OPEN ACCESS
IOP Publishing
Journal of Physics D: Applied Physics
J. Phys. D: Appl. Phys. **52** (2019) 013001 (48pp)
https://doi.org/10.1088/1361-6463/aad926

Topical Review

The 2019 materials by design roadmap

Kirstin Alberi[1], Marco Buongiorno Nardelli[2], Andriy Zakutayev[1], Lubos Mitas[3], Stefano Curtarolo[4,5], Anubhav Jain[6], Marco Fornari[7], Nicola Marzari[8], Ichiro Takeuchi[9], Martin L Green[10], Mercouri Kanatzidis[11], Mike F Toney[12], Sergiy Butenko[13], Bryce Meredig[14], Stephan Lany[1], Ursula Kattner[15], Albert Davydov[15], Eric S Toberer[16], Vladan Stevanovic[16], Aron Walsh[17,18], Nam-Gyu Park[19], Alán Aspuru-Guzik[20,21], Daniel P Tabor[20], Jenny Nelson[22], James Murphy[23], Anant Setlur[23], John Gregoire[24], Hong Li[25], Ruijuan Xiao[25], Alfred Ludwig[26], Lane W Martin[27,28], Andrew M Rappe[29], Su-Huai Wei[30] and John Perkins[1]

[1] National Renewable Energy Laboratory, Golden, CO 80401, United States of America
[2] University of North Texas, Denton, TX, United States of America
[3] North Carolina State University, Raleigh, NC, United States of America
[4] Duke University, Durham, NC, United States of America
[5] Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany
[6] Energy Storage and Distributed Resources Department, Lawrence Berkeley National Laboratory, Berkeley, CA, United States of America
[7] Department of Physics, Central Michigan University, Mt. Pleasant, MI 48859, United States of America
[8] Theory and Simulation of Materials, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
[9] University of Maryland, College Park, MD, United States of America
[10] Materials Measurement Science Division, National Institute of Standards and Technology, Gaithersburg, MD, United States of America
[11] Northwestern University, Evanston, IL, United States of America
[12] SLAC, Menlo Park, CA, United States of America
[13] Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, United States of America
[14] Citrine Informatics, Redwood City, CA, United States of America
[15] Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD, United States of America
[16] Colorado School of Mines, Golden, CO, United States of America
[17] Department of Materials, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom
[18] Department of Materials Science and Engineering, Yonsei University, Seoul 03722, Republic of Korea
[19] School of Chemical Engineering, Sungkyunkwan University (SKKU), Suwon 440-746, Republic of Korea
[20] Department of Chemistry and Department of Computer Science, University of Toronto, Canada
[21] Vector Institute for Artificial Intelligence, Toronto, Canada
[22] Department of Physics and Centre for Plastic Electronics, Imperial College London, London, United Kingdom
[23] Functional Materials Laboratory, GE Global Research, Niskayuna, NY, United States of America
[24] California Institute of Technology, Pasadena, CA, United States of America
[25] Institute of Physics, Chinese Academy of Sciences, Beijing, People's Republic of China

[26] MEMS Materials, Institute for Materials, Ruhr-University Bochum, Bochum, Germany
[27] Department of Materials Science and Engineering, University of California, Berkeley, CA, United States of America
[28] Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States of America
[29] Department of Chemistry, University of Pennsylvania, Philadelphia, PA, United States of America
[30] Beijing Computational Science Research Center, Beijing 100193, People's Republic of China

E-mail: Kirstin.Alberi@nrel.gov

Received 19 February 2018, revised 25 June 2018
Accepted for publication 9 August 2018
Published 24 October 2018

Abstract

Advances in renewable and sustainable energy technologies critically depend on our ability to design and realize materials with optimal properties. Materials discovery and design efforts ideally involve close coupling between materials prediction, synthesis and characterization. The increased use of computational tools, the generation of materials databases, and advances in experimental methods have substantially accelerated these activities. It is therefore an opportune time to consider future prospects for materials by design approaches. The purpose of this Roadmap is to present an overview of the current state of computational materials prediction, synthesis and characterization approaches, materials design needs for various technologies, and future challenges and opportunities that must be addressed. The various perspectives cover topics on computational techniques, validation, materials databases, materials informatics, high-throughput combinatorial methods, advanced characterization approaches, and materials design issues in thermoelectrics, photovoltaics, solid state lighting, catalysts, batteries, metal alloys, complex oxides and transparent conducting materials. It is our hope that this Roadmap will guide researchers and funding agencies in identifying new prospects for materials design.

Keywords: density functional theory, materials genome initative, materials design, high-throughput methods, energy applications

(Some figures may appear in colour only in the online journal)

Contents

1. Introduction

Kirstin Alberi[1], *Marco Buongiorno Nardelli*[2] *and Andriy Zakutayev*[1]

[1] National Renewable Energy Laboratory, Golden, CO 80401, United States of America
[2] University of North Texas, Denton, TX, United States of America

Advances in renewable and sustainable energy technologies critically depend on our ability to design materials with the optimal properties for each individual application. Computational methods have accelerated materials design efforts through rapid and comprehensive prediction of materials stability and properties. A very simplistic metric for assessing the rise of computational materials efforts is the total number of materials that have been 'predicted' (which does not capture the extent or diversity of the calculated properties). As schematically shown in figure 1(a), the number of theoretically predicted materials in computational materials property databases, including AFLOW, the Open Quantum Materials Database and the Materials Project (10^4–10^6), is now comparable to the number of experimental entries in crystallographic databases ($\sim 10^5$). Perhaps even more importantly, increased accessibility to the computed properties has also sped up experimental research and development of new functional materials for a wide range of applications. Acceleration of materials by design research is evidenced by the nearly exponential growth in the number of publications on materials design, shown in figure 1(b), where breakthroughs were facilitated by the development of user friendly *ab initio* codes (mid-90s) and automation of these codes to run high-throughput computations (>2010). Yet, for all its recent successes, the materials by design concept is relatively new and has the potential for further expansion and impact.

The purpose of this Roadmap is to present an overview of the current state of computational materials prediction approaches, corresponding advanced synthesis and characterization methods, and the application of these computational and experimental techniques to various energy relevant technologies. Future challenges and opportunities that must be addressed to improve materials by design approaches are also discussed. We have asked leading researchers in each of these areas to weigh in on these issues and provide their perspectives and visions for the advancement of the materials by design field. The covered topics include computational techniques, validation of the results, materials databases, materials informatics, high-throughput combinatorial methods, advanced characterization approaches, as well as materials design issues in thermoelectrics, photovoltaics, solid state lighting, batteries, metal alloys, complex oxides and transparent conducting materials.

A unifying theme of many of the contributions to this collection is the need for high-throughput computational and experimental techniques as a foundation for the materials by design paradigm, as well as methods to exploit synthesis and manufacturing processes for new materials. Nowadays, we possess

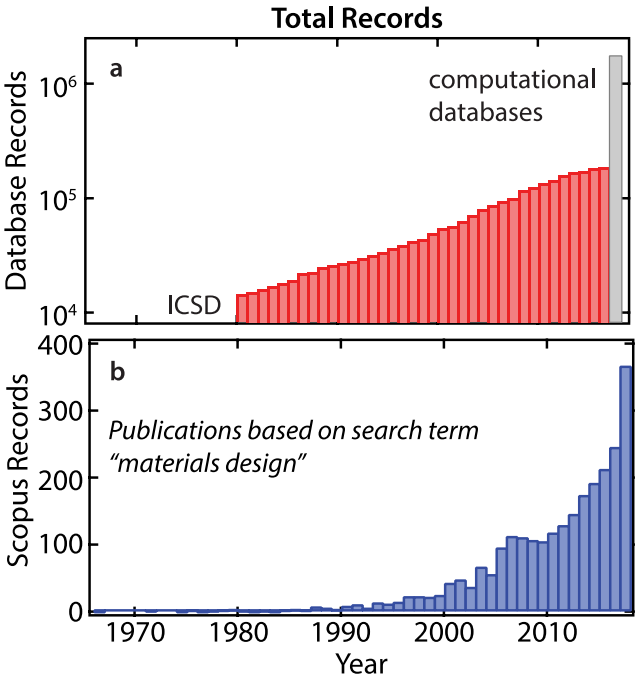


Figure 1. (a) Total number of compounds contained within the Inorganic Crystal Structure Database (ICSD) and computational databases. These values do not reflect the extent of the information in each entry. (b) The number of publications returned in from a Scopus search using query terms 'materials design' and constraining the search to exclude irrelevant results (e.g. furniture, textiles, bridges, etc).

the ability to efficiently generate and manage large amounts of computational data in open repositories, facilitating access to a plethora of calculated properties and functions of millions of different materials. Computational efforts that go beyond predicting the thermodynamic stability of a material and provide additional calculations of electronic structure, properties and even optical spectra of diverse material systems are becoming increasingly important and valuable. Similar large data repositories of experimentally measured properties are less common but would be needed to benchmark and supplement the computations. From here, we envision innovative ways to interrogate the big data space through data mining, machine learning, autonomous systems and artificial intelligence techniques. We emphasize that all of these techniques must work together to realize the full potential of the materials by design approach. Another common theme of several contributions to this Roadmap is the need for *in situ* and operando measurement techniques to derive deeper scientific insight into material synthesis processes.

A simplified example of a materials design process that can be used to accelerate materials transfer from computer simulations to lab bench and consumer products is illustrated in figure 2. Theoretical challenges range from intelligent optimization algorithms that predict candidate material compositions and structures to the exploitation of the appropriate descriptors of functional properties. Experimental needs include accelerated synthesis of the most promising candidates and advanced characterization of these materials. Finally, application requirements involve validation of the measured or calculated properties, improved synthesis routes,

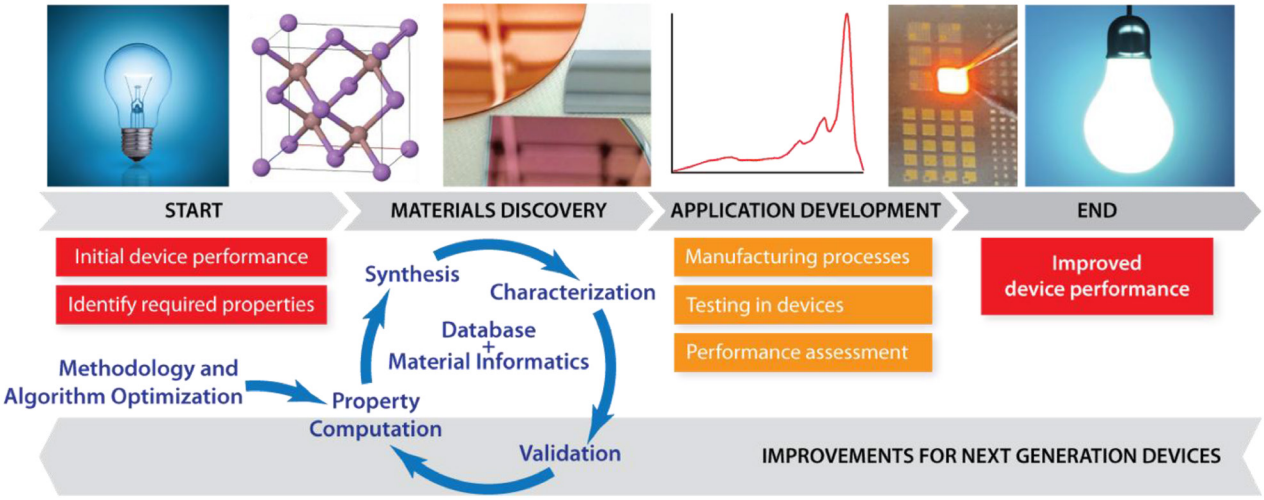


Figure 2. Schematic of the materials by design approach.

testing of the materials in devices, and the clear articulation of desired materials properties needed for prediction of next generation materials candidates in the next cycle.

The resulting Roadmap is broadly divided in two main sections on Methods and Applications. In the Methods section, we review advances and challenges in three areas: computation of materials properties beyond the current standard, novel experimental techniques for materials design and discovery, and the curation and use of digital data. In the Applications section, we provide a snapshot of the current issues and trends in materials design in areas ranging from semiconductors to batteries to structural materials. Each application may present its own specific material design challenges to overcome, but the general materials design approach is expected to be germane to all of them. Another relevant issue is how to rapidly and efficiently implement such material design approaches at laboratory prototyping and even industrial manufacturing scales.

We hope that this Roadmap will provide a concise yet comprehensive review of a fast-growing field of materials design, one that has the potential to shape the global economy and human well-being for years to come.

2. Data generation beyond standard DFT for high-throughput applications

Marco Buongiorno Nardelli[1] *and Lubos Mitas*[2]

[1] University of North Texas, Denton, TX, United States of America
[2] North Carolina State University, Raleigh, NC, United States of America

Status. The last decade has established that a combination of first principles theoretical computations in synergy with experimental investigation is a powerful foundation for the discovery of new materials, new functions, and new design concepts in a multi-disciplinary effort that encompasses the development of transformative computational tools, unprecedented data analysis approaches, and systematic interaction with experimental discovery and validation.

All the existing materials property databases derived from computation are based on density functional theory (DFT) in the local density (LDA) or generalized gradient (GGA) approximations. Although the reproducibility of results in density functional calculations of solids has by now been an established fact [1], much less documented at the scale of large materials databases is the *veracity* of the quantities that are calculated. *Accurate* prediction of the electronic properties of materials at a low computational cost has been a major challenge in *ab initio* computational materials science from the first applications of DFT in the early 80s to the current advanced high-throughput frameworks. Despite the enormous success of DFT in describing many physical properties of real systems, the method is crippled by the presence of a correlation term that represents the difference between the true energy of the many-body system of the electrons (only formally known) and the approximate energy that we can compute. The next step beyond DFT is based on GW theory that provides perturbational improvements, in particular for band gaps that are crucial for many applications [2].

At present, *many-body approaches*, such as quantum Monte Carlo (QMC) methods, are becoming used more broadly for many key *energy differences*, such as fundamental and optical gaps, cohesions, energy orderings of various structures and defect energetics [3–5]. Very recently, QMC has reached even finer energy scales, such as differences between different magnetic states or dissociations of non-covalently bonded systems with subchemical accuracy (0.1 kcal mol^{-1}) [3, 6]. For reliable description of *spectral properties* and response functions, the methods of choice are based on dynamical mean field theory (DMFT) that offers insights beyond perturbative corrections and enables one to also study electronic phase changes, such as metal–insulator and magnetic transitions. *Finite temperature effects* that are crucial for functions of real materials are often studied by a combination of DFT electronic structure and molecular dynamics approaches. Many-body alternatives, such as path integral Monte Carlo, are still under development and so far have been applied mostly to systems with light elements. Despite all of these promising advances, databases of tools and calculations from many-body methods are basically non-existent at present, as a result of still very intense developments and the diversity of ideas that are being pursued (see below for very recent progress in this direction).

