laszlo R. Seress, Johannes Hachmann and Alán Aspuru-Guzik, "The Harvard organic photovoltaic dataset," *Scientific Data* 3 (2016), DOI: 10.1038/sdata.2016.86.

30. Hachmann et al. 2011, op. cit.

31. A. J. Lawson, J. Swienty-Busch, T. Géoui and D. Evans, "The Making of Reaxys—Towards Unobstructed Access to Relevant Chemistry Information," in *ACS Symposium Series* 1164 (2014): 127–148, Chapter DOI: 10.1021/bk-2014-1164.ch008.

32. *Reaxys* (website), accessed 2017, www.reaxys.com/.

33. Rafael Gómez-Bombarelli, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams and Alán Aspuru-Guzik, "Automatic chemical design using a data-driven continuous representation of molecules," *Cornell University Library*, January 6, 2017, arXiv:1610.02415v2 [cs.LG].

34. David K. Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alan Aspuru-Guzik and Ryan P. Adams, "Convolutional Networks on Graphs for Learning Molecular Fingerprints," in the proceedings of Advances in Neural Information Processing Systems 28 (NIPS 2015), papers.nips.cc/paper/5954-convolutional-networks-on-graphs-for-learning-molecular-fingerprints.

35. Benjamin Sanchez-Lengeling, Carlos Outeiral, Gabriel L. Guimaraes and Alan Aspuru-Guzik, "Optimizing distributions over molecular space. An Objective-Reinforced Generative Adversarial Network for Inverse-design Chemistry (ORGANIC)," Harvard University, Chem Rxiv, August 17, 2017, doi.org/10.26434/chemrxiv.5309668.v3.

36. Edward O. Pyzer-Knapp,a Gregor N. Simmb and Alán Aspuru Guzik, "A Bayesian approach to calibrating high-throughput virtual screening results and application to organic photovoltaic materials," *Materials Horizons* 3, Issue 3 (2016) 226–233, DOI: 10.1039/C5MH00282F.

37. R. Ramakrishnan and O. A. von Lilienfeld, "Machine Learning, Quantum Chemistry, and Chemical Space," Chapter 5 in *Reviews in Computational Chemistry* 30, Abby L. Parrill and Kenny B. Lipkowitz, Eds. (John Wiley & Sons, Inc., 2017), Chapter DOI: 10.1002/9781119356059.ch5.

38. Steven Chu, Yi Cui and Nian Liu, "The path towards sustainable energy," *Nature Materials* 16 (2017): 16–22, DOI:10.1038/nmat4834.

39. Jarad A. Mason, Mike Veenstra and Jeffrey R. Long, "Evaluating metal–organic frameworks for natural gas storage," *Chemical Science* 5 (2014): 32–51, DOI: 10.1039/c3sc52633j.

3 The Urgent Need for a Materials Acceleration Platform

Throughout history, humans have developed energy technologies to maximize the energy available for human consumption and increase productivity. This drive has led to transitions from wood burning to the use of coal, oil, and natural gas, whose extraction by machinery became far more efficient than harvesting trees through human innovation. The development of nuclear power required scientific understanding to harness the high-energy content of nuclear fuel, while tackling persistent challenges in safety, cost, and waste handling. With GHG emission reductions in mind, the current shift to clean energy is the most challenging and potentially most rewarding energy transition to date.

Advanced materials represent today about 50% of the manufacturing cost of clean energy and are expected to increase to 80% in the near future, according to the Energy Materials Industrial Research Initiative (EMIRI) [1]. Thus, a successful transition to clean energy requires the development of new, high-performance, low-cost materials that are resilient, safe for humans and the environment, recyclable, and use abundant elements so that they can be deployed globally. This challenge is not merely an engineering problem; it requires fundamental new scientific advances to design and organize matter from the atomic scale to the systems scale. The scale of the research is on the order of multiple Manhattan project efforts, i.e., enormous and requiring more human effort than is currently plausible.

This transition can be accelerated by leveraging new ways of doing science. The groundwork has already been completed due to the creation of many simple materials with tremendous functionality, such as silicon-based materials that, after 60 years of R&D, have led to solar cells on roofs and billions of transistors in hand-held electronics. Traditional, human researcher-centric methods, however, are now insufficient for providing the materials innovations needed to deliver ubiquitous clean energy. We have reached the limits of any one human

mind to assimilate the level of detail and magnitude of data that are now produced by experiments and computational simulations. Computer simulations with current computer power are approaching the limit to predict the properties of matter at the relevant length and time scales. We have also reached the limit of any single research group to own and operate the variety of synthetic and characterization equipment needed to completely understand these materials. Moreover, we are approaching the time limit for deploying worldwide ubiquitous clean energy technologies before the environmental damage becomes irreversible.
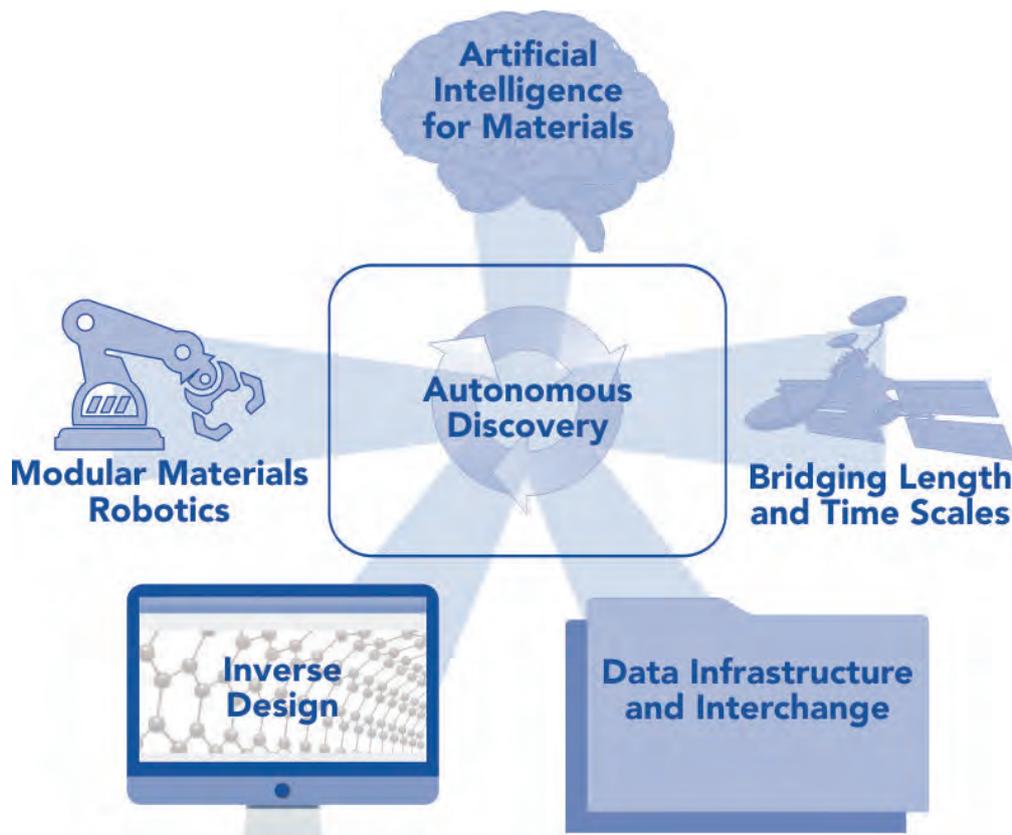

Figure 3.1 Six Grand Goals of the Materials Acceleration Platform.

It is time for a materials innovation revolution that demands a fundamental change in how materials research, scientific discovery, and technology development are conducted. This is not unlike the transition in nuclear physics research: it began in the 1920s and 1930s with small, tabletop, single-laboratory experiments, was accelerated by massive mission-oriented public investments such as the Manhattan project, and is now dominated by giant, internationally funded, collaborative particle accelerators and supercolliders.

The Materials Acceleration Platform, or MAP, aims to reduce the materials development cycle from 10 to 20 years to 1 or 2 years. It builds on recent scientific breakthroughs and the ability to program machines to assist the design of materials, moving away from the drudgery of Edisonian discovery methods. Such breakthroughs allow scientists to order whole families of new materials based on desired properties

without tedious hours of trial and experiments in the lab. The ability of AI to sift through vast quantities of data can deliver new scientific insights to humans. The development of new AI-assisted theoretical methods can speed up the simulation and design of new materials. Achieving this, however, requires development of machine intelligence and materials synthesis capabilities beyond the scale of any human team.

MAP features six integrated priority research areas, the Six Grand Goals, which can be approached and achieved more effectively and faster through strong international cooperation (Figure 3.1):

1. **Closing the Loop in Autonomous Discovery and Development:** "Self-driving laboratories" that autonomously design, perform, and interpret experiments are needed to discover new materials. Creating and deploying autonomous laboratories that can perform this closed-feedback-loop discovery and development process would be the culmination of all the other goals.

2. **Artificial Intelligence for Materials:** Autonomous research relies on reasoning, decision making, and creativity. The particular scale and details of theoretical, computational, synthetic, and characterization evidence in materials research require the establishment of this new branch of AI. National and international research organizations can facilitate an integrated computer and materials science research effort to develop algorithms that mimic, and then supersede, the intellect and intuition of expert materials scientists.

3. **Modular Materials Robotics:** To accommodate evolving materials demands and the ever-expanding breadth of clean energy technologies, autonomous laboratories must remain nimble and motivate a modular approach to the development of materials science automation. The elegant representation of techniques and materials as modular building blocks fosters human-machine communication and simplifies the path to materials exploration beyond the bounds of known materials.

4. **Inverse Design**: Materials innovation by an autonomous laboratory can be seeded and accelerated by conceiving novel materials compositions or structures that can meet specific requirements. Inverse design enables automated generation of candidate materials designed to meet the performance, cost, and compatibility requirements of a given clean energy technology.

5. **Bridging Length and Time Scales:** Materials consist of atoms, connected by bonds and arranged at the nano, micro, and macro scales—a variation in length scale akin to going from the width of a human hair to the diameter of Earth. Light absorption occurs in femtoseconds, chemical bonds are broken and formed in picoseconds, and syntheses and characterizations require microsecond- to hour-long experiments. Materials that are stable for decades are needed, an equally daunting breadth of scale. Although there are appropriate scientific theories

for each of these length and time scales, systematic methods of connecting results and ideas across these scales would enable transformative discoveries.

6. **Data Infrastructure and Interchange:** Innovation relies on communication and appropriate representation of both data and the knowledge obtained from data. This poses a substantial challenge to the international research community to join forces in establishing and populating a materials data infrastructure. The resulting product, which would embody an understanding of materials beyond that attainable by an individual scientist or even a team of scientists, would enable and enhance autonomous laboratories.

References for Chapter 3

1. Fabrice Stassin, "The Role of Materials Research & Innovation for European Growth & Competitiveness," Energy Materials Industrial Research Initiative (EMIRI), October 2017.

2. Alan Aspuru-Guzik, Roland Lindh and Markus Reiher, "The Matter Simulation (R)evolution," Harvard University, *ChemRxiv*, November 21, 2017, doi.org/10.26434/chemrxiv.5616115.v1.

4 Goal #1: Closing the Loop in Autonomous Discovery and Development

An autonomous platform for materials discovery would exploit the "feedback" from measurements and guide new synthesis and characterization experiments or simulations, i.e., enable closed-loop research or a self-driven laboratory. It would ultimately converge to an optimal materials design for target functionality, while autonomously generating and validating hypotheses of the underlying physical and chemical phenomena. The overarching challenge of autonomous materials research combines (i) AI-based predictive theory of new materials and their properties, (ii) autonomous robotic systems for synthesis and experimental data collection, (iii) data analytics such as feature extraction, (iv) machine learning-based classification and regression of the results, and (v) decision modules to drive optimal experimental design for subsequent experimental iterations. For autonomous discovery and development, these elements need to be integrated into a closed-loop platform for designing and performing experiments, which can ultimately create new materials to meet society's needs for clean energy.

Such an ambitious program cannot be accomplished all at once. Instead, it requires incremental progress as materials-related AI algorithms and robotic technologies are developed and integrated. These algorithms should first be applied to simple closed-loop optimization and discovery schemes. The experience built up by many such optimizations can inform higher-level AI programs that can design experiments with progressively greater autonomy. *Autonomy* refers to the ability to assimilate results to make decisions in high-dimensional parameter space, which at this scale can surpass human intuition. By contrast, *automation* of experimental and computational tools does not require AI; it only follows commands from autonomous control systems. Automation reduces the cycle-time, but autonomy closes the loop.

When optimizing or discovering new materials, human intuition currently drives experiment design. The resulting data sets tend to be clustered, sparse, and incomplete, especially since humans favour inclusion of

"successful" data and do not report data of "failed" experiments. The comprehensive incorporation of all data, successful or failed, is one of the strengths of machine learning approaches to AI for materials. With sufficient data, an AI-driven machine can choose the next step in experiments or simulations more efficiently than humans, speeding up the optimization of a given property.

In closed-loop materials design and discovery, autonomy can emerge at multiple levels. At the lowest level, AI decides the next step in an optimization loop. Such optimization can be implemented in the theoretical domain (e.g., what is the best composition of ternary oxide to simulate that is expected to yield a particular functionality?) or in the experimental domain (e.g., what is the best reactor condition to yield faster crystalline growth?). At a higher level, AI decides how to design materials, such as selecting among hierarchical structures of disparate building blocks. At the highest level, AI can develop chemical insights from raw data and correlations, drawing information from vastly different types of properties, e.g., electrical, mechanical, optical, electrochemical, thermal, or structural. AI in the loop can confirm or negate mooted hypotheses on the operative physical and chemical phenomena more efficiently than conventional approaches.

4.1 OPPORTUNITIES

A key challenge in materials discovery, particularly in the organic materials domain, is the limited availability of automated or closed-loop synthetic tools. Synthetic organic molecules are an important form of matter in a huge range of human activities. Organic molecules are used as medicines, functional coatings, energy storage, and energy harvesting materials, to name a few applications. Despite the widespread applications, synthesizing novel organic molecules is an activity reserved for highly trained specialists. The limited pool of experts trained in synthetic chemistry creates a significant bottleneck for researchers in other fields interested in using synthetic molecules to achieve their R&D aims. To drive discovery and development of new materials for clean energy, materials researchers need access to synthetic molecules much faster and on a much larger scale than currently possible.

Characterization of materials is critical to the discovery process and must be integral to autonomous systems. Currently, a plethora of instrumental techniques, data analysis software or models, and interpretation or visualization tools are used, depending on several variables, such as sample type, instrument availability, and/ or user experience. Rarely are these processes automated or coupled between different techniques. This presents practical limitations on the speed of the materials development loop, but there is ample room for intense R&D efforts to solve these issues with characterization.

To date, characterization has used a technique-centric approach, in which sample surfaces are prepared and analyzed independently in individual, technique-specialized laboratories. This approach has proven viable for research activities producing limited amounts of samples. However, analyzing vast quantities of samples, such

as those produced by automated synthesis techniques, could be more challenging. As such, a new characterization paradigm is needed in which sample analysis is performed in an automated fashion driven by machine learning and AI. There are two types of automation loops, broadly described as "optimization" and "discovery" (Section 4.4.4). The distinction between them is not sharp, and many investigations would include aspects of both. What is clear, however, is that both approaches need similar growth and characterization tools, but their workflows could differ owing to different time scales of the characterization techniques.

4.2 BENEFITS AND IMPACT

Achieving this goal could result in the creation of a shared international framework of autonomous materials synthesis and characterization research systems available to industry, academia, and national labs for rapid testing of hypotheses. Such systems would manage data by internally adopting laboratory information management that automatically harvests as-taken data and metadata into a community database and provides tools for the dissemination of data, methods, and analysis results. Samples would be created and transported around each facility, in a fully autonomous manner with the machines suggesting the next measurement point to be taken locally and globally through integration of data taken in a worldwide network. Closing the loop could also result in Moore's law-like behaviour, but in this case the rate of discovery would be doubled every two years [1], which would enable substantial developments in new materials for energy, where multiple competing properties have to be optimized.

On the individual lab scale, success in this goal would ultimately lead to a bench-top device that is capable of autonomous synthetic chemistry. Its first iteration would integrate the process of designing the synthetic route to the target molecule, validating the routes and optimization of the resulting synthetic steps using integrated analytical technologies. The device would also purify and characterize the target molecule against the molecular properties desired. It would be able to learn from the results of each experiment to further refine the AI that designs the synthetic routes and the algorithms that are used to optimize the individual chemical steps.

Going beyond this, the long-range impact could be the development of an autonomous tool that designs for function instead of target molecule. The current paradigm for materials discovery is to design a molecule, then make the molecule, and then test the molecule to see if it has the target properties. The ultimate goal, however, is not the specific structure of the molecule, but obtaining of a material with the target properties. The ideal device would begin with a description of the target properties and then iterate on molecular architectures until it achieves these properties.

Closing the loop would also result in a paradigm where samples are no longer replicated or transferred between separate technique-oriented facilities, but become the centre; the techniques would be merged in national-scale facilities dedicated to creating carefully curated sample arrays or libraries that are shared

among co-located techniques. To ensure that the characterizations have the most information, the robotics could transfer the sample in tailored environments, such as controlled atmosphere, ultra-high vacuum, temperature, etc. It is unlikely that one facility could encompass the three classes of materials that are the focus of this document (i.e., organic, inorganic, composite) along with others. However, several facilities with similar organization, but with tools and environments optimized for each class, could be developed. Some common questions for all materials classes could lead to overlap in the characterization methods.

4.3 RELATIONSHIP TO THE OTHER GOALS OF THE MATERIALS ACCELERATION PLATFORM

While the techniques required for closing the loop must typically be tailored for a particular class of materials or target technology, any establishment of an autonomous laboratory must inevitably build upon the contributions of the other five goals. A holistic ontology for materials data, as described in Goal 6 (Data Infrastructure and Interchange), is necessary to enable facile and meaningful communication between the feedback loop components. A key step is to connect data and information from characterization and simulations to synthetic conditions. This includes retaining metadata from the synthesis for the experimental sample, to match with characterization and simulation, and to enable continuous optimization. Furthermore, including experiments as an integral component requires setting of tolerances for the uncertainty and quality of data. For example, samples of the same material grown in different laboratories may differ due to variations in unavoidable experimental conditions and may deviate from nominal compositions or differ in the exact identification of the phases.

To integrate materials data in closed-loop discovery systems, the data stored must go beyond the raw formats to include pipelines for feature extraction and automated processing of characterization output. For example, extraction of domain sizes, heterogeneity, microstructure, and other parameters from micrographs (e.g., scanning electron microscopy, SEM; transmission electron microscopy, TEM; atomic force microscopy, AFM; scanning tunnelling, microscopy, STM) is often performed on an ad hoc basis in a particular research group. However, tools are often not shared or connected directly in machine-friendly data formats to calculated data or experimental conditions. Therefore, along with the data, materials informatics tools are expected to help accelerate the materials discovery process within an integrated feedback loop.

Inverse design (Goal 4) could address a common roadblock in seeding an autonomous experiment: assembling an initial list of target materials. On a practical level, the adaptation of autonomous laboratories for a given problem would rely on the successful development of modular materials robotics (Goal 3). To the extent that a quickly executed single step of an autonomous loop covers a breadth of length and time scales (Goal 5), advancements in Goal 2 would also be needed to effectively close the loop. While this autonomous discovery goal involves substantial integration of the

other goals, the integration, management, and communication of the components comprise a unique set of basic research challenges whose development would be accelerated through shared resources and effective dissemination of know-how.

Decisions and research planning based on the data could initially be performed with existing AI optimization algorithms. To further capitalize on the autonomous infrastructure and enable autonomous laboratories to explore materials space beyond the confines of previous research, the reasoning and intuitive decision making of AI would be necessary.

4.4 MATERIALS-SPECIFIC CHALLENGES AND APPROACHES

4.4.1 Inorganic materials

Existing automation and high-throughput discovery platforms, both commercial and single-lab "home-built" systems, have been developed for high-throughput experimental (HTE) synthesis of thin films, nanomaterials, metal-organic frameworks, membranes [2,3,4], and more. Industry has widely adopted HTE methods for inorganics, particularly with the cost-oriented objective to "fail fast" [5]. From this work, new materials for medical stents, catalysts to reduce pollutant release, new refrigerant materials, and novel alloys for next-generation jet engines have emerged.

For example, in the area of functional oxides, accelerated photocatalyst discovery was recently demonstrated by combining high-throughput experiments with computational screening for the discovery of solar fuels photoanodes [6]. These multi-functional inorganic materials must simultaneously exhibit a broad set of properties. The primary feedback loop demonstrated by this project was refining theory and experiment techniques so that each helped inform the other and addressed materials functionalities that were particularly difficult to screen by other methods. With the refined screening pipeline in place, from a computational screening of approximately 3,000 ternary oxides, 47 materials were identified as potentially promising photoanodes, from which 17 were successfully synthesized, and 16 were identified as photoanodes that could enable fuel generation using sunlight.

This suite of discoveries represents a hallmark achievement that demonstrates the importance of closing the loop. Its success, despite an incomplete closing of the loop (screening was performed via serial filtering) and lack of automated feedback between experiments and computations, highlights the transformative impact of advancements in these areas and realization of a fully closed loop. In another example, Balachandran *et al.* demonstrated the use of a feedback loop in finding materials with desired elastic properties, with minimal iterations in optimization [7].

The current state of the art tends towards non-interchangeable, non-standardized, and application-specific approaches [8]. While there are examples of active-learning integration into autonomous systems [9], most implementations need further development of automation and integration with synthesis, AI, and data collection. Of the few existing inorganic materials knowledge extraction platforms (CombiView, Oak

Ridge National Laboratory (ORNL) group, Joint Center for Artificial Photosynthesis (JCAP) [10,11]), even fewer are open access/open source. Ideally, data taken and analyzed on the fly are referenced against open databases, both experimental (e.g., ICSD) and theoretical (e.g., OQMD [12], NOMAD [13], AiiDA [14], Materials Project [15], AFLOW [16]). Potential contradictions are also automatically flagged for deeper experimental investigation and theoretical validation. Developing this type of software requires databases to expose public application programming interfaces (APIs) that are structured to facilitate use by software engineers from a variety of backgrounds and development environments.

For future improved, modular, automated implementations, it is important to recognize that the discovery and optimization of target functional inorganic materials span a multi-dimensional space, covering functionality, stability, interface, and morphology metrics. A key decision for any autonomous research paradigm is translating hypothesis and design into metrics for measuring success. For example, in an effort to discover new Heusler compounds, Oliynyk *et al.* [17] parametrized the chemical formula for over 50 different real-valued and categorical features that could be calculated directly from the formula.

HTE characterization techniques have been developed for a broad range of inorganic materials functionalities including ferroelectric, magnetic, energy storage, catalysis, and corrosion resistance. However, existing platforms tend to be application specific and expensive. Making decisions about which techniques to include (e.g., Raman; x-ray diffraction, XRD; x-ray absorption spectroscopy, XAS), depending on the design metrics space, is a clear challenge of high-throughput characterization [18]. Another challenge for all inorganic materials is the automated investigation of a given material's microstructure, a key need in understanding performance. Nano and atomic structure is confirmed only by electron microscopy, which is difficult to automate. Advances in modular robotics would enable on-site, on-the-fly sample transfer between synthesis and characterization, accelerating an already accelerated process for materials discovery. Advances in on-the-fly machine learning during synthesis and characterization are required to maximize the amount of knowledge gained per experimental point and reduce instrument "dead time."

4.4.2 Organic materials

Synthetic chemistry is a series of discrete steps: identify the target molecule, determine the available building blocks, design the synthesis of the target from the available building blocks, carry out the synthetic experiments, analyze the results, and iterate until the synthesis of the target molecule is achieved. Once the target is synthesized, it must be purified and characterized against the target properties sought. The cycle then repeats with the identification of the next target and the process iterates until the target properties are achieved. Each step can be automated to accelerate the process and, critically, make it available to non-experts. See Section 6.4.2 for an expanded discussion of issues related to the scope and modularity of an automated synthesis machine.

Technologies enabling autonomous organic synthesis are rapidly advancing towards the goal of a truly autonomous discovery machine. Although many of the pieces required for such a machine exist today, they are not yet fully connected. Initiatives, such as DARPA's Make-It program [19] and European effort Dial-a-Molecule [20], have made advances in this area.

The automation of synthetic design has a long history of efforts to leverage computers to design synthetic routes, from LHASA in the 1960s to programs such as ChemPlanner [21] and Chematica today [22]. Although the utility of these tools has greatly improved over the years, they are still tools for experts and disconnected from experimental outcomes. None has yet been integrated into an automated synthesis platform where the AI can receive feedback from experiments to enhance the training set for the AI with real data [23].

The automation of the mechanics of chemical synthesis has been accomplished as well. For a few structural classes, this automation has existed for decades. See Section 6.4.2 for specific examples, but, for the most part, they involve biological molecules. The pharmaceutical industry has invested heavily in this area with companies such as Eli Lilly creating fully automated robotic platforms for rapid synthesis of drug-like molecules [24]. While these parallel efforts are impressive, their capital costs are cost prohibitive and they largely operate in a parallel synthesis mode that may not be the ideal platform for a truly autonomous synthesis machine. More recently, there have been efforts to develop an automated synthesis machine that can perform a broad range of reactions to assemble pharmaceutical drugs through a multi-step synthetic process carried out in small footprint flow chemical platforms (Figure 4.1) [25,26,27].

Figure 4.1 A reconfigurable system for continuous production and reformulation of APIs. [Source: ref 25]

Automation of the optimization of synthetic processes has also seen significant advances in the past decade. For some automated reaction optimization technologies, computer algorithms are used to optimize multiple reaction parameters to achieve a target goal, typically the yield or purity profile of a specific molecule. An important application is the ability to use an inline analytical technology coupled to an automated flow chemistry synthesis platform to allow for real-time iterative optimization of a synthetic reaction (Figure 4.2) [28,29].
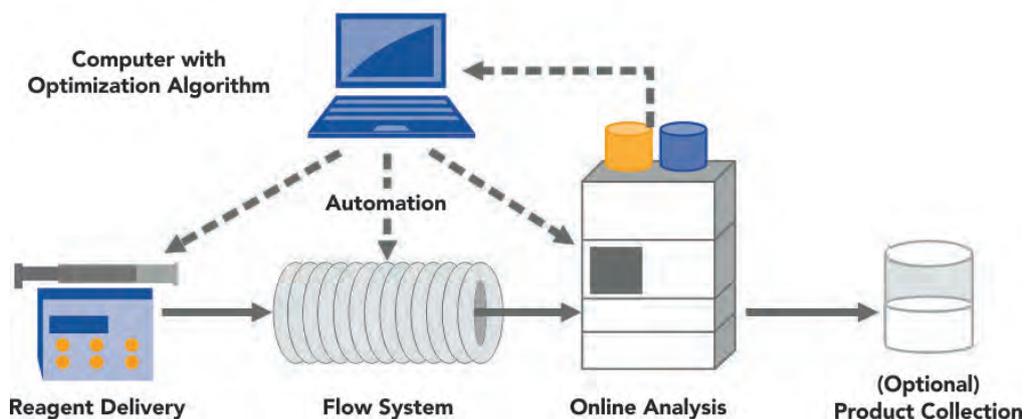
Figure 4.2 Workflow for automated feedback optimization. [Source: ref 28]

An important component of self-optimizing systems is the online and inline analysis tools available to characterize the results of the automated reactions. Spectroscopic tools, such as NMR, infrared (IR), and Raman, along with high-performance liquid chromatography (HPLC), have been successfully applied to enable automated optimization, particularly in flow chemical systems [30,31]. Relatively simple and static algorithms are used in all these automated optimization examples; enabling learning algorithms would yield improvements.