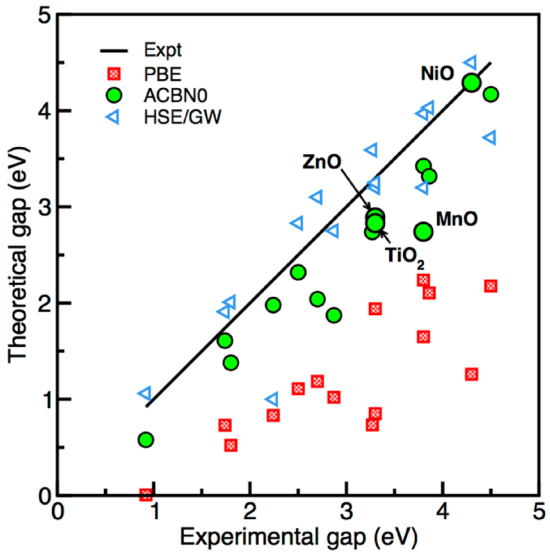


Figure 3. Performance of ACBN0 for a number of transition metal oxides and chalcogenides. In the figure, we compare the experimental band gap with the one obtained by PBE (red), ACBN0 (green) and HSE or GW (blue). For a complete discussion, see [7].

Current and future challenges. The key to achieving significant breakthroughs rests on our ability to efficiently integrate all the components in a seamless constructive cycle and in particular one development of innovative theoretical methods and tools beyond the state-of-the-art DFT approaches, which are fast, robust and amenable to high-throughput (HT) computation.

In this respect, we see many distinct but parallel requirements:

1. The development and validation of *novel functionals* to improve accuracy of traditional DFT; verification of data for complex materials systems with strong electron localization and correlation; development of novel computational algorithms to evaluate exchange energy in hybrid density functional for HT applications.
2. The inclusion of methodologies beyond DFT for the generation of materials data towards chemical and subchemical accuracy, such as QMC and DMFT, which are crucial for increasing the accuracy of calculations for energetics, as well as spectral properties that are needed for building significantly more accurate data sets for both equilibrium and non-equilibrium conformations.
3. The development of procedures for fast *computational characterization* of materials properties, such as: calculation of transport (both regular and anomalous) properties; development of efficient methodologies for the simulation of theoretical spectroscopies in the broadest energy range and with maximum accuracy and high computational efficiency.

4. The development of *materials modelling and prediction software* to match the scaling challenges posed by the ever-evolving hardware architectures and accelerated hybrid computer systems. This effort requires a substantial redesign of software and algorithms to efficiently take advantage of the increased hardware power.
5. Another important direction is expansions of calculations to nonzero temperatures both by DFT+ molecular dynamics approaches, as well as by many-body treatments based on thermal density matrices, such as path integral Monte Carlo.
6. A further important goal is the many-body treatment of spin, spin–orbit effects and relativity in general for heavy element materials, including fast characterization of phases with topological properties.

Advances in science and technology to meet challenges. A key challenge for current DFT is the accurate description of materials with strong electron localization and correlation. Work to address this challenge and at the same time maintain a competitive computational cost must continue beyond the existing efforts (see, for instance, figure 3 [7, 8]). There are important developments in many-body methods, such as the new generation of pseudopotentials from correlated treatments [9], second-quantized QMC approaches based on auxiliary fields and/or stochastically sampled excited state expansions, as well as finite temperature many-body calculations [4, 5]. Another direction of intense study is the use of stochastic methods for responses and time-dependent phenomena both in DMFT and QMC approaches. The databases for many-body methods as well as for the storage of results from many-body calculations have been getting significant attention very recently and several initiatives have been formed that aim to establish such repositories in a systematic open source/open data framework for both many-body codes and data for broad use [9].

Concluding remarks. The next leap in building reliable data will encompass several important aspects. There is a continuing effort to push the limits of accuracy for materials in key directions: energy differences for systems in equilibrium and non-equilibrium atomic conformations; explorations of non-ideal or composite systems, such as imperfect crystals with defects and impurities; 1D and 2D systems on substrates; organic–inorganic and cluster-based structures. Much better quality of data and the inclusion of spin-dependent interactions in many-body methods is highly important for finer energy scales, such as magnetic, topological or exotic electronic phases and heavy atom systems. Materials functions at nonzero temperatures and therefore better and more accurate description of many quantities at finite temperatures are highly desirable. Almost all materials operate in some type of response regime and therefore a better understanding of responses and time-dependent phenomena is another important goal—systematic description of such phenomena for materials is still very difficult in general. Since materials research is a vast 'universe' of phenomena and spans a huge range of observed quantities, the diversity of approaches is of paramount importance to address all of these important challenges.

Acknowledgments

M B N acknowledges support from DOD-ONR (N00014-13-1-0635, N00014-11-1-0136, N00014-15-1-2863) and the Texas Advanced Computing Center at the University of Texas, Austin. L M acknowledges support by the DOE BES MSED Computational Materials Sciences Program and DOE Center for Predictive Simulation of Functional Materials for the research on new generation of pseudopotentials and the DOE DE-SC0012314 grant for the research on development of spin–orbit in QMC.

3. Computational infrastructures for data generation

Anubhav Jain[1] *and Stefano Curtarolo*[2,3]

[1] Energy Storage and Distributed Resources Department, Lawrence Berkeley National Laboratory, Berkeley, CA, United States of America
[2] Duke University, Durham, NC, United States of America
[3] Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany

Status. New materials have historically been designed through intuition and experimentation. However, the high cost, long times, and manual effort required for experimental study have always served as major barriers to this process. In the last two decades, however, advancements in theory, computing hardware, and numerical algorithms have converged to provide new simulation-based methods for investigating materials that are fast, cost-efficient, and scalable to millions of materials. In particular, density functional theory (DFT) calculations, which solve the electronic structure of a material with few adjustable parameters, are now routinely run in a 'high-throughput' mode [10] in which researchers are able to evaluate thousands or even millions of materials on supercomputers with little intervention (figure 4). Today, there exist several examples in which such computational techniques have identified new functional materials that have subsequently been confirmed by experiments [11].

Furthermore, computational data sets can be shared through one of several online databases [12], such as the Materials Project (www.materialsproject.org) or AFLOWlib (www.aflow.org). These resources contain millions of computational 'measurements' of materials properties, such as formation enthalpy, electronic band structure, and elastic moduli, that can be systematically searched and that are constantly expanding in scope. The ability to rapidly generate reliable materials data in this manner improves every year as computing costs decrease, theoretical methods to study materials become more accurate, and the software to apply these techniques becomes more powerful and accessible to a larger audience.

Today, many research groups regularly employ high-throughput computing to screen materials libraries for functional applications. However, a major opportunity for the future is to incorporate techniques from the fields of data mining and statistical learning to the analysis of materials data. The arrival of large-scale computational data generation infrastructures has created the potential to develop a new science of 'materials informatics' [13]. It is possible that entirely new ways of developing chemical rules and thinking about materials behaviour will result from the marriage of simulation data with machine learning advancements, thereby adding a new dimension to the traditional methods of materials design.

Current and future challenges. There exist both fundamental and practical challenges in data generation through simulation. For example, developing physical theories that are amenable to computation and that achieve high accuracy across

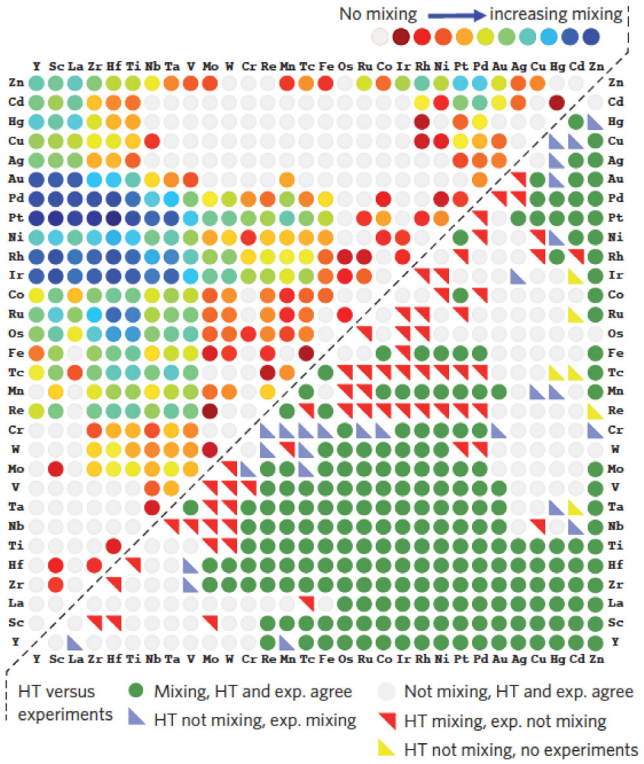


Figure 4. High-throughput computational analysis of binary alloy formation as compared with known experimental data. This serves an example of how computational data generation can rapidly 'map' a chemical space. Reprinted by permission from Macmillan Publishers Ltd: Nature Materials [10], Copyright 2013.

materials with very diverse electronic structures (such as metals, semiconductors, ceramics) is an extremely difficult task. In this regard, the DFT approach, pioneered by Kohn and Sham, and for which Kohn would later receive a Nobel prize, serves as a very good starting point. Even simple approximations to DFT can produce accurate results across many materials classes, with discrepancies in accuracy between computation and experiment being as low as a few percent. However, certain materials classes (such as strongly-correlated systems) and certain materials properties (such as excited-state properties, e.g. band gap or optical spectrum) are poorly modelled with the typical DFT approaches. Thus, major research efforts worldwide are being devoted to developing methods that improve the accuracy of the method. For example, frameworks to automate the QMC method (despite its very high computational cost) are now actively being developed and tested [14, 15].

A second fundamental challenge relates to the scaling of the computational effort needed for the computation in relation to the system size in electrons. Today, it is routine to calculate the properties of systems with unit cells of a few hundred atoms, but the poor N^3 scaling of DFT methods with system size means that systems with thousands or tens of thousands of atoms are either inaccessible or require specialized treatment. DFT methods today are largely limited to systems of low or intermediate complexity and approaches to either improve the scaling of DFT methods or to 'glue together' different modelling techniques through multiscale modelling also form a current major research topic.

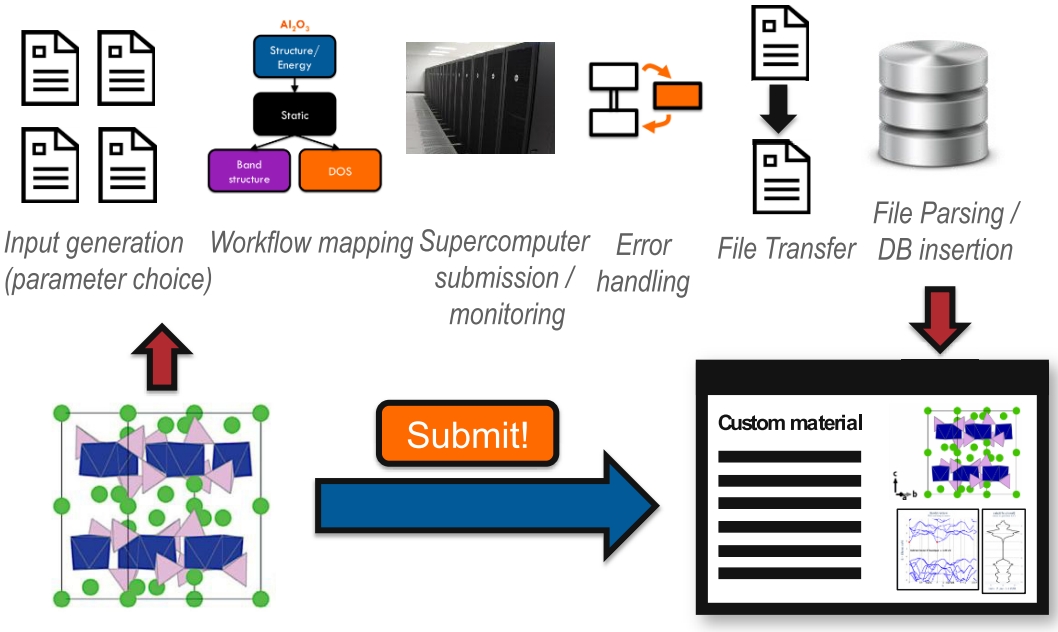


Figure 5. Schematic of some of the steps needed to execute a simulation (top route). Today, the existence of computational data generation frameworks essentially makes this process as simple as clicking a button or calling a single function (bottom route).

There also exist important practical considerations that must be addressed when generating large amounts of simulation data. For example, one must evaluate the various trade-offs in computational cost, complexity, and accuracy when determining the level of theory at which to perform calculations. Each simulation that is executed must undergo a complex sequence of steps (figure 5) including generation of input files, execution on specialized supercomputing centers, and error handling. Furthermore, many materials properties require chaining together dozens of such simulations in workflows with complex dependencies. One must be able to track millions of simulations and files and be able to quickly access any result. Here, advancements in software have greatly progressed in the last decade such that these practical aspects of data generation, once a large undertaking, can be handled by individual researchers.

Advances in science and technology to meet challenges. A summary of the fundamental challenges for DFT calculations has been previously covered [16]. Here, we summarize some of the major active worldwide efforts in developing computational data generation frameworks.

The Automated Interactive Infrastructure and Database for Computational Science (AiiDA) platform [17], developed by the European NCCR-MARVEL collaboration and written in Python, is available to researchers as open-source and aims to assist researchers with the ADES (automation, data, environment, sharing) components of data generation. A major benefit of the AiiDA platform is that one can perform rich searches over a database of calculation workflows, thus introspecting many features of the computations both programmatically and visually.

The Automatic Flow (AFLOW) platform [18], developed by the AFLOW.org consortium and written in C++ and python is available as a free download. AFLOW assists users in many aspects of simulations, from generating materials models (e.g. from common prototypes) to performing simulation sequences and correcting errors that occur. This all-in-one nature makes many powerful tools and analyses available to users and can be used either alone or in combination with other tools. The AFLOW platform has been used to create the AFLOW database accessible at www.aflow.org. Several interface libraries are available for using the framework in python workflows. Among them are the symmetry analysis (AFLOW-SYM), phonon and thermal transport (AFLOW-APL), disordered analysis (AFLOW-POCC), and machine learning automation (AFLOW-ML, aflow.org/aflow-ml). The consortium has also standardized a cloud-language for complex data analysis and retrieval (AFLUX).

The Atomic Simulation Environment (ASE) library [19], first developed at Denmark Technical University and written in Python, is available to researchers as open-source. It was one of the first high-level interfaces to simulation software and has since expanded to include a host of useful capabilities. ASE can for example help build complex models, such as surfaces, and is unique in that it allows interchanging the specific DFT theory calculator (software) used to perform the calculation.

The atomate library [20], developed by the Materials Project collaboration and written in Python, is available as open-source. The atomate library uses several underlying libraries also developed by the same collaboration (e.g. pymatgen, FireWorks, and custodian) to perform a range of actions, such as creating sophisticated materials models, managing workflows on supercomputing centers, and providing error correction. Atomate implements many common materials workflows and was used to create the Materials Project database, available at www.materialsproject.org.

Collectively, these frameworks are greatly expanding the audience for computational data generation.

Concluding remarks. Advancements in computational data generation have provided researchers with a new toolkit and a new avenue with which to address materials design problems. With a few exceptions, these high-throughput techniques have only been applied for about a decade or so and it is likely that some of the most important advancements in the field are yet to come. In particular, addressing fundamental challenges in achieving high accuracy and in modelling large, realistic systems remain formidable topics for future work. Similarly, extracting knowledge from large materials data sets through machine learning techniques is still in its infancy. Nevertheless, the ability to quickly generate data on a library of materials of interest and to share these results with collaborators worldwide has already changed the way in which many researchers, experimentalists and theories alike, are conducting materials design studies and has led to a new collective, collaborative method for applying theory to materials.

Acknowledgments

A J acknowledges funding from the Materials Project Center through Grant No. KC23MP through the US Department of Energy, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under Contract No. DE-AC02 05CH11231. S C acknowledges funding from DOD-ONR (N00014-15-1-2863, N00014-17-1-2090, N00014-16-1-2583, N00014-17-1-2876).

4. Verification and validation for electronic-structure databases

Marco Fornari[1] *and Nicola Marzari*[2]

[1] Department of Physics, Central Michigan University, Mt. Pleasant, MI 48859, United States of America
[2] Theory and Simulation of Materials, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Status. Electronic-structure calculations have had a profound influence on the development of computational materials science, especially thanks to the relative efficiency and accuracy of density-functional theory (DFT). The 'Materials Genome Initiative', launched by President Obama in 2011, has given worldwide visibility to this effort, and the task of developing novel materials has started to leverage queryable databases whose content is exploited to accelerate the discovery process. Large sets of experimental and theoretical data, built on the continuous effort of selected research groups, are now being curated, organized, and reconsidered for purposes beyond personal repositories. Because of their homogeneity in terms of format, results from DFT calculations were among the first data made publicly available to complement or expand existing databases of experimental crystal structures, such as those collected in the Inorganic Crystal Structure Database (ICSD), the Crystallographic Open Database (COD), and the Pauling File.

To the best of our knowledge, a list of electronic-structure databases includes the Materials Project (materialsproject.org), the Computational Materials Repository (cmr.fysik.dtu.dk), the Open Quantum Materials Database (oqmd.org), the Open Materials Database (httk.openmaterialsdb.se), the Theoretical Crystallographic Open Database (www.crystallography.net/tcod), the Materials Mine (www.materials-mine.com), the NREL Materials Database (materials.nrel.gov), the Automatic FLOW repository (aflow.org), the Materials Cloud (materialscloud.org), and the Novel Materials Discovery Repository (nomad-repository.eu); these allow, with different licenses, to download selected records, or in some cases even the entire repository. The records that are accessible are usually generated with standard plane-wave pseudo-potential electronic structure codes; mostly with VASP (www.vasp.at) [21], or more recently with Quantum ESPRESSO (www.quantum-espresso.org) [22] and other electronic-structure codes. They usually include input files to establish some amount of reproducibility for the calculations. Typical quantities that are reported in the databases are relaxed geometries of crystal structures, together with total energies, band structures, and densities of states.

In most cases, data generation has been performed for specific projects and the properties included in the materials records may vary greatly, even within the same repository. In addition, due to the specific research goals that drove the calculations, the overall quality of the data has not been extensively assessed. It is thus assumed that the data are 'good enough' for the specific research goal, although this approach hinders the ability to further use the data in unrelated data driven research, and often even the same calculations for the same structure performed with the same code can have significant discrepancies. In order to force consistent quality among the records within a repository, several groups have opted to establish what the calculation parameters are that should guarantee reliable results across structural and chemical variations [23]. Other groups have performed systematic testing on selected systems aiming to provide stricter transferability criteria and improve, for example, the quality of the pseudo-potential calculations against all-electron data ([1, 24] and http://materialscloud.org/sssp/).

In this Roadmap, we touch upon the efforts toward the concepts of verification and validation assuming the following definitions:

- Verification efforts are aimed at assessing that the calculations have been performed correctly, and provide the theoretical results that are expected—e.g. there is one single theoretical value for the lattice constant of crystalline silicon within the LDA approximation to DFT, even if no one knows what it is with an accuracy greater than a few parts per thousand. This effort comprises establishing and assessing the quality of the calculations in terms of the input parameters, from energy cutoffs to k-point sampling to the convergence thresholds, the presence of bugs, approximate numerical methods, and so on. In this context, the major and most difficult challenges involve hidden bugs, and the use of the pseudo-potential approximation.
- Validation involves comparing the theoretical calculations and experimental measurements in order to quantify the predictive value of the theory—for this, one needs also to carefully assess, for example, the condition at which the comparisons are made (environmental conditions, such as temperature, degree of imperfections in the experimental sample, the role of quantum nuclear effects not considered in the theory, etc.).

Current and future challenges. Two main aspects must be stressed when discussing the path toward verification: the first one involves the definition of quality standards for the calculated quantities, the second focuses on tools needed to verify the records already available in the electronic-structure databases.

The first systematic effort of verification in the world of solid-state calculations has been performed by Lejaeghere *et al* [1], which assessed the reproducibility of DFT calculations of elemental solids across a variety of electronic-structure codes and different libraries of pseudo-potentials. It is noteworthy that this was done more than 50 years after the introduction of DFT; the computational chemistry community started such an effort much earlier. The variety of the computed properties available in electronic structure databases, however, complicates this task. The total energy is the least sensitive quantity to numerical errors, thanks to the variational principles, and it has already been pointed out that, for example, verified band structures may not translate to agreement on vibrational properties [25] and that a more careful analysis

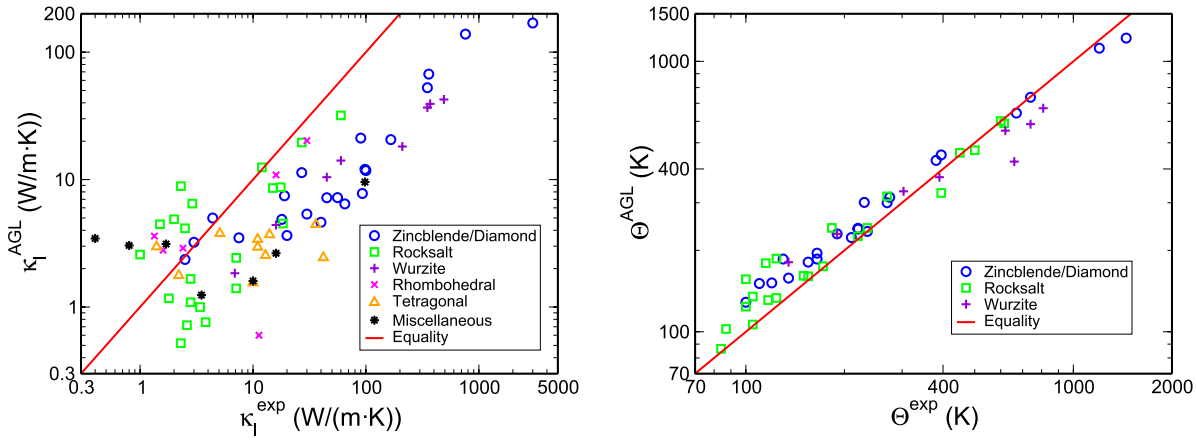


Figure 6. Validation study of the AGL computational method [27] applied to the lattice thermal conductivity and the Debye temperature. Theoretical predictions of simpler harmonic properties, such as the Debye temperature, have larger predictive values. The calculations of thermal transport coefficients that must include accurate treatment of the anharmonic contribution provide more scattered results. Reproduced with permission from Cormac Toher.

must be conducted at least when using density-functional perturbation theory [26].

Tools to retrieve/compare/assess records in selected electronic structure data repositories are often available as a dedicated REST application program interface (API); however, only recently were search APIs made available to perform preliminary verification tasks on a large set of data. Rose *et al* [27] have used their search API to analyse the convergence of variable cell relaxation for all the structures contained in the ICSD.

Validating theoretical predictions requires synergy with experimental databases and, although it has been commonly done on single materials, only a few examples of systematic comparison are available. Aiming to validate DFT formation energies, Kirklin *et al* [28] have found agreement in 86% of the 89 cases considered. Toher *et al* [29] investigated thermo-mechanical properties, such as bulk and shear moduli, Debye temperatures, and Gruneisen parameters and assessed the reliability of different computational approaches compared with 74 experimentally characterized systems (see figure 6).

Advances in science and technology to meet challenges. Several major issues remain: (1) establishing transferable standards and protocols to assess the predictive value of electronic-structure data, (2) further development and implementation of software tools for automatic verification, (3) establishing community test cases ([30] and https://galligroup.uchicago.edu/Research/hybrid_functionals.php#tables) that contribute to develop high-quality standard datasets, (4) building synergies between theoretical and experimental databases for validation, (5) expanding the number of computed quantities in order to facilitate direct comparison with experiments, (6) defining and computing universally 'reliability scores' to provide direct information regarding data quality to database users.

A recent and notable effort has been that of the OptiMade API, which intends to add a compatibility layer to access data from different repositories. Such an effort is still in progress but could potentially help to address some of the difficulties in verifying electronic-structure data. The definition of validation protocols is even more difficult, since it involves a much more diverse universe of techniques, formats, and details.

Concluding remarks. Several independent repositories of electronic structure data based on DFT are currently publicly accessible. The data contains millions of computed properties that can be used for machine learning and more. Verification standards, however, are missing, with limited curation of data; validation has been rarely addressed and typically on very small subsets. This rapidly growing area of research dedicated to verification and validation must be expanded, aiming at community definitions of accepted standards for accuracy, and well-defined protocols and tools for the calculations.

Acknowledgments

M F acknowledges collaboration with the AFLOW Consortium (www.aflow.org) under the sponsorship of DOD-ONR (Grants N000141310635 and N000141512266). N M acknowledges support from the Swiss National Centre of Computational Design and Discovery of Novel Materials NCCR MARVEL.

5. High-throughput (combinatorial) experimental methods for materials design/discovery

Ichiro Takeuchi[1] *and Martin L Green*[2]

[1] University of Maryland, College Park, MD, United States of America
[2] Materials Measurement Science Division, National Institute of Standards and Technology, Gaithersburg, MD, United States of America

Status. The high-throughput (combinatorial) approach to materials discovery enables synthesis and screening of a large number of different alloys or compounds simultaneously. Early incarnations of the high-throughput strategy appeared in the 1960s where co-deposition of thin films was used to generate composition spreads of ternary metallic alloys and functional materials, such as ferromagnets and luminescent materials. Despite some success, early efforts did not lead to widespread adoption largely because of the lack of tools for rapid characterization, as well as computers and automated measurement techniques. High-throughput materials exploration truly came of age in the early 1990s following the advent of combinatorial chemistry in biochemical fields and in the pharmaceutical arena [31]. The early 90s was also the era when the need for rapid, systematic investigation of new materials was first recognized by the materials science community, following the discovery of high temperature superconductors. Since then, the combinatorial approach has become an accepted methodology in almost all areas of materials science [32]. Combinatorial catalysis is a large field practiced by academic and industrial labs alike in tackling a variety of homogeneous and heterogeneous catalytic reactions with applications in production of chemicals, refinery operations and environmental protection [33]. Polymeric materials also represent a vast target, including formulations for coating, tissue engineering, and polymerization catalysts [34]. While stoichiometric control and the search for new compositions with enhanced physical properties is the most common mode of combinatorial investigation, microstructure and processing control through fine-tuning of myriad synthesis parameters is equally important. In the arena of functional materials, the investigation has been increasingly focused on energy-related materials, such as battery electrodes, fuel cell electrolytes, photovoltaic materials, and thermoelectric materials [35].

Recent advances in computational materials science and data science are an exciting development. Integration of computational and theoretical predictions of materials with the experimental combinatorial approach can signal a new chapter in materials discovery, and such efforts are underway in multiple fronts.

Current and future challenges. The history of the combinatorial approach is paved with a series of technical challenges that the community has endured over the years. In the early days, the synthesis posed the initial test: is it really possible to make hundreds to thousands of compositionally varying samples in a single experiment in a controlled manner? The answer depends on the topic, synthesis technique, and the extent of composition variation one attempts to map on a given library. While it is enticing to apply the approach to the latest exotic and exciting topics, the cardinal rule is that one needs to be able to reliably synthesize the correct benchmark composition on a corner of the combinatorial library. To this end, one needs to critically evaluate the library design taking into account simultaneously achievable ranges of synthesis parameters (composition, temperature, atmosphere, etc) on a library.

The second set of challenges was in the form of high-throughput screening tools [36]. Because the high-throughput methodology presented a new way to measure materials properties, it often required a major instrumentation effort to develop new tools for local, rapid and accurate characterization on libraries comprised of small quantities of materials. Techniques based on scanning probe microscopy (SPM) have been effective, and MEMS and electronic device array configurations have also proven to be powerful platforms [37]. For instance, SPM techniques have been used to measure ferroelectric properties, magnetic properties, and piezoelectric properties. Micromachined cantilever arrays have been used for high-throughput detection of martensitic transformation for shape memory alloys, hydrogen storage materials, and magnetostrictive materials. While some properties, such as electrochemical catalysis and the latent heat in caloric materials, are inherently difficult to quantitatively capture by high-throughput experimentation, researchers have made great strides in streamlining the screening techniques of virtually all physical and chemical properties [35].

Various types of measurement data, generated from library characterization in large quantities, have always presented a challenge for the community. In recent years, the issue of how to manage (curate and analyze) large, heterogeneous data sets has come to the fore. Some national laboratories have taken on this challenge and have successfully set up curated databases for high-throughput experimentation. Good examples are the online data handling systems developed at NIST (https://mgi.nist.gov/materials-data-curation-system) and NREL (https://htem.nrel.gov/). Given that there are also now enormous amounts of computed materials properties available from theoretical work, the situation calls for an integrated approach to designing theory-guided combinatorial experiments and performing holistic data processing and mining.

Advances in science and technology to meet challenges. Effective integration of experimental and computational high-throughput approaches can serve as an engine to drive materials discovery in a variety of fields. In order for the integration to be seamless, frequent feedback loops are needed between theory and experiment (figure 7). Combinatorial experiments can be used to rapidly validate theoretical predictions of new compounds within targeted yet broad composition ranges. Experimental data, systematically generated from libraries, can in turn be used to build new theoretical models for further predictions. It would be ideal to have such an integrated engine on a flexible data-handling platform, which includes a repository containing both experimental and computational data. It is also important that the data-handling

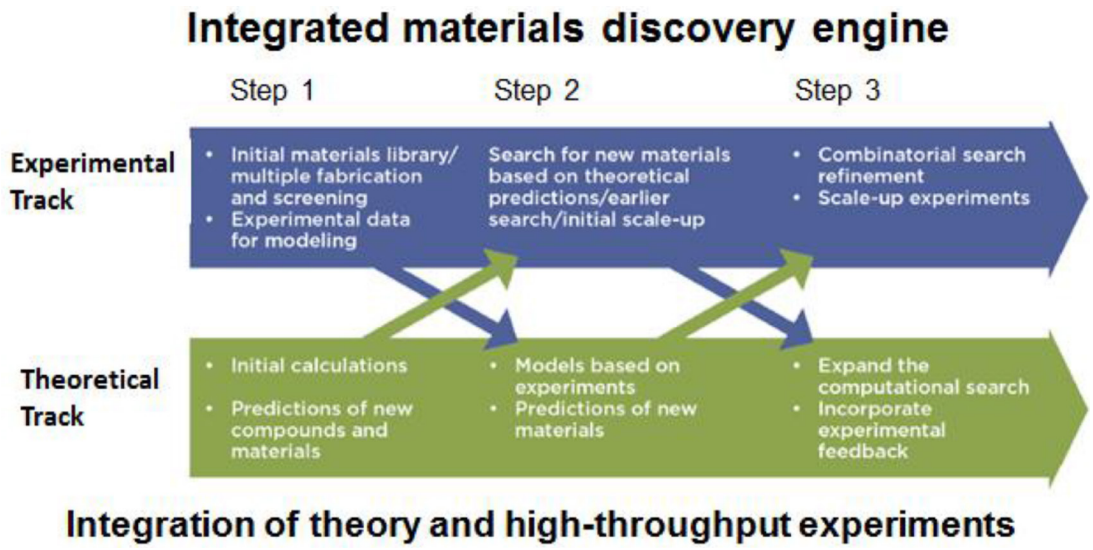


Figure 7. Integration of high-throughput experimentation and theory. Effective coupling of the experimental and theoretical tracks, both carried out in high-throughput manners, can facilitate materials discovery. The key is to have as many connection points between the two tracks as possible. We call this coordinated effort the 'Integrated Materials Discovery Engine'.