4.3.3 Nanomaterials and composites

Synthesis of nanomaterials has a vast array of potential processing conditions, such as temperature, pressure, pH, and composition, which complicates the potential applications of AI and informatics methods. Descriptors capturing synthesis and processing steps are essential for autonomous research in these systems. Nanocomposites also lack structural characterization tools that fully capture the relevant structure/property/processing relationships. Development of these tools would benefit the autonomous discovery of such materials. Additionally, the goals of synthesis for nanomaterials and nanocomposites cover multiple property dimensions, processability, process robustness, and cost-effectiveness. The vastness of the parameter space makes the challenge overwhelming for the current, human-driven research processes.

Chan *et al.* have built WANDA, a robot for high-throughput nanocrystal synthesis, which pioneered automated nanoparticle synthesis [32]. As mentioned in Chapter 2, the state of the art in closed-loop autonomous synthesis is ARES, which uses AI and closed-loop, iterative experimentation to learn to grow carbon nanotubes at targeted rates [33]. It is the first autonomous research robot for materials development.

Closed-loop autonomous research systems for nanomaterials and nanocomposites synthesis are especially useful because of the vast experimental parameter space and large number of potential operative physical and chemical phenomena. Going forward, the approach would be to build upon the state of the art to encompass more nanomaterials and nanocomposites, as well as to achieve fundamental understanding resulting from the closed-loop research systems. For example, the

transition from carbon nanotubes to scaled production would directly affect the clean energy materials goals because applying carbon nanotubes to lightweighting transportation vehicles would improve power transmission efficiency, greatly reducing energy consumption. Additionally, the transformation of hydrocarbons (e.g., from conversion of captured CO_2) to carbon nanomaterials would effectively sequester the carbon indefinitely.

For nanocomposites, developing closed-loop autonomous research systems would create the ability to understand and control their synthesis at unprecedented levels, paving the way for multi-functional, holistic property suites at scaled, cost-effective rates. Investing in the synthesis of nanomaterials and composites would result in the building of more autonomous synthesis robotics, and eventually generalize the method to a wide range of nanomaterials.

4.4.4 Universal approaches to characterization

Two kinds of loops must be considered for closed-loop platforms: optimization loops and discovery loops. *Optimization loops* typically focus on a functionality (e.g., oxygen evolution rate of a catalyst [34], band gap of a photovoltaic material, electronic conductivity for thermoelectric materials, ionic conductivity of electrolyte materials), which is closely tied to a direct characterization. It is beneficial if the characterization tool can be coupled directly to the synthesis process as both aim to narrow the composition/structure degrees of freedom.

Discovery loops focus on complex systems whose structure-function relationships are difficult to predict and attempt to explain what these relationships are in a meaningful way. In such complex systems, there are many internal degrees of freedom that can lead to emergent properties and phenomena. Because each internal degree is probed separately, there may not be clear ties between individual measurements and the ultimate functionality. Therefore, multiple measurements are needed at each sample preparation stage. Discovery algorithms need as much guidance as possible to narrow the phase space for sample growth. Machine learning with both theoretical and experimental inputs is expected to play a leading role.

There are universal characterization techniques associated with all classes of materials: scanning probe microscopies for atomic-scale information, x-ray nanoprobes for spatially resolved chemical and electronic information, lasers and x-ray lasers for time-resolved response, and tools integrated with the sample growth environments to characterize growth dynamics.

To take advantage of these and other characterization techniques on a large scale, robotic tools need to be developed and new algorithms need to be created. These algorithms need to incorporate machine learning and AI.

The historical approach to characterize materials is optimized for nominally homogeneous materials. However, when dealing with heterogeneous samples, the

value of this approach is diminished because it is by nature statistical. If the probes lack sufficient spatial resolution, or are not applied to the same region of the same samples, the information available from direct correlations is lost. Furthermore, the samples are no longer in the same environment or in the same stage of their lifetimes. Nominally identical, but distinct, samples may have been prepared under similar, but not identical, conditions and therefore further weaken the statistical strength of the enterprise. This greatly weakens the correlations between techniques, and wastes a great deal of time (Figure 2.2).

Even more important is the fact that future discoveries are likely to be found in the morphological and combinatorial modifications of materials. These would surely require side-by-side comparison of tailored samples with identical and known provenance. Creation and understanding of such sample arrays would create a future bottleneck, and replicating these arrays in multiple laboratories would not be generally possible or practical. This is especially true because many valuable techniques are surface-sensitive, or the samples themselves are dominated by surface effects.

A principal challenge in autonomous characterization is the diverse time scales and information quantities available from the disparate characterization techniques, which can vary from ultra-fast (femtosecond lasers) [35], to over seconds simple measurements (conductivity, reflectivity), to a substantial part of a day (scanning probe microscopies). While the more time-consuming characterization techniques may inhibit a high-throughput approach, they may offer a tremendously rich data set that can be leveraged in integrating AI-based predictive systems in autonomous discovery. Establishing hierarchical measurement workflows based on the time and complexity of techniques would therefore be useful.

Figure 4.3 shows such a workflow, where first, a large quantity of samples is analyzed in an automatic manner for a specific functionality that is faster to measure, after which a reduced number of highly contrasting samples is subjected to more detailed analysis (middle section of the graph) to investigate the relationship between preparation conditions and composition. This information is used to zoom in on the promising samples that are subsequently subjected to in-depth and more time-consuming analysis to understand the material with tools that yield the most important atomic, energy, momentum, spin, and/or time resolution. Theoreticians can then use this knowledge to refine predictive models. The left side of the workflow focuses on the optimization process, which can be optimized for high-speed machine-controlled operation. The right side focuses on the "why", which, on its slower time scale, is better suited to the human side of the process.

Centralization is both a significant challenge and an opportunity in the characterization of materials in the automated discovery process. How can we establish and promote an intermutual environment to foster international collaboration? How can we avoid duplication of effort to design uniform hardware and software standards for
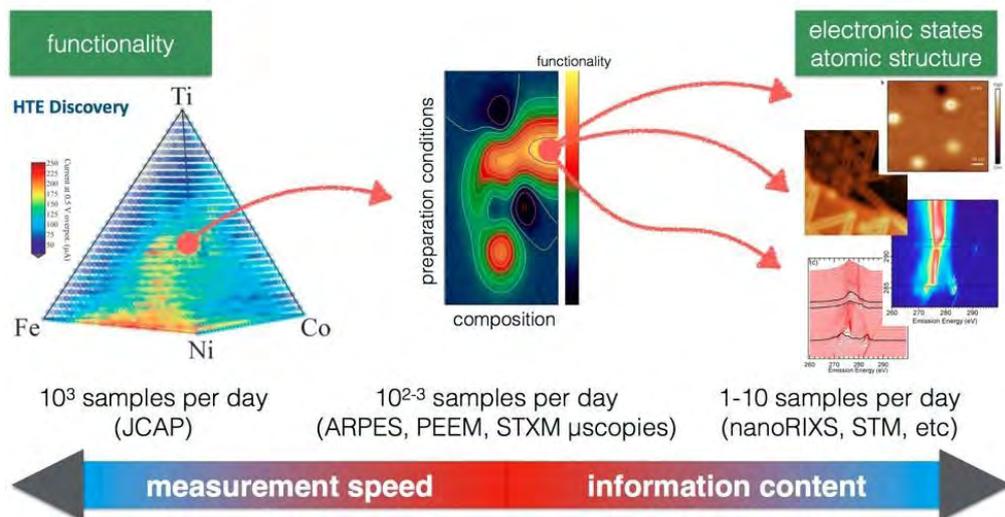
Figure 4.3 A typical progressive pathway to both optimization and understanding of material functionality begins with left, a high-throughput screening of promising material families (theoretical and/or experimental) to narrow down the phase space, followed by a more in-depth, narrower search of phase space to optimize structural/functional characteristics. The final step (right) is focused on gaining understanding through the slowest, most information-rich techniques with spatial, energy, momentum, and spin resolution. [Source: ref 36, and based on original works presented in refs 37,38,39,40]

exchanging samples between characterization tools, data archiving, etc.? These challenges can be addressed by establishing international-scale user facilities focused on automated materials characterization. Establishing centralized facilities of this nature, however, raises new challenges:

• Where should such facilities be located?

• What are the access modes for centralized facilities?

• How can we manage the optimum use of the individual growth/characterization chambers?

• Are there intellectual property (IP) concerns?

The emergence of a portfolio of centralized and decentralized organizations of different scales is expected to address these questions. Investment in centralized institutions such as existing national labs would not only integrate specialized probes (e.g., synchrotron x-rays or high-resolution TEM), but also leverage the vast engineering resources for development of uniform software and hardware tools for managing the characterization. These include the engineering of complex robotics, the modular components, algorithms, etc. An example of a nascent effort in this direction is the HTE materials collaboratory, a centre without walls, which combines access to world-class HTE and AI tools across multiple institutions to accelerate materials discovery [41]. Private-sector organizations would also participate and invest, in the same way that organizations such as Transcriptic and Emerald Cloud Lab provide services in synthetic biology, biochemistry, and pharmaceutical chemistry. Once this approach is fully developed through a concerted and centralized effort, these same tools can then be deployed in dispersed laboratories.

References for Chapter 4

1. George I. Seffers, "Scientists Pick AI for Lab Partner: Robotic research systems may foster a boom in scientific knowledge," *Signal*, September 1, 2017, www.afcea.org/content/scientists-pick-ai-lab-partner.

2. Martin L. Green, Ichiro Takeuchi and Jason R. Hattrick-Simpers, "Applications of high throughput (combinatorial) methodologies to electronic, magnetic, optical, and energy-related materials," *Journal of Applied Physics* 113, Issue 23, Number 231101 (2013), DOI: 10.1063/1.4803530.

3. Emory M. Chan, Chenxu Xu, Alvin W. Mao, Gang Han, Jonathan S. Owen, Bruce E. Cohen and Delia J. Milliron, "Reproducible, High-Throughput Synthesis of Colloidal Nanocrystals for Optimization in Multidimensional Parameter Space," *Nano Letters* 10, Issue 5 (2010): 1874–1885, DOI: 10.1021/nl100669s.

4. Jonathan K. Bunn, Richard Z. Voepel, Zhiyong Wang, Edward P. Gatzke, Jochen A. Lauterbach and Jason R. Hattrick-Simpers, "Development of an Optimization Procedure for Magnetron-Sputtered Thin Films to Facilitate Combinatorial Materials Research," *Industrial & Engineering Chemistry Research* 55, Issue 5 (2016): 1236–1242, DOI: 10.1021/acs.iecr.5b04196.

5. Daniel J. Arriola, Edmund M. Carnahan, Phillip D. Hustad, Roger L. Kuhlman and Timothy T. Wenzel, "Catalytic Production of Olefin Block Copolymers via Chain Shuttling Polymerization," *Science* 312, Issue 5774 (2006): 714-719, DOI: 10.1126/science.1125268.

6. Qimin Yan, Jie Yu, Santosh K. Suram, Lan Zhou, Aniketa Shinde, Paul F. Newhouse, Wei Chen, Guo Li, Kristin A. Persson, John M. Gregoire and Jeffrey B. Neaton, "Solar fuels photoanode materials discovery by integrating high-throughput theory and experiment," *Proceedings of the National Academy of Sciences* 114, Number 12 (2016): 3040–3043, DOI: 10.1073/pnas.1619940114.

7. Prasanna V. Balachandran, Dezhen Xue, James Theiler, John Hogden and Turab Lookman, "Adaptive Strategies for Materials Design using Uncertainties" Scientific Reports 6 (2016), DOI: 10.1038/srep19660.

8. Jason R. Hattrick-Simpers, John M. Gregoire and A. Gilad Kusne, "Perspective: Composition–structure–property mapping in high-throughput experiments: Turning data into knowledge," *APL Materials* 4, Issue 5 (2016), DOI: 10.1063/1.4950995.

9. Aaron Gilad Kusne, Tieren Gao, Apurva Mehta, Liqin Ke, Manh Cuong Nguyen, Kai-Ming Ho, Vladimir Antropov, Cai-Zhuang Wang, Matthew J. Kramer, Christian Long and Ichiro Takeuchi, "On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets," *Scientific Reports* 4 (2014), DOI: 10.1038/srep06367.

10. Sergei V. Kalinin, Evgheni Strelcov, Alex Belianinov, Suhas Somnath, Rama K. Vasudevan, Eric J. Lingerfelt, Richard K. Archibald, Chaomei Chen, Roger Proksch, Nouamane Laanait, and Stephen Jesse, "Big, Deep, and Smart Data in Scanning Probe Microscopy," *ACS Nano* 10, Issue 10 (2016): 9068–9086, DOI: 10.1021/acsnano.6b04212.

11. Santosh K. Suram, Yexiang Xue, Junwen Bai, Ronan Le Bras, Brendan Rappazzo, Richard Bernstein, Johan Bjorck, Lan Zhou, R. Bruce van Dover, Carla P. Gomes, and John M. Gregoire, "Automated Phase Mapping with AgileFD and its Application to Light Absorber Discovery in the V–Mn–Nb Oxide System," *ACS Combinatorial Science* 19, Issue 1 (2017): 37–46, DOI: 10.1021/acscombsci.6b00153.

12. Scott Kirklin, James E Saal, Bryce Meredig, Alex Thompson, Jeff W Doak, Muratahan Aykol, Stephan Rühl & Chris Wolverton, "The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies," *npj Computational Materials* 1, Number 15010 (2015), DOI: 10.1038/npjcompumats.2015.10.

13. "NOMAD Repository," The NOMAD Laboratory (website), accessed 2017, nomad-repository.eu/.

14. G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari and B. Kozinsky, "AiiDA: Automated Interactive Infrastructure and Database for Computational Science," *Computational Materials Science* 111 (2016): 218–230, DOI: 10.1016/j.commatsci.2015.09.013.

15. A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Materials* 1, 11002 (2013), DOI: 10.1063/1.4812323.

16. S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo and O. Levy, "AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations," *Computational Materials Science* 58 (2012): 227–235, DOI: 10.1016/j.commatsci.2012.02.002.

17. A. Oliynyk, E. Antono, T. Sparks, L. Ghadbeigi, M. Gaultois, B. Meredig, A. Mar, "High-throughput machine-learning-driven synthesis of full-heusler compounds," *Chemistry of Materials* 28, Issue 20 (2016): 7324–7331, DOI: 10.1021/acs.chemmater.6b02724.

18. Fang Ren, Ronald Pandolfi, Douglas Van Campen, Alexander Hexemer, and Apurva Mehta, "On-the-Fly Data Assessment for High-Throughput X-ray Diffraction Measurements," *ACS Combinatorial Science* 19, Issue 6 (2017): 377–385, DOI: 10.1021/acscombsci.7b00015.

19. Anne Fischer, "Make-It," Defense Advanced Research Projects Agency (DARPA), accessed 2017, www.darpa.mil/program/make-it.

20. Victor Sans and Leroy Cronin, "Towards dial-a-molecule by integrating continuous flow, analytics and self-optimisation," *Chemical Society Reviews* 45, Issue 8 (2016): 2032–2043, DOI: 10.1039/c5cs00793c.

21. "Wiley ChemPlanner," Wiley, accessed 2017, www.chemplanner.com.

22. S. Szymkuć, E. P. Gajewska andT. Klucznik. "Computer-Assisted Synthetic Planning: the End of the Beginning," *Angewandte Chemie International Edition* 55 (2016), DOI: 10.1002/anie.201506101.

23. Nicholas J. Browning, Raghunathan Ramakrishnan, O. Anatole von Lilienfeld, and Ursula Roethlisberger, "Genetic Optimization of Training Sets for Improved Machine Learning Models of Molecular Properties," *The Journal of Physical Chemistry Letters* 8, Issue 7 (2017): 1351–1359, DOI: 10.1021/acs.jpclett.7b00038.

24. Alexander G. Godfrey, Thierry Masquelin and Horst Hemmerle, "A remote-controlled adaptive medchem lab: an innovative approach to enable drug discovery in the 21st Century," *Drug Discovery Today* 18, Issues 17–18 (2013): 795–802, DOI: 10.1016/j.drudis.2013.03.001.

25. A. Adamo, R. L. Beingessner, M. Behnam, J. Chen, T. F. Jamison, K. F. Jensen, M. C. Monbaliu, A. S. Myerson, E. M. Revalor, D. R. Snead, T. Stelzer, N. Weeranoppanant, S. Y. Wong and P. Zhang, "On-demand continuous-flow production of pharmaceuticals in a compact, reconfigurable system," *Science* 352 (2016): 61–67, DOI: 10.1126/science.aaf1337.

26. Steven V. Ley, Daniel E. Fitzpatrick, Rebecca M. Myers, Claudio Battilocchio and Richard. J. Ingham, "Machine-Assisted Organic Synthesis," *Angewandte Chemie International Edition* 54 (2015):10122–10136, DOI: 10.1002/anie.201501618.

27. Richard J. Ingham, Claudio Battilocchio, Daniel E. Fitzpatrick, Eric Sliwinski, Joel M. Hawkins and Steven V. Ley, "A Systems Approach towards an Intelligent and Self-Controlling Platform for Integrated Continuous Reaction Sequences," *Angewandte Chemie International Edition* 54 (2014): 144 –148, DOI: 10.1002/anie.201409356.

28. Brandon J. Reizman and Klavs F. Jensen, "Feedback in Flow for Accelerated Reaction Development," *Accounts of Chemical Research* 49, Issue 9 (2016): 1786–1796, DOI: 10.1021/acs.accounts.6b00261.

29. B. J. Reizman, Y-.M. Wang, S. L. Buchwald and K. F. Jensen, "Suzuki–Miyaura cross-coupling optimization enabled by automated feedback," *Reaction Chemistry & Engineering* 1, Issue 6 (2016): 658-666, DOI: 10.1039/c6re00153j.

30. David C. Fabry, Erli Sugiono and Magnus Rueping, "Self-Optimizing Reactor Systems: Algorithms, On-line Analytics, Setups, and Strategies for Accelerating Continuous Flow Process Optimization," *Israel Journal of Chemistry* 54, Issue 5 (2014): 341–350, DOI: 10.1002/ijch.201300080.

31. D. C. Fabry, E. Sugionoa and M. Rueping, "Online monitoring and analysis for autonomous continuous flow self-optimizing reactor systems," *Reaction Chemistry & Engineering* 1, Issue 2 (2016): 129–133, DOI: 10.1039/c5re00038f.

32. Chan et al., op. cit.

33. Pavel Nikolaev, Daylond Hooper, Frederick Webber, Rahul Rao, Kevin Decker, Michael Krein, Jason Poleski, Rick Barto and Benji Maruyama, "Autonomy in materials research: a case study in carbon nanotube growth," *npj Computational Materials* 2 (2016), DOI: 10.1038/npjcompumats.2016.31.

34. K. A. Stoerzinger, O. Diaz-Morales, M. Kolb, R. R. Rao, R. Frydendal, L. Qiao, X. R. Wang, N. B. Halck, J. Rossmeisl, H. A. Hansen, T. Vegge, I. E. L. Stephens, M. T. M. Koper and Y. Shao-Horn, "Orientation-Dependent Oxygen Evolution on RuO_2 without Lattice Exchange," *ACS Energy Letters* 2 (2017): 876–881, DOI: 10.1021/acsenergylett.7b00135.

35. G. Pariente, V. Gallet, A. Borot, O. Gobert and F. Quéré, "Space–time characterization of ultra-intense femtosecond laser beams," *Nature Photonics* 10 (2016): 547–553, DOI:10.1038/nphoton.2016.140.

36. *Solving Scientific Challenges with Coherent Soft X-Rays: Workshop report on early science enabled by the Advanced Light Source Upgrade,* Advanced Light Source (ALS-U) Lawrence Berkeley National Laboratory and U.S. Department of Energy, January 18–20, 2017, als.lbl.gov/wp-content/uploads/2017/08/ALS-U-Early-Science-Workshop-Report-Full.pdf.

37. John M. Gregoire, Chengxiang Xiang, Xiaonao Liu, Martin Marcin and Jian Jin, "Scanning droplet cell for high throughput electrochemical and photoelectrochemical measurements," *Review of Scientific Instruments* 84, 024102 (2013), DOI: 10.1063/1.4790419.

38. Dillon Wong, Jairo Velasco Jr, Long Ju, Juwon Lee, Salman Kahn, Hsin-Zon Tsai, Chad Germany, Takashi Taniguchi, Kenji Watanabe, Alex Zettl, Feng Wang and Michael F. Crommie, "Characterization and manipulation of individual defects in insulating hexagonal boron nitride using scanning tunnelling microscopy," *Nature Nanotechnology* 10 (2015): 949–953, DOI: 10.1038/nnano.2015.188.

39. Sara Barja, Sebastian Wickenburg, Zhen-Fei Liu, Yi Zhang, Hyejin Ryu, Miguel M. Ugeda, Zahid Hussain, Zhi-Xun Shen, Sung-Kwan Mo, Ed Wong, Miquel B. Salmeron, Feng Wang, Michael F. Crommie, D. Frank Ogletree, Jeffrey B. Neaton and Alexander Weber-Bargioni, "Charge density wave order in 1D mirror twin boundaries of single-layer $MoSe_2$," *Nature Physics* 12 (2016): 751–756, DOI: 10.1038/nphys3730.

40. Ruimin Qiao, Qinghao Li, Zengqing Zhuo, Shawn Sallis, Oliver Fuchs, Monika Blum, Lothar Weinhardt, Clemens Heske, John Pepper, Michael Jones, Adam Brown, Adrian Spucces, Ken Chow, Brian Smith, Per-Anders Glans, Yanxue Chen, Shishen Yan, Feng Pan, Louis F. J. Piper, Jonathan Denlinger, Jinghua Guo, Zahid Hussain, Yi-De Chuang and Wanli Yang, "High-efficiency in situ resonant inelastic x-ray scattering (iRIXS) endstation at the Advanced Light Source," *Review of Scientific Instruments* 88, 033106 (2017), DOI: 10.1063/1.4977592.

41. National Institute of Standards and Technology (NIST) Materials Genome Initiative (MGI), "High-Throughput Experimental Materials Collaboratory," *NIST MGI: Gateway to Materials Genome Information* (website), last updated February 18, 2015, mgi.nist.gov/htemc.

5 Goal #2: Artificial Intelligence for Materials

AI is a broad field of computer science with the general goal of making machines intelligent. Within the past decade, AI has achieved breakthroughs in computer vision, speech, language translation, and natural language understanding. AI-based computer systems have surpassed humans for complex tasks such as image recognition, cancer detection, and games (e.g., Chess, Go, Poker, Jeopardy, video games). Active areas of AI research include self-driving cars and the understanding of human speech. This chapter introduces Artificial Intelligence for Materials (AI4M) as a Grand Goal at the intersection of AI and materials research.

Machine learning, a subfield of AI that has exploded in recent years, involves statistical algorithms that improve with experience. Support vector machines, kernel methods, and neural networks are among the methods used for machine learning. Deep neural networks are perhaps the largest active research area due to the applications that they enable. These types of algorithms benefit from and enhance automation. As robotics transforms design, manufacturing, and transportation — constituting a modern industrial revolution — achieving the analogous materials discovery revolution requires AI and, in particular, the emulation of human scientific intuition, reasoning, and decision making.

Humans have traditionally enjoyed reasoning and intuition capabilities that far exceed those of machines. As AI erodes these barriers, materials science must not only adopt the state of art in AI, but also push the frontier of AI to mimic, and then supersede, the scientific intuition and decision making of an expert materials scientist. The AI4M goal exemplifies the need to advance beyond available machine learning algorithms in two primary ways: (i) elevating machine learning from generating data models to generating human-understandable explanations, inferences, and conclusions; and (ii) enabling autonomous reasoning about these outcomes and prior data to generate an actionable research plan. These accomplishments are critical for conducting basic science where

knowledge and understanding take precedence over quantitative results, rather than an optimization or an image recognition task for which a data model can be sufficiently scored by its goodness-of-fit or convergence rate.

5.1 OPPORTUNITIES

The current challenges in realizing AI4M range from technical to organizational. The necessary AI algorithms cannot be derived or adapted from the present AI field, largely because development of these algorithms did not consider materials science. Algorithm inputs for materials science still look less like big data and more like sparse, heterogeneous data that can only be understood in the context of the chemical and physical laws that constrain the accessible portions of the vast, multi-dimensional materials parameter space. AI for materials must be enlightened by these chemical and physical laws. This constitutes an entirely new field of research that can only be pioneered at the intersection of the AI and materials communities, with top researchers from both working side by side. The need for collaboration requires an evolution of the organizational and funding mechanisms of stakeholders and research institutions. New mechanisms for cross-discipline education are also needed to enable an informed research community to develop multidisciplinary research programs that can simultaneously advance the materials and AI fields.

Machine learning provides an opportunity to blend together two broad classes of models. First, *discriminative* models are powerful ways of representing correlations in the data. For example, the chemistry and materials community has a long history of success in developing quantitative structure-property prediction models. More recently, advanced kernel methods and neural networks have allowed for advances in the field.

Second, *generative* models allow for the generation of samples from probability distributions. Traditional methods, such as the Monte Carlo method, have been used for some time. Other methods, specifically autoencoders and generative adversarial networks, have rapidly advanced other areas of machine learning for the generation of artificial data. Examples include computer-generated art and video. These methods are relevant to AI4M as they employ a latent-space representation, i.e., a relatively low dimensional vector space to represent the problem of interest. The points in this space can be clustered according to properties of interest, and can lead to the understanding and generalization of concepts from the materials data.

Generating models that not only give the correct predictions, but also lead to new understandings of materials chemistry and physics is an important challenge in machine learning. The overall problem of generating human-interpretable machine learning results is an active area in computer science research (e.g., Fairness, Accountability, and Transparency in Machine Learning (FAT/ML) initiative [1], DARPA's Explainable AI program [2]). Attaining "understanding" of materials phenomena requires translation of knowledge across different physical models and types of computational and experimental data; therein lie not only the initial barriers to

development of materials-specific AI algorithms, but also the opportunities for these algorithms to supersede human capabilities of deriving understanding from data. Moreover, the many external constraints, such as experimental cost, hazards, desirability, and regulatory concerns, must be appropriately incorporated into material-specific AI. In the short term, human-machine cooperation may be the most effective way to capture the tacit knowledge, as demonstrated in a recent application to the optimization of nuclear fusion experiments [3]. Materials-specific software platforms that facilitate rapid and broad collection of human expertise, e.g., via crowdsourcing, are required for building the datasets needed to train AI4M to reproduce scientific expertise [4].

5.2 BENEFITS AND IMPACT

AI4M can enable breakthroughs and materials solutions for clean energy technologies that are unimaginable/unattainable using current approaches on any time scale. Although the materials science community's achievements to date are laudable, human creativity and intellect may be limiting our communal progress in providing materials that enable clean energy technologies — allowing AI4M to disrupt cost and performance learning curves. Materials-specific advancements in AI would make automated research "better" as well as "faster". That is, autonomous laboratories would not only exceed human capabilities in speed and efficiency, but also in creativity and problem solving, which hinge upon successful development of AI4M. In the same way that a brainstorming session of a team of scientists generates more creative ideas than from an individual alone, a community of enlightened AI algorithms could provide a suite of complementary research paths that follow different lines of reasoning to accelerate innovation.