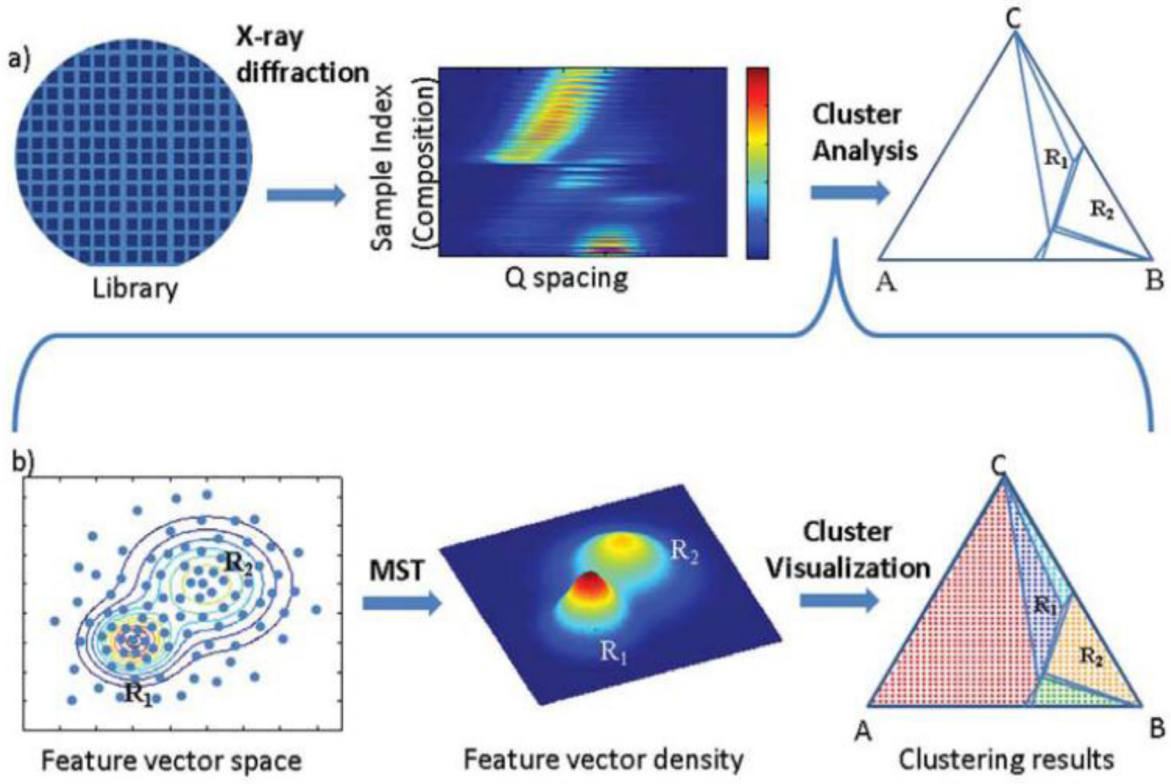


Figure 8. Combinatorial experimentation and data analysis flow. (a) X-ray diffraction data are taken from a thin-film composition spread wafer mapping a ternary (A–B–C) compositional phase diagram created by co-sputtering. The diffraction data are then analyzed using cluster analysis to produce a potential structural phase distribution diagram, identifying separated phase regions. (b) The mean shift theory (MST) as the machine learning algorithm is applied here: feature vectors are produced for each sample on a combinatorial library. Each sample is projected into the feature vector space—shown here as 2D and unitless for ease of visualization, and the feature vector density is correlated to an underlying probability density function (PDF) for each 'hidden' classification, which in this case are assumed to be two separated different phase regions R1 and R2. PDF analysis is performed using MST-based mode detection, and all samples from the same PDF are clustered together. Reprinted by permission from Macmillan Publishers Ltd: Scientific Reports [39], Copyright 2014.

platform has access to a variety of existing literature databases. The goal is to carry out datamining on such legacy databases to help delineate composition-structure-property relationships, as well as to derive models for predicting new compounds which can serve as the basis for designing new library experiments [38].

As discussed elsewhere in this article, it is increasingly becoming clear that machine learning can play a major role in several aspects of this endeavour. Because combinatorial experimentation can generate a large amount of data from a single library, researchers have been relying on machine learning to quickly decipher the underlying trends in complex sets of data. For instance, unsupervised machine learning can be used to rapidly separate a large number of diffraction patterns into different clusters (figure 8). For a composition spread library, the clusters nominally correspond to regions of the same crystal structures [39]. Machine learning is also actively used to streamline the efforts in computational materials science. Here, the goal is to quickly identify proxy descriptors to simplify the calculations and minimize computational time and resources. In this manner, machine learning can be used to curtail the amount of expensive and time-consuming *ab initio* simulations, which need to be carried out for a project. There are also proposed efforts to use machine reading and machine vision to comb through volumes of journal articles in order to automatically build databases based on previous literature. Proper threading of the results of the various data-centric tasks is then crucial for effective operation of the integrated engine for materials discovery.

Concluding remarks. High throughput (combinatorial) materials science started as a natural extension of developments in the pharmaceutical industry, but it has evolved into a versatile approach applicable to a wide variety of materials systems. Because any materials design project requires actual materials synthesis and validation, high-throughput experimentation serves as a *sine qua non* of any systematic materials discovery and development effort [40]. Moving forward, the key to continued success of the integration of the high-throughput experimentation and theory is to close the gap through data-driven activities.

Acknowledgments

We acknowledge many past and present colleagues we have worked with over the years. In particular, we acknowledge A Gilad Kusne and Jason Hattrick-Simpers for many useful discussions. I T is funded by ONR N00014-13-1-0635.

6. Advanced *in situ* and synchrotron based methods for materials design/discovery

Mercouri Kanatzidis[1] *and Mike Toney*[2]

[1] Northwestern University, Evanston, IL, United States of America
[2] SLAC, Menlo Park, CA, United States of America

Status. Exploratory synthesis has been a key strategy in the past several decades that has yielded many of the important new materials we study and use today. The level of predictability in this admittedly highly successful approach is generally low because the reaction mechanisms, as well as how and when phases form, is not known and are challenging to predict within the present theoretical frameworks. In this successful 'Edisonian' paradigm, one predetermines a given set of reaction conditions (e.g. time, temperature, heating, cooling rates) and waits for completion to isolate and identify any formed compounds. There is a general lack of awareness ('blind synthesis') of what has occurred during the reaction and when phases form and this hinders our ability to identify and make new materials or to devise successful synthetic processes for desired and targeted materials. This is particularly pronounced for synthesis of metastable compounds which often have very desirable functionality, since such phases often appear transiently and unpredictably during a synthesis. As a result, the design and synthesis of metastable materials remains largely empirical.

Current and future challenges. Recently, a number of new *in situ* based approaches that allow us to 'see' *all forming phases* in the course of a variety of synthesis reactions have become of interest. The *in situ* approach uses x-ray diffraction to monitor the reaction to capture signatures of new phases as they form even when they are transient and short lived. The results published so far are very encouraging because entire new phases have been observed to form on reactions which had been missed in previous investigations on the same system. Because all crystalline phases are revealed during the reaction in this approach, we call it '*panoramic synthesis*'. For example, this approach has been used for flux reactions, hydrothermal growth, and nanoparticle formation [41, 42]. Along with these experimental developments, the theory is making rapid progress in advancing computationally-driven predictive synthesis of inorganic materials, through concepts such as remnant metastability (i.e. during synthesis, metastable phases that form as end products are remnants of phases that were thermodynamically stable during particle growth) [43].

In the future, it will be a crucial challenge to implement complementary *panoramic synthesis* experimental probes (such as Raman spectroscopy, total x-ray scattering and x-ray absorption spectroscopy) that can 'see' amorphous phases which can form important intermediates during synthesis. Likewise, it will be important to augment the structure-based (diffraction) measurements with local imaging, such as electron microscopies (see [44] for an example). These efforts should be carefully compared and integrated into computationally-driven predictive synthesis both to test and refine these theories. Finally, expanding *panoramic synthesis* into other spaces (e.g. electrodeposition, hydro- and solvento-thermal synthesis [45, 46], high pressure synthesis, and rapid thermal processing [47]) will broaden the applicability.

Advances in science and technology to meet challenges. Here, we give two short examples of recent advances in the application of *panoramic synthesis* to illustrate the advances that can address the challenges for advanced experimental methods for materials discovery and design.

The first illustration of the power of panoramic synthesis was in the systems K–Cu–S and K–Sn–S. A schematic of the *in situ* capillary furnace we designed to investigate phase formation during flux reactions is shown in figure 9. This experiment generates x-ray diffraction maps that reveal the complex real-time phase relationships in the reaction [48]. These experiments revealed surprisingly more phases that had been found in conventional reactions. The diffraction patterns collected while heating and cooling during this reaction are given in figure 9(a). The panoramic synthesis showed additional crystalline phases that formed and then disappeared by the end of the reaction [48, 49].

First, we see the signatures of the reagent metal and polysulfide phases, but upon heating, low-2θ peaks appear in the diffraction data (red region). This real-time information (prior to any analysis) clearly shows that ternary K–Cu–S phases form early in the reaction, as observed in figure 9(a). Continued heating leads to the disappearance of all Bragg peaks (the black region in figure 9(a)). At this point, the formed ternary sulfides have dissolved completely into the molten polysulfide salt. After cooling, low angle Bragg peaks again showed the presence of ternary phases (green region in figure 9(a)). The structures of the occurring phases are closely related and shown in figure 9(b).

The *in-situ* monitoring in the reactions of Cu with K_2S_3 and Cu with K_2S_5 produced not only different phases $K_3Cu_8S_6$, KCu_3S_2 and $K_3Cu_4S_4$, but also generated key information of when they formed during the course of the reaction, how long they lasted and what the final product was. If this reaction were to be performed *ex-situ*, no evidence of the formation of KCu_3S_2 and $K_3Cu_4S_4$ would exist—only $K_3Cu_8S_6$ would remain [50].

Another successful example is the new phases discovered using *in-situ* synchrotron x-ray diffraction studies in the Cs/Sn/P/Se system [50]. The diffraction data was translated into phase fraction versus temperature. Seven known crystalline phases were observed to form on warming in the experiment: Sn, Cs_2Se_3, Cs_4Se_{16}, Cs_2Se_5, $Cs_2Sn_2Se_6$, $Cs_4P_2Se_9$, and $Cs_2P_2Se_8$. Six unknown phases were also detected; using the *in-situ* x-ray data as a guide, three of them were isolated and characterized *ex-situ*. These are $Cs_4Sn(P_2Se_6)_2$, α-$Cs_2SnP_2Se_6$, and $Cs_4(Sn_3Se_8)[Sn(P_2Se_6)]_2$. $Cs_4(Sn_3Se_8)[Sn(P_2Se_6)]_2$ is a 2D compound that behaves as an n-type doped semiconductor below 50 K and acts more like a semimetal at higher temperatures.

A second illustration relates to the development of Pt nanostructures and shows the power of complementary

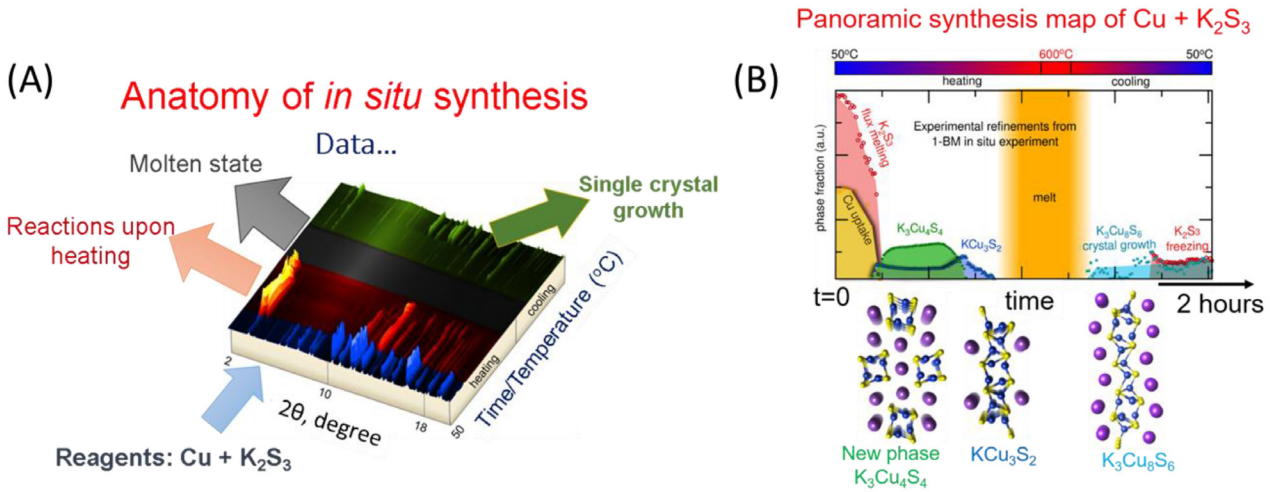


Figure 9. (a) Phase formation revealed in during *in situ* reaction monitoring using x-ray diffraction. These data maps of the molten and crystalline regions can be constructed to show all crystalline phases formed. (b) Panoramic map of the Cu-K_2S_3 reaction. New phases form on heating, with different phases forming upon cooling. Similarities between phases imply the ability to tailor linkages and topology.

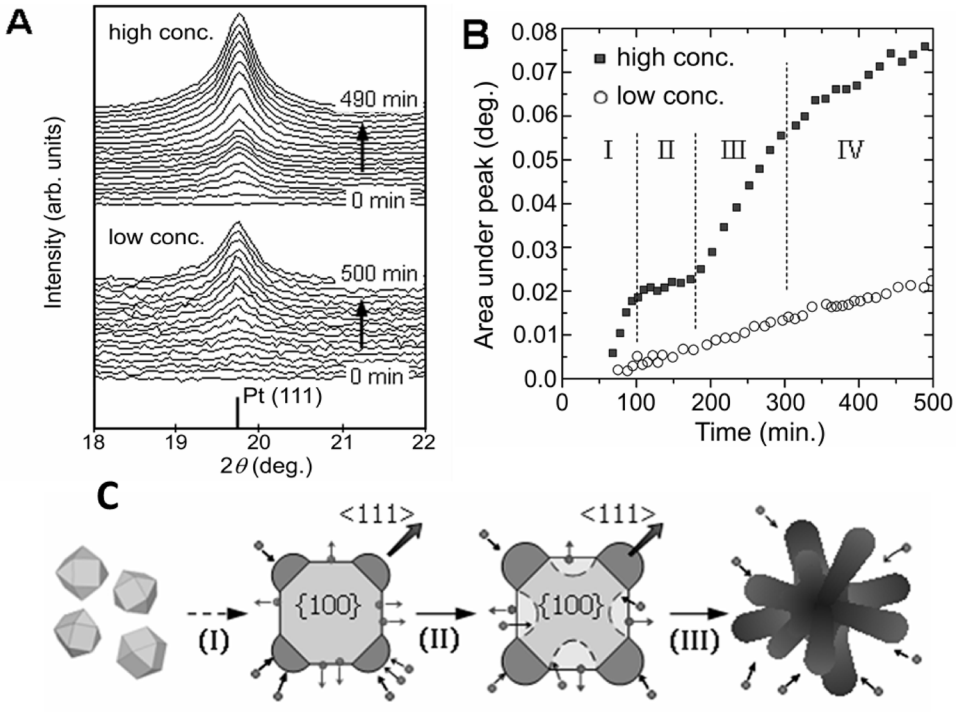


Figure 10. *In-situ* synthesis of Pt nanostructures (70 °C, 200 kPa H_2) at low and high concentrations of stabilizing agent. (a) Time-resolved x-ray diffraction; (b) time evolution of area under the Pt(1 1 1) peak with growth stages I–IV denoted. (c) Schematic illustration of proposed Pt nanoparticle growth mechanism for stages (I–III). Pt monomers first nucleate into cuboctahedral nuclei (I), and then grow into single-crystalline quasi-octapods (II). Growth of the arms of the quasi-octapods, coupled with selective etching on the edges and centers of facets, leads to the formation of etched-octapods (II). These processes continue and transform the nanocrystals to porous nanocrystals.

techniques [44]. The synthesis involved thermal decomposition of a Pt precursor under a hydrogen atmosphere in the presence of a stabilizing agent (at low and high concentrations). Figure 10(a) shows the development of the Pt(1 1 1) x-ray diffraction peak area with growth time; for the low concentration reaction, this increases approximately linearly. However, for the higher concentration these data are more complex and allow for the identification of four distinct growth regimes (labeled I–IV). *Ex-situ* transmission electron microscopy (TEM) of Pt nanoparticles allow identification of the nanoparticles shapes and size, showing a fascinating evolution from a compact cubic morphology (I) to quasi-octapods (II), etched-octapods (III) to porous nanocrystals (IV). From the diffraction and TEM, a detailed growth model is developed as illustrated in figure 10(c). This example demonstrates the utility and complementarity of TEM and XRD for revealing nanostructure growth mechanisms. A future challenge is to develop a predictive capability for

nanostructure synthesis—not only phase but also particle shape and size.

Concluding remarks. Progress has been made over the past decade in developing a rational, predictive understanding of exploratory synthesis, but much remains to be accomplished to enable extensive applications of this approach. The challenge of accelerated and predictable materials discovery will be met with increasing success if we can achieve the organization of new knowledge coming from these new approaches so it can be more effectively taken advantage of. For example, reactivity patterns under specific reactions conditions may have a general scope and could be used to classify reaction and reaction types so that they can be used as synthetic tools for materials discovery. We have described an *in-situ*, '*panoramic synthesis*' approach that, when coupled with progress in computational predictive synthesis, will help enable the widespread adaptation of predictive synthesis. This will profoundly accelerate the discovery and development of new functional materials.

Acknowledgments

M G K gratefully acknowledges support from the National Science Foundation Grant DMR- 1708254, and MFT gratefully acknowledges support from the Center for Next Generation Materials by Design: Incorporating Metastability, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences under Award # DE-AC36-08GO28308.

7. Materials informatics

Sergiy Butenko[1] *and Bryce Meredig*[2]

[1] Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, United States of America
[2] Citrine Informatics, Redwood City, CA, United States of America

Status. The objective of materials informatics [51], or data-driven materials science, is to use a set of powerful tools from data mining, machine learning, and mathematical optimization to systematically reveal materials processing-structure-property-performance (PSPP) relations. Once uncovered, these PSPP relations can drive the predictive discovery and design of novel materials and optimized manufacturing processes.

The shift toward data-driven discovery is becoming broadly prevalent in modern research and is referred to as the fourth paradigm of science [52]. This term, coined by Jim Gray in 2007, reflects the historical developments in scientific methods, beginning with empirical science (first paradigm), giving a rise to theoretical science (second paradigm), enabling computational science (third paradigm), and ultimately paving the way for data-driven science. While fields such as biology have embraced the fourth paradigm for some time, it is a comparatively new concept in materials science [53].

The rapid ascent of materials informatics coincides with the 2011 launch of the US Materials Genome Initiative (MGI) [54], which explicitly elevated the role of digital data and related software tools in the materials research enterprise. Since the MGI announcement, materials informatics have driven a series of laboratory discoveries of materials and processing routes, in areas ranging from thermoelectrics [55] to hydrothermal synthesis [56]. Further, the community has seen a rapid increase in research articles wherein various materials informatics-based models of PSPP relations are constructed.

A generic materials informatics workflow is shown in figure 11. The analysis starts with data extraction and preprocessing, which is used to identify and select the key components of the data set. The reduced data set is further examined for relationships between the components of interest. The discovered relationships are utilized to generate the so-called inverse and forward models, the former of which can be used to design materials with desired properties, whereas the latter are used for predictive analytics [53]. Experiments and computer simulations based on theoretical models are used to generate new data for the materials databases, thus closing the loop.

Current and future challenges. Below, we describe five key challenges that hinder broader application of materials informatics.

Data heterogeneity and siloing. The datasets characterizing materials and their properties are of a diverse nature, come from a wide variety of sources (e.g. myriad different experimental and simulation techniques) with different levels of accessibility, and are stored in many formats. Materials data tend to exist across many scattered 'small data' silos, making systematic mining more difficult.

Lack of consistent metadata. The generation and collection of materials data are associated with numerous uncertainties and sources of error that may not be easily detectable, making the quality of data difficult to verify. This issue is frequently exacerbated by a lack of metadata necessary for precisely replicating the experiment or a simulation used to obtain the data [57].

Inverse materials PSPP models to search materials design space. Forward models try to predict the structure of a material based on the processing used, properties based on the structure, and performance based on properties. On the other hand, inverse models aim to determine the material design parameters that would yield materials with desired properties and performance. In general, the forward modelling problem is easier than the inverse problem, yet the inverse problem is more relevant for materials discovery.

Novel representations of materials for informatics applications. Representing materials concepts (e.g. crystal structure, chemical composition, or microstructure) as computational objects for input to analysis algorithms is an essential prerequisite for materials informatics. An example emerging representation strategy is describing materials as networks, which could reveal relationships and connections between materials and potentially identify multiple materials that have the same or similar properties or are otherwise related according to some criteria.

Advances in science and technology to meet challenges. Below, we highlight two particular science and technology development goals that promise to be fruitful areas of exploration for the materials informatics community.

Explicit integration of experimental data, computational data, and materials theory to enable multiscale modelling. In the well-established integrated computational materials engineering framework, powerful *individual* PSPP models have been developed over time, but deep integration is lacking between these models, experimental results, and established theory. Materials informatics are a promising integrator of these varied sources of 'signal' on the behavior of materials. This capability is especially important given the distribution of materials data across many small, isolated data silos as described above.

Similarity metrics for materials. One of the central scientific questions arising in materials informatics is a systematic way of determining quantitative metrics characterizing the level of similarity between pairs of materials. Addressing this question would help advancing methods for understanding PSPP relations and would enable the use of network analysis techniques for exploring both local and global properties of systems of materials. The metrics of interest could potentially be derived from first-principles computations, e.g. based on electronic density of states or projections of localized basis sets [58, 59], or atomic coordination environments [60]. Then one could represent the entire space of materials as an extended network of interacting entities, where the connections between individual materials are based on pairwise similarities in their properties derived from first principles. This would allow us to take advantage of powerful network analysis methods, which exploit the

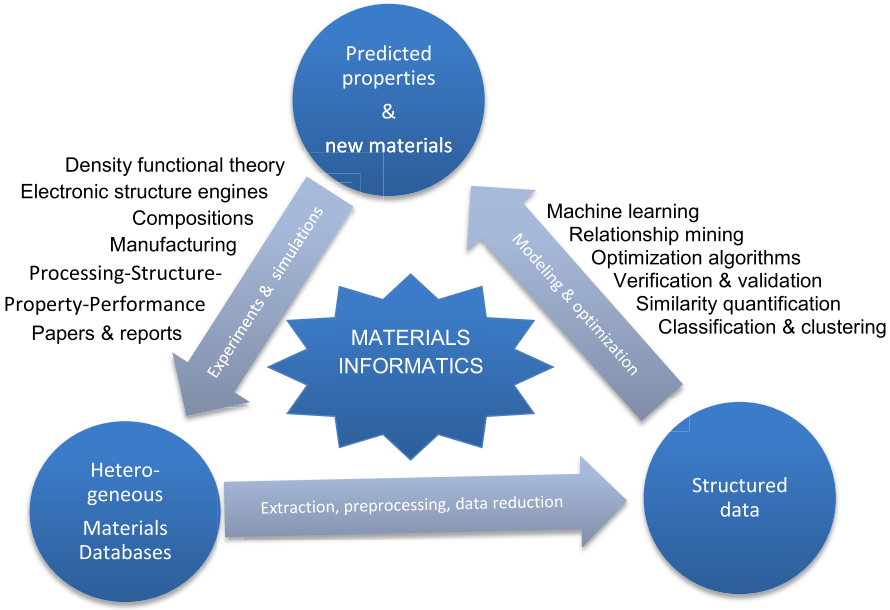


Figure 11. Generic materials informatics workflow.

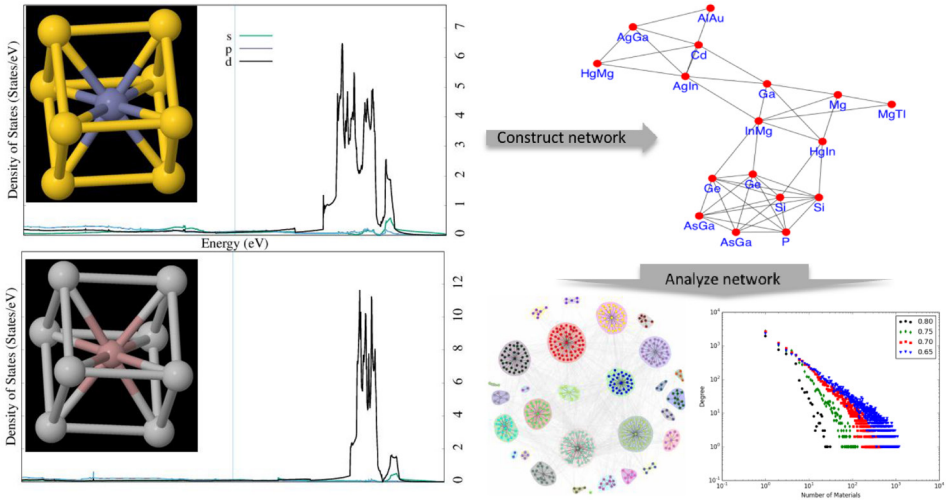


Figure 12. Network analysis approach in materials informatics. DoS functional are used to define a similarity metric (left), a network is constructed based on the considered similarity metric (top right), and structural properties of the resulting network are analysed (bottom right).

use of graphs or networks as a convenient tool for modelling relations in large datasets. In this general framework, certain elements of a dataset of interest are thought of as vertices, and the pairwise relations between different elements are described by edges, yielding a network representation of the underlying complex system. With respect to the materials databases, several conceptually different network representations could be utilized, which would provide alternative vantage points for exploring myriad materials data from a systems perspective. For example, clusters in these networks would correspond to materials with similar properties. Furthermore, one could develop optimization models aiming to find the best subsets of materials according to a given objective function (see figure 12).

Concluding remarks. Materials informatics are a key enabler of the MGI, as well as related international efforts such as Japan's Materials Research by Information Integration Initiative (MI^2I), and the faster development of higher-performance materials. The focus of this emerging field is on algorithmic approaches that would advance our understanding of processing-structure-property-performance relations. Developing a cross-disciplinary collaborative culture that would allow integrating the experimental, computational and applied sides of materials science in developing advanced data mining solutions is essential. Defining reasonable quantitative similarity metrics for pairs of materials could lead to significant advances in classification of materials and navigating the ever-expanding search space for new materials. The reader is referred to recent survey articles [51, 61] for further information on advances and challenges in materials informatics.

Acknowledgments

The first author acknowledges the support by DOD-ONR (N00014-13-1-0635) grant.

8. Computational prediction and experimental realization of new materials

A Zakutayev and S Lany

National Renewable Energy Laboratory, Golden, CO 80401, United States of America

Status. The discovery of novel materials and the control of their properties are key drivers for technological innovations. This observation is particularly true for electronic and optoelectronic materials, which have fueled the information technology revolution, and on which the hopes for the advanced energy revolution rest. Historically, materials discovery has been a serendipitous endeavor. For the past century, materials chemists have been synthesizing numerous solid-state compounds for different reasons and at different times. Their findings are documented in crystallographic databases such as the Inorganic Crystal Structure Database (ICSD), and that of the International Centre for Diffraction Data (ICDD). While the entries in these databases count in the hundreds of thousands, surprisingly little other than the crystal structure is known for most of these materials. Starting from the crystal structures as input, high-throughput first principles calculations based on density functional theory (DFT) and post-DFT methods provide an ever-increasing number of calculated properties, made available in online databases like https://materialsproject.org/, http://www.aflowlib.org/, http://oqmd.org/, https://materials.nrel.gov/, and others. High throughout experimental materials property databases are also emerging (e.g. http://htem.nrel.gov/)

While extensive, the crystallographic databases are by no means complete. The availability of synthesis methods and preferences of researchers and funding agencies have emphasized some chemical spaces over the others, leaving white spots where plausible materials may exist but are not presently reported. To unearth these 'missing materials', computational searches are now being performed to predict their structure, and accompanying experimental efforts are underway to either verify or falsify their stability. Several broad conclusions can be drawn from such studies. First, there is no doubt that the search space is vast, considering the combinatorial explosion of candidate materials with the number of involved elements and their possible ratios. This is especially true when including metastable structures and non-stoichiometric compositions in materials search. Taking into account this second point, it also becomes increasingly clear that the 'convex hull' criterion (thermodynamic stability with respect to other structures and compositions made of the same elements) is too narrow to judge whether a potential new material would be possible to synthesize. These conclusions reflect the challenges faced by materials discovery discussed next.