As this field develops, progress can be tracked analogously to other implementations of AI, where landmark achievements in game playing become achievements in clean energy as AI4M-conceived materials outperform those proposed by human scientists to advance clean energy technology. For example, IBM's Watson defeated Jeopardy! Champions, and Google's AlphaGo defeated the world's best Go players only later to lose to DeepMind, which has never received any human training [5]. To extend the analogy, in the wake of AlphaGo's dominance, artificially conceived strategies have inspired top Go players, demonstrating how AI can expand the collective mindset of a human community. In this way, AI4M would not only identify new high-performance clean energy materials, but also educate humans by demonstrating new innovation strategies and new understanding of materials chemistry and physics.

5.3 RELATIONSHIP TO THE OTHER GOALS OF THE MATERIALS ACCELERATION PLATFORM

Achieving this goal both depends upon and enables the success of the other five goals. In many ways, it focuses on developing the "central brain" that learns from a variety of data streams to make informed decisions. Therefore, success in this area would provide immediate tangible results and solutions in the other goal areas.

Reasoning and AI model interpretation offer a direct insight into and facilitate inverse design (Goal 4). AI also provides a necessary technology to facilitate a dramatic breakthrough in modular materials robotics (Goal 3) and bridging length and time scales (Goal 5).

AI reasoning capabilities are required to fully capitalize on the infrastructure described in the chapters on autonomous discovery and development (Goal 1) and data infrastructure and interchange (Goal 6). In some ways, AI4M complements these goals because it is ultimately tasked with solving problems not readily addressable by robotics and data-driven approaches.

5.4 MATERIALS-SPECIFIC CHALLENGES AND APPROACHES

5.4.1 Inorganic materials

For solid-state materials, a good example of the need for AI4M to overcome a current research barrier is the automated generation of phase diagrams from characterization data, particularly XRD patterns [6,7]. Machine learning algorithms developed in other fields have been effectively adapted and deployed for data reduction and providing phase diagram-like information for simple materials systems [8,9,10,11]. In addition to the wealth of unexplored materials space, some existing datasets are resistant to interpretation and understanding by both state-of-the-art algorithms and expert scientists, the latter struggling to conceptualize the full dataset due to the inability of 2D and 3D visualization to sufficiently capture the data structure in multi-dimensional composition spaces. While computers are less intimidated by the dimensionality of the problem, the algorithms do not hold the prior knowledge and reasoning skills of the scientists. Integrating human and machine intelligence can overcome the barriers in this field to accelerate identification of new inorganic materials. This is particularly important for clean energy technologies that require new multi-functional materials.

 The design of synthesis and characterization experiments is particularly challenging for inorganic materials due to the breadth of complementary synthesis techniques, the expansive parameter space within each technique, and the need to tailor choice and parameterization of characterization experiments based on the synthesis. Since these decisions are made using reasoning based on the deep experience of experts, autonomous laboratories will be limited in their efficiency and ability to explore materials space until AI4M is integrated with the experiment automation. Beyond the ability to choose an appropriate synthesis method for a desired material, AI4M must in the long run also be able to propose previously unexplored approaches, such as the ingenious 2010 Nobel Prize-winning realization that exfoliation provides single-layer graphene. To the extent that machine learning is a box, it will be challenging to teach a machine to think outside of the box; however, the greatest opportunities lie in overcoming this challenge.

One example of the state of the art in this area is the use of machine learning to predict previously unknown similarities between inorganic materials (Figure 5.1),

then to use these similarities to rapidly predict the properties of materials. For example, AFLOW consortium data make it possible to find similarities that can predict important properties relating to light absorption and mechanical strength (e.g., metal/insulator classification, band gap energy, bulk/shear moduli, Debye temperature, heat expansion coefficient, heat capacities), which can be used to predict these properties for virtually any new inorganic crystalline solid [12]. In turn, this prediction can be used to extract simple heuristic rules that explain "why" a particular material has these properties, in terms of thermodynamic properties of the atoms and chemical bonds comprising the material. These simple rules not only reduce the time for estimating properties from >1000 hours/material to <0.1 second/material, but also provide human-interpretable insight and facilitate inverse design (Goal 4, Chapter 7).
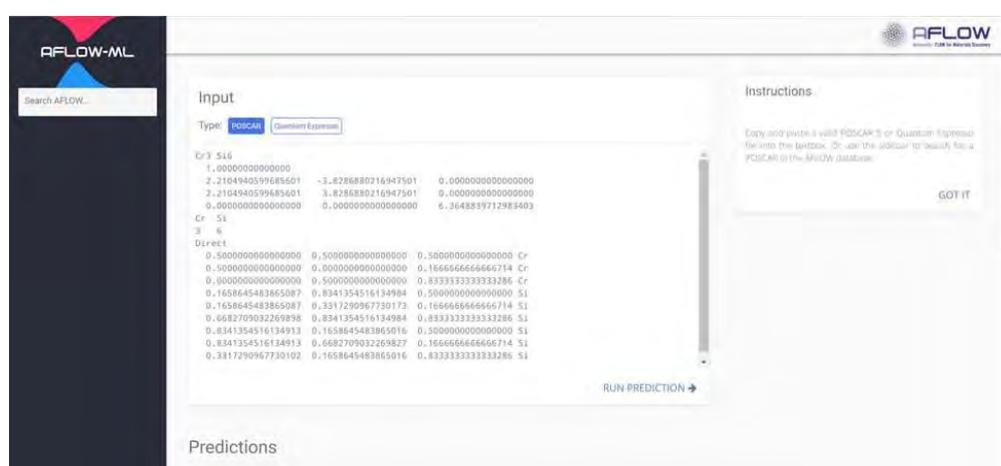
Figure 5.1 A screenshot of user-friendly machine learning web app at http://aflow.org/aflow-ml to predict properties of inorganic materials.

A second example is the discovery of new materials for catalysis. Starting from a set of detailed density functional theory simulations of catalyst surfaces and their interaction with the substrates, it was possible to use machine learning methods to extract the key quantities that predict the catalytic activity. To date, these key quantities are generalizations of existing theoretical models of catalysis, such as finding a relationship to the d-band centre [13] and d-band width [14] of the material, coordination number of the surface catalytic atoms [15], and atomic-orbital specific coordination numbers of those surface atoms [16]. These relationships "explain" the detailed simulation results in terms of existing theories, and also serve as simple rules-of-thumb for the design of new materials. Moving forward, the challenges in this field include (i) using developing AI methods that can use this type of information to plan the most promising "next move" in unexplored chemical space that will yield enhanced performance, and (ii) developing machine learning methods that can go beyond existing catalysis concepts.

5.4.2 Organic materials

Quantum mechanics computation, combined with the advent of supercomputing, has drastically increased our ability to understand and identify materials computationally. AI4M offers a new revolution in computational efficiency and efficacy with two main high-impact avenues: (i) enhancing the speed and accuracy of calculations for a family of materials, and (ii) using data and understanding of different methods to design and automatically execute computation campaigns.

In the first area, phenomenal success has been recently demonstrated in the computational modelling of organic molecules through adoption of machine learning methods. Instead of performing expensive quantum mechanics calculations on large molecular screening libraries, calculations for a subset of the library were used to train a more efficient machine learning model for discovering organic light-emitting diodes (LEDs) [17]. Neural networks have also been used to learn an appropriate quantum mechanical model of organic molecules that, when incorporated into mass computation, offered a million-fold increase in computational efficiency (Figure 5.2) [18]. For fast calculations to generate scientific knowledge, the calculation accuracy must be optimized and quantified. Machine learning methods such as Bayesian statistics have also enabled rapid progress in this area, by identifying when fast calculations can retain the accuracy of more time-consuming methods and by learning how to identify and correct systematic errors in the calculations [19, 20, 21, 22, 23, 24].
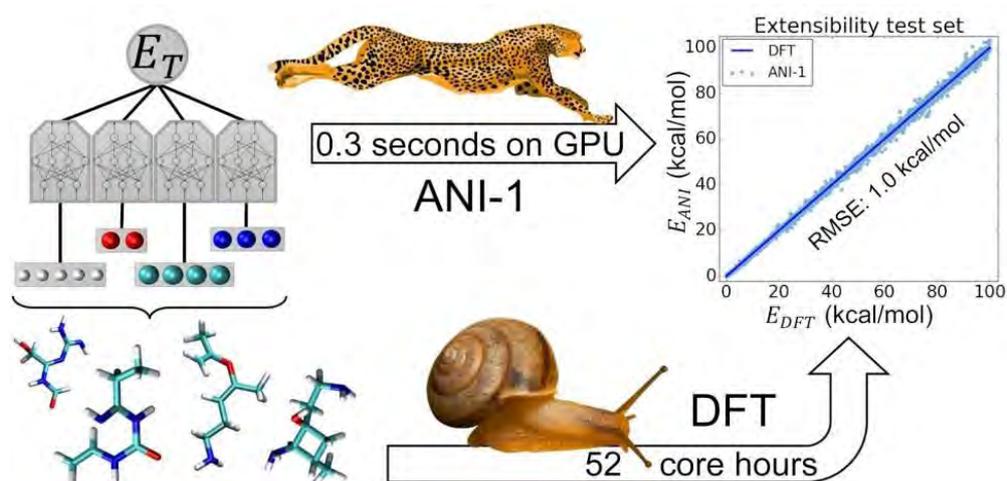

Figure 5.2 Comparison between speed and accuracy of neural network versus DFT in energy prediction for organic molecules.

With these improvements in place, human decisions on which calculations to perform and what we have learned from the results will increasingly become a bottleneck, requiring advancement in the second area noted above where AI emulates human decision making. A model called ORGAN (Objective-Reinforced Generative Adversarial Networks) is capable of generating organic molecules as SMILES (a text representation) from scratch by employing a generative adversarial network

[25]. ORGAN can be trained to produce more diverse molecules through a reward mechanism that highly resembles classical conditioning in psychology, which can be good at producing outliers. Variational autoencoder representations are another recent development widely applied to sentence and image generation [26]. These methods attempt to map discrete molecules to a continuous space. Once in this continuous space, it is possible to move around in the chemical space and find new molecules outside of the training set. In the example above, the authors found two molecules that scored better than any of the molecules in their training set.

Another key pillar of AI in organic functional materials is determining how to synthesize the promising molecules generated by theory. Similar to Google Maps that help humans navigate possible directions to a destination, novel AI algorithms can provide routes for organic synthesis planning (Figure 5.3). Recent work has addressed this problem for the exploration of the reaction space [27], as well as for targeting specific molecular structures [28, 29, 30].
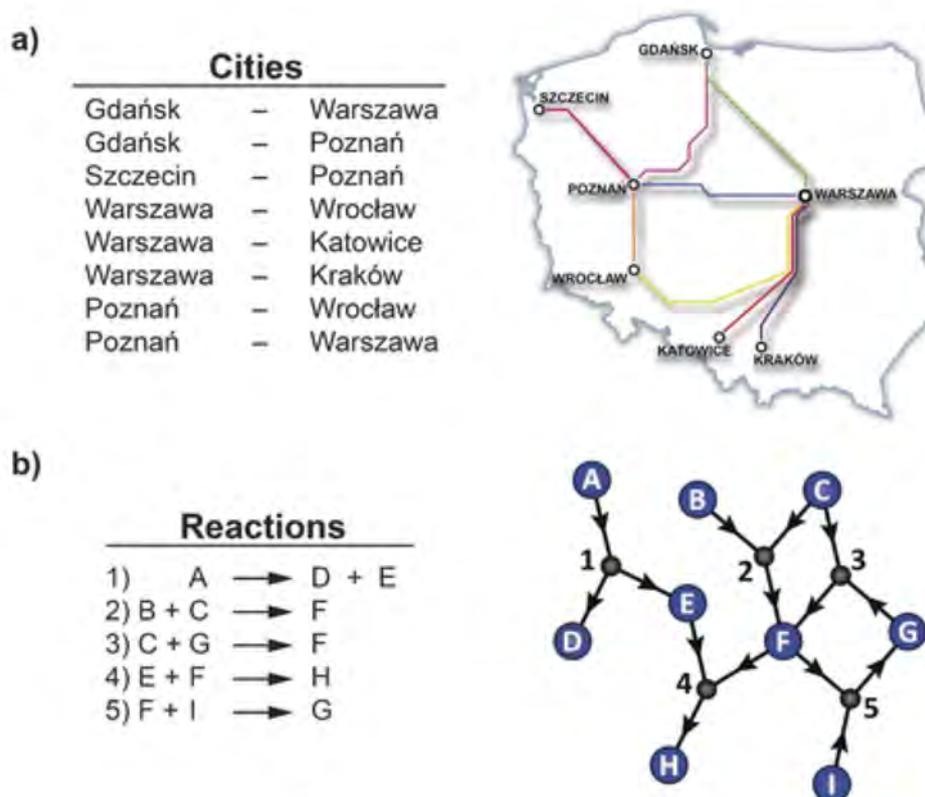
Figure 5.3 (a) Finding the shortest path between cities connected by a set of roads is a challenging computer science problem, although efficient algorithms exist for finding nearly optimal solutions. (b) Optimizing the series of chemical reactions needed to synthesize a new organic molecule is an analogous problem. Individual types of molecules correspond to "cities" and the reactions that interconvert them correspond to "roads"; often these are "one-way streets" because the reactions are irreversible. [Source: adapted from ref 31]

AI algorithms can potentially discover unknown chemical reactions and build a large network of reaction as well as navigate a chemist through the known set of potential pathways, from reactant to product, in an optimal way. Further algorithm development can take this one step further. Using the generative methods described

above, machines can suggest completely new reactions, opening new pathways in the synthesis network or, in the language of Google Maps, create new roads and highways to reach new destinations, i.e., novel materials.

5.4.3 Nanomaterials and composites

An early proof of principle that AIs can learn from data and generate human-understandable scientific hypotheses has been shown for the synthesis mechanisms of organic-inorganic hybrid materials with applications to catalysis and non-linear optical mechanisms [32]. After demonstrating that the machine learning model "knew" more than human experts (as measured by its success in predicting the conditions needed to make new materials), the model was used to generate artificial data that could be used to make a surrogate model — a "model of the model" — that expresses the machine's decision-making process in terms of human-readable descriptions of the reactant properties and reaction conditions (see Figure 5.4).

This process revealed three previously unknown hypotheses, which have subsequently been verified in the laboratory, about the experimental conditions and properties needed to produce these materials. Future work should focus on (i) extending these types of explanations to include more general types of experimental descriptions and use various forms of characterization and computational results as justifications, and (ii) closing the loop by allowing AIs to plan future experiments that optimally test these hypotheses under the direction of human experts. Both of these require the development of materials-specific AI methods that can present evidence convincingly to materials scientists and chemists.

Machine learning has already been used for microstructure [33] and thermal properties optimization of stanene nanostructures [34]. While exciting developments are taking place in this field, the key challenges remain unresolved. Algorithms must be developed to make autonomous decisions to discover new materials rather than optimize properties out of known ones. Most of the studies so far have shown the importance of machine learning for specific properties and materials systems. Standardized methods and best practices for using AI tools for nanocomposite design and optimization are still lacking. Furthermore, the reliability of developed methods for new applications, properties, and materials systems has not been assessed properly. A limited set of machine learning tools cater to the prediction of morphologies and kinetic behaviour such as phase transformation. Addressing this challenge requires better theoretical methods to resolve time-scale issues in atomistic modelling (Goal 5, Chapter 8) of polymers and composite materials. Machine learning tools do not yet utilize sophisticated multi-scale methods that could resolve the kinetics challenge. Finally, since current tools are not open source or integrated into open source packages, they are not easily accessible. As Goal 6 (Chapter 9) envisions open data access, open access to AI4M algorithms and concepts is critical for meeting MAP's ambitious vision.
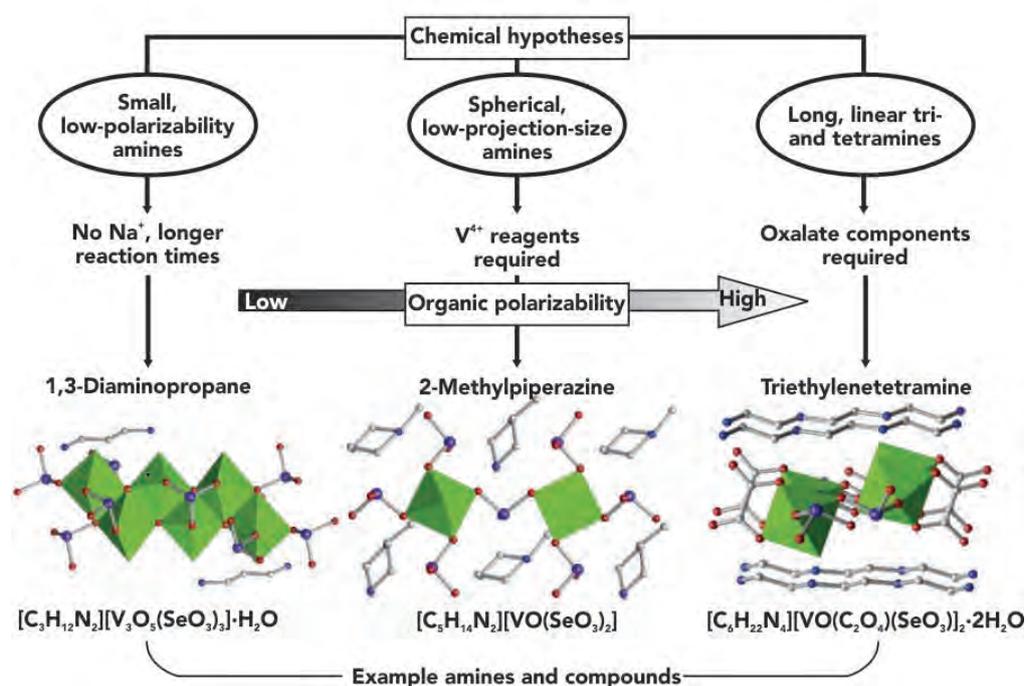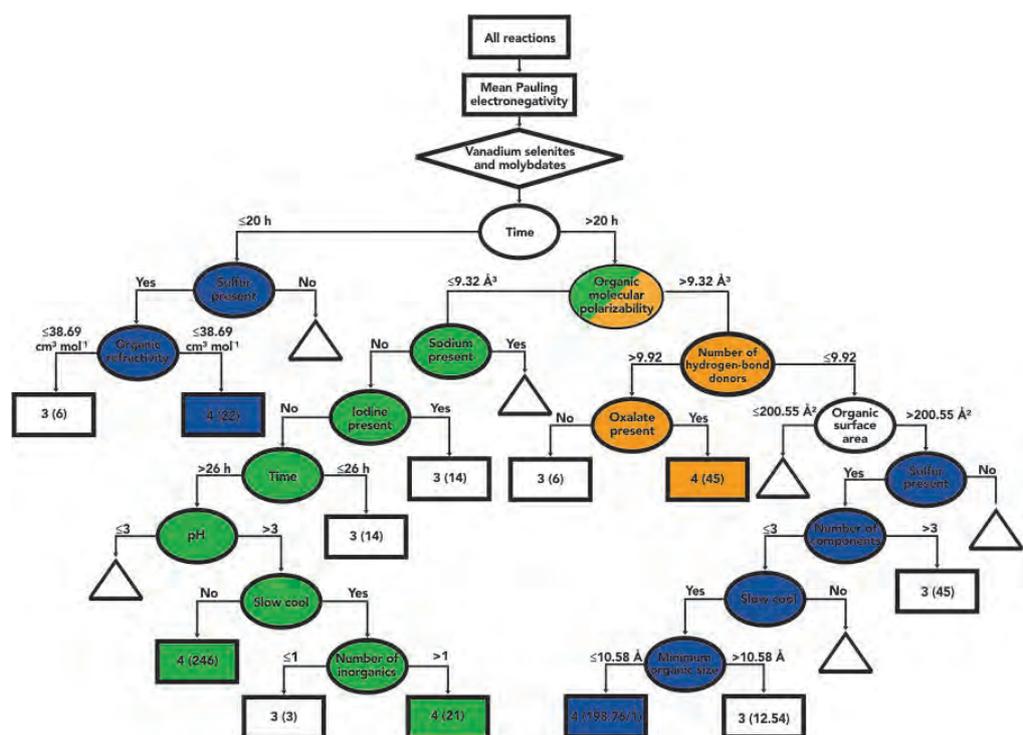
Figure 5.4 Teaching humans what AIs know about the conditions needed to synthesize new materials. Although the AI uses a model that is too complicated to be interpreted by humans, it can be used to create a "model of a model" that represents its knowledge as a tree of "if-then" decisions in terms of chemical properties and reaction conditions. (Left Inset) Human experts can read the colour-coded path of decisions that lead to successfully making a new material, and from this obtain new, previously unknown hypotheses about the reaction formation conditions (shown schematically in the right inset). [Source: ref 32]

References for Chapter 5

1. *Fairness, Accountability, and Transparency in Machine Learning* (FAT/ML) (website), accessed 2017, www.fatml.org/.

2. David Gunning, "Explainable Artificial Intelligence (XAI)," Defense Advanced Research Projects Agency (DARPA), accessed 2017, www.darpa.mil/program/explainable-artificial-intelligence.

3. E. A. Baltz, E. Trask, M. Binderbauer, M. Dikovsky, H. Gota, R. Mendoza, J. C. Platt and P. F. Riley, "Achievement of Sustained Net Plasma Heating in a Fusion Experiment with the Optometrist Algorithm," *Scientific Reports* 7, Number 6425 (2017), DOI:10.1038/s41598-017-06645-7.

4. Michael Nielson, Reinventing Discovery: The New Era of Networked Science (Princeton University Press, 2011).

5. Demis Hassabis, Dharshan Kumaran, Christopher Summerfield and Matthew Botvinick, "Neuroscience-Inspired Artificial Intelligence," *Neuron* 95, Issue 2 (2017): 245-258, DOI: 10.1016/j.neuron.2017.06.011.

6. Jason R. Hattrick-Simpers, John M. Gregoire and A. Gilad Kusne, "Perspective: Composition–structure–property mapping in high-throughput experiments: Turning data into knowledge," *APL Materials* 4, Issue 5 (2016), DOI: 10.1063/1.4950995.

7. L. Ward, R. Liu, A. Krishna, V. I. Hegde, A. Agrawal, A. Choudhary and C. Wolverton, "Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations," *Physical Review B* 96, Issue 2, Number 024104 (2017), DOI: 10.1103/PhysRevB.96.024104.

8. Yuma Iwasaki, A. Gilad Kusne and Ichiro Takeuchi, "Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries," *npj Computational Materials* 3, Number 4 (2017), DOI: 10.1038/s41524-017-0006-2.

9. C. J. Long, D. Bunker, X. Li, V. L. Karen, and I. Takeuchi, "Rapid identification of structural phases in combinatorial thin-film libraries using x-ray diffraction and non-negative matrix factorization," *Review of Scientific Instruments* 80, 103902 (2009), DOI: 10.1063/1.3216809.

10. A. G. Kusne, D. Keller, A. Anderson, A. Zaban and I. Takeuchi, "High-throughput determination of structural phase diagram and constituent phases using GRENDEL," *Nanotechnology* 26 (2015), DOI: 10.1088/0957-4484/26/44/444002.

11. Santosh K. Suram, Yexiang Xue, Junwen Bai, Ronan Le Bras, Brendan Rappazzo, Richard Bernstein, Johan Bjorck, Lan Zhou, R. Bruce van Dover, Carla P. Gomes, and John M. Gregoire, "Automated Phase Mapping with AgileFD and its Application to Light Absorber Discovery in the V–Mn–Nb Oxide System," *ACS Combinatorial Science* 19, Issue 1 (2017): 37–46, DOI: 10.1021/acscombsci.6b00153.

12. Olexandr Isayev, Corey Oses, Cormac Toher, Eric Gossett, Stefano Curtarolo and Alexander Tropsha, "Universal fragment descriptors for predicting properties of inorganic crystals," *Nature Communications* 8 (2017), DOI: 10.1038/ncomms15679.

13. Xianfeng Ma, Zheng Li, Luke E. K. Achenie and Hongliang Xin, "Machine-Learning-Augmented Chemisorption Model for CO_2 Electroreduction Catalyst Screening," *The Journal of Physical Chemistry Letters* 6, Issue 18 (2015): 3528–3533, DOI: 10.1021/acs.jpclett.5b01660.

14. Juhwan Noh, Jaehoon Kim, Seoin Back and Yousung Jung, "Catalyst design using actively learned machine with non-ab initio input features towards CO_2 reduction reactions," *Cornell University Library*, September 14, 2017, arXiv:1709.04576 [cond-mat.mtrl-sci].

15. Federico Calle-Vallejo, David Loffreda, Marc T. M. Koper and Philippe Sautet, "Introducing structural sensitivity into adsorption–energy scaling relations by means of coordination numbers," *Nature Chemistry* 7 (2015): 403–410, DOI:10.1038/nchem.2226.

16. Xianfeng Ma and Hongliang Xin, "Orbitalwise Coordination Number for Predicting Adsorption Properties of Metal Nanocatalysts," *Physical Review Letters* 118, 036101 (2017), DOI: 10.1103/PhysRevLett.118.036101.

17. Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, David Duvenaud, Dougal Maclaurin, Martin A. Blood-Forsythe, Hyun Sik Chae, Markus Einzinger, Dong-Gwang Ha, Tony Wu, Georgios Markopoulos, Soonok Jeon, Hosuk Kang, Hiroshi Miyazaki, Masaki Numata, Sunghan Kim, Wenliang Huang, Seong Ik Hong, Marc Baldo, Ryan P. Adams and Alán Aspuru-Guzik, "Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach," *Nature Materials* 15 (2016): 1120–1127, DOI:10.1038/nmat4717.

18. J. S. Smith, O. Isayev and A. E. Roitberg, "ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost," *Chemical Science* 8 (2017): 3192–3203, DOI: 10.1039/C6SC05720A.

19. Felix Brockherde, Leslie Vogt, Li Li, Mark E. Tuckerman, Kieron Burke and Klaus-Robert Müller, "By-passing the Kohn-Sham equations with machine learning," *Cornell University Library*, June 15, 2017, arXiv:1609.02815 [physics.comp-ph].

20. Rune Christensen, Heine A. Hansen and Tejs Vegge, "Identifying systematic DFT errors in catalytic reactions," *Catalysis Science & Technology* 5 (2015): 4946-4949, DOI: 10.1039/C5CY01332A.

21. J. Wellendorff, K. T. Lundgaard, A. Mogelhoj, V. Petzold, D. D. Landis, J. K. Norskov, T. Bligard and K. W. Jacobsen, "Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation," *Physical Review B* 85, Issue 23, 235149 (2012), DOI: 10.1103/PhysRevB.85.235149.

22. Christensen et al., op. cit.

23. Rune Christensen, Heine A. Hansen, Colin F. Dickens, Jens K. Nørskov and Tejs Vegge, "Functional Independent Scaling Relation for ORR/OER Catalysts," *The Journal of Physical Chemistry C* 120, Issue 43 (2016): 24910–24916, DOI: 10.1021/acs.jpcc.6b09141.

24. Andrew A. Peterson, Rune Christensen and Alireza Khorshidi, "Addressing uncertainty in atomistic machine learning," *Physical Chemistry Chemical Physics* 19, Issue 18 (2017): 10978–10985, DOI: 10.1039/c7cp00375g.