Current and future challenges. The biggest current challenge in systematic materials discovery is the vastness of chemical space where materials can occur. In general, a 'material' is defined by its constituent chemical elements, their relative composition (stoichiometry), and the atomic structure, which

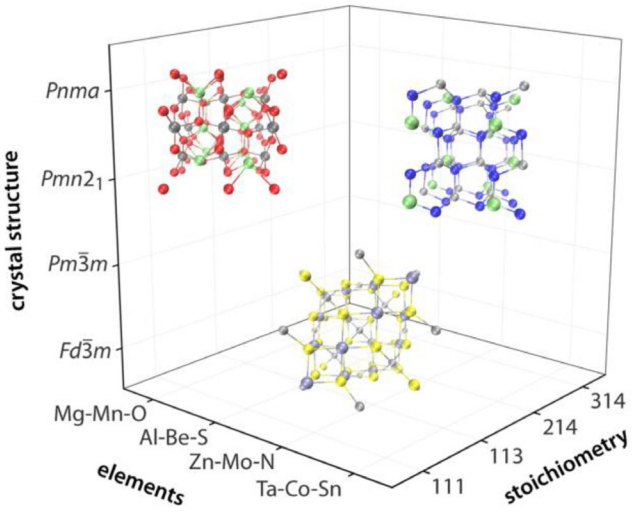


Figure 13. Simplified illustration of the vast multi-dimensional materials discovery space, showing different possible elements, stoichiometries, and structures.

can be depicted in 3D as shown in figure 13. Numerous approaches and tools to predict crystal structures from first principles are available [62], but they are often limited to the materials with a small number of elements, to formula units with small integer indices, and to unit cells with a small number of formula units. In fact, stability and properties of real materials often depend on the non-ideal structures that cannot be described by their primitive cells, such as defects or disorder, and, ultimately, the meso- and microstructure. One theoretical approach to screen for many possible elements is to constrain the search to one or a few chemical stoichiometries (e.g. ABX, ABX_3, A_2BX_4, and so forth), and restrict the possible structures to all known prototypes (e.g. spinel, olivine, etc, for A_2BX_4) [63]. An alternative approach is to select a constrained number of elements and structures, and then screen many possible low index stoichiometries [64]. To aid both of these approaches, simplified stability descriptors [65] can help to identify search spaces where new materials are likely to be discovered.

One of the biggest future challenges in materials discovery is to go beyond the search for thermodynamic ground state compounds. Metastability [43] comes in many flavors, including polymorphs, thermochemically unstable materials, solid solutions, non-stoichiometric compounds, hierarchical and low-dimensional materials. Unlike the case of ground states, which are universally defined by free energy minimization, computational discovery of metastable materials can no longer be agnostic to the synthesis approach (figure 14). Thus, the synthesizability of the predicted candidate materials would have to be emphasized more; alternatively, materials searches should be tailored to the capabilities of specific synthesis approaches. For example, non-equilibrium synthesis of metastable heterostructural semiconductor alloys using physical vapor deposition methods can be enabled by novel phase diagram behavior that is not observed in conventional solid solutions [66]. Such materials discovery on a continuous composition scale is distinct from the more common search for

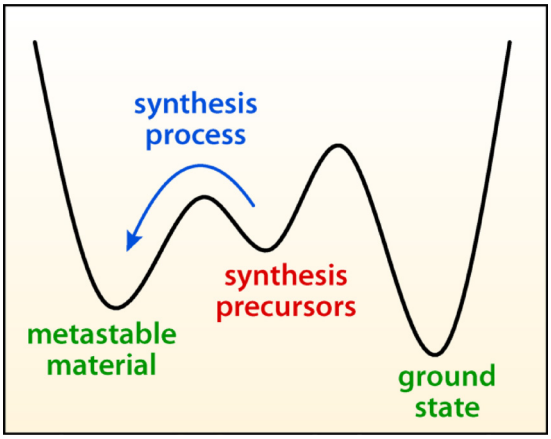


Figure 14. Schematic illustration of metastable materials and how their synthesis from higher energy precursors can be favored compared to ground state structures.

discrete stoichiometric compounds, and poses new challenges to computational prediction and experimental realization of new materials.

Advances in science and technology to meet challenges. High-throughput experimentation methods can be used to quickly cover both chemical space and process parameters. For example, the growth of sample libraries with continuous composition spreads and temperature gradients provide large amounts of synthesis data from a single deposition. In order to connect computational materials predictions to such non-equilibrium synthesis techniques, it may be possible to map process parameters onto 'effective' thermodynamic variables. For example, effective non-equilibrium chemical potentials accessible during the synthesis can be used to describe the formation of thermochemically metastable materials [67]. Also, finite temperature effects must be reconsidered in metastable materials. Atomic disorder induced by kinetic limitations during growth can be converted into an effective temperature [68], which can be much higher than the actual temperature. Therefore, such an effective temperature influence can be vastly stronger than the free energy contributions, e.g. due to atomic vibrations in the thermodynamic equilibrium state. Advancing the understanding of how these descriptors vary between different materials and synthesis parameters will enable the computational prediction of materials within their accessible range of effective thermodynamic variables.

The next step is to define effective kinetic variables that facilitate a predictive atomistic modelling of synthesis processes. Since metastable materials result from the inhibition of the equilibration of certain processes, modelling of synthesis requires identification of variables that describe appropriate kinetic constraints. For example, it is experimentally known that surface diffusion is usually faster than bulk diffusion for thin film growth. Creating the corresponding theoretical models tailored to this synthesis constraint is facilitated by developing problem-specific model Hamiltonians with similar (or ideally higher) accuracy as DFT, allowing an efficient Monte-Carlo or molecular dynamics sampling for specified non-equilibrium descriptors [69]. Experimentally, the development of new *in situ* techniques for monitoring synthesis and processing of materials would be an important advance for validation of computational models. Using synchrotron radiation, it can be shown that many new metastable phases can be present as reaction intermediates and absent from the reaction products [48]. The adoption of such *in situ* techniques on a smaller scale in research labs would therefore help to accelerate the discovery of metastable materials.

Concluding remarks. Materials discovery is branching out to capture the opportunities of a wide range of different synthesis approaches and their capabilities to access a spectrum of metastable materials. The definition of 'materials' being discovered is generalized beyond the Daltonian compositions and the corresponding crystallographic primitive cells. Future material discoveries will include metastable compounds, solid solutions, defect- and disorder-enabled materials, and low-dimensional structures. Furthermore, it remains a great challenge to concurrently discover new materials and design their properties. Predicting and synthesizing new materials is difficult enough that property calculations and measurements for the discovered new materials often come as an afterthought. The truly simultaneous search for new materials and their properties may be enabled by genetic algorithms and machine learning, if it is possible to train them to significant accuracy, and scale them to the vast chemical space of materials discovery.

Acknowledgments

Writing this chapter was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, as part of the Energy Frontier Research Center 'Center for Next Generation of Materials by Design: Incorporating Metastability' under Contract No. DE-AC36-08GO28308 to Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory.

9. Predicting synthesizability

Albert V Davydov and Ursula R Kattner

Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD, United States of America

Status. Advancements in multiscale multi-physics computational materials design have led to the accelerated discovery of advanced materials for energy, electronics and engineering applications [70]. For many common bulk materials, synthesizing and processing procedures are reasonably well established. This also applies to modelling tools that can be utilized for the understanding of phenomena occurring in these procedures. However, theoretical approaches have limited analytical power for predicting viable synthetic routes towards making entirely new materials. The knowledge about growth mechanisms, free-energy landscape and dynamics of chemical and physical processes during synthesis is quite limited. This uncertainty is exemplified in figure 15(a) by showing multiple pathways for crystallization from the solution, where a mechanism of forming bulk crystal depends on the interplay between thermodynamic and kinetic factors [71]. Therefore, the state-of-the-art in *materials design* needs to be complemented with substantial efforts in advancing the field of *synthesis design*. To increase the predictive ability of material synthesizability, it is necessary to define both equilibrium and out-of-equilibrium descriptors that control synthetic routes and outcomes. The key metrics include free-energy surfaces in multidimensional reaction variables space (e.g. activation energies for nucleation and formation of stable and metastable phases in figures 15(b) and (c)), composition, size and structure of the initial and emerging reactants, and various kinetic factors, such as diffusion rates of reactive species and the dynamics of their collision and aggregation.

Current and future challenges. To identify and quantify key descriptors towards predictable synthesis design, it is essential to integrate (i) exploratory synthesis and (ii) *in-situ* process monitoring with (iii) computational design of synthetic routes.

(i) Challenges of experimental exploratory synthesis are associated with the complexity of chemistries and reaction routes that depend on the interplay between equilibrium and out-of-equilibrium processes. Crystalline material growth methods, which span from condensed matter synthesis (all-solid-state synthesis and crystallization from melt or solution) to physical or chemical deposition from vapour (sputtering, e-beam deposition, pulsed laser deposition, atomic layer deposition, chemical vapour deposition), often proceed at non-equilibrium conditions, e.g. in highly supersaturated media, at ultra-high pressure, or at low temperature with suppressed species diffusion. Identification of chemical evolution reactions and the associated physical processes followed by their 'equilibrium versus metastable' classification is extremely difficult but is an essential step towards assessing material synthesizability. An illustration of possible reaction pathways to realize stable and metastable states of material is illustrated in figure 15(c), where highly non-equilibrium synthetic routes are superimposed on a generalized phase diagram [72].

(ii) Developing *in-situ* multi-probe measurements to capture important steps along the synthetic route is critical to make the synthesis design and its validation more efficient. For all-solid-state synthesis, it is important to develop high spatial and temporal resolution 3D tomographic mapping of phase evolution. The same applies for developing in-line diagnostics for solid growth under extreme environments, including synthesis in supercritical fluids, at extreme pressures, temperatures, photon/radiation fluxes or electromagnetic fields. This is noteworthy since real-time multi-probe diagnostics generating massive sets of data, which need to be promptly utilized in a closed-loop-feedback with synthesis, data curation protocols and machine learning techniques, need to be advanced.

(iii) On the modelling side, the idea of extending computational material discovery to *in-silico* synthesis design is still in its nascent state. Assessment of equilibrium and dynamic key variables for predicting the lowest activation energies and fastest routes for fabricating targeted material remains to be exceptionally challenging. The availability of data needed for modelling of new materials and processes poses another challenge.

Advances in science and technology to meet challenges. The challenge of operating in the multidimensional space of material fabrication can be addressed by integrating exploratory synthesis with multimode dynamic process monitoring to define key growth process parameters. Experimental synthesis and *in-situ* measurements should be further integrated with computational tools to enable robust predictive synthesis of materials with tailored properties. This unified 'experimental/*in-situ*/*in-silico*' synthesis concept is emphasized in the Department of Energy report [73] with a focus on materials for energy, including experimentally verified design of novel thermoelectric and battery materials, metal nanoparticle catalysts, and transparent conducting oxides.

To address emerging materials needs, exploratory synthesis is focusing more and more on metastable, hybrid, and hierarchical structures, such as thin film heterostructures, nanoparticle superlattices, and core–shell nanostructures. For example, the core–shell nanowire in figure 16(a) demonstrates how thermodynamically favoured phase separation in a GaAsSb alloy can be suppressed by strain from the GaAs shell layer [74]. Similarly, a metastable rock-salt structure in the SnSe thin film in figure 16(b) can be stabilized by depositing it epitaxially on a suitable substrate [75].

Advances in *in-situ* diagnostics include the application of multi-probe optical spectroscopies and neutron/x-ray scattering and diffraction for real-time process monitoring, e.g. for crystal growth from melt [42], roll-to-roll solution drying of organic photovoltaic films, solvothermal synthesis of metal-organic frameworks, etc. In addition, *in-situ* scanning probes

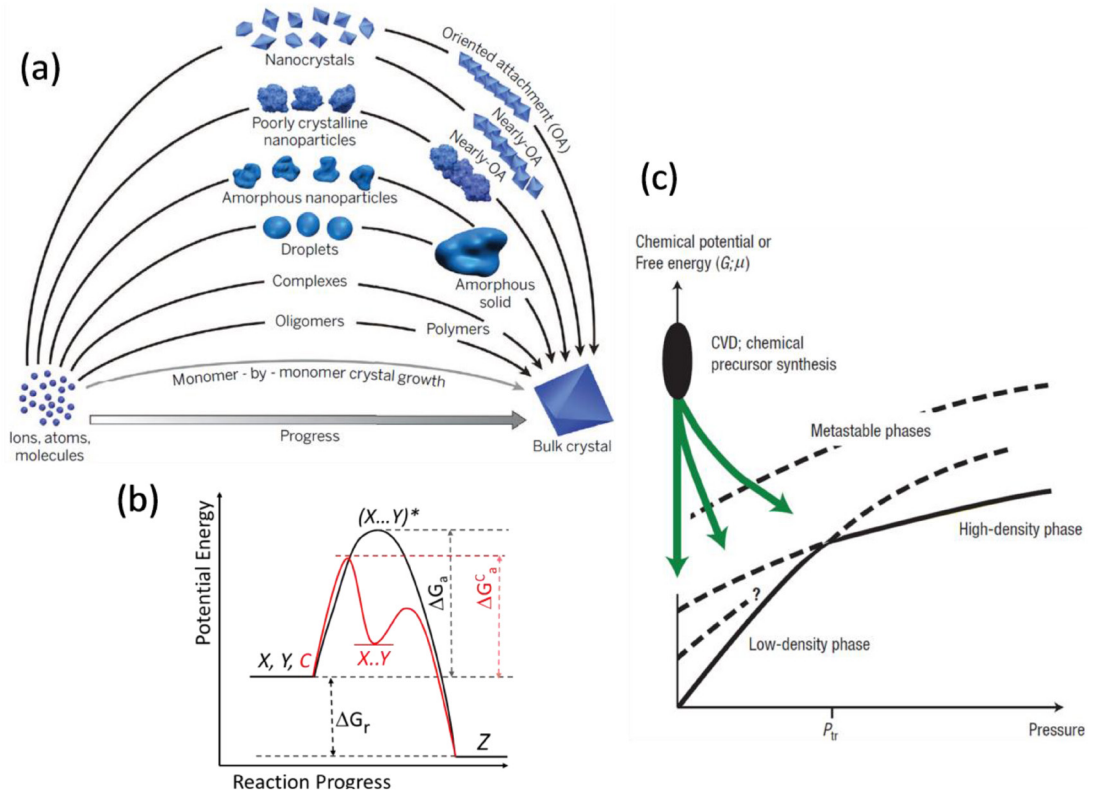


Figure 15. (a) Crystal growth model, 'Crystallization by Particle Attachment' (CPA), shows multiple pathways of crystal growth from solution. Unlike a classical monomer-by-monomer growth model (gray curve), CPA operates with higher-order species (black curves) and involves the interplay between thermodynamic factors and reaction dynamics. From [71]. Reprinted with permission from AAAS. (b) Potential energy profile from reactants X and Y to product Z with and without catalyst C. $(X\ldots Y)^*$ is high-energy transition state, $X...Y$ is a metastable product; ΔG_r is the Gibbs energy for the $X + Y \rightarrow Z$ reaction; ΔG_a^C and ΔG_a are energies for activating transition states with and without catalyst C, respectively. (c) Generalized free-energy—pressure phase diagram with superimposed synthetic routes (bold green arrows) for obtaining metastable phases. Reprinted by permission from Macmillan Publishers Ltd: Nature Materials [72], Copyright 2002.

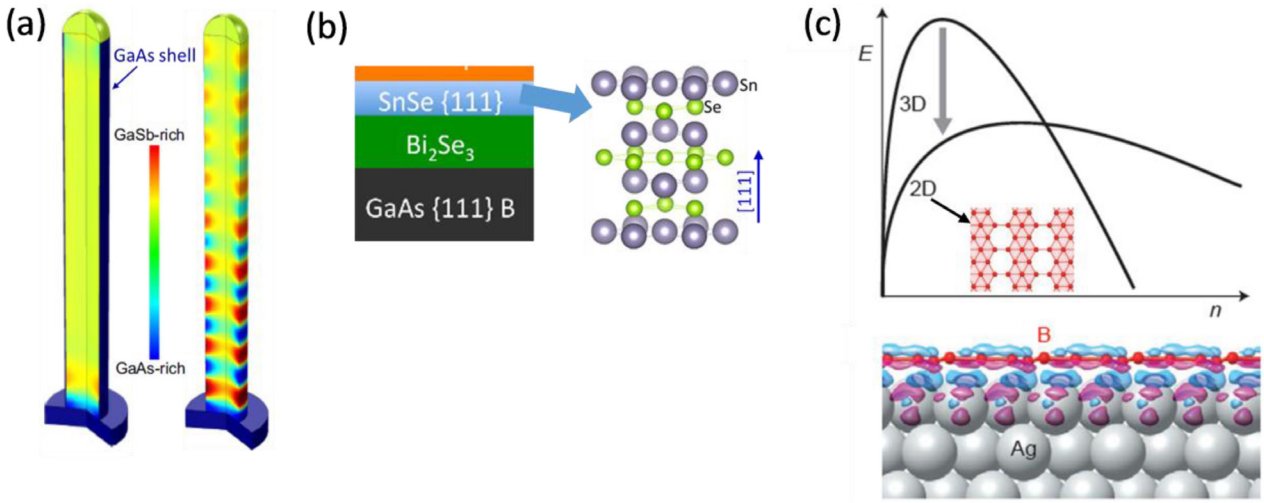


Figure 16. (a) GaAsSb semiconductor nanowire with (left) and without (right) GaAs shell. GaAs shell suppresses GaAsSb phase segregation, while the alloy without shell decomposes into GaSb-rich (red) and GaAs-rich (light blue) alternating segments. Reprinted from [74], Copyright 2017, with permission from Elsevier. (b) Crystal structure of topological insulator SnSe in its metastable rock-salt structure, stabilized by low-temperature molecular beam epitaxy on a GaAs substrate. Reproduced from [75]. CC BY 4.0. (c) (Top) The energies of the growing nuclei versus the number of atoms, E(n), show how the substrate steers the synthesis from the 3D towards 2D route by suppressing the nucleation barrier; (bottom) computed charge density shows how the Ag substrate donates electrons (from pink to blue) to the boron layer to stabilize its 2D structure. Reprinted by permission from Macmillan Publishers Ltd: Nature Chemistry [78], Copyright 2016.

and electron microscopies can provide direct insight into synthetic phenomena with atomic scale resolution [76].

Theory-guided data science has shown great potential for discovery and design in diverse scientific disciplines [77]. A recent example of theory-guided synthesis is shown in figure 16(c): *ab initio* modelling has predicted a new metastable allotrope of 2D boron, a.k.a. 'borophene', and suggested a synthetic route via epitaxial deposition on a metal substrate, which was subsequently validated by the experiment [78]. Efficient *in silico* synthesis of new materials requires the availability of data. A need for reliable data makes the integration of experiments, computation and theory imperative and machine learning and artificial intelligence methodologies will be needed to fill modelling and data gaps.

Concluding remarks. Even though the prediction of material synthesizability is an extremely challenging task, advances in modelling, *in-situ* measurements and increasing computational power will pave the way for it to become a reality. *In-silico* design of advanced materials will have to combine theory guided data science with statistical and theoretical computational methods. However, it is an open question whether it will be possible even with the most advanced modelling and simulation techniques to predict completely unknown pathways for synthesizability. For example, is an additional crystal growth route possible other than those shown in figure 15(a)? The development of techniques and tools to propose the most efficient synthetic pathways will remain one of the major challenges for predicting new material synthesizability.

Acknowledgments

The authors acknowledge the Material Genome Initiative funding allocated to the National Institute of Standards and Technology (NIST).

10. Thermoelectric materials discovery

Vladan Stevanovic and Eric S Toberer

Colorado School of Mines, Golden, CO, United States of America

Status. Achieving the widespread use of thermoelectric generators for direct heat-to-electricity power conversion critically relies on novel, better performing, and less costly thermoelectric materials [79, 80]. The vital role that new materials play is best witnessed by a recent, nearly three-fold improvement in the efficiency of thermoelectric generators spawned exclusively by the discovery of new materials classes (see figure 17) [81]. As a result, a new research paradigm emerged about a decade ago: computational screening of large chemical spaces in searching for new and even better thermoelectric materials. Following the pioneering work of Madsen [82], several groups made significant contributions to developing and applying computational tools to assess the transport properties of solids, both charge carrier and heat transport, in a manner amenable to high-throughput computational screening [83–86].

If judged by the number of new and experimentally validated candidate materials, the success of high-throughput searches has been limited so far. This is largely due to (i) the challenges associated with predicting transport properties of materials and (ii) the slow, serial nature of experimental validation. However, a few materials and material classes that have been successfully experimentally validated [79, 80] demonstrate the potential of computationally guided searches in advancing thermoelectric material discovery. These include materials previously not anticipated for thermoelectric performance (e.g. n-type Zintl compounds), suggesting the power of computation to lift us away from our assumptions. Calculations have also passed a critical milestone: we are now consistently able to retrospectively discover known materials without explicit experimental input. This success likely stems from the development of combined experimental and computational learning sets that are complementary in the properties they address.

To date, computational searches have predominantly considered known, previously synthesized materials (i.e. Inorganic Crystal Structure Database) with unknown charge carrier and phonon transport properties. Venturing into completely new material systems, including stoichiometric compounds and their alloys as well as the metastable structures, has yet to be done on a large scale, but the potential return on investment may be worth the effort.

Current and future challenges and opportunities. Similar to other material searches, the large search space size, coupled with the desire to accurately predict material properties, represents a significant challenge. As an illustration, figure 18 shows how only a very small fraction out of tens of thousands of known compounds have actually been experimentally characterized for thermoelectric performance. In combination with the complexity of the theory of transport phenomena and the required computational resources to *quantitatively predict*

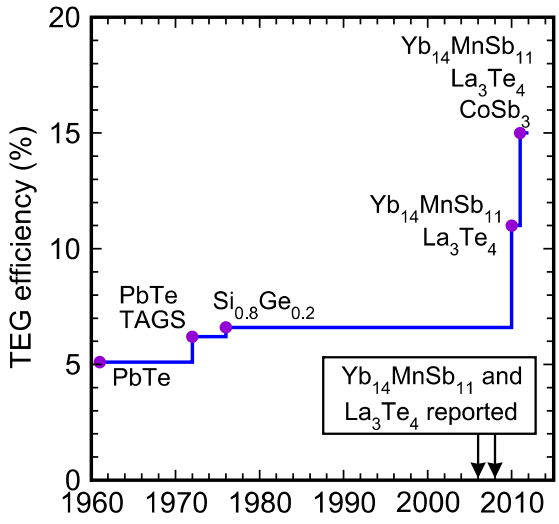


Figure 17. Time evolution of the efficiency of thermoelectric generators (TEG). Recent discoveries of new thermoelectric materials have resulted in an almost threefold increase in TEG efficiency after a 30 year long period of stagnation. [81] 2014 © TMS 2014. With permission of Springer.

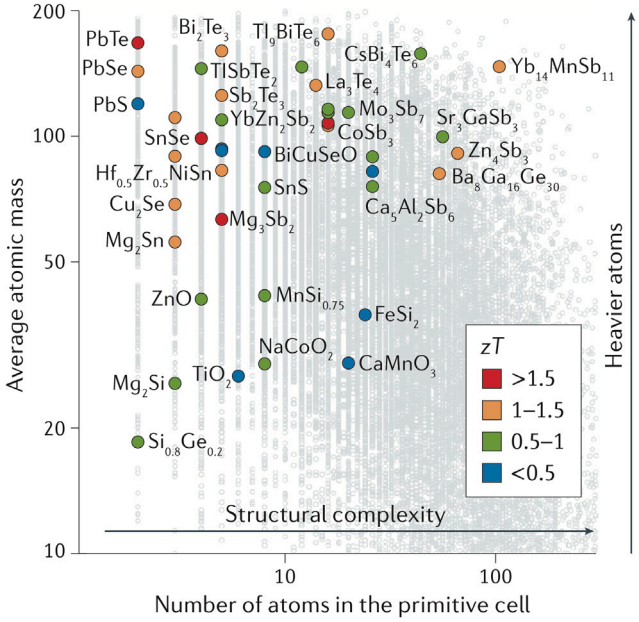


Figure 18. Relatively few materials have been characterized for their thermoelectric performance. About 44 systems (shown in color) out of the ~40 000 crystalline, stoichiometric and ordered metal–nonmetal compounds (gray) from the inorganic crystal structure database (ICSD) have their thermoelectric figure of merit zT reported in the literature (color coded). Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Materials [80], Copyright 2017.

the potential for the thermoelectric performance of a single material (orders of magnitude more expensive than density functional theory), computationally guided searches for new thermoelectrics may at first seem intractable.

However, the size of the search space also represents the biggest opportunity! The vastness of possible chemistries, both known and unknown, practically ensures the existence

of new, game-changing materials for any given application. The problem is then reduced to how to find the 'needle in the haystack' and not whether 'the needle is in the haystack', which is an important simplification. Second, it is critical to note that for the purpose of identification of new promising materials, it is sufficient to *estimate* relevant properties instead of *accurately predict* their absolute values. Although ideally one would prefer the latter, as long as the chemical trends are correctly reproduced, the ranking of different materials and identification of promising candidates can be reliable. This is what actually allows the screening of large chemical spaces and is the basis for a number of approaches and/or approximations that have been devised and employed in computational searches for new thermoelectrics. As a result, a number of databases providing predictions of transport properties of materials have emerged in the last decade (see [79, 80] and the references therein).

The main weakness of all these approaches is the focus on intrinsic materials properties and the assumption that semiconductors can be doped to a given charge carrier type (n or p) and carrier concentration. Many systems are not dopable at all and/or exhibit strong doping asymmetry favoring only one charge carrier type; thus, incorporating dopability assessment into computational searches is critical. Concerning experiments, the serial nature of material synthesis and characterization is another big challenge limiting accelerated materials discovery. As the reliability of computational predictions is largely probabilistic, high-throughput experiments are required to accurately assess the success rate of various approaches and provide the feedback loop to the theory about the feasibility of different approximations that are employed.

Advances in science and technology to meet challenges. In relation to predicting/assessing the dopability of materials, the good news is that the theory of defects in semiconductors and its computational implementations have evolved to a point where it is possible to accurately predict both the intrinsic and extrinsic defect chemistry and associated doping levels [87]. This includes advances in predicting materials stability and phase equilibria, which are an integral part of defect calculations. Moreover, recent successful automation of defect calculations [88] demonstrates the maturity of defect theory and its potential for large-scale applications. Yet, predicting the dopability of semiconductors is still far away from being 'black boxed'; it is a relatively tedious process requiring an in depth domain knowledge. The solution to these obstacles is in revealing deeper relationships between the defect chemistry and dopability on one side, and the chemical composition and crystal structure on the other. These relationships are presently either unknown or fairly qualitative.

Concerning the high-throughput experimentation, synthesis techniques are required that yield near-equilibrium samples with a form factor appropriate for accurate high throughput measurements. Jointly satisfying these requirements is presently not achievable with combinatorial thin film growth; advances in the high throughput synthesis of free-standing, dense monoliths would be enabling. Such a development would have cross-cutting implications for other bulk functional material searches. Given a high throughput experimental synthesis infrastructure, challenges remain in linking the computational descriptors with the experimental observables. For example, defects and dopants that may drive electronic and thermal properties are challenging to characterize robustly due to their low concentrations. Likewise, scattering sources and strengths are difficult to deconvolute from transport measurements. Strategies to proceed with while in an information-limited regime will thus be critical.

Concluding remarks. Thermoelectric materials discovery has come a long way in the last decade, from being guided predominantly by intuition and serendipity to the point where guidance is complemented by high-throughput calculations. At this point in time, it is safe to say that the computational challenges associated with assessing the potential of semiconductors for thermoelectric applications from the intrinsic (bulk) materials properties have largely been overcome. The remaining (grand) challenge that is still obstructing computational identification of truly game-changing new thermoelectrics is the assessment of dopability of candidate materials. Given the maturity and previous success of the defect theory and its computational implementations, there is, in our mind, little doubt that the dopability of semiconductors will be conquered and the true potential of computations in guiding thermoelectric materials discovery will be fully realized. More nascent is the development of high throughput bulk syntheses to complement these advances in computation. Given an effective computational framework coupled to such a high throughput synthesis, there is the opportunity for active learning within a machine learning context to further accelerate materials discovery.

Acknowledgments

Writing this article was supported by the NSF DMR program under award 1729594.

11. Perovskite photovoltaics

Aron Walsh[1,2] *and Nam-Gyu Park*[3]

[1] Department of Materials, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom
[2] Department of Materials Science and Engineering, Yonsei University, Seoul 03722, Republic of Korea
[3] School of Chemical Engineering, Sungkyunkwan University (SKKU), Suwon 440-746, Republic of Korea

Status. Metal halide perovskites form a large family of compounds ranging from small bandgap semiconductors to wide bandgap dielectrics [89]. The light-to-electricity conversion efficiency of metal halide solar cells now exceeds 22% for champion laboratory-scale devices [90]—following pioneering efforts on perovskite-sensitized [93, 94] and durable solid-state perovskite solar cells [93, 94]—that is comparable to mature thin-film photovoltaic technologies.

The compositional flexibility of the perovskite ABX_3 structure type allows for the control of chemical and physical properties over a wide range, including lattice constants, phase stability, optical bandgaps, charge carrier confinement, and defect processes. While the prototype hybrid organic-inorganic perovskite is methylammonium lead iodide ($CH_3NH_3PbI_3$), the highest performing compounds are multi-component mixtures, e.g. $(CH_3NH_3)_{1-x}(CH(NH_2)_2)_xPbI_{3-y}Br_y$ [90]. As our understanding of the fundamental structure-property relationships of halide perovskites increases, many opportunities arise to design novel materials and composites with enhanced properties, new device architectures with improved performance, and to explore alternative application domains including light emission, heat conversion, chemical sensing, information storage, spintronics, and radiation (γ and x-ray) detectors. The halide perovskites represent a vast playground for functional materials discovery (see figure 19 for some examples).