25. Gabriel Lima Guimaraes, Benjamin Sanchez-Lengeling, Carlos Outeiral, Pedro Luis Cunha Farias, Alán Aspuru-Guzik, "Objective-Reinforced Generative Adversarial Networks (ORGAN) for Sequence Generation Models," *Cornell University Library*, November 29, 2017, arXiv:1705.10843 [stat.ML].

26. Rafael Gómez-Bombarelli, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams and Alán Aspuru-Guzik, "Automatic chemical design using a data-driven continuous representation of molecules," *Cornell University Library*, January 6, 2017, arXiv:1610.02415v2 [cs.LG].

27. Jennifer N. Wei, David Duvenaud and Alán Aspuru-Guzik, "Neural Networks for the Prediction of Organic Chemistry Reactions," *ACS Central Science* 2, Issue 10 (2016): 725–732, DOI: 10.1021/acscentsci.6b00219.

28. Bowen Liu, Bharath Ramsundar, Prasad Kawthekar, Jade Shi, Joseph Gomes, Quang Luu Nguyen, Stephen Ho, Jack Sloane, Paul Wender and Vijay Pande, "Retrosynthetic Reaction Prediction Using Neural Sequence-to-Sequence Models," *ACS Central Science* 3, Issue 10 (2017): 1103–1113, DOI: 10.1021/acscentsci.7b00303.

29. Marwin Segler, Mike Preuss and Mark P. Waller, "Towards 'AlphaChem': Chemical Synthesis Planning with Tree Search and Deep Neural Network Policies," *Cornell University Library*, January 31, 2017, arXiv:1702.00020 [cs.AI].

30. Marwin H. S. Segler and Prof. Mark P. Waller, "Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction," *Chemistry: A European Journal* 23, Issue 25 (2017): 5966–5971, DOI: 10.1002/chem.201605499.

31. S. Szymkuć, E. P. Gajewska andT. Klucznik. "Computer☐Assisted Synthetic Planning: the End of the Beginning," *Angewandte Chemie International Edition* 55 (2016), DOI: 10.1002/anie.201506101.

32. P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, "Machine-learning-assisted materials discovery using failed experiments," *Nature* 533 (2016): 73–77.

33. Ruoqian Liu, Abhishek Kumar, Zhengzhang Chen, Ankit Agrawal, Veera Sundararaghavan and Alok Choudhary, "A predictive machine learning approach for microstructure optimization and materials design," *Scientific Reports* 5, Number 11551 (2015), DOI: 10.1038/srep11551.

34. Mathew J. Cherukara, Badri Narayanan, Alper Kinaci, Kiran Sasikumar, Stephen K. Gray, Maria K. Y. Chan and Subramanian K. R. S. Sankaranarayanan, "*Ab Initio*-Based Bond Order Potential to Investigate Low Thermal Conductivity of Stanene Nanostructures," *The Journal of Physical Chemistry Letters* 7, Issue 19 (2016): 3752–3759, DOI: 10.1021/acs.jpclett.6b01562.

6 Goal #3: Modular Materials Robotics

Today, synthesizing materials is an activity reserved for highly trained specialists. Identifying synthetic pathways is intellectually different from producing material variations in an effort to develop critical structure-property relationships and create new materials. Even among experts, the level of methodological specialization for synthesis and characterization limits a researcher's ability to explore all possible materials for satisfying an application-specific functionality. Thus, large portions of materials systems remain unexplored, not due to an inability to imagine islands of promising functionality, but rather to an inability to apply the appropriate tools to scale these lofty barriers. Decoupling methodological expertise from materials discovery would enable researchers to fully explore materials space to target functionality. Even if this were achieved, the lack of scale in synthesis and characterization would prevent rapid discovery in novel materials spaces. Therefore, a flexible and scalable infrastructure is necessary to translate the understandings developed in the other five goals into systems that design, synthesize, characterize, and iterate until discovery.

This third goal envisions a modular robotic system composed of building blocks with standardized interfaces for source management, synthesis, and characterization. In effect, these building blocks would serve as the basis for a hypothesis testing facility that could rapidly retool for new problems in the same vein as modern factories. The modular nature of these building blocks would enable the flexibility necessary to couple the most appropriate synthesis methodologies with the most appropriate characterization capabilities. Standardizing interfaces, which is key to decoupling methodological expertise from discovery, would ensure that an application-specific researcher could traverse the full landscape of materials structure and chemistry. Decoupling of methodological expertise from discovery would not invalidate domain specialists, but rather refocus their efforts on codifying and embedding their intuition into the modular blocks, which would in turn enable application researchers to rapidly converge on the appropriate parameters to fully leverage each technique.

6.1 OPPORTUNITIES

The current generation of synthesis and characterization is not suited for modularity or flexibility. While certain industries, such as semiconductors, have managed to develop large-scale production systems with several heterogeneous interleaved steps, these have been the result of decades of development and billions in investment. Further, the proliferation of vendor-specific interfaces creates a myriad of technical specifications that must be supported for every new capability. This generates a large overhead for breakthroughs in fundamental synthesis technologies due to the need for large development efforts to enable broad interest and support. The same complications are compounded in characterization where proprietary formats and interfaces limit or even deter interoperability. These R&D issues need to be addressed by creating new tools that do not have these systematic limitations.

Many of the general challenges surrounding synthesis revolve around the development of new technologies that can provide the necessary flexibility and modularity for an adaptive system. This requires an earnest effort to develop generic interfaces and protocols that many groups can use to develop tools and techniques. Data formats used for characterization are often proprietary or cryptic, making integration a near impossibility. While standards are not required, incentives are needed to motivate industry to work together on the components of a modular robotic system. Although significant resources are poured into robotics and automation in other fields, their development focuses on orthogonal necessities such as bipedal motion or large-scale stabilization. These, however, are not the most pressing needs for well-structured research settings that are at the limits of technologies, such as manipulators or precision and reproducibility in position.

6.2 BENEFITS AND IMPACT

Similar fields that have successfully deployed modular and flexible architectures for synthesis have experienced explosive growth in discovery, development, and interest. For example, flexible and modular synthesis and characterization infrastructures in the manufacturing of integrated circuits have led to a scalable platform for discovery and production that enables our modern technological world. In fact, modular and flexible production systems are a cornerstone of developing a technology from a lab scale into an industry, ranging from simple products such as chemicals and source materials to complex devices such as pharmaceuticals, cars, airplanes, and even rockets. The fields that now support modern society depend on the ability to make this transition. Recently, the development of automated production tools in proteomics and genomics has made synthetic proteins and polynucleotides widely available and created tremendous impact in biology and medicine. Similarly, modular materials robotics would enable the new generation of materials scientists to rapidly and efficiently develop materials for targeted applications.

A successful modular robotic system would also democratize materials synthesis and characterization, enabling a broader community of non-experts to participate in

the discovery process. In many ways, it would mirror the rapid explosion of interest and activity in additive manufacturing with the proliferation of three-dimensional (3D) printing via the maker movement. This type of organic growth naturally lends itself to self-directed ecosystems that enable all users to leverage not only the infrastructure, but also the expertise of participants. It could potentially lead to an exponential increase in materials discovery and deployment for clean energy goals.

6.3 RELATIONSHIP TO THE OTHER GOALS OF THE MATERIALS ACCELERATION PLATFORM

This goal focuses on the most interdisciplinary and crucial step to translate theoretical materials exploration into actual discovery. Development of modular materials robotics would test hypotheses from AI (Goal 2), develop the understanding necessary to successfully apply inverse design (Goal 4), and bridge length and time scales (Goal 5) into tangible results. Over time, it would evolve into the autonomous discovery loop (Goal 1) described in Chapter 4. Appropriate selection of the building blocks that are inputs to the automated synthesis device would be a critical issue. From the characterization, MAP would enable intelligent selection of pathways based on optimization (based on AI or other methods) to achieve the precise synthesis of the target compounds. Such a development would then allow autonomous materials selection and optimization for a particular application (Goal 4).

6.4 MATERIALS-SPECIFIC CHALLENGES AND APPROACHES

6.4.1 Inorganic materials

One of the largest and most successful efforts in modular materials robotics is the design and optimization of semiconductor deposition clusters that enabled the onset of the silicon age. The semiconductor industry has maintained the trend of Moore's law for more than 50 years through targeted capital investment driven by industry-agreed milestones. The requirements of device density, complexity, and low-defect density have long dictated that humans cannot possibly produce wafers with sufficient yield to make the technologies competitive. To ameliorate this, the semiconductor industry developed automated deposition tools in which a wafer is inserted into the cluster and, several hours later, millions of semiconductor gate stacks are ejected. These systems are modular and interoperable so that they can be repurposed based on the needs of the overarching industry (e.g., photovoltaics versus power electronics) and tailored to the needs of each stakeholder (e.g., Intel versus Micron). In fact, most post-processing steps that lead to the eventual device are also automated for similar reasons. The pioneering efforts and investments of the semiconductor industry in automating the assembly and testing of heteromaterial systems provide a strong basis for automated materials discovery used by the materials community over the last 20 years [1]. (See Figure 6.4.1 for an example with inorganic thin films.)

This bedrock of experience is both a valuable resource for understanding how to automate synthesis and a significant barrier to generating a general paradigm for
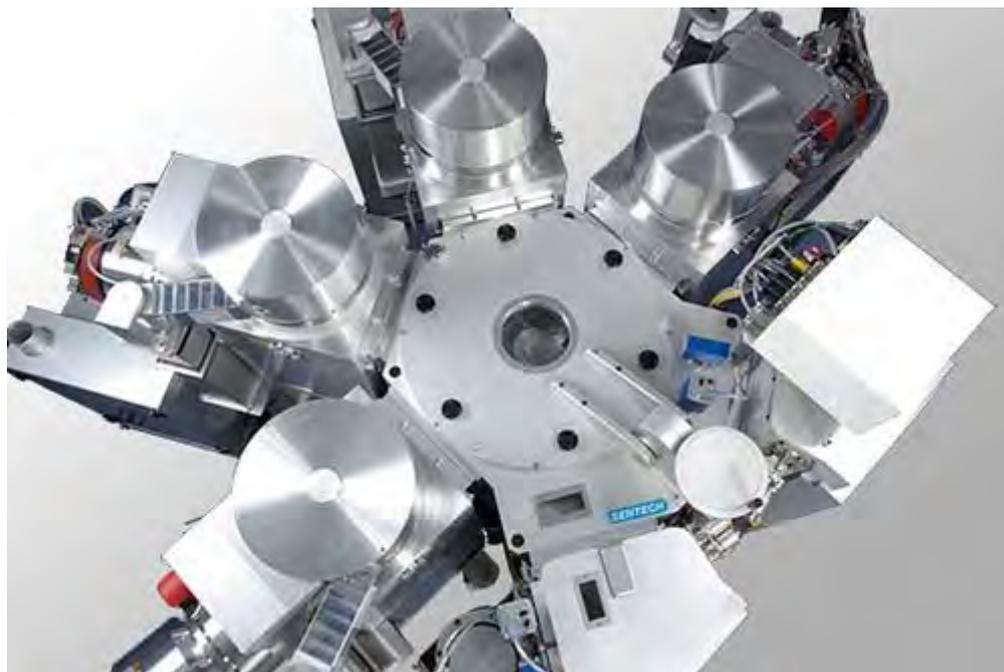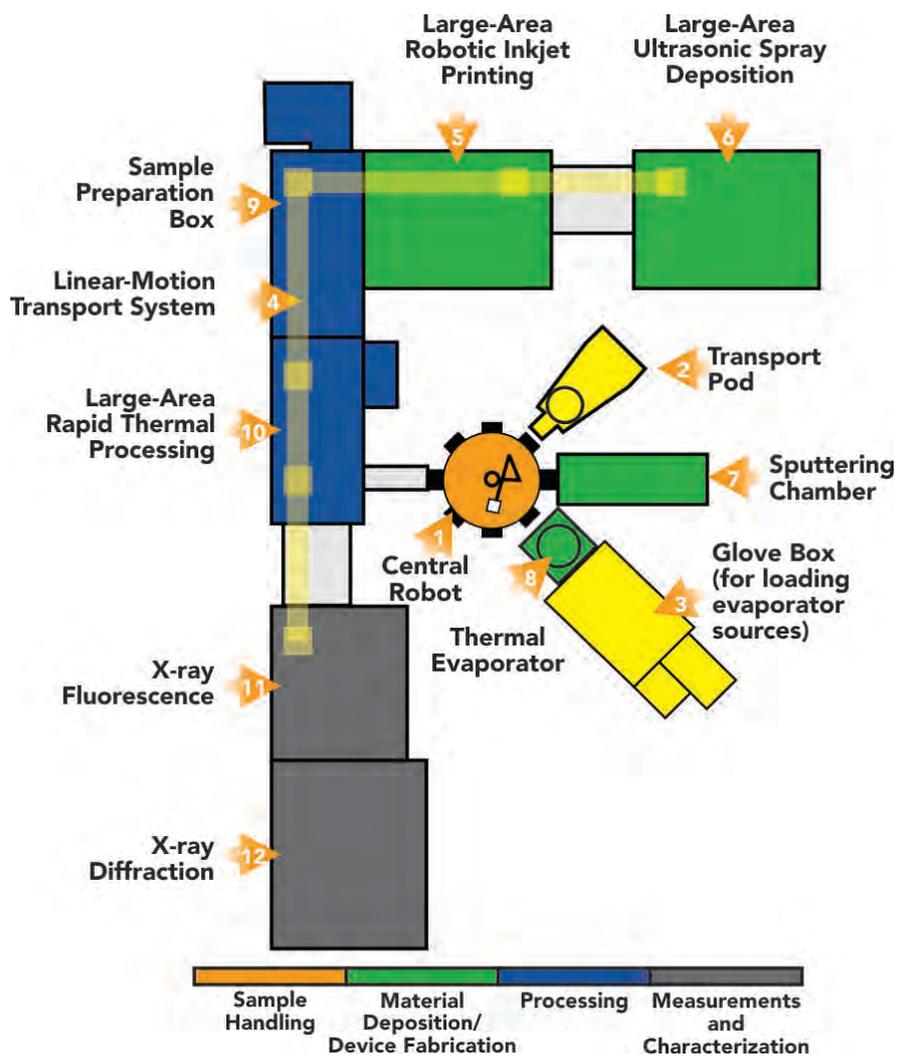
Figure 6.1 Cluster tools for high-throughput synthesis of inorganic thin films. [Source: refs 2,3]

automating synthesis. The wafer geometry dictates much of the design criteria from manipulators to synthesis chambers to characterization methodologies. Moving away from this one overarching constraint is difficult because much of the available sample manipulation technology in research also depends on wafer geometry. General manipulator systems that provide the precision necessary for extremely precise and reproducible characterization would enable other technologies to rapidly automate and bring down the cost of auxiliary technologies for further development in synthesis automation.

A key goal for automated synthesis infrastructure is a capability to self-optimize based on targeted characterization metrics. One example is if deposition clusters could detect errors, create actionable information, and provide autonomous corrections to changes in device quality during the deposition. Triaging anomalous device/synthesis failure during production represents large expenditures for both the semiconducting device and tool fabrication industries. Creating the capability for autonomous on-the-fly deposition optimization would be a game-changing technology.

Just as the development of new materials can improve the performance of existing technologies, a range of transformational new technologies is possible if materials can be developed with properties that make them economically viable. Thermoelectric materials could make it possible to harvest electricity from heat-flow or move heat from the flow of electricity. Magnetocaloric materials could enable super-efficient heating and cooling of buildings, if materials with the right combination of properties were identified [4]. Both these classes of materials lack effective scale-up mechanisms for synthesis and characterization. Scaling is a key challenge to the employment of any new material, but particularly so if industry lacks experience with that specific class, chemistry, geometry, etc. Automated facilities could be a key de-risking technology to industry adoption.

Many opportunities exist for new materials to contribute to energy efficiency and, correspondingly, many approaches to developing the necessary materials. However, the approaches and labs that pursue them are largely unconnected today. An enormous range of synthesis methods is applied to inorganic materials, including techniques that take place in the solid, liquid, vapour, and plasma phases. In some cases, the methods are necessarily batch processes and not easily automated. While it is possible, for example, to contemplate the creation of robotics to undertake mass-customized tasks such as glass-blowing for the encapsulation of materials to make single crystal specimens for detailed characterization, it is more profitable to address processes that are immediately amenable to automation. New paradigms for solving the same synthesis problems may be the only way to automate a task such as glass-blowing.

6.4.2 Organic materials

Organic molecules are used as medicines, functional coatings, energy storage, and energy harvesting, to name but a few applications. To design a universal organic materials synthesis machine is a massive undertaking due to the huge number of possible molecular structures. An important research question is to determine the minimal space of molecular architecture that can address the broadest range of properties, i.e., the functional degeneracy of organic molecules. If a high degree of functional diversity can be achieved with a small set of structural diversity or from a set of molecules that can all be obtained using a small set of reactions, the complexity of the design of an automated device is greatly reduced. Research is needed to determine the structural space that would enable the broadest functional space.

The automation of chemical synthesis has been accomplished for a few structural classes. Peptide synthesis was first automated in the 1960s through the development of solid-phase peptide synthesis technology. More recently, the synthesis of polynucleotides and oligosaccharides has been automated. New methods are being developed to automate the synthesis of polyketide-type products. For all these methods, the products produced are built up from the iterative use of a single reaction with structurally related starting materials. There have been efforts to develop an automated synthesis machine that can perform a broad range of reactions to assemble pharmaceutical drugs through a multi-step synthetic process [5]. The ability to automate the mechanics of synthesis has been even extended to the synthesis of biological molecules such as DNA and proteins [6].

Any high-throughput research pipeline that takes new functional organic compounds to a fully optimized and tested device integration must incorporate automated materials characterization tools to achieve a time-efficient feedback loop. Advanced characterization tools integrated with modular materials robotics, such as spectroscopies (e.g., NMR, infrared, ultra-fast spectroscopy), are essential to close the loop of materials design and automated synthesis.

6.4.3 Nanomaterials and composites

Nanomaterials, such as nanoparticles and nanowires, are an intriguing class of organic or inorganic materials with dimensions of between 1 and 100 nanometres (nm). The nanometre is a critical length scale because it lies between the typical dimensions of molecules (0.1 nm) and those of bulk materials (>100 nm). Consequently, physical properties such as color, melting point, chemical reactivity, and magnetism often vary as the size of nanometre-scale materials decreases. Furthermore, new properties can emerge by tailoring not only nanoparticle size and morphology, but also by positioning chemically diverse atoms, domains, and functional groups at precise locations within such materials. For example, the specific location of individual dopant atoms within inorganic nanocrystals directly leads to novel physical properties, enabling new energy-saving smart window technologies [7], new

materials for hydrogen storage [8], and catalysts for generating fuel from sunlight [9]. Nanostructured materials of this sophistication must be programmed by multi-step chemical synthesis, where heterogeneous components are assembled in a particular order, each under specific conditions. Much like the modular synthesis of peptides and organic molecules, the atomic structure of inorganic nanomaterials can be controlled by the regulation of a few key synthesis variables in sequence.

To design next-generation, multi-functional nanomaterials and nanocomposites with an impact on energy applications, it is essential to establish modular, precision-tunable, and robust protocols for their synthesis. The major challenge in designing and fabricating such complex nanomaterials is that many different synthetic methods exist for a given nanomaterial, and synthetic parameters such as reagent concentration and growth temperature are numerous and highly interdependent. Understanding these dependencies and optimizing synthetic pathways is difficult because the chemical mechanisms for nanomaterial growth are not well defined. When nanoparticles are combined with disparate materials into nanocomposites (e.g., metal nanoparticles in a polymer matrix), fabrication methods become even more complex and time- and resource-intensive. Typical approaches to produce bulk nanocomposites (e.g., precipitation of secondary phases from metastable solid solutions) lack precision control over the distribution of phases and/or are limited in compositional versatility. Therefore, new low-cost and simple strategies to produce bulk nanocomposites must be developed [10].

Modular robotics has the potential to enable high-throughput synthesis and evaluation of these materials to gather and trend data. These automated workflows can potentially find the limits of a given synthetic method in producing a nanomaterial with a given chemical or physical property. The ability to "fail fast" using high-throughput, automated synthesis would enable rapid evaluation of different synthesis recipes for achieving a target material. Convergence on a single and robust protocol would discourage proprietary recipes and allow faster entry of newly discovered nanomaterials into technology space.

Early automation of nanoparticle synthesis has been achieved in continuous flow reactors often in millifluidic or microfluidic volumes [11,12]. In flow reactors, reagent concentrations are controlled digitally by varying the flow rates of intersecting streams of precursor solutions. Reaction times are controlled by varying the flow rates through known lengths of tubing, which are heated across programmed temperature profiles. The reactions are monitored spectroscopically by placing optical probes along or at the end of the tubing. Rapid optimization of nanoparticle properties, such as luminescence color and color purity of CdSe nanoparticles, has been demonstrated using feedback between online spectroscopy, reagent flow rates, and reactor temperatures. The advantage of flow reactors is that they can be fabricated modularly and inexpensively. However, their use is restricted to well-behaved, homogeneous reactions in which a limited number of reagents are used. This precludes their use for screening a large number of parameters.

The batch synthesis of nanoparticles in parallel arrays offers a more flexible platform for automated nanocrystal synthesis. However, due to the more extreme reaction conditions, only a few custom-developed robots have been used to automate the synthesis, characterization, and high-throughput screening of nanoparticles. The Molecular Foundry, a U.S. DOE nanoscience user facility, pioneered the automated, high-throughput synthesis of inorganic nanocrystals, introducing its first-generation robot, WANDA, in 2008 [13]. (See Figure 6.2 for a photograph of HERMAN, WANDA's successor.) WANDA and other synthesis robots perform typical chemistry operations such as dispensing reagents, heating and stirring reactions, and sampling aliquots into arrays of 8 to 96 reactor wells. Modular characterization tools such as high-throughput microplate readers measure the absorption, luminescence, x-ray diffraction, Raman scattering, and dynamic light scattering of products synthesized by upstream robots.

Figure 6.2 HERMAN. Robotic deck of HERMAN, the High-through Experimentation Robot for Multiplexed Automation of Nanochemistry. HERMAN, the successor of WANDA, is used to synthesize colloidal inorganic nanomaterials and nanocomposites at the Molecular Foundry (Lawrence Berkeley National Laboratory, USA). Each module of HERMAN (horizontally distributed deck elements) performs fundamental operations such as weighing solid reagents, dispensing liquid solutions, heating and stirring chemical reactions, sampling products over time, and purifying products, all inside an inert glovebox atmosphere.

Automated nanoparticle synthesis workflows provide researchers with a set of robust protocols to rapidly test multiple theory-guided hypotheses via systematic variation of nanocrystal properties or reaction conditions. Users of WANDA, drawn from an international base of researchers, have generated significant insights into the mechanisms of nanocrystal growth [14, 15] and identified new materials for thermoelectric [16], sensing [11], and electronic [18] applications. These automated synthesis workflows have increased the reproducibility and throughput of nanocrystal syntheses by an order of magnitude [19], facilitating the high-quality production of semiconductors (CdSe), metals (Ag), and oxides (ZnO) (Figure 6.3). Rapid iteration with robots has allowed the realization of precise control over nanomaterial parameters such as size [20], shape [21], color [22,23], emission intensity [24], and thermal conductivity [25].

Despite the successes of automating nanomaterials synthesis with robotics, these approaches still require significant human interaction and have not yet integrated AI into autonomous, closed-loop workflows. Specifically, the examples referenced above
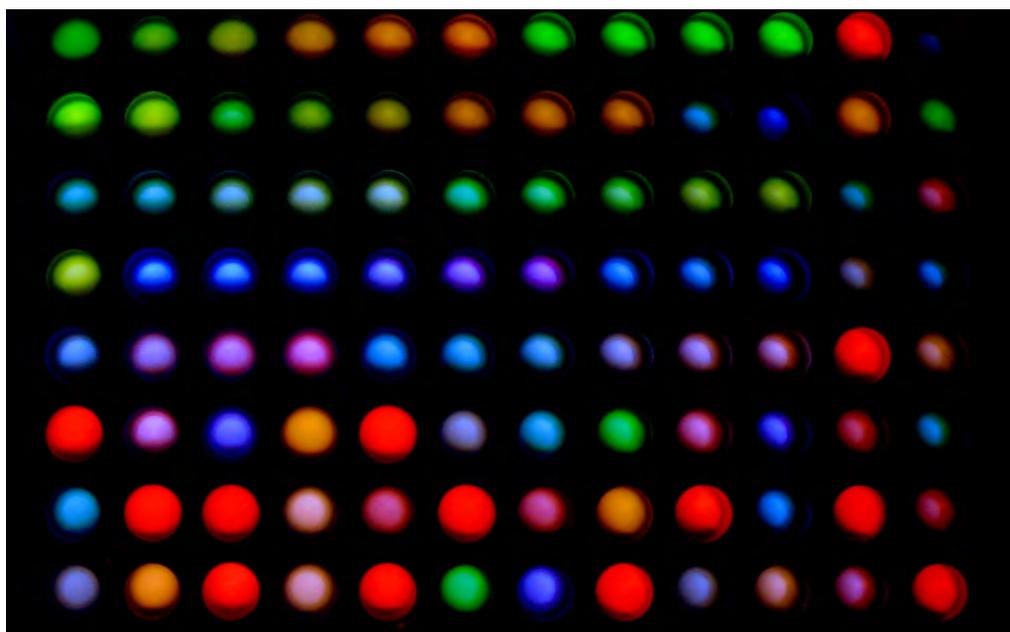
Figure 6.3 Microplate. Photoluminescence micrograph of robotically synthesized cadmium selenide and cadmium telluride nanoparticle solutions in a 96-well microplate under ultraviolet illumination. The color of each nanoparticle reaction (each well) varies with their size and growth conditions. [Credit: Emory Chan, Molecular Foundry]

largely harness combinatorial strategies that inefficiently test every combination of parameters, resulting in only a small fraction of optimized "hits." Despite the accelerated iteration cycles provided by high-throughput methods, the permutations of synthesis variables number in the billions or trillions — far too many to test with even the most efficient robot. While several approaches have leveraged physical models to narrow the number of required experiments, physical models are often sparse in the early stages of the development of a new nanomaterial. Consequently, robotic nanomaterials synthesis would benefit significantly from integration with feedback from AI algorithms that rely on statistical models early in research campaigns, while evolving into physically insightful models at later stages.

Practically, the main obstacle to the autonomous discovery of nanomaterials is the inability to verify 3D structure and internal composition at the same rate that the products can be synthesized in high-throughput robotics. Electron microscopy is a critical characterization tool because the properties of nanomaterials are inherently linked to their size and shape, which can only be resolved by electron microscopes. However, standard electron microscopes are only able to characterize 1 to 5% of the samples currently synthesized by high-throughput robotics because the repetitive act of loading, focusing, and imaging large numbers of nanoparticle samples is prohibitively time-consuming (ca. 5 samples/h, 200 nanoparticles/sample). Additionally, while traditional TEM imaging gives a wealth of information (e.g., size, shape, faceting), the resulting micrographs are 2D projections that mask 3D morphology and spatial variations in composition, which are critical to understanding of multi-component heterostructures. Although automated "critical-dimension" scanning electron microscopes are available for the semiconductor

industry and a handful of automated transmission electron microscopes have been custom built for biomedical research, these rare and expensive tools have not been integrated modularly into a robotic synthesis and screening workflow. Ultimately, the automation of 2D and 3D TEM characterization modules to match the production rate of automated synthesis workflows would dramatically accelerate the discovery of atomically engineered nanomaterials for clean energy and catalysis.

References for Chapter 6

1. M. L. Green, C. L. Choi, J. R. Hattrick-Simpers, A. M. Joshi, I. Takeuchi, S. C. Barron, E. Campo, T. Chiang, S. Empedocles, J. M. Gregoire, A. G. Kusne, J. Martin, A. Mehta, K. Persson, Z. Trautt1, J. Van Duren, and A. Zakutayev, "Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies," *Applied Physics Reviews* 4, Issue 1, 011105 (2017), DOI: 10.1063/1.4977487.

2. Cluster tool diagram (website), accessed 2017, https://www.nrel.gov/pv/assets/images/ap-diagram.jpg.

3. Cluster tool image (website), accessed 2017, http://www.sentech.com/files/1/0/0/10708.jpg.

4. K. A. Gschneidner and V. K. Pecharsky, "Thirty years of near room temperature magnetic cooling: Where we are today and future prospects," *International Journal of Refrigeration* 31 (2008): 945–961, DOI: 10.1016/j.ijrefrig.2008.01.004.

5. Junqi Li, Steven G. Ballmer, Eric P. Gillis, Seiko Fujii, Michael J. Schmidt, Andrea M. E. Palazzolo, Jonathan W. Lehmann, Greg F. Morehouse and Martin D. Burke, "Synthesis of many different types of organic small molecules using one automated process," *Science* 347, Issue 6227 (2015) 1221–1226, DOI: 10.1126/science.aaa5414.

6. Craig Venter in Nature Biotechnology 35, 672–675 (2017) doi:10.1038/nbt.3859

7. Anna Llordés, Guillermo Garcia, Jaume Gazquez and Delia J. Milliron, "Tunable near-infrared and visible-light transmittance in nanocrystal-in-glass composites," *Nature* 500 (2013): 323–326, DOI: 10.1038/nature12398.

8. Ki-Joon Jeon, Hoi Ri Moon, Anne M. Ruminski, Bin Jiang, Christian Kisielowski, Rizia Bardhan and Jeffrey J. Urban, "Air-stable magnesium nanocomposites provide rapid and high-capacity hydrogen storage without using heavy-metal catalysts," *Nature Materials* 10 (2011): 286–290, DOI: 10.1038/nmat2978.

9. Vismadeb Mazumder, Youngmin Lee and Shouheng Sun, "Recent Development of Active Nanoparticle Catalysts for Fuel Cell Reactions," *Advanced Functional Materials* 20 (2010): 1224–1231, DOI: 10.1002/adfm.200902293.

10. Maria Ibáñez, Zhishan Luo, Aziz Genç, Laura Piveteau, Silvia Ortega, Doris Cadavid, Oleksandr Dobrozhan, Yu Liu, Maarten Nachtegaal, Mona Zebarjadi, Jordi Arbiol, Maksym V. Kovalenko and Andreu Cabot, "High-performance thermoelectric nanocomposites from nanocrystal building blocks," *Nature Communications* 7, Number 10766 (2016), DOI: 10.1038/ncomms10766.

11. Emory M. Chan, Richard A. Mathies, and A. Paul Alivisatos, "Size-Controlled Growth of CdSe Nanocrystals in Microfluidic Reactors," *Nano Letters* 3, Issue 2 (2003): 199–201, DOI: 10.1021/nl0259481.

12. Samuel Marre and Klavs F. Jensen, "Synthesis of micro and nanostructures in microfluidic systems," *Chemical Society Reviews* 39 (2010): 1183–1202, DOI: 10.1039/B821324K.

13. Emory M. Chan, Chenxu Xu, Alvin W. Mao, Gang Han, Jonathan S. Owen, Bruce E. Cohen and Delia J. Milliron, "Reproducible, High-Throughput Synthesis of Colloidal Nanocrystals for Optimization in Multidimensional Parameter Space," *Nano Letters* 10, Issue 5 (2010): 1874–1885, DOI: 10.1021/nl100669s.

14. Jonathan S. Owen, Emory M. Chan, Haitao Liu and A. Paul Alivisatos, "Precursor Conversion Kinetics and the Nucleation of Cadmium Selenide Nanocrystals," *Journal of the American Chemical Society* 132, Issue 51 (2010): 18206–18213, DOI: 10.1021/ja106777j.

15. Michael D. Clark, Sanat K. Kumar, Jonathan S. Owen and Emory M. Chan, "Focusing Nanocrystal Size Distributions via Production Control," *Nano Letters* 11, Issue 5 (2011): 1976–1980, DOI: 10.1021/nl200286j.

16. Joseph P. Feser, Emory M. Chan, Arun Majumdar, Rachel A. Segalman and Jeffrey J. Urban, "Ultralow Thermal Conductivity in Polycrystalline CdSe Thin Films with Controlled Grain Size," *Nano Letters* 13, Issue 5 (2013): 2122–2127, DOI: 10.1021/nl400531f.

17. Jacob D. Kilbane, Emory M. Chan, Christian Monachon, Nicholas J. Borys, Elizabeth S. Levy, Andrea D. Pickel, Jeffrey J. Urban, P. James Schuck and Chris Dames, "Far-field optical nanothermometry using individual sub-50 nm upconverting nanoparticles," *Nanoscale* 8, Issue 22 (2016): 11611–11616, DOI: 10.1039/c6nr01479h.

18. Saahil Mehra, Amy Bergerud, Delia J. Milliron, Emory M. Chan, and Alberto Salleo, "Core/Shell Approach to Dopant Incorporation and Shape Control in Colloidal Zinc Oxide Nanorods," *Chemistry of Materials* 28, Issue 10 (2016): 3454–3461, DOI: 10.1021/acs.chemmater.6b00981.

19. Chan et al. 2010, op. cit.

20. Chan et al. 2010, op. cit.

21. Mehra et al., op. cit.

22. Chan et al. 2010, op. cit.

23. Emory M. Chan, Gang Han, Joshua D. Goldberg, Daniel J. Gargas, Alexis D. Ostrowski, P. James Schuck, Bruce E. Cohen, and Delia J. Milliron, "Combinatorial Discovery of Lanthanide-Doped Nanocrystals with Spectrally Pure Upconverted Emission," *Nano Letters* 12, Issue 7 (2012): 3839–3845, DOI: 10.1021/nl3017994.

24. Daniel J. Gargas, Emory M. Chan, Alexis D. Ostrowski, Shaul Aloni, M. Virginia P. Altoe, Edward S. Barnard, Babak Sanii, Jeffrey J. Urban, Delia J. Milliron, Bruce E. Cohen and P. James Schuck, "Engineering bright sub-10-nm upconverting nanocrystals for single-molecule imaging," *Nature Nanotechnology* 9 (2014): 300–305, DOI: 10.1038/NNANO.2014.29.

25. Feser et al., op. cit.

7 Goal #4: Inverse Design

The current approach to materials discovery centres on the use of human intuition as the starting point and driver for desired functionality. That intuition is the basis for a search on structure, composition, and properties, which, in a linear fashion, reduce into hypotheses, testing methodologies, and iterative processes that update the initial intuition until the desired target functionality is achieved. Computational screening has accelerated the "guess and check" process, but it will always be hampered by the limits of this Edisonian paradigm.

Inverse design, as the name implies, inverts the paradigm by starting with the final goal and searches for an ideal materials solution to a particular clean energy challenge. It is closely related to the generative models described in Chapter 5. Inverse design represents the "Holy Grail" for autonomous materials discovery, as it connects the desired properties to specific compositions, structures, and materials as targets to synthesize, characterize, and test. This requires the design process to incorporate the necessary physics that describes the desired functionality. In this case, the researcher creates the machinery for discovery. Once enacted, it can self-optimize to identify the pathway to the desired goal. The inverse design process (see Figure 7.1) is currently not autonomous or automated, except for special instances in specific systems. Moreover, it has not been integrated across disciplines.

7.1 OPPORTUNITIES

The fundamental challenges in realizing generic inverse design strategies emanate from the strong correlation of steps involved in conventional design processes. Several paradigm changes are required to ensure that this ideology does not become a wrapper around a still inefficient Edisonian process, but rather disrupts the perceptions of materials discovery.

Defining appropriate building blocks is the most important capability that is still lacking. This describes a wide array of efforts from developing appropriate fingerprints for structure-property relationships
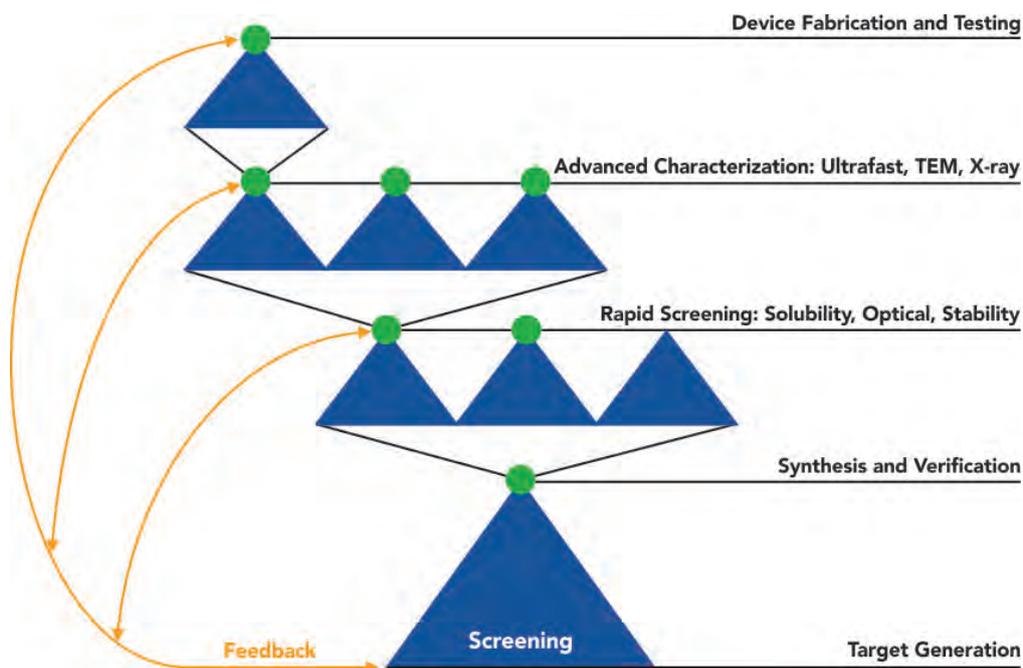
Figure 7.1 Schematic representation of a computational inverse materials design process. After devising design principles and generating a set of candidates that satisfy them, theory and experiment are employed in a high-throughput manner to rapidly screen the candidates down to a short list. Subsequent advanced characterization is conducted and all aspects of the approach are integrated to aid in further rational inverse design. [Source: Based on ref 1]

to appropriate building blocks for synthesis and characterization. These efforts are all disparate, resulting in little to no crossover, which is critical to take a single concept and describe its functionality, synthesizability, and characterizability — the cornerstone of inverse design. Much of this is due to a lack of appropriate paradigms for many of the processes currently utilized in synthesis and characterization. Simple reaction models are often used to understand dynamic processes that involve many variables. This oversimplification in order to create tractable models for human brains has resulted in a dearth of complex data necessary for machines to develop the same character of understanding.

Even if all the data and models actually existed, there is very little understanding on how researchers should mix structured and unstructured data. Methods of incorporating the unknown are necessary to have a "built-in" ability to discover transformational materials that do not exist in the current models of understanding, e.g., the discovery of quasi-crystals. There is an opportunity to expand the inverse design process to take materials discovery into new regions of understanding, opening up all possible functionality, rather than only what we currently grasp.

Existing computational methods rely on a limited palette of building blocks (e.g., a few different types of atoms) that can only be placed into a limited grid of spaces. One method of sidestepping this limitation is to develop methods to intelligently select these building blocks at different levels of complexity to suit the targeted goal. This idea also extends into synthesis and characterization. Often in synthesis, notions of stability are used to filter down candidate materials, while many

breakthrough materials are non-equilibrium in a global sense. Key to surmounting this challenge is incorporating fingerprints that represent synthesis to serve as proxies for complex dynamical considerations that could alleviate complexity and make the inverse design process tractable. These fingerprints would be synthesis, chemistry, structural, and domain specific to ensure that a variety of dynamical considerations, from kinetics of atomic motion to macroscopic degradation such as corrosion and fatigue, could be adequately incorporated into the search.

7.2 BENEFITS AND IMPACT

Inverse design assists in properly defining the problem space for many areas of clean energy research. Currently, targeted functionality focuses on individual properties or processes believed to be problematic or promising for overall functionality. Focusing on final goals allows complex problems, such as finding optimal non-platinum group metal catalysts [2], to be fully modelled and optimized before initial testing. Once automated, the machinery begins testing possible materials, compositions, and structures. The resulting iterative process is Pareto-optimal given the amount of existing data already present and the ability to glean as much information as possible from characterization. This same methodology can accelerate discovery in a variety of functionalities key to energy production and use, ranging from thermoelectrics for better waste heat recovery to novel solar cells for improved light collection efficiencies and better dielectrics for more efficient power conversion [3,4,5].

More generally, inverse design processes can improve the efficiency of existing infrastructure for synthesis and characterization. Fingerprints are not limited to fundamental materials building blocks, but rather any combination of variables that define the search space for an outcome. Thus, an inverse design process can be used to direct the most appropriate characterization conditions instead of searching for optimal parameters via brute force. Similarly, in synthesis, the best parameters would be directed rather than randomly searching for them, reducing the waste and ensuring that the global optimum is achieved rather than local optima, as is often the case when a human is directly optimizing a process. This would enable crucial scaling necessary for a materials breakthrough to translate from a lab-scale curiosity to an industrially made and applied breakthrough.

7.3 RELATIONSHIP TO THE OTHER GOALS OF THE MATERIALS ACCELERATION PLATFORM

Inverse design is key to developing not only novel materials, but also to materials understanding in general. While AI4M (Goal 2) would develop the techniques necessary to translate data into models and accelerate sub-processes within discovery, inverse design would build the "why" of connecting structure, environment, processing parameters, and any fundamental building block for targeted functionality and outcomes. This would be done by ingesting the trove of data generated by modular materials robotic (Goal 3) to synthesize and characterize materials as directed by inverse design. Understanding length and time scales (Goal 5) is key to bounding

building blocks to the appropriate scope and ensuring consideration of all aspects of materials parameters, from atomic to engineering scale.

7.4 MATERIALS-SPECIFIC CHALLENGES AND APPROACHES

7.4.1 Inorganic materials

The relatively underdeveloped description of inorganic materials structures represents one of the foundational challenges for inverse design in this space. In contrast to organic materials, in which the fundamental building units (molecules) have well-defined placements of the atoms, many functional inorganic materials are intrinsically disordered or partially inhomogeneous alloys at a variety of length scales. Thus, inorganic-specific inverse design methods must be capable of evaluating and optimizing acceptable distributions of possible microscopic atomic configurations and structures. Some efforts to tackle this challenge focus on understanding the effects of disorder or deconvoluting the disorder into its own set of building blocks such as describing disorder as a macroscopic solid solution of defects that can be uniquely defined. This relies on conventional intuition as to the limits of these models. Accessing these text-based models is currently outside the realm of automated design models. However, it could be a valuable source of structure and data for future inorganic inverse design efforts and alleviate the issue that these data are not machine encoded as they often are interspersed in literature as text.

Basic inverse design methodology has already been applied to several challenges in inorganic semiconductor materials [5]. For example, silicon and germanium are the most widely used materials for producing solar cells, but they are inherently inefficient at absorbing light due to their indirect band gaps. In principle, specially designed alternating layers of silicon and germanium (Si/Ge heterostructures) can have a direct band gap that more efficiently absorbs light, thus reducing the amount of material — and hence the cost — of solar cells. Here, the modular building blocks are Si and Ge atoms, which can be arranged in an astronomical number of possible stoichiometries and configurations, e.g., solid solutions, mixtures, layers, varying layer thicknesses. Zunger and co-workers demonstrated that low-computational-cost proxy models could be used to assess the stability of the arrangements and light absorption properties, and genetic algorithms used to direct the search for spatial arrangements with this property [6]. A similar approach can be used to design layers in Si/Ge nanowires with optimal solar light-absorbing properties [7], and even Si/Ge materials with optimal performance for quantum computers [8].

Similarly, the performance of dye-sensitized solar cell (DSSC) devices is critically dependent on the electronic structure of electrode materials. Moot et al. recently used the current best known electrode, NiO, as a proxy for all the target functionalities of effective DSSC electrodes [9]. The authors used a similarity search to find the closest neighbour to the nickel oxide in a numeric vector space that encoded band structure diagram. The fabricated $PbTiO_3$ DSSC devices exhibited the best performance in an aqueous solution, showing remarkably high fill factors compared
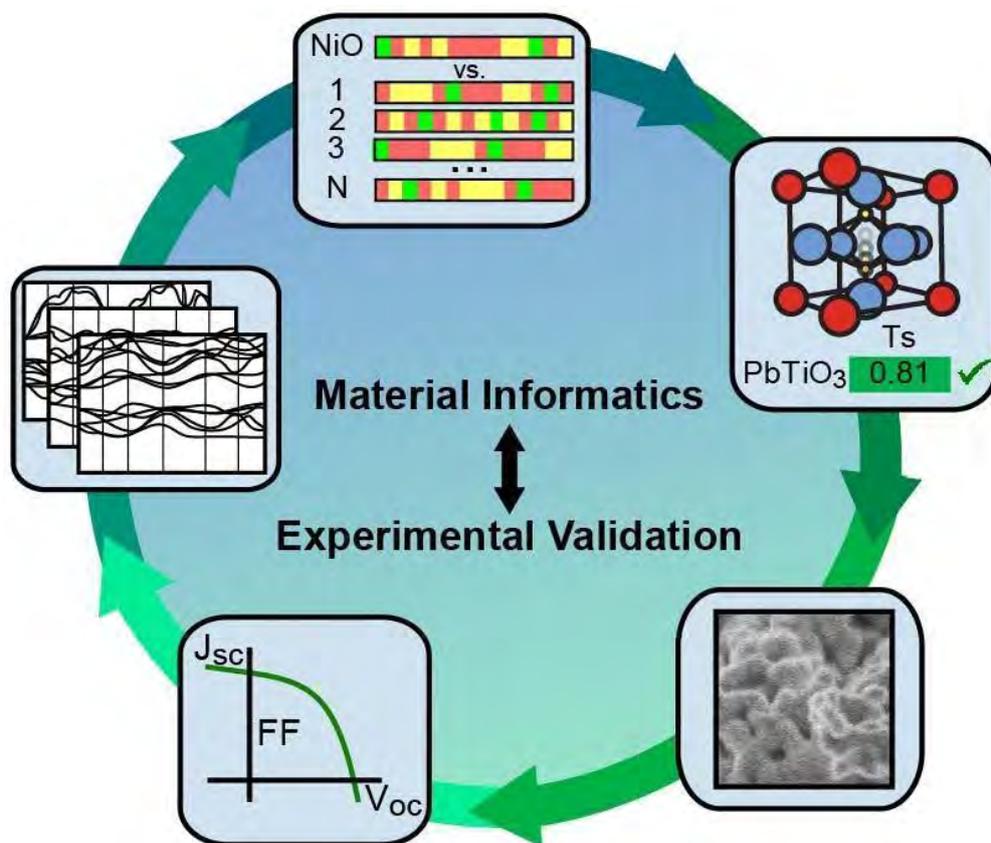
Figure 7.2 A schematic illustration of an informatics-driven loop enabling design of novel DSSC material. [Source: ref 9]

with typical photocathode systems. Such integration of informatics driving method and experiment allowed the creation of a full circle feedback loop (Figure 7.2).

In both these examples, human intuition still played a key role in inverse design. Fully automated inverse design would require minimal definition from humans on structure-property relationships and find the most appropriate connections based on the available data. Nevertheless, these examples showcase paradigms that can be templated with simple substitutions, reducing the human involvement in decision making, and expanding scope from simple structural and property models to fully integrated atomistic- to engineering-scale descriptions.

7.4.2 Organic materials

Inverse design strategies have also been used for organic materials challenges including designing organic LED materials [10] and photovoltaic materials (Figure 7.3) [11]. New machine learning techniques, such as variational autoencoder representations [12] and generative adversarial networks (Chapter 5) [13], allow for the direct generation of molecular structures from optimizations in an abstract property space, without the need to enumerate explicit atom positions.

A key challenge is the ability to handle more than the structure of single molecules and encompass the full spatial-scale, non-equilibrium arrangements of the molecules
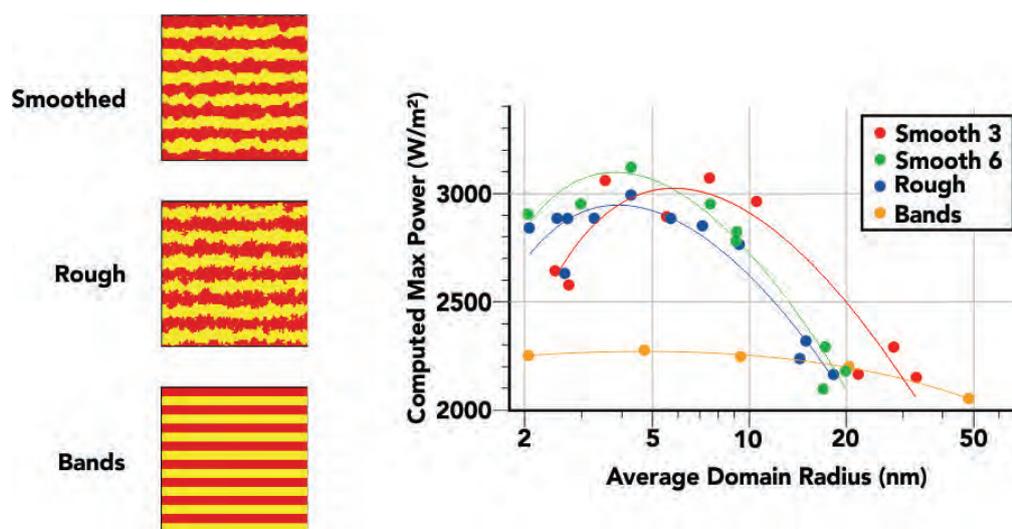
Figure 7.3 Example of inverse design for improved power output in organic photovoltaic materials by increased interfacial roughness. [Source: Dr. Geoffrey Hutchison, University of Pittsburgh]

that emerge from the intermolecular interactions between the molecules. At present, the fundamental theory of the self-assembly processes that give rise to non-equilibrium/metastable arrangements in organic materials and organic block co-polymers is not developed [14]. The need for further basic research efforts in this field is pressing. One promising area of impact is the design of morphologies for improved charge transport [15] or inverse design to optimize combined ion and charge transport in supercapacitors and polymer-based batteries.

7.4.3 Nanomaterials and composites

Inverse design of nanomaterials and composites for clean energy applications has yet to reach the same level of maturity as the inorganic and molecular fields described above, in part due to the large structural phase space. Examples of inverse design of nanomaterials can be found in related areas, e.g., nanophotonics, where general inverse design algorithms that directly incorporate fabrication constraints have been developed and applied successfully [16]. Generic macroscopic descriptors are also scarce because simple concepts such as particle size, which are useful to describe some nanomaterials, fail to encompass the true complexity of structures and to fully describe their functionality, e.g., surface termination and chemical ordering, faceting, compositional heterogeneity, etc.

 Nanoalloy catalysis provides an illustrative example of predicting structure and composition of nanomaterials for clean energy applications. Atomic-scale simulations, machine learning, and genetic algorithms are now being used to accelerate the prediction of structure and chemical ordering of nanoparticles. However, this is typically constrained to analyzing only a fixed size and stoichiometry. Figure 7.4 shows the prediction of global optimum configuration for 2 nm CuNi nanoalloys, i.e., an icosahedral NiCu core-shell structure [17]. Generalizing this process towards true inverse design is currently underway, e.g., identification of optimal stoichiometries to generate surface compositions predicted to yield high catalytic
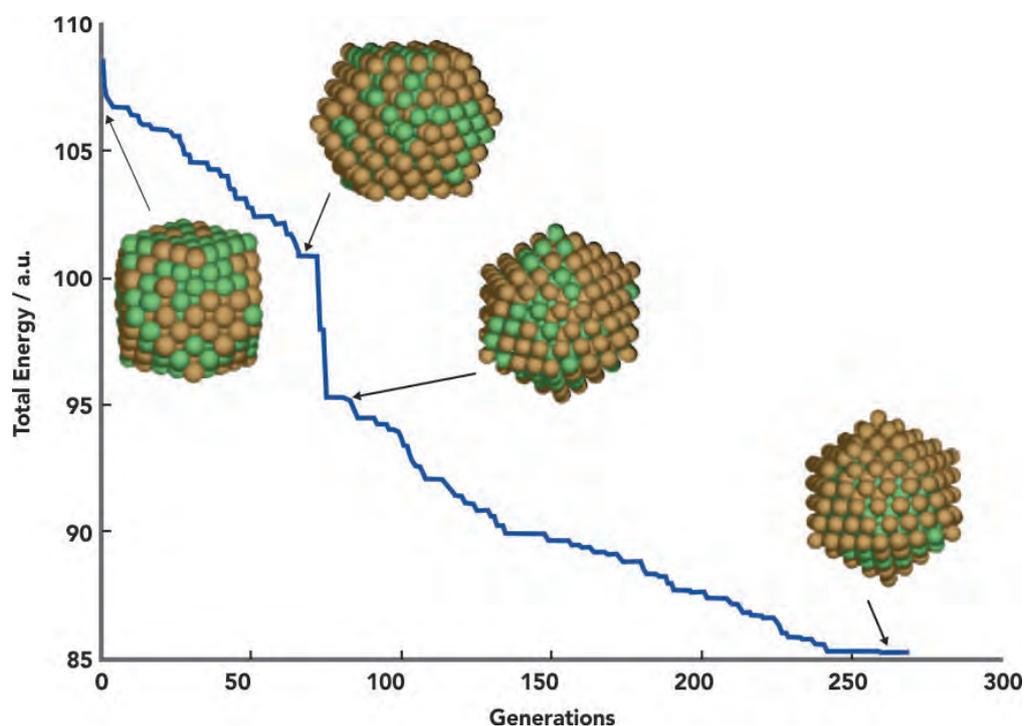
Figure 7.4 Generic algorithms for prediction of the global optimum icosahedral Ni@Cu core-shell configuration for 2 nm CuNi nanoalloys. [Source: ref 17]

activity and selectivity, such as nanoalloys for electrochemical reduction of CO_2 [18] or larger Au nanoparticles on carbon supports [19].

The next step towards inverse nanocatalyst design is identifying the specific particle size, composition, and structure that yield the highest total particle activity. This is already within reach through use of materials descriptors and scaling relations as a fingerprint of the catalytic activity. Using a Gaussian process surrogate model trained on adsorption energies and group additivity fingerprints, Nørskov et al. recently demonstrated that it is possible to predict the most important reaction step on a simple metal surface to be calculated explicitly without computationally demanding electronic structure theory [20]. An inverse design scenario for nanoparticle catalysts would also need to include the ability to assess, differentiate, and generate multiple possible reaction pathways on a wide distribution of ill-defined, under-coordinated, and strained sites, which remains beyond existing techniques.

References for Chapter 7

1. Center for Inverse Design (website), accessed 2017, *http://www.centerforinversedesign.org/approach.html*

2. Sang Moon Kim, Chi-Yeong Ahn, Yong-Hun Cho, Sungjun Kim, Wonchan Hwang, Segeun Jang, Sungsoo Shin, Gunhee Lee, Yung-Eun Sung and Mansoo Choi, "High-performance Fuel Cell with Stretched Catalyst-Coated Membrane: One-step Formation of Cracked Electrode," *Scientific Reports* 6, Number 26503 (2016), DOI: 10.1038/srep26503.

3. Millicent B. Smith and Josef Michl, "Singlet Fission," Chemical Reviews 110, Issue 11 (2010): 6891–6936, DOI: 10.1021/cr1002613.

4. Chun-Te Wu, Wen-Pin Liao and Jih-Jen Wu, "Three-dimensional ZnO nanodendrite/nanoparticle composite solar cells," *Journal of Materials Chemistry* 21, Issue 9 (2011): 2871–2876, DOI: 10.1039/c0jm03481a.

5. I. E. Castelli and K. W. Jacobsen, "Designing rules and probabilistic weighting for fast materials discovery in the Perovskite structure," *Modelling and Simulation in Materials Science and Engineering* 22, Number 055007 (2014), DOI: 10.1088/0965-0393/22/5/055007.

6. Mayeul d'Avezac, Jun-Wei Luo, Thomas Chanier and Alex Zunger, "Genetic-Algorithm Discovery of a Direct-Gap and Optically Allowed Superstructure from Indirect-Gap Si and Ge Semiconductors," *Physical Review Letters* 108, Number 027401 (2012), DOI: 10.1103/PhysRevLett.108.027401.

7. Lijun Zhang, Mayeul d'Avezac, Jun-Wei Luo and Alex Zunger, "Genomic Design of Strong Direct-Gap Optical Transition in Si/Ge Core/Multishell Nanowires," *Nano Letters* 12, Issue 2 (2012): 984–991, DOI: 10.1021/nl2040892.

8. Lijun Zhang, Jun-Wei Luo, Andre Saraiva, Belita Koiller and Alex Zunger, "Genetic design of enhanced valley splitting towards a spin qubit in silicon," *Nature Communications* 4, Number 2396 (2013), DOI: 10.1038/ncomms3396.

9. Taylor Moot, Olexandr Isayev, Robert W. Call, Shannon M. McCullough, Morgan Zemaitis, Rene Lopez, James F. Cahoon and Alexander Tropsha, "Material informatics driven design and experimental validation of lead titanate as an aqueous solar photocathode," Materials Discovery 6, Issue C (2016), DOI: 10.1016/j.md.2017.04.001.

10. Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, David Duvenaud, Dougal Maclaurin, Martin A. Blood-Forsythe, Hyun Sik Chae, Markus Einzinger, Dong-Gwang Ha, Tony Wu, Georgios Markopoulos, Soonok Jeon, Hosuk Kang, Hiroshi Miyazaki, Masaki Numata, Sunghan Kim, Wenliang Huang, Seong Ik Hong, Marc Baldo, Ryan P. Adams and Alán Aspuru-Guzik, "Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach," *Nature Materials* 15 (2016): 1120–1127, DOI:10.1038/nmat4717.

11. Johannes Hachmann, Roberto Olivares-Amaya, Adrian Jinich, Anthony L. Appleton, Martin A. Blood-Forsythe, László R. Seress, Carolina Román-Salgado, Kai Trepte, Sule Atahan-Evrenk, Süleyman Er, Supriya Shrestha, Rajib Mondal, Anatoliy Sokolov, Zhenan Baod and Alán Aspuru-Guzik, "Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry – the Harvard Clean Energy Project," Energy & Environmental Science 7, Issue 2 (2014): 698, DOI: 10.1039/C3EE42756K.

12. Rafael Gómez-Bombarelli, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams and Alán Aspuru-Guzik, "Automatic chemical design using a data-driven continuous representation of molecules," *Cornell University Library*, January 6, 2017, arXiv:1610.02415v2 [cs.LG].

13. Benjamin Sanchez-Lengeling, Carlos Outeiral, Gabriel L. Guimaraes and Alan Aspuru-Guzik, "Optimizing distributions over molecular space. An Objective-Reinforced Generative Adversarial Network for Inverse-design Chemistry (ORGANIC)," *ChemRxiv*, chemrxiv.org/articles/ORGANIC_1_pdf/5309668.

14. Marcus Müller and De-Wen Sun, "Process-directed self-assembly of block copolymers: a computer simulation study," *Journal of Physics: Condensed Matter* 27, Number 19 (2015), DOI: 10.1088/0953-8984/27/19/194101.

15. Ryan S. Gebhardt, Pengfei Du, Olga Wodo and Baskar Ganapathysubramanian, "A data-driven identification of morphological features influencing the fill factor and efficiency of organic photovoltaic devices," *Computational Materials Science* 129 (2017): 220–225, DOI: 10.1016/j.commatsci.2016.12.020.

16. Alexander Y. Piggott, Jan Petykiewicz, Logan Su and Jelena Vučković, "Fabrication-constrained nanophotonic inverse design," *Scientific Reports* 7, Number 1786 (2017), DOI: 10.1038/s41598-017-01939-2.

17. Steen Lysgaard, David D. Landis, Thomas Bligaard and Tejs Vegge, "Genetic Algorithm Procreation Operators for Alloy Nanoparticle Catalysts," *Topics in Catalysis* 57 (2014): 33–39, DOI: 10.1007/s11244-013-0160-9.

18. Steen Lysgaard, Jón S. G. Mýrdal, Heine A. Hansen and Tejs Vegge, "A DFT-based genetic algorithm search for AuCu nanoalloy electrocatalysts for CO_2 reduction," *Physical Chemistry Chemical Physics* 17 (2015): 28270–28276, DOI: 10.1039/c5cp00298b.

19. Tao Cheng, Yufeng Huang, Hai Xiao and William A. Goddard III, "Predicted Structures of the Active Sites Responsible for the Improved Reduction of Carbon Dioxide by Gold Nanoparticles," *The Journal of Physical Chemistry Letters* 8, Issue 14 (2017): 3317–3320, DOI: 10.1021/acs.jpclett.7b01335.

20. Zachary W. Ulissi, Andrew J. Medford, Thomas Bligaard and Jens K. Nørskov, "To address surface reaction network complexity using scaling relations machine learning and DFT calculations," *Nature Communications* 8, Number 14621 (2017), DOI: 10.1038/ncomms14621.

Goal #5: Bridging Length and Time Scales

Materials systems frequently demand understanding and control of properties that span wide ranges of length and time scales. For example, mechanical properties of materials depend on atom-atom adhesion at the nanoscale, crack formation and grain-boundary interactions at the mesoscale, and stress/strain relations at the macroscopic scale. In addition to determining bulk material properties, a range of different length scales dominate the function of complete devices (Figures 8.1-8.3), and complex interfacial interactions between combinations of materials are crucial to the design and function of most devices, such as solar cells, batteries, organic light-emitting devices, and others. Such length scales can range over 10 orders of magnitude, from sub-nanometre to multiple metres or larger. Similarly, event time scales for materials can range from femtoseconds, e.g., in the case of photoexcitation in solar cells, to years and decades in the case of long-term photodegradation and fatigue. Since important phenomena at vastly different scales are frequently interrelated, methods that can bridge all relevant scales consistently are needed to achieve a comprehensive and predictive understanding of material properties.
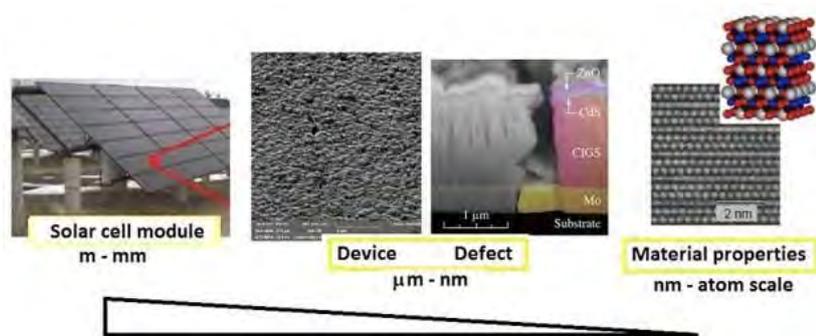
Figure 8.1 Range of length scales that must be considered to understand the performance characteristics of a device (left), down to its material components (right). [Source: ref 1]
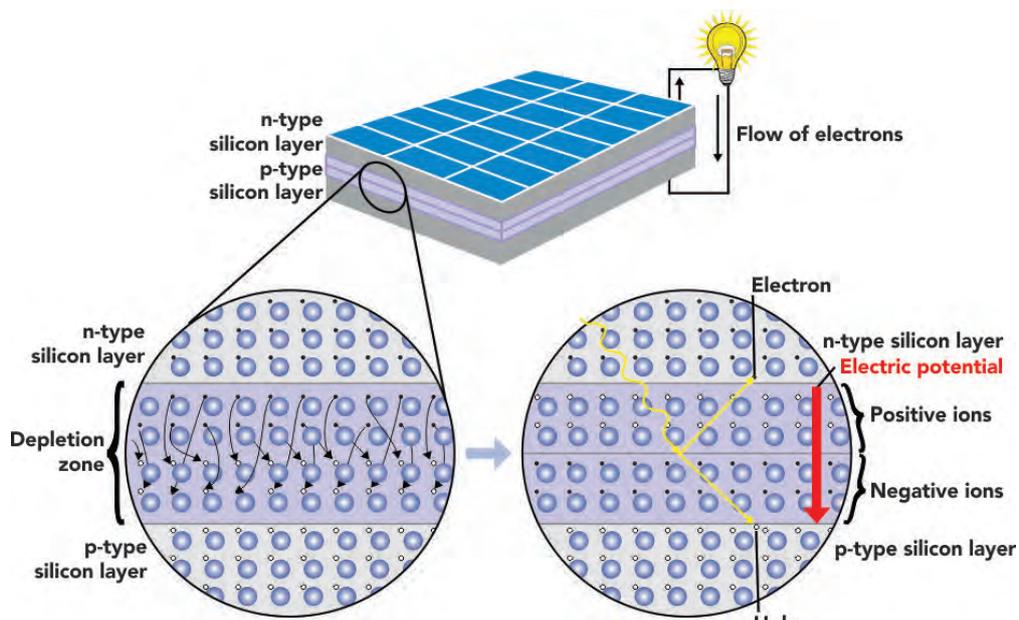
Figure 8.2 The importance of interfacial interactions between materials for a device. [Source: ref 2]

A powerful toolbox of experimental and computational techniques currently exists with each providing a view of a small range of length scales over a small range of time scales, for perhaps one set of properties of a single class of materials. These techniques, however, are generally not integrated into a single platform. Instead, the current approach is to manually patch together different methods and data from different scales using human intuition.

Computationally, techniques for specific length and time scales include atomic-level quantum mechanical calculations, nanoscale classical molecular dynamics simulations, and mesoscale and macroscopic continuum field methods (Figure 8.4). Such techniques are usually implemented in separate software programs and written in different programming languages, which are not easily integrated. Parameterization and validation of theoretical models and methods are often narrowly scoped and unautomated. Model development to bridge scales is usually
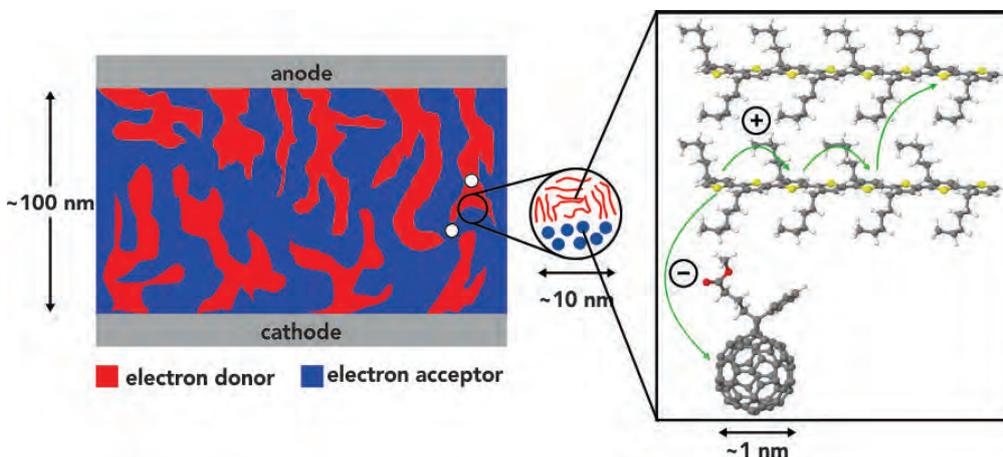
Figure 8.3 The performance of a flexible organic solar cell depends on electronic processes that occur on a wide range of length scales, down to the atomic level.

undertaken on a case-by-case basis, such as in the development of multi-scale coarse-grained (MS-CG) molecular dynamics models from atomistic models [3]. Hybrid-resolution methods, such as quantum mechanics/molecular mechanics (QM/MM) [4], and particle/continuum [5] approaches that can bridge some scales have been developed, but current implementations do not allow for general application or automation. Different computational methods are adapted for different scales, but no general method exists for partitioning space or time between different scales to be treated by different techniques. Some progress, however, has been made, such as in the mapping of atomistic to coarse-grained molecular models [6]. General-purpose software for some methods/scales does not exist or is still in the early stages of development, as in the case of field-theoretic or dynamical (classical) density functional theory simulations of mesoscale to macroscopic properties [7]. A general, integrated theoretical framework and computational platform for treating all relevant length and time scales to simulate and predict material properties does not exist, nor do data or software standards for interfacing software and experiments designed to treat different scales.

To date, most theory and simulation have been devoted to calculating equilibrium properties, whereas materials and devices in energy applications often operate out-of-equilibrium. Basic theories of non-equilibrium processes are much less
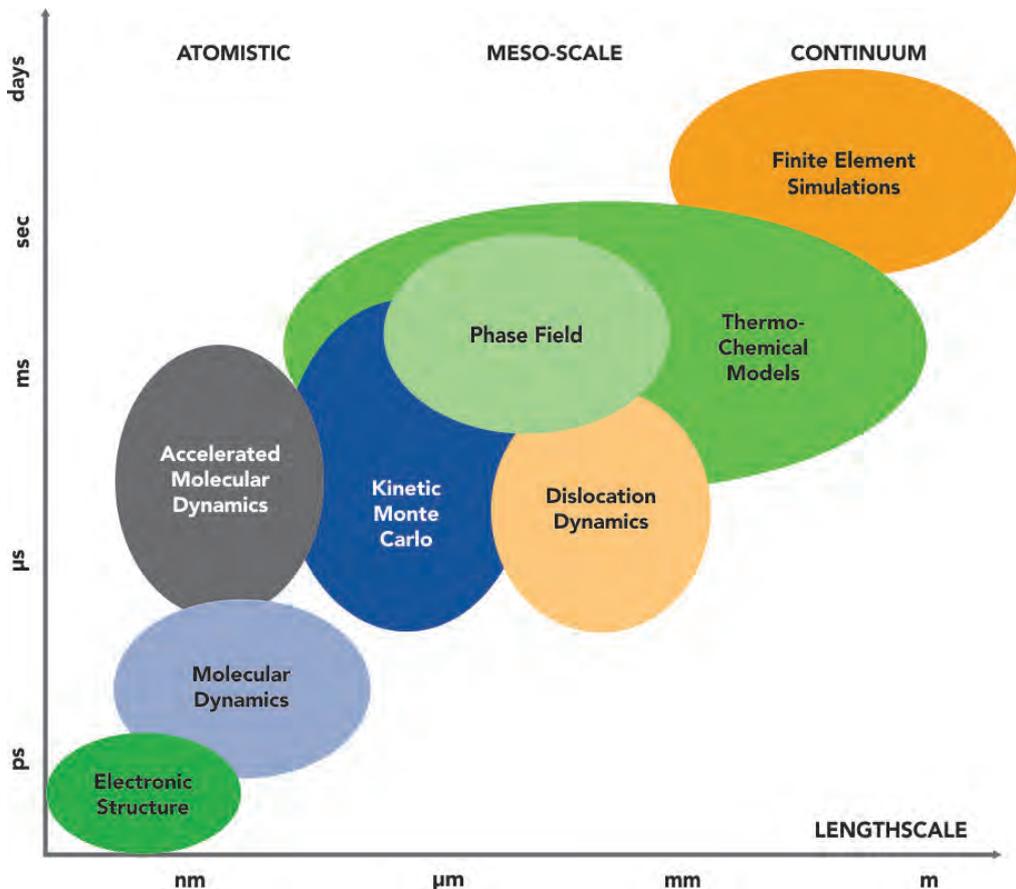
Figure 8.4 Breadth of different time and length scales encompassing various levels of theory and interactions. Multi-scale theoretical and computational methods used for materials model development and computer simulations. [Source: ref 14]

developed, and fewer computational algorithms for addressing non-equilibrium, time-dependent, or excited-state behaviour are available compared with equilibrium properties, particularly for phenomena that span many scales [8].

As in theory, the few examples of multi-scale characterization in recent literature focus on single systems and single techniques that may have sufficient resolution to bridge one set of length scales [9,10,11]. Similarly, tools for chemical and structural mapping of surfaces and interfaces are beginning to emerge, but still lack standards for use and interpretation of results [12,13]. Finally, very few examples of general experimental approaches to characterize non-equilibrium processes and metastability in materials exist in the current literature, which is in part likely due to gaps in theoretical approaches to the same problem.

8.1 OPPORTUNITIES

The development of new, automated experimental and computational bridging tools that give a holistic view of a material is critical to accelerating the design of better functional materials and devices. Improved data standards and software interfaces between simulation codes and between experimental instruments would enable the integration of tools adapted to different scales. A combination of fundamental theoretical basic research, software development, and advances in machine learning could allow time and length scales to be systematically and reliably partitioned between models of different resolution, making possible more efficient, accurate, and automated multi-scale or hybrid-resolution simulations of materials. AI tools could also help to bridge both spatial and time scales, e.g., in the sampling of rare events [15].

Simulation methods could be made more computationally tractable by using AI techniques, such as machine-learned surrogate models for computationally expensive density functional calculations, which can be trained on accurate atomistic dynamics and then applied to longer length and time scales relevant to mesoscale or bulk properties [16]. Integrated experimental and computational tools would enable a feedback loop between experimental characterization/analysis and theoretical refinement and prediction across length and time scales, facilitated by AI tools for decision making. New AI approaches to feedback, decision making, and prediction could enable autonomous analysis and rapid identification of gaps in understanding and necessary further feedback. Computational methods combined with AI techniques could help determine the proper scale, morphology, or formulation for a device once an optimal ("Goldilocks") material is known.

Many important functionalities depend on many-body interactions, which govern the lifetime of excitations in materials, and also energy transfer between the electronic and vibrational excitations. For example, since electron-phonon coupling affects both the electronic conductivity and the phonon spectrum of materials, it is a key underlying property of thermoelectric materials [17,18]. The fundamental electronic and vibrational excitations interact because they have similar energy scales, but

the wavelengths can be quite different since they are typically long for acoustic phonons and short for electrons in metals. Thus, the different excited states interact differently with their surroundings, suggesting new schemes to tailor the electron-phonon coupling by hierarchically structuring the materials on the nanoscale.

Synchrotron x-ray probes provide an ideal way to characterize these effects because the x-ray wavelengths available at synchrotron sources cover phenomena occurring from the atomic to the mesoscopic scale. Available techniques include scattering, which is sensitive to short- and long-range structures, and photoelectron spectroscopy, which is sensitive to electronic structure, chemical composition, and electronic lifetimes due to many-body interactions. Furthermore, x-rays can be tuned to atomic energy levels, making these techniques simultaneously sensitive to chemical state. Finally, synchrotron beams can now be focused to nanoscale dimensions for probing individual structures.

In the next 10 years, synchrotrons are expected to evolve towards fully coherent x-ray sources, opening up new approaches. When coherent x-rays scatter from samples, the inhomogeneity is encoded in the scattered wavefronts, and this information can be analyzed with spatial, chemical, magnetic, and temporal resolution.

From the temporal point of view, chemical processes bridge spatial scales from the atomic/ultra-fast (bond breaking and formation) to the mesoscale (diffusion) to the higher-level structures on the microscale. Coherent soft x-ray wavelengths cover the entire range of length scales, provide chemical sensitivity, and scale from seconds to nanoseconds. These tools will revolutionize the information that can be obtained from heterogeneous, multi-step chemical processes.

It may be possible to develop a system that uses synchrotron x-rays as the critical characterization tool in an optimization loop that explores the morphology of materials. Such a system could be used to optimize new materials for various applications. However, to maximally leverage available knowledge of chemical characterization, advanced characterization techniques should be more closely integrated. Schemes for sharing, storing, and relating data collecting for characterization at the many length and time scales are critical for the success of such a program. In addition, surface and interfacially sensitive characterization techniques are not well developed enough to integrate cleanly in such a system and fundamental research into their development is needed. Finally, the framework for integrated characterization must be adaptable to understanding the impact of defects and impurities, since these are critical to the performance of many energy materials, e.g., catalysts and photo-absorbers.

8.2 BENEFITS AND IMPACT

Improving multi-scale modelling and characterization across both length and time scales would dramatically benefit materials design for a wide range of clean energy applications, including energy storage (from electron-transfer events to materials

fatigue and decay) and solar energy conversion (from individual photo-excitations to photodegradation). Such tools would accelerate discovery and improve predictions for inverse design. Rather than waiting for years to determine if a new material will be stable, predicting this long-term behaviour during short experiments would be possible. The deeper and broader fundamental understanding of materials (including heterogeneous, disordered, and non-equilibrium phases, interfaces, and defects), and of how properties on different length and time scales relate would enable new, emergent structure-processing-property relationships to be predicted, would lead to new physics and chemistry (e.g., multi-functional or active materials) that could be exploited for energy applications. New in situ approaches for characterization of materials and devices across scales would enable more appropriate characterization of critical properties relevant to performance under operation conditions. The development of an integrated multi-scale platform for modelling and characterization would also create a more multidisciplinary community and foster relationships and collaboration schemes, enabling and accelerating complementary work.

8.3 RELATIONSHIP TO THE OTHER GOALS OF THE MATERIALS ACCELERATION PLATFORM

Achieving this goal requires advances in several other goals. Modular materials robotics (Goal 3) must be able to perform experiments that provide information across many length and time scales as raw input for validating the new theoretical methods required. Interfacing experimental and computational tools seamlessly across scales requires new standards and protocols for data infrastructure (Goal 6) to store and exchange information. Developing AI for materials (Goal 2) supports the advancement of the necessary conceptual developments.

Achieving this goal would provide an essential building block for many of the other goals. More efficient experimental and computational tools for bridging scales would accelerate materials characterization and prediction, thereby speeding up the design cycles needed for autonomous discovery and development (Goal 1). AI for materials (Goal 2) needs to make inferences and provide explanations that span length and time scales. A better fundamental understanding of how interactions on small length and time scales influence large-scale behaviour would enable more accurate predictions of functional material properties, such as device performance, stability, and durability needed for inverse design (Goal 4). Finally, advances in knowledge of relationships across scales could lead to improvements in how data infrastructure (Goal 6) is organized.

8.4 MATERIALS-SPECIFIC CHALLENGES AND APPROACHES

8.4.1 Inorganic materials

The weakest link in our ability to relate nanoscale properties to macroscale device performance is in the understanding of how defects and interfacial phenomena relate to the final outcome. Understanding specifically where, how, and when defects in materials form is a second question directly related to multi-scale

phenomena, with applications ranging from the design of more durable aircraft alloys to semiconductor chips in modern computers. For example, the length of solid-electrolyte interface is critically related to both the lifetime and performance of lithium-ion batteries. Understanding the formation of this layer at multiple length scales and understanding the time scales of its nucleation and longer-term consequences is critical to designing more efficient energy storage materials [19,20]. In another example, perovskites $Ba_{1-x}Nd_xCuO_2$ and $CaCuO_2$ are both insulating, but they can only be combined to form a high temperature (high Tc) superconductor in super lattice form [21].

Understanding how to interface materials to optimize device performance, or whether it is necessary to do so, is also important. For example, significant effort has been invested in discovering new photoelectrocatalysts, but many of the most successful materials for solar fuel conversion are multi-component systems that combine the functionalities of their components [22,23,24]. Databases and AI could aid the design and optimization of such multi-component materials and interfaces. For example, a systematic understanding of how electrons, holes, or substrates migrate in and out of materials is necessary to a more effective design in photovoltaics and photocatalysts. The time scales relevant to these applications range from femtoseconds (e.g., light absorption) to seconds (e.g., substrate flux).

8.4.2 Organic materials

Understanding molecular shape and packing and how these scale to nanoscale and microscale materials partitioning and morphology, and the resulting effects on electronic structure and transport properties, presents an important challenge [25]. It is unclear how to translate microscopic intermolecular interactions to predict and design bulk properties. What is the overlap of organic structural degeneracy with macroscopic observable properties? This relates to the fundamental question raised in Section 6.4.2: what is the minimal set of organic chemical space that can be exploited to understand and predict all possible physical properties? To address this question, integrated analytical tools are needed to characterize the organic material both during synthesis and for under load, i.e., when being utilized. This helps address the synthesis of a novel material with a targeted property, and to ensure that life-cycle and material robustness are tested early in the discovery phase. Integration and correlative analysis of multiple analytical data sets present a multifaceted challenge that may be intractable using conventional analysis. AI-enabled machine learning provides a method of delineation and identification of data trends [26].

8.4.3 Nanomaterials and composites

Nanomaterials and composites are inherently heterogeneous both in composition and organization. Functional properties are typically sensitive to interactions between molecular components, as well as to material interfaces, defects, and disorder on a range of length scales. For example, organic molecules affixed to quantum dots suspended in a liquid crystal environment modify the interfacial properties between the solid and the liquid medium, modulating dispersion and driving self-assembly (Figure 8.5) [27]. Ultimately, this governs how individual components of the liquid crystal organize themselves and, therefore, controls function. The heterogeneous nature of nanomaterials (ligand/particle) or composites requires theoretical understanding and characterization data covering both molecular components and nanoscale domains and interfaces. Time-dependent phenomena (e.g., nucleation and aging) that span a huge range of scales are crucial to understanding and controlling the synthesis of nanomaterials and composites, as well as to ensuring long-term stability in applications. Understanding time-dependent behaviour is also important for designing novel active materials. Understanding of heterogeneity, disorder, and interfaces inherent to nanomaterials and composites would be enhanced by the development of new, automated tools to study buried, multi-phase interfaces (e.g., organic-inorganic, solid-liquid, solid-gas, triple-phase), i.e., translating 2D to 3D imaging and translating local to ensemble properties. Currently, no tools exist for chemically mapping surfaces. Heterogeneity and disorder also present a challenge for the reliable and reproducible synthesis of nanomaterials. Methods for characterizing and controlling heterogeneity and disorder on a range of scales are important for designing specific materials for specific applications.
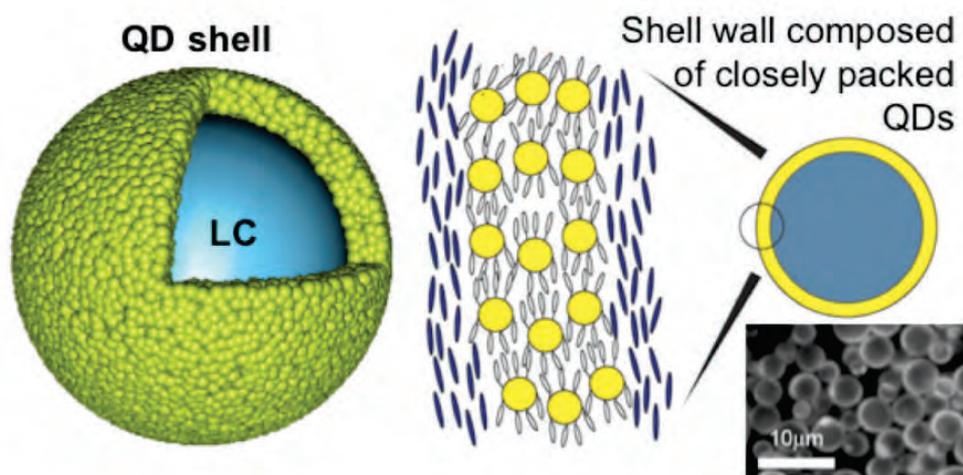
Figure 8.5 Nanoparticle quantum dot (QD) "shells" arranged around a liquid crystal (LC) core. The arrangement and functionality of the QDs in the shell are dictated by interactions between nanoparticle surfaces, which may be functionalized by ligand adsorption, and the LC phase. [Source: ref 28]

References for Chapter 8

1. Pavel Škarvada, Lubomír Škvarenina, Pavel Tománek, Dinara Sobola, Robert Mack☐, Jitka Brüstlová, Lubomír Grmela, Steve Smith, "Multiscale experimental characterization of solar cell defects," *Proceedings: 20th Slovak-Czech-Polish Optical Conference on Wave and Quantum Aspects of Contemporary Optics* Volume 10142 (2016), DOI: 10.1117/12.2259884.

2. American Chemical Society (ACS), "How a Solar Cell Works," ChemMatters Online, accessed 2017, www.acs.org/content/acs/en/education/resources/highschool/chemmatters/past-issues/archive-2013-2014/how-a-solar-cell-works.html.

3. M. G. Saunders and G. A. Voth, "Coarse-graining methods for computational biology," *Annual Review of Biophysics* 42 (2013): 73–93, DOI: 10.1146/annurev-biophys-083012-130348.

4. A. Warshel and M. Levitt, "Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme," *Journal of Molecular Biology* 103 (1976): 227–249, DOI: 10.1016/0022-2836(76)90311-9.

5. Ronald E. Miller and E. B. Tadmor, "A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods," *Modelling and Simulation in Materials Science and Engineering* 17, Number 053001 (2009), DOI: 10.1088/0965-0393/17/5/053001.

6. Anton V. Sinitskiy, Marissa G. Saunders and Gregory A. Voth, "Optimal Number of Coarse-Grained Sites in Different Components of Large Biomolecular Complexes," *The Journal of Physical Chemistry B* 116, Issue 29 (2012): 8363–8374, DOI: 10.1021/jp2108895.

7. Akash Arora, Jian Qin, David C. Morse, Kris T. Delaney, Glenn H. Fredrickson, Frank S. Bates and Kevin D. Dorfman, "Broadly Accessible Self-Consistent Field Theory for Block Polymer Materials Discovery," *Macromolecules* 49, Issue 13 (2016): 4675–4690, DOI: 10.1021/acs.macromol.6b00107.

8. Marcus Müller and Juan J. de Pablo, "Computational Approaches for the Dynamics of Structure Formation in Self-Assembling Polymeric Materials," *Annual Review of Materials Research* 43 (2013): 1-34, DOI: 10.1146/annurev-matsci-071312-121618.

9. H. Simons, A. King, W. Ludwig, C. Detlefs, W. Pantleon, S. Schmidt, F. Stöhr, I. Snigireva, A. Snigirev and H. F. Poulsen, "Dark-field X-ray microscopy for multiscale structural characterization," *Nature Communications* 6 (2015), DOI: 10.1038/ncomms7098.

10. A. Junay, S. Guézo, P. Turban, S. Tricot, A. Le Pottier, J. Avila, S. Ababou-Girard, P. Schieffer and F. Solal, "Effective Metal Top Contact on the Organic Layer via Buffer-Layer-Assisted Growth: A Multiscale Characterization of Au/Hexadecanethiol/n-GaAs(100) Junctions," *The Journal of Physical Chemistry C* 120, Issue 42 (2016): 24056–24062, DOI: 10.1021/acs.jpcc.6b05729.

11. Nicolas Besnard, Aurélien Etiemble, Thierry Douillard, Olivier Dubrunfaut, Pierre Tran-Van, Laurent Gautier, Sylvain Franger, Jean-Claude Badot, Eric Maire and Bernard Lestriez, "Multiscale Morphological and Electrical Characterization of Charge Transport Limitations to the Power Performance of Positive Electrode Blends for Lithium-Ion Batteries," *Advanced Energy Materials* 7 (2017), DOI: 10.1002/aenm.201602239.

12. D. R. Baer, D. J. Gaspar, P. Nachimuthu, S. D. Techane and D. G. Castner, "Application of surface chemical analysis tools for characterization of nanoparticles," *Analytical and Bioanalytical Chemistry* 396 (2010): 983–1002, DOI: 10.1007/s00216-009-3360-1.

13. Gorgi Kostovski, Udayakumar Chinnasamy, Sasani Jayawardhana, Paul R. Stoddart and Arnan Mitchell, "Nanoimprinting: Sub-15nm Optical Fiber Nanoimprint Lithography: A Parallel, Self-aligned and Portable Approach," *Advanced Materials* 4 (2011), DOI: 10.1002/adma.201002270.

14. Marius Stan, "Discovery and design of nuclear fuels," *Materials Today* 12, Issue 11 (2009): 20–28, DOI: 10.1016/S1369-7021(09)70295-0.

15. Tim Mueller, Aaron Gilad Kusne and Rampi Ramprasad, "Machine Learning in Materials Science: Recent Progress and Emerging Applications," Chapter 4 in *Reviews in Computational Chemistry* 29 (2016), Abby L. Parrill and Kenny B. Lipkowitz Eds. (John Wiley & Sons, Inc., 2017), Chapter DOI: 10.1002/9781119148739.ch4.

16. J. S. Smith, O. Isayev and A. E. Roitberg, "ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost," *Chemical Science* 8 (2017): 3192–3203, DOI: 10.1039/C6SC05720A.

17. X. Zianni, "Effect of electron-phonon coupling on the thermoelectric efficiency of single-quantum-dot devices," *Physical Review B* 82, 165302 (2010), DOI: 10.1103/PhysRevB.82.165302.

18. Heng Wang, Yanzhong Pei, Aaron D. LaLonde and G. Jeffrey Snyder, "Weak electron–phonon coupling contributing to high thermoelectric performance in n-type PbSe," *Proceedings of the National Academy of Sciences* 109, Issue 25 (2012): 9705–9709, DOI: 10.1073/pnas/1111419109.

19. A. A. Tahmasbia, T. Kadyka and M. H. Eikerling, "Statistical Physics-Based Model of Solid Electrolyte Interphase Growth in Lithium Ion Batteries," *Journal of the Electrochemical Society* 164, A1307 (2017), DOI: 10.1149/2.1581706jes.

20. J. Vetter, P. Novák, M. R. Wagner, C. Veit, K.-C. Möller, J. O. Besenhard, M. Winter, M. Wohlfahrt-Mehrens, C. Vogler and A. Hammouche, "Ageing mechanisms in lithium-ion batteries," *Journal of Power Sources* 147, Issue 1 (2005): 269–281, DOI: 10.1016/j.jpowsour.2005.01.006.

21. B. Freelon, A. Augustsson, J.-H. Guo, P. G. Medaglia, A. Tebano and G. Balestrino, "Electron Correlation and Charge Transfer in $[(Ba_{0.9}Nd_{0.1})CuO_{2+\square}]_2/[CaCuO_2]_2$ Superconducting Superlattices," *Physical Review Letters* 96, Number 017003 (2006), DOI: 10.1103/PhysRevLett.96.017003.

22. Marcel Schreier, Laura Curvat, Fabrizio Giordano, Ludmilla Steier, Antonio Abate, Shaik M. Zakeeruddin, Jingshan Luo, Matthew T. Mayer and Michael Grätzel, "Efficient photosynthesis of carbon monoxide from CO_2 using perovskite photovoltaics," *Nature Communications* 6, Number 7326 (2015), DOI: 10.1038/ncomms8326.

23. Joseph H. Montoya, Linsey C. Seitz, Pongkarn Chakthranont, Aleksandra Vojvodic, Thomas F. Jaramillo and Jens K. Nørskov, "Materials for solar fuels and chemicals," *Nature Materials* 16 (2017): 70–81, DOI: 10.1038/nmat4778.

24. M. Cerny, M. Sob, J. Pokluda and P. Sandera, "Ab initio calculations of ideal tensile strength and mechanical stability in copper," *Journal of Physics: Condensed Matter* 16, Issue 7 (2004): 1045–1052, DOI: 10.1088/0953-8984/16/7/004.

25. Geoffrey R. Hutchison, Mark A. Ratner and Tobin J. Marks, "Electronic Structure and Band Gaps in Cationic Heterocyclic Oligomers. Multidimensional Analysis of the Interplay of Heteroatoms, Substituents, Molecular Length, and Charge on Redox and Transparency Characteristics," *The Journal of Physical Chemistry B* 109, Issue 8 (2005): 3126–3138, DOI: 10.1021/jp046579v.

26. S. B. Ghugare, S. Tiwary, V. Elangovan and S. S. Tambe, "Prediction of Higher Heating Value of Solid Biomass Fuels Using Artificial Intelligence Formalisms," *BioEnergy Research* 7, Issue 2 (2014): 681–692, DOI: 10.1007/s12155-013-9393-5.

27. Andrea L. Rodarte, Blessing H. Cao, Harmanpreet Panesar, Ronald J. Pandolfi, Makiko Quint, Lauren Edwards, Sayantani Ghosh, Jason E. Hein and Linda S. Hirst, "Self-assembled nanoparticle micro-shells templated by liquid crystal sorting," *Soft Matter* 11, Issue 9 (2015): 1701–1707, DOI: 10.1039/C4SM02326A.

28. Ibid.

9 Goal #6: Data Infrastructure and Interchange

Autonomous materials discovery requires the involvement of many disparate communities ranging from materials science, chemistry, physics, and engineering to AI and robotics. Each community has a unique terminology and set of standards. Further, the rich and rapidly growing data on materials and their property, synthesis, and characterization remain largely uncurated, unintegrated, and unexplored due to the compartmentalization of this information, e.g., such data remain in the primary research lab. The full potential of the effective exploitation of both textual data and published chemical data often is untapped, mostly due to the lack of tools and focused methods to curate, integrate, and transform the data into new, experimentally testable hypotheses.

Integration of these communities, as well as of databases, hardware, and software, is critical to an efficient materials design ecosystem where tools developed in one discipline or geographical region are easily understood and integrated with another. Similar to the International Union of Pure and Applied Chemistry (IUPAC) in the chemistry community, or the World Wide Web Consortium (W3C) standards groups for HTML, a centralized stakeholder group is needed to set common languages, categories, and standards in which data are captured and retained, software and hardware are developed, and components communicate with each other. This group must be connected to both funding agencies and publishing bodies to ensure uniform application of standards, deposition of data, and maintenance of repositories. A 2017 report funded by the U.S. National Science Foundation (NSF) highlights these issues [1].

This final goal of MAP aims to capitalize on the timely opportunity to standardize and collect the community's increasing amount of digital materials data on synthesis and characterization of materials for clean energy generation and conversion. The vision is of an open source, online, international knowledge base that contains computational data, characterization, and synthetic procedures and outcomes. Data

on poor or failed experiments or samples would be deposited along with data on successful experiments.

9.1 OPPORTUNITIES

As the essential input to AI and machine learning, data directly influence the predictive capability envisioned for MAP. Integrating existing and future materials data, originating from theory, synthesis, and characterization, could therefore accelerate the materials discovery process. Given the depth, breadth, and heterogeneity of materials property and process data, standardized data formats, data exchange protocols, and workflows are needed. Building schema conversion and migration tools could also leverage these protocols to facilitate an open data infrastructure accessible both to users and machine learning algorithms. This section identifies some of the opportunities that could help close the gap in data integration and knowledge sharing.

A major challenge in this area is to develop protocols for communication between robotic hardware and software for automating synthesis and characterization. Synthesis and characterization workflows should be codified and stored. These protocols also need to preserve metadata and provenance of all datasets. A new control mechanism for data quality and integrity could include, for example, benchmarking calculations when using large calculated data sets as a main input to machine learning. This should preferably be done using high-quality experimental data. Enabling image data contributions to the platform, for example, from optical or electron microscopy, would create an opportunity to implement image processing machine techniques in materials discovery processes. An ambitious vision for research transparency, enabled by this data storage and referencing infrastructure, would be to include navigable links for individual data points contained within the figures in journal articles so that readers can see the particular synthesis, processing, characterization, and/or computational process leading to that assigned value. This would require a "sample-oriented" archiving scheme rather than the "compound-oriented" archiving schemes used by most existing databases. Emerging standards, such as the Allotrope initiative [2], are beginning to address these points and deserve continued support and extension.

Incentivizing data contribution and participation could help facilitate the data infrastructure described above. In particular, contributions from individual groups, analogous to crowdsourcing, would drive expansion and large-scale use of integrated materials datasets. Participation could be achieved by requiring compliance with the mandates of funding agencies or journals or by giving researchers easy-to-use data tools to help them store, analyze, visualize, and share data while, at the same time, providing credit attribution. A number of efforts are attempting to do this, e.g., MPContribs [3], Citrination [4,5], Materials Data Curation System (MDCS) [6,7,8,9], Materials Data Facility (MDF) [10,11], Materials Commons [12,13]. Some of these platforms currently cater more to using machine learning and AI (Citrination), while others focus on data management (MPContribus, MCDS, MDF, Materials Commons), particularly large data (MDF).

Searchability tools that provide end-to-end tracking of individual contributions to synthesis, characterization, and workflow development could also encourage contributions. Industrial researchers should be encouraged to form pre-competitive data sharing collaborations, similar to existing initiatives for pharmaceutical development. In short, to grow data infrastructure, incentives to participate should be offered for all types of research. Beyond capturing future contributions of research groups, a large corpus of experimental and characterization data, particularly in areas such as synthesis, is embedded in publications and awaits conversion to machine-readable, standardized knowledge, for example, via language processing as demonstrated recently [14].

9.2 BENEFITS AND IMPACT

Integration across systems, databases, and communities is necessary to achieve the accelerations discussed in the previous section. Improving access to existing results would significantly decrease duplicate experiments, saving time and resources. A shared, integrated data platform would transform the science of materials, enabling greater productivity, less repetition of failed experiments or calculations, rapid identification of new areas of "design space", and better informed research decisions. Building a large repository of materials and their properties would allow known materials to be used for new applications. Combining this with machine-readable descriptions of the experimental plans that successfully lead to these materials would suggest new directions for both human- and AI-directed experimentation. It is important that these databases also include unsuccessful reactions. Knowledge of "failures" can both limit redundant attempts and support efforts to redesign experiments and techniques to increase the likelihood of a successful outcome. Furthermore, their inclusion in data enables training of more accurate predictive AI models [15]. Thus, this type of database has a central role in fostering constructive collaboration and can be a data resource for the other five goals.

Simultaneously, this data infrastructure would increase reproducibility and traceability of new results. Well-documented, more universally standardized characterization methods would lead to increased confidence (or uncertainty quantification). Finding relevant information about techniques, applications, and tools would become less time-intensive. For example, improved data interchange would improve interoperability of instruments between vendors and techniques. Common standards would facilitate communication between distinct disciplines, including synthesis, characterization, and simulation. Finally, increasing the amount and quality of data available would advance machine learning and AI approaches.

9.3 RELATIONSHIP TO THE OTHER GOALS OF THE MATERIALS ACCELERATION PLATFORM

The creation of a robust data infrastructure containing common language, categories, and standards provides opportunities that bridge all the goals. From an inverse design (Goal 4) perspective, which starts with the final property goal to find a target

material structure, the first step is access to materials' characteristics, or property sets, with interchange of data formats across fields. A major component of bridging length and time scales (Goal 5) is uniting data across different computational tools, which would benefit from standardized data storage, availability, and exchange protocols. Similarly, autonomous discovery and development (Goal 1) require retaining and interconverting data and metadata at all stages of the research progress to match characterization and simulation, enabling continuous optimization. For synthetic production from modular materials robotics (Goal 3), retaining data and metadata related to particular samples, including starting materials, processing conditions, purification, quality, and time of production, is needed for integrated end-to-end autonomous research. Similarly, the development of AI for materials (Goal 2) requires reliable data, including both positive and negative results.

9.4 MATERIALS-SPECIFIC CHALLENGES AND APPROACHES

9.4.1 Inorganic materials

Current inorganic material databases have different focus areas, including computational ones, such as the Materials Project (Figure 9.1), OQMD, NOMAD, AiiDA, and AFLOW, and experimental ones, such as ICSD, NIST, NIMS, and others. At the scale of narrowly describing a material via its structure and composition, some databases offer importing and exporting. However, generalization of standardized formats to other properties (e.g., morphology, mechanical properties, electronic structure, synthesis) is essential for more effective integration and use of available inorganic material data. Functional properties of inorganic materials derive from their internal hierarchical structures. These structures rely on the development and understanding of bulk properties and interfaces. The interfaces can be treated as building blocks for higher order heterostructures. The prerequisite for a machine learning-guided search of the space of heterostructures is a centralized catalogue of recipes to create these interfaces, and knowledge of their properties, to enable higher-level optimization loops.
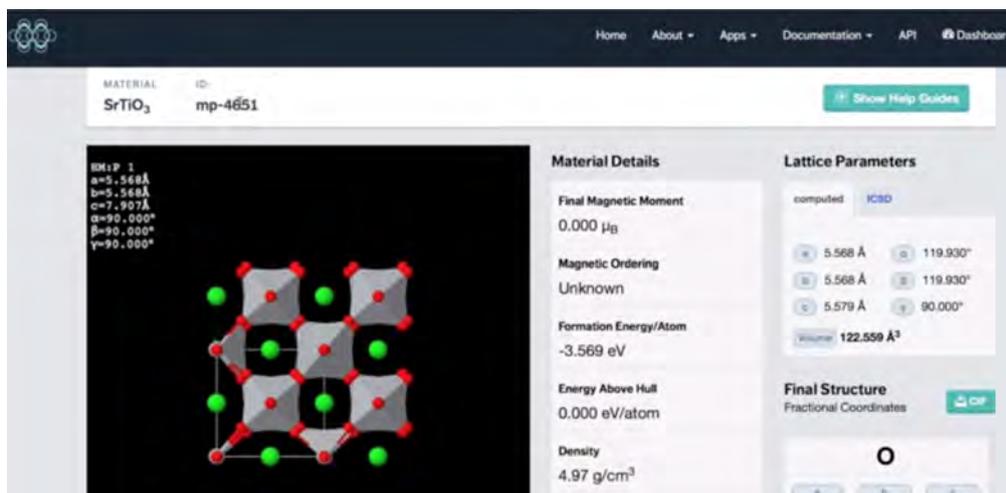
Figure 9.1 Example of Materials Project repository for inorganic materials, including visualization of structure and properties. [Source: https://materialsproject.org/]

9.4.2 Organic materials

Currently, many open repositories (see Figure 9.2 for two examples), often in the organic computational quantum chemical space, have little connection to synthesis or characterization data, or to pharmaceutical compounds. Some proprietary and commercial repositories (e.g., internal industrial data) do not allow interconversion or automated queries. Data types include reactions (Reaxys, SciFinder); molecular geometries (e.g., crystallography through CCSD, Crystallographic Open Database); spectroscopy (IR, NMR, Mass Spec); and density functional calculations. While standard molecular identifiers, such as the IUPAC International Chemical Identifier (InChI) strings, can connect molecular data, more connections are needed, even between open databases. A large quantity of proprietary data in organic molecules and polymers (e.g., solubilities, melting points, spectra) could be contributed in a pre-competitive manner if proper incentives were designed.

Figure 9.2 Examples of molecular repositories including (top) Pitt Quantum Repository [Source: https://pqr.pitt.edu/] and (bottom) Harvard Clean Energy Project (Molecular Space) [Source: http://www.molecularspace.org].

In contrast to other areas of materials science, for organic materials some protocols and methods have been established in the cheminformatics community for data quality, consistency, and interoperability [16,17,18]. Although there are some barriers to polymer applications, standardization tools could be adapted from other organic material applications [19].

9.4.3 Nanomaterials and composites

Nanomaterials and composites have unique challenges for representation and description of structures, properties, and processing. By definition, they incorporate a broad set of materials. Quantitative metrics for heterogeneity and order at multiple length scales are inherently difficult to define. The metrics of nanomaterials and composites have their own domain-specific problems, methods of capture, and current methodologies, which do not necessarily translate to compilation in current databases. However, it is crucial that researchers can draw on property descriptions and new insights to realize their ultimate performance. Currently, ad hoc approaches are used to compile and analyze the data and insights obtained from research studies to design promising new materials. Human researchers access separate publications, databases, and communities, and new materials targets are guided by intuition. Success in this field requires a framework for describing and quantifying order and lack of order. The nanoinformatics community has developed a roadmap towards this goal [20].

The incompatibility of data types with current repositories has led to limited integration of the results of research studies. The creation of a set of standards would allow the seamless drawing of insights across components and result in a massive acceleration in discovery and understanding of energy materials. Further, the integration of data would permit autonomous systems and AI programs to identify trends, understand behaviours, and predict materials targets, thus enabling complex multi-dimensional research decisions for human-robot research teams (Figure 9.3).
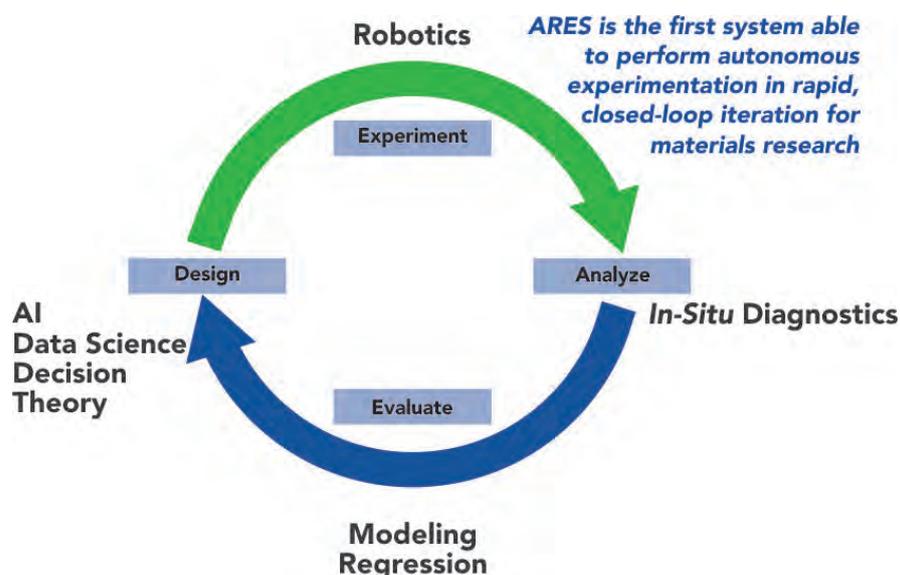

Figure 9.3 Example of integrated autonomous research, enabled by data interchange across synthesis/processing, characterization, and analysis.

References for Chapter 9

1. The Minerals, Metals & Materials Society (TMS), "Building a Materials Data Infrastructure: Opening New Pathways to Discovery and Innovation in Science and Engineering," www.tms.org/mdistudy.

2. Allotrope Foundation, "Solution," *Allotrope Foundation* (website), accessed 2017, www.allotrope. org/solution.

3. Patrick Huck, Dan Gunter, Shreyas Cholia, Donald Winston, Alpha N'Diaye and Kristin Persson, "User Applications Driven by the Community Contribution Framework MPContribs in the Materials Project," *Concurrency and Computation: Practice and Experience* 28, Issue 7 (2015): 1982–1993, DOI: 10.1002/cpe.3698.

4. Citrine, "Citrine's public platform for data-driven materials and chemistry research," accessed 2017, citrination.com/.

5. Jordan O'Mara, Bryce Meredig and Kyle Michel, "Materials Data Infrastructure: A Case Study of the Citrination Platform to Examine Data Import, Storage, and Access," *JOM* 68, Issue 8 (2016): 2031–2034, DOI: 10.1007/s11837-016-1984-0.

6. Alden Dima, Sunil Bhaskarla, Chandler Becker, Mary Brady, Carelyn Campbell, Philippe Dessauw, Robert Hanisch, Ursula Kattner, Kenneth Kroenlein, Marcus Newrock, Adele Peskin, Raymond Plante, Sheng-Yen Li, Pierre-François Rigodiat, Guillaume Sousa Amaral, Zachary Trautt, Xavier Schmitt, James Warren and Sharief Youssef, "Informatics Infrastructure for the Materials Genome Initiative," *JOM* 68, Issue 8 (2016): 2053–2064, DOI: 10.1007/s11837-016-2000-4.

7. NIST GMI, "Materials Data Curation System," *NIST MGI: Gateway to Materials Genome Information* (website), last updated February 18, 2015, mgi.nist.gov/materials-data-curation-system.

8. NIST GMI, "Facilitating the Development of Modular Data Models in Materials Science," *NIST MGI: Gateway to Materials Genome Information* (website), last updated February 18, 2015, mgi.nist.gov/ facilitating-development-modular-data-models-materials-science.

9. Mark D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data* 3, Number 160018 (2016), DOI: 10.1038/sdata.2016.18.

10. *Materials Data Facility* (MDF) (website), accessed 2017, materialsdatafacility.org/.

11. B. Blaiszik, K. Chard, J. Pruyne, R. Ananthakrishnan, S. Tuecke and I. Foster, "The Materials Data Facility: Data Services to Advance Materials Science Research," *JOM* 68, Issue 8 (2016): 2045–2052, DOI: 10.1007/s11837-016-2001-3.

12. Brian Puchala, Glenn Tarcea, Emmanuelle. A. Marquis, Margaret Hedstrom, H. V. Jagadish and John E. Allison, "The Materials Commons: A Collaboration Platform and Information Repository for the Global Materials Community," *JOM* 68, Issue 8 (2016): 2035–2044, DOI: 10.1007/s11837-016-1998-7.

13. *Materials Commons: Projects, Published Data* (website), accessed 2017, materialscommons.org.

14. Edward Kim, Kevin Huang, Alex Tomala, Sara Matthews, Emma Strubell, Adam Saunders, Andrew McCallum and Elsa Olivetti, "Machine-learned and codified synthesis parameters of oxide materials," *Scientific Data* 4 (2017), DOI: 10.1038/sdata.2017.127.

15. Paul Raccuglia, Katherine C. Elbert, Philip D. F. Adler, Casey Falk, Malia B. Wenny, Aurelio Mollo, Matthias Zeller, Sorelle A. Friedler, Joshua Schrier and Alexander J. Norquist, "Machine-learning-assisted materials discovery using failed experiments," *Nature* 533 (2016): 73–76, DOI: 10.1038/ nature17439.

16. Noel M O'Boyle, Michael Banck, Craig A. James, Chris Morley, Tim Vandermeersch and Geoffrey R. Hutchison, "Open Babel: An open chemical toolbox," Journal of Cheminformatics 3: 33 (2011), DOI: 10.1186/1758-2946-3-33.

17. Denis Fourches, Eugene Muratov and Alexander Tropsha, "Curation of chemogenomics data," *Nature Chemical Biology* 11, Issue 8 (2015): 535, DOI: 10.1038/nchembio.1881.

18. Denis Fourches, Eugene Muratov and Alexander Tropsha, "Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research," *Journal of Chemical Information and Modeling* 50, Issue 7 (2010): 1189–1204, DOI**:** 10.1021/ci100176x.

19. Ibid.

20. Diana de la Iglesia, Stacey Harper, Mark D. Hoover, Fred Klaessig, Phil Lippell, Bettye Maddux, Jeffrey Morse, Andre Nel, Krishna Rajan, Rebecca Reznik-Zellen and Mark T. Tuominen, "Nanoinformatics 2020 Roadmap," *InterNano* (website) (2011), eprints.internano.org/607/.
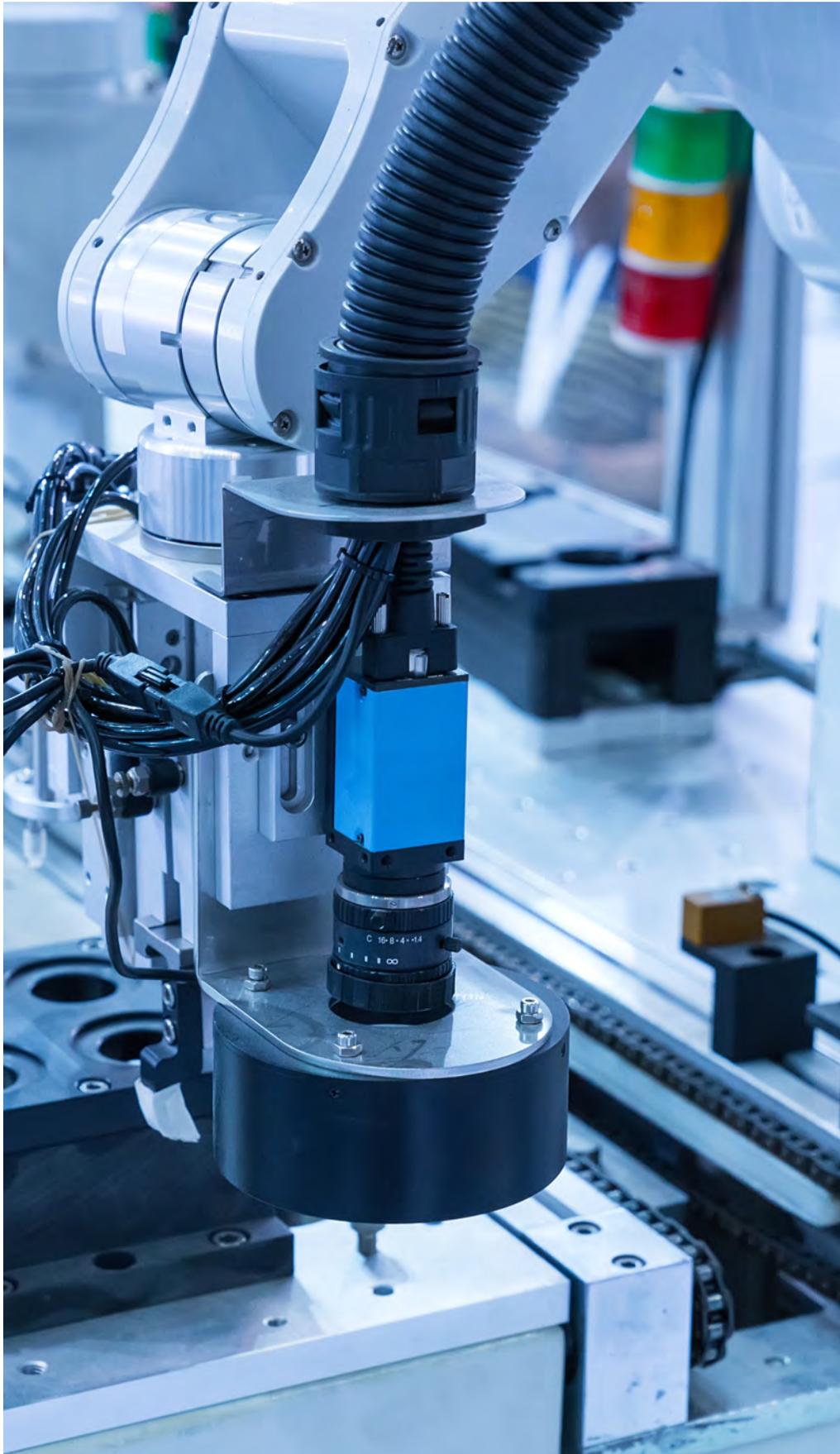
10 Conclusion

Achieving the urgent transition to a low-carbon economy requires the development of new, high-performance, low-cost materials that are safe for humans and the environment, recyclable, and use abundant elements so that they can be deployed globally. Today, advanced materials represent about 50% of the manufacturing cost of clean energy, and this is expected to increase to 80% in the near future. This challenge is not merely an engineering problem. New fundamental scientific advances are required to design and organize matter from the atomic scale to the systems scale. Therefore, the cross-cutting goal of the Clean Energy Materials Innovation Challenge is crucial: to accelerate the pace of materials discovery and develop new high-performance, low-cost clean energy technology solutions.

"As the discovery processes for advanced materials accelerate, the potential economic benefits will grow. Thus, private-sector stakeholders that enter early will presumably have a first-mover advantage: the know-how to adjust and gain a larger share of these growing benefits."

– Dr. Hermann Tribukait

This report, and the integrated materials acceleration approach it proposes, is intended to provide policy-makers and other stakeholders with direction and ideas for research investments and for soliciting and supporting projects that take advantage of these opportunities (see sidebar quote). For example, MI members and stakeholders may choose to draw on the R&D opportunities presented to launch collaborative research and technology development initiatives, align, or even refocus their national programs to incorporate these research directions. They may also invest in these R&D opportunities on their own, perhaps by launching Requests for Applications, or pursue these efforts collaboratively with other MI members via bilateral or even multilateral agreements. These collaborations may be in consortia with the private sector and other stakeholders.

The report may also inform the development of a multidisciplinary workforce for the discovery, production, and integration of advanced materials. The ultimate success of these efforts would depend on the support and leadership of all stakeholders, including academia, governments, industry, non-governmental organizations (NGOs), foundations, and multilateral organizations. The continuous support of R&D initiatives, such as the Materials Genome Initiative, can aid the development and deployment of a discovery workforce that is ready for the challenges ahead [1,2]. Undoubtedly, international coalitions built around particular topics would advance the proposed agenda and produce results more effectively.

This final chapter briefly presents the report's conclusions, beginning with the opportunities and near-term advances in the six priority areas of the Materials Acceleration Platform. It then emphasizes the importance of communicating the benefits of investing in materials science and of educating tomorrow's students in multidisciplinary materials design. The chapter closes with some final reflections.

10.1 THE INTEGRATED MATERIALS ACCELERATION PLATFORM

Chapters 4 through 9 outlined the Six Grand Goals of the Materials Acceleration Platform (MAP) and the opportunities, challenges, and promising research approaches related to each. To make the proposed integrated platform a reality by 2030, significant advances must be made in the short term in each of the priority areas. The following approaches can provide tangible benefits and improvements in the near future.

- *Artificial Intelligence for Materials*: AI models for materials that include constraints, including constraints dictated by physics for theoretical predictions of properties as well as experimental constraints, should continue to be developed. Development and increased usage of novel machine learning algorithms and architectures, specifically built to accelerate materials discovery, can result in much better performing models for materials, especially in "low-data" scenarios. Purpose-built AI tailored for materials discovery is needed to provide breakthroughs.

Engaging Students

Embedding multidisciplinary experiences at all levels at universities would educate students and promote programs that support interdisciplinary research, spanning climate and energy policy, science (e.g. materials science, computer science), technology, engineering, and mathematics (STEM).

One mechanism to boost the engagement of both high school and undergraduate students to the field of materials discovery would be to implement a transformative program similar to the International Genetic Engineered Machines (IGEM) competition in synthetic biology. This initiative engages students from a wide range of scientific disciplines and backgrounds. Students could be further engaged through student exchange programs and international competitions with recognized prizes.

design. Continued development of discriminative models will also be key due to the accelerated rate for which properties can be predicted.

- *Bridging Length and Time Scales:* Automating the current state of the art in multi-scale simulation will make simulations more accessible to general users. Machine learning methods must be integrated into all aspects of multi-scale simulation; integrating characterization tools that bridge length and time scales must also be developed.

- *Data Infrastructure and Interchange:* An important step is to make decisions on data standards and begin developing user-friendly tools for existing data repositories that naturally incentivize user participation. Continued development of methods that deal well with heterogeneous data is also essential.

- *Closing the Loop:* Designing and developing automated platforms that integrate in situ analysis of a material, either during synthesis and/or after fabrication, into a single unit will solve some of the current analytical bottlenecks.

Ultimately, all six areas of MAP can be contained in a single facility, housing "well-to-wheel" materials discovery operations similar to the operational structure of Tesla's manufacturing facilities, including the Gigafactory. Autonomous facilities (or "materials factories") will accelerate the discovery process by increasing the pace of experiments performed, harnessing and storing controlled data sets, and seamlessly integrating these with computational predictions and machine learning algorithms with robotic and automation tools. Thus, traditionally siloed competence areas, such as simulation, characterization, and synthesis, can be integrated by emerging computer science, machine learning, data infrastructure, and robotic tools. Operations can be tightly tied together, enabling researchers to quickly receive and incorporate feedback from adjacent processes to make both high-level decisions and necessary adjustments to optimize design and development. The proposed platform could also quantify uncertainty in materials discovery to help decide whether a particular set of experiments is likely to yield promising results. It could also elevate human creativity away from the details of materials synthesis and calculation, towards a higher-level control of goals-driven design of theoretical and experimental studies, where it is most needed.

10.2 COMMUNICATING THE BENEFITS OF INVESTING IN MATERIALS DISCOVERY

The importance of materials research and its critical role in the global engineering challenges ahead of us need to be communicated clearly and effectively, to students, policy-makers, global leaders, investors, and the general public. A high-profile communications program that engages the public would help ensure long-term support and funding for this mission-oriented initiative by motivating the general public and stakeholders to act decisively and promote and support this R&D initiative.

Public engagement efforts could be enhanced by creating strong awareness of past successes and exciting breakthroughs enabled by public funding of materials research and development. Telling relatable "stories" about how new materials discoveries have improved living standards and produced substantial societal benefits would increase awareness and support, including for efforts dedicated to training future generations in the quest for better technologies.

One example is the benefits health research to which the general public and stakeholders can personally relate. In this context, the priority would be to highlight the advances in areas that connect directly to societal improvements, such as medicine, information technologies, and transportation—advances that would not have been possible without crucial contributions from the fundamental sciences: physics, chemistry, and materials. For example, MRI (materials resonance imaging) relies on the inventions of NMR and superconductivity, and their practical implementation on highly functional materials such as magnets and superconductors. Although these materials are probably best known for their benefits to medicine, they could also enable a breakthrough in the energy sector. Today, superconductors only operate at a very low temperature, which incurs tremendous cost. If materials with high-temperature superconductivity were designed and discovered, the resulting cables would radically improve efficiency in power transmission and distribution. Developing new materials with these properties would revolutionize the electric grid, facilitate wider use of renewable energy resources and result in significant cost-savings from lower transmission losses.

From solar cells, to lithium-ion batteries and electric vehicles to microchips and the shale gas revolution, there are many more examples of scientific breakthroughs and new materials that can help communicate the relevance and impact of materials discovery. The benefits of developing an integrated, end-to-end materials platform, such as MAP, can be the subject of many more success stories.

10.3. FINAL REFLECTIONS

The need to accelerate the transition to a low-carbon economy is a *global* challenge that requires a *global* response. Moreover, this challenge also represents a tremendous opportunity to create wealth and improve the quality of life by making clean energy widely affordable and accessible. The proposed integrated Materials Acceleration Platform (MAP) represents a major opportunity to drive innovations that crosscut all clean energy technologies.

MAP integrates materials science with supercomputers, machine learning, and robotics to identify and develop new high-performance, low-cost materials that can be used to develop better and clean breakthrough energy technologies. This, in turn, will accelerate the transformation of the energy sector, decarbonizing the global economy. Eventually, MAP would include encoding of the physical properties of materials, along with more complex data such as manufacturing-related data, allowing for rapid inquiries. As the computers themselves become faster and more powerful, incorporating them in the development of clean energy technologies could drive an "acceleration of acceleration" phase.

The immense challenge of developing MAP requires deep international collaboration and long-term support and investment. This involves an evolution of the organizational and funding mechanisms of stakeholders and research institutions. Many of the scientific challenges associated with discovering novel materials are pre-competitive, i.e., they historically and naturally belong in the open scientific literature. As such, they could become the foundation for fruitful international multidisciplinary collaborations and research programs, bringing together leading scientists in academia and industry, engineers, and thought leaders, and inspiring the best and brightest students from around the world. The resulting dramatic acceleration in materials discovery and development, and the effect on clean energy technologies, would generate widespread social and economic benefits, transforming industries beyond the energy sector.

References for Chapter 10

1. Materials Genome Initiative, "The First Five Years of the Materials Genome Initiative: Accomplishments and Technical Highlights," *Materials Genome Initiative* (website), August 2, 2016, www.mgi.gov/sites/default/files/documents/mgi-accomplishments-at-5-years-august-2016.pdf.

2. M. L. Green, C. L. Choi, J. R. Hattrick-Simpers, A. M. Joshi, I. Takeuchi, S. C. Barron, E. Campo, T. Chiang, S. Empedocles, J. M. Gregoire, A. G. Kusne, J. Martin, A. Mehta, K. Persson, Z. Trautt1, J. Van Duren, and A. Zakutayev, "Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies," *Applied Physics Reviews* 4, Issue 1, 011105 (2017), DOI: 10.1063/1.4977487.

Acknowledgments

DISCLAIMER

The content of this report reflects the views and opinions of the participating workshop experts and report authors and do not necessarily state or reflect those of their respective government, institution, or any agency thereof or of Mission Innovation members.

Workshop Participants

Name	Affiliation	Role
Aburto Anell, Jorge Arturo	Instituto Mexicano del Petróleo	Observer
Alexander-Katz, Alfredo	Massachusetts Institute of Technology	Panel Chair
Alvarez, Gerardo		Observer
Alvarez-Franyutti, Gerardo	Mexichem	Observer
Alyamani, Ahmed	King Abdulaziz City for Science and Technology	MI Representative
Amador-Bedolla, Carlos	Universidad Nacional Autónoma de México	Local Chair
Andrade, Arturo	Universidad Nacional Autónoma de México	Observer
Antes, Matt	Energetics	Staff
Arjona, Diego		Observer
Aspuru-Guzik, Alán	Harvard University and CIFAR	Workshop Co-Chair
Aykol, Muratahan	Toyota Research Institute	Panel Chair
Barrueta, Nancy	Universidad Nacional Autónoma de México	Observer
Bárzana, Eduardo	Universidad Nacional Autónoma de México	Observer
Becker, Jill S.	02139 Inc	Participant
Beltrán, Leonardo	Secretaría de Energía	Invited Speaker
Beltrán Zarza, Norma Angélica	Instituto Nacional de Electricidad y Energías Limpias	Observer
Bentley, Stephen	Cardiff University	Observer
Berlinguette, Curtis	University of British Columbia and CIFAR	Participant
Bernstein, Alan	CIFAR	Invited Speaker
Bio, Matthew	Snapdragon Chemistry	Participant
Burke, Marty	University of Illinois at Urbana Champaign	Invited Speaker
Calva, Nacx	Energía Mx	Observer
Camacho, Emilio	California Energy Commission	Observer
Carrillo, Arturo	Sostener	Observer
Castañón, Tatiana Romero	Instituto Nacional de Electricidad y Energías Limpias	Observer
Ceder, Gerbrand	University of California, Berkeley	Invited Speaker
Cernuschi, Federico	Ricerca sul Sistema Energetico	MI Representative
Chan, Emory	Lawrence Berkeley National Laboratory	Panel Chair
Clapper, Maureen	U.S. Department of Energy	MI Representative
Cogdell, Richard	Glasgow University	Participant
Contreras, César		Observer
Contreras, Rafael		Observer
Cruz, Rosalba		Observer
de Pablo, Juan	University of Chicago	Invited Speaker
Des Rosiers, Frank	Natural Resources Canada	Invited Speaker
Dodson, Jennie	U.K. Department of Business, Energy and Industrial Strategy	MI Representative
Domenella, Sergio	Universidad Nacional Autónoma de México	Observer
Durrant, Paul	U.K. Department of Business, Energy and Industrial Strategy	MI Representative
Dwaraknath, Shyam	Lawrence Berkeley National Laboratory	Participant

E. Heinonen, Jarmo	Tekes	Observer
Early, Bryan	California Energy Commission	Observer
Fabrizio, Monica	CNR ICMATE	Participant
Fernández Rodríguez, Daniel	Instituto Nacional de Electricidad y Energías Limpias	Observer
Fischer, Peter	Lawrence Berkeley National Laboratory	Participant
Flores, Jose Roberto	Instituto Nacional de Electricidad y Energías Limpias	Observer
Flores-Leonar, Martha	Universidad Nacional Autónoma de México	Participant
Flores-Moreno, Roberto	Universidad de Guadalajara	Observer
Garcia, Karl	Universidad Nacional Autónoma de México	Observer
Gieb, Martin	European Commission	MI Representative
Golden, Ariel	Student Energy Summit	Observer
Gómez, Fernando	World Economic Forum	Participant
González, Francisco Arturo López	Instituto Nacional de Electricidad y Energías Limpias	Observer
González, Abelardo	ITAER	Observer
Grageda, Federico A.	Centro de Desarrollo y Estrategia Empresarial S.C.	Observer
Gregoire, John	California Institute of Technology	Panel Chair
Guillén, Mario Román Díaz	Instituto Nacional de Electricidad y Energías Limpias	Observer
Hasmath, Reza	CIFAR	Observer
Hattrick-Simpers, Jason	National Institute of Standards and Technology	Panel Chair
Hein, Jason	University of British Columbia	Panel Chair
Holst Jørgensen, Birte	Technical University of Denmark	MI Representative
Huang, David M.	University of Adelaide	Participant
Hutchison, Geoffrey	University of Pittsburgh	Panel Chair
Isayev, Olexandr	University of North Carolina, Chapel Hill	Participant
Jaffer, Shaffiq	TOTAL SA	Participant
Juaristi, Eusebio	CINVESTAV	Participant
Jung, Yousung	Korea Advanced Institute of Science and Technology	Panel Chair
Kanellis, Pamela	CIFAR	Staff
Kendall, Michaela	Adelan	Observer
KIM, Bong Gyun	Korea Institute of Energy Technology Evaluation and Planning	Observer
King, Alex	Critical Materials Institute	Participant
Kiviaho, Jari	VTT Technical Research Centre of Finland	Participant
Kozdras, Mark	Natural Resources Canada	MI Representative
Kreisbeck, Christoph	Harvard University	Participant
Laguna, Humberto	Universidad Nacional Autónoma de México	Participant
Lambert, Juliette	Centro de Innovación Aplicada en Tecnologías Competitivas	Participant
Lindh, Roland	Uppsala University	Participant
Løvvik, Ole Martin	SINTEF Materials and Chemistry	Participant
Magadán, Jose Manuel Martínez	Instituto Mexicano del Petróleo	Observer

Maldonado, José-Luis	Centro de Investigaciones en Óptica	Observer
Malo Tamayo, José María	Instituto Nacional de Electricidad y Energías Limpias	Observer
Maruyama, Benji	U.S. Air Force Research Laboratory	Participant
Molina, Mario	Universidad Nacional Autónoma de México	Invited Speaker
Montone, Amelia	ENEA	MI Representative
Montoya, Joseph	Lawrence Berkeley National Laboratory	Participant
Morales, José Luis	Universidad Nacional Autónoma de México	Observer
Mosby, Anna	Energetics	Staff
Noguez, Cecilia	Universidad Nacional Autónoma de México	Participant
Odom, Susan	University of Kentucky	Participant
Olmedo, Jose		Observer
Ortiz, Carlos	Secretaría de Energía	Invited Speaker
Pai, Ranjith Krishna	Indian Ministry of Science and Technology	MI Representative
Palma, Julio	Pennsylvania State University	Participant
Peón Peralta, Jorge	Universidad Nacional Autónoma de México	Observer
Persson, Kristin	University of California, Berkeley, LBNL	Workshop Co-Chair
Pijpers, Joep	Instituto Nacional de Electricidad y Energías Limpias	Participant
Pulido, Natalia	Tecnológico de Monterrey, Campus Monterrey	Observer
Rejón, Leonardo	Instituto Nacional de Electricidad y Energías Limpias	Observer
Roberts, Jeff	Lawrence Livermore National Laboratory	Participant
Roch, Loïc	Harvard University	Participant
Rocha Rinza, Tomás	Universidad Nacional Autónoma de México	Observer
Rodas Grapain, Arturo	Instituto Nacional de Electricidad y Energías Limpias	Observer
Ross, Michael	University of California, Berkeley	Participant
Rotenberg, Eli	Lawrence Berkeley National Laboratory	Participant
Ruiz-Morales, Yosadara	Instituto Mexicano del Petroleo	Observer
Saikin, Semion	Harvard University	Participant

Saldaña Talvitie, Elli	Universidad Nacional Autónoma de México	Staff
Salinas Bravo, Víctor Manuel	Instituto Nacional de Electricidad y Energías Limpias	Observer
Sanchez, Rafael		Observer
Sánchez, Mariano	Universidad Nacional Autónoma de México	Participant
Sargent, Edward	University of Toronto and CIFAR	Participant
Schrier, Joshua	Haverford College	Panel Chair
Schuster, Frédéric	Commissariat à l'Énergie Atomique	Observer
Segalman, Rachel	University of California, Santa Barbara	Panel Chair
Serrano, Siria	Harvard University	Staff
Sfeir, Matthew	Brookhaven National Laboratory	Participant
Simon, Horst	Lawrence Berkeley National Laboratory	Invited Speaker
Singh, Chandra Veer	University of Toronto	Observer
Solis-Ibarra, Diego	Universidad Nacional Autónoma de México	Panel Chair
Soto, Maritza	Sostener	Staff
Spadaccini, Christopher	Lawrence Livermore National Laboratory	Participant
Stassin, Fabrice	Energy Materials Industrial Research Initiative/Umicore	Observer
Subramaniam, Velumani	CINVESTAV	Participant
Tabor, Daniel	Harvard University	Participant
Tao, Andrea	University of California, San Diego	Participant
Torres, Ramon Carlos	Instituto Nacional de Electricidad y Energías Limpias	Observer
Tribukait-Vasconcelos, Hermann	Mexico Energy R&D Innovation Funds	Executive Chair
van Grondelle, Rienk	Vrije Universiteit Amsterdam and CIFAR	Participant
Vegge, Tejs	Technical University of Denmark	Participant
Villegas, Alejandro	Sostener	Staff
Warren, Jeff	Simon Fraser University	Participant
Wilhelm, Martin	Geopolis Energy Partners	Observer
Williams, Daryl	Imperial College	Participant
Zeron, Paulino		Observer

Acronyms

2D	Two-Dimensional
3D	Three-Dimensional
AC	Alternating Current
AFLOW	Automatic Flow (database)
AFM	Atomic Force Microscopy
Ag	Silver
AI	Artificial Intelligence
AI4M	Artificial Intelligence for Materials
AiiDA	Automated Interactive Infrastructure and Database for Computational Science
ANI	Accurate NeurAl networK engINe for Molecular Energies (ANAKIN-ME or ANI)
API	Application Programming Interface
ARES	Autonomous Research System
ARPES	Angle-Resolved Photoemission Spectroscopy
Ba	Barium
BEIS	Department for Business, Energy & Industrial Strategy (United Kingdom)
C	Carbon
Ca	Calcium
CCSD	Cambridge Crystallographic Structural Database
Cd	Cadmium
CEO	Chief Executive Officer
CH_4	Methane
CIFAR	Canadian Institute for Advanced Research
CO_2	Carbon Dioxide
COP21	21st Conference of the Parties
COP22	22nd Conference of the Parties
Cu	Copper
DARPA	U.S. Defense Advanced Research Projects Agency
DC	Direct Current
DFT	Density Functional Theory
DNA	Deoxyribonucleic Acid
DOE	U.S. Department of Energy
DSSC	Dye-Sensitized Solar Cell
EMIRI	Energy Materials Industrial Research Initiative
eV	Electron-Volt(s)
FAT/ML	Fairness, Accountability, and Transparency in Machine Learning
g	Gram(s)
Ge	Germanium
GHG	Greenhouse Gas
GSP	Genome Sequencing Program (NHGRI)
H	Hydrogen
h	Hour(s)
HERMAN	High-throughput Experimentation Robot for the Multiplexed Automation of Nanochemistry
high Tc	High-Temperature (superconductivity)
HPLC	High-Performance Liquid Chromatography
HTE	High-Throughput Experimental
ICSD	Inorganic Crystal Structure Database
iGEM	The International Genetically Engineered Machine
InChI	International Chemical Identifier
IP	Intellectual Property
IR	Infrared
IUPAC	International Union of Pure and Applied Chemistry
JCAP	Joint Center for Artificial Photosynthesis

kg	Kilogram(s)
kJ	Kilojoule(s)
kWh	Kilowatt-Hour(s)
LBNL	Lawrence Berkeley National Laboratory
LC	Liquid Crystal
LED	Light-Emitting Diode
LHASA	Logic and Heuristics Applied to Synthetic Analysis (computer program)
Li	Lithium
MAP	Materials Acceleration Platform
MDCS	Materials Data Curation System
MDF	Materials Data Facility
MI	Mission Innovation
mol	Mole(s)
MP	Materials Project
MRI	Materials Resonance Imaging
MS-CG	Multi-Scale Coarse-Grained
N	Nitrogen
Natl. Lab	National Laboratory
Nd	Neodymium
NGO	Non-Governmental Organization
NHGRI	U.S. National Human Genome Research Institute
Ni	Nickel
NIMS	National Institute for Materials Science
NIST	National Institute of Standards and Technology
nm	Nanometer(s)
NMR	Nuclear Magnetic Resonance
NOMAD	Novel Materials Discovery
NSF	National Science Foundation (United States)
O	Oxygen
OQMD	Open Quantum Materials Database
ORGAN	Objective-Reinforced Generative Adversarial Networks
ORNL	Oak Ridge National Laboratory
Pb	Lead
PEEM	Photoemission Electron Microscopy
PL	Photoluminescence
PV	Photovoltaic(s)
QD	Quantum Dot
QM/MM	Quantum Mechanics/Molecular Mechanics
R&D	Research and Development
RIXS	Resonant Inelastic X-ray Scattering
Se	Selenium
SEM	Scanning Electron Microscopy
SENER	Mexican Ministry of Energy (Secretaría de Energía)
Si	Silicon
SMILES	Simplified Molecular-Input Line-Entry System
STM	Scanning Tunnelling Microscopy
STXM	Scanning Transmission X-ray Microscopy
Te	Tellurium
TEM	Transmission Electron Microscopy
Ti	Titanium
UNAM	Universidad Nacional Autónoma de México
W3C	World Wide Web Consortium
WANDA	Workstation for Automated Nanomaterials Discovery and Analysis
Wh	Watt-Hour(s)
XAS	X-ray Absorption Spectroscopy
XRD	X-ray Diffraction
Zn	Zinc