Current and future challenges. The science and technology of halide perovskite solar cells has developed rapidly over the past decade. These compounds were first treated as photoactive dyes deposited on a scaffold of TiO_2 [91, 92]. It took time to recognise that the materials were themselves semiconductors with the ability to conduct photogenerated electrons and holes. It was then found that they could also conduct ions, giving rise to slow hysteresis in the current–voltage response of solar cells [95, 96]. Despite a vast literature of thousands of publications concerning halide perovskites, there is still a large number of outstanding challenges, ranging from understanding the fundamental materials properties to physical processes on a device scale. These include:

- *Local crystal structure*—there is evidence that the local structure of halide perovskites has lower symmetry than the average spacegroup symmetry measured using standard Bragg diffraction techniques [97]. What is not known is the correlation lengths and lifetimes of local domains and how they interact with mobile carriers in operating solar cells.

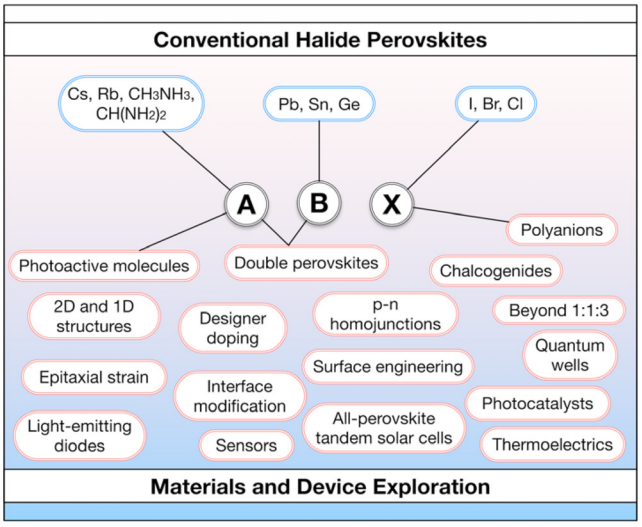


Figure 19. Illustration of the materials and device innovation space for the halide perovskite family.

- *Role of ferroelectricity*—there is substantial debate around the presence of polarisation domains in halide perovskites, in part because it is hard to separate lattice polarisation from effects due to mobile charges (electrons and ions). The literature currently contains many conflicting reports.
- *Point defect engineering*—all current solar cells are based on an intrinsic (undoped—low carrier concentration) perovskite layer with selective electron and hole electrical contacts. There have been no convincing reports of (robust) p- or n-doped halide perovskites, which would open a wide application space in optoelectronic technologies.
- *Extended lattice defects*—very little is known about the atomic configurations and electronic structure of grain boundaries, dislocations, interfaces, and surfaces of perovskites. Effective passivation of extended defects, in particular suppressing interface recombination events, could enhance device performance towards the theoretical limit of ~30% for bandgap of about 1.6 eV.
- *Chemical stability and breakdown*—many halide perovskites react with oxygen and water. Progress has been made with surface treatments [98] and physical encapsulation, but low-cost and robust approaches to achieving perovskite devices with long-term stability under realistic environments would represent a major breakthrough.
- *Pb-free compositions*—although Pb is a low cost and relatively abundant element, there is motivation for exploring element substitution, while maintaining beneficial photovoltaic properties. The isoelectronic replacement of Pb by Sn or Ge is problematic (reactive M^{2+} ions), so a route of active current investigation is double (mixed metal) perovskites, which have stability and electronic issues that need to be overcome.
- *Photophysics of solar cells*—in halide perovskites, photogenerated electrons and holes recombine slowly and hot states have anomalously long lifetimes. There are currently conflicting experiments and models, but control of

these processes could be used to realise hot carrier solar cells with efficiencies beyond the single-junction limit of ~30% light to electricity.

Advances in science and technology to meet challenges. For materials synthesis, the thin-film deposition of halide perovskites is dominated by solution-processing, with a growing number of vapour-processing studies being reported. The growth of higher quality thin-films on a wider range of substrates could enable better materials characterisation. In particular, epitaxial hetero-interfaces and perovskite homo-interfaces would allow a number of the challenges outlined above to be addressed, and the testing of new device architectures including all-perovskite p-n junctions, high-efficiency tandem solar cells, quantum wells and field-effect transistors.

Materials theory and simulation have played an important role in the understanding of perovskite technologies. The limitations of static band structure calculations on small unit cell representations is now recognised. Multi-scale methods are required to span the range of length and time scales necessary to describe the connection between structural disorder and dynamics with electron–hole generation, transport and recombination in solar cells. Furthermore, relativistic effects and electron–phonon coupling cannot be ignored; more research is required to understand the role of spin–orbit coupling and associated Rashba–Dresselhaus effects on the macroscopic physical and device behaviour.

Concluding remarks. Halide perovskites represent fertile ground for materials exploration. Now that high-efficiency photovoltaic devices have been realised, there is an opportunity to revisit the intriguing materials science of these compounds. Solving the challenges outlined in this section will require reliable and quantitative data on well-defined materials, with the close collaboration between theory, simulation and experiment. An improved understanding of the chemistry and physics of halide perovskites is essential to enable rational design of new functional materials that can provide similar technological breakthroughs.

Acknowledgments

A W acknowledges support from the Royal Society and the Leverhulme Trust. N G P is grateful for the support from the National Research Foundation of Korea (NRF) under Contract No. NRF-2012M3A6A7054861 (Global Frontier R&D Program on Center for Multiscale Energy System.)

12. Organic semiconductors: PV and LEDs

Daniel P Tabor[1], *Alán Aspuru-Guzik*[1,2] *and Jenny Nelson*[3]

[1] Department of Chemistry and Department of Computer Science, University of Toronto, Toronto, Canada
[2] Vector Institute for Artificial Intelligence, Toronto, Canada
[3] Department of Physics and Centre for Plastic Electronics, Imperial College London, London, United Kingdom

Status. Organic semiconductor materials have been the subject of intense research over more than 20 years because of their potentially tuneable properties, ease of processing, abundance and low cost. Many optoelectronic applications are based on inorganic semiconductors, but the range of stable crystalline semiconductors is limited and the *ab initio* design of new ones is limited by the strong dependence of properties on crystal structure and the difficulty in predicting new crystalline materials from an atomistic level. Design of organic functional materials, however, can be reduced to consideration of the molecular (or monomeric) level and the intermolecular interactions. Although the latter do influence material behaviour, the key optoelectronic properties are typically captured by examination of a single molecular unit or pairs of neighbouring units. Organic molecules can be thought of as comprised of building blocks that have clear structure-property relationships, making rational inverse design possible.

Here, we focus first on two applications: organic photovoltaics (OPVs) and organic light-emitting diodes (OLEDs). Each of these employs π-conjugated molecules, and depends on the intermolecular transport and transfer of charge carriers and the absorption or emission of light. OPVs are an appealing alternative to the dominant silicon technology because of straightforward fabrication, low cost, low weight, choice of colour and device flexibility. Due to these advantages, research accelerated during the 2000s (figure 20); the current record for conversion efficiency for a single junction OPV is over 13% [99], enabled by the recent development of high-performance organic acceptor materials other than fullerene derivatives. OLED research began in the late 1970s and polymer OLED research accelerated in the 1990s after the discovery of electroluminescence from conjugated polymers [100]. Blue OLEDs have surged due to the discovery of thermally-assisted delayed fluorescence, which gives access to higher luminescent efficiency by allowing both singlet and triplet excitons to emit light [101]. OLEDs have recently entered consumer markets as energy-efficient, high-contrast ratio display materials. Further advances could lead to cheaper displays with longer lifetimes. In this Roadmap, we will review the experience gained in materials development for OPVs and OLEDs and consider how this can assist the design of other organic functional materials, including organic redox flow battery (ORFB) electrolytes, organic photocatalysts, and organic thermoelectrics.

Current and future challenges. Device efficiency (power conversion efficiency of solar cells and luminous efficacy of OLEDs) remains a challenge. With OPVs, whilst the design rules concerning the energetics of component materials are well known, the precise role of and means to control film microstructure are still poorly understood. Processability of organic semiconductors comes at the cost of structural disorder and associated disorder in site energies and charge transfer rates, penalising efficiency [102]. Local ordering can benefit charge transport and pair separation, but can also introduce traps. Phase segregation in binary systems in OPVs is critical to performance but is still challenging to control by design.

Another persistent challenge in organic semiconductors is the operational stability of the device. For instance, OPVs have substantially shorter lifetimes than silicon-based photovoltaic devices [103]. Instabilities can come from a variety of sources, including photo-oxidation, electrochemical stress and morphological instabilities of thin films (via phase segregation and heterogeneous crystallisation). Samsung cited long term stability as a reason for their shift from OLED to QLED (quantum dot LED) development for televisions. Stability is also an issue in ORFBs, so a strategy that solves the stability problem for organic semiconductors may shed light and allow for similar methodologies to emerge in related materials. Overall, stability has been relatively under-researched to date compared to other properties and a detailed understanding of structure-stability relationships is lacking.

Additionally, scientists and engineers in these fields need to learn more about the nature of chemical space of these materials. Without constraints, chemical space is massive, estimated to be 10^{80} for organic molecules. By determining the minimum number of starting materials that are needed to cover all of the relevant parts of functional materials space, materials development efforts can be further focused. By analogy, it has been demonstrated that only about 5000 building blocks are needed to synthesize ~70% of small-molecule natural products [104]. Researchers in organic functional materials need to discover the corresponding number for their field and the most relevant degrees of freedom for their particular properties of interest. Determining these properties of chemical space will assist in the accelerated, rational development of new materials that are competitive with inorganic materials.

Advances in science and technology to meet challenges. Nearly all of the above challenges can be met via the efficient, rational exploration of chemical space, both theoretically and experimentally. We will focus on the case of OPVs, but these principles translate to other materials. With hindsight, the necessary structural features for some OPV properties would have been straightforward to calculate; theory is an excellent tool for calculating donor–acceptor (or push–pull) structures for low optical gap, electron-poor or electron-rich units to control ionization potential and electron affinity, side chain structure and positioning and backbone torsion to control crystallinity, and searching for molecules with low conformational phase space to limit energetic disorder.

Other properties, such as mobility or phase separation, are harder to predict because of more complex dependence of properties on multiple degrees of freedom. Here, materials identification can be accelerated by identifying intermediate properties, for example, isotropy in electronic coupling

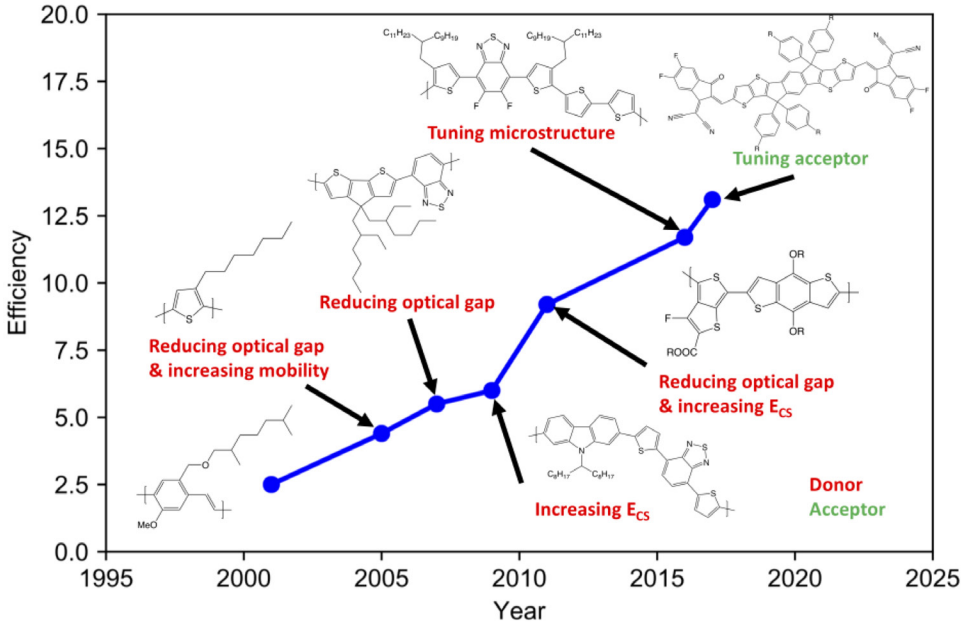


Figure 20. A timeline depicting the strategies used in materials development for OPV. As understanding of the relationships between material properties and device performance developed, increasingly sophisticated strategies were used to improve device performance. The results of years of studies, mostly by trial and error, have produced a set of design principles, many of which are relatively straightforward to implement using calculations and which can be used in screening to accelerate the discovery of new materials.

(considering both sign and amplitude) is beneficial for high charge carrier mobilities [105]; calculated solubility parameters or molecular-dynamics simulations of binaries [106] could help predict phase behaviour. Although such approaches could not predict new materials, calculating these more accessible quantities can reduce the design effort by screening potential winners from losers. Similarly, identifying the most important structural degrees of freedom for a given property can reduce the conformational phase space.

Virtual screening methods for organic materials have become increasingly useful over the past decade, with large-scale studies conducted to discover new molecules for OLEDs, OPVs, photocatalysts, thermoelectrics and ORFBs [101, 107]. From an experimental perspective, accelerating the synthesis and characterization can be done through adoption of high-throughput methods and robotics. Similar to theory, experimental methods can also employ advances in machine learning. A platform was recently used to optimize carbon nanotube growth based on on-the-fly characterization via Raman spectroscopy [108] and highly porous organic materials have been discovered aided by computational design [109]. The stability problem, in particular, needs significantly more characterization data to identify decomposition pathways. With this information, such pathways could be also screened for virtually. All of these advances also need to be underpinned by the adoption of better data management standards, where negative results are made available to virtual and experimental screening systems.

Concluding remarks. Research in organic semiconductors has moved into an era where principles learned from years of experiments can be employed by theorists to rationally design new materials. Going forward, theorists need to devise new techniques to compute more complex properties of organic semiconductors. Tighter feedback between experimentalists and theorists, aided by the continued development of machine learning methods, can accelerate the inverse design of the next generation of materials. The lessons learned from research in OPVs and OLEDs could also be used in other organic functional materials, including electrolytes for organic redox flow batteries, organic photocatalysts, and organic thermoelectrics.

Acknowledgments

D P T and A A-G acknowledge support through NSF STC Center for Integrated Quantum Materials (CIQM), NSF Grant DMR-1231319. AA-G is supported by CIFAR Grant No. BSE-ASPU-162439-CF. J N acknowledges the support of the EPSRC via Grant No. EP/P005543/1 and the ERC via Advanced Grant Action No. 742708.

13. Materials for solid-state lighting applications

Kirstin Alberi[1], *James Murphy*[2] *and Anant Setlur*[2]

[1] National Renewable Energy Laboratory, Golden, CO 80401, United States of America
[2] Functional Materials Laboratory, GE Global Research, Niskayuna, NY, United States of America

Status. Solid-state lighting (SSL) exploits electroluminescence processes from semiconductors to produce light more efficiently than heated filament or gas sources. White light is typically produced by pairing a blue LED with a down-conversion material, which re-emits absorbed blue photons across the rest of the visible spectrum (see figure 21). Advanced approaches can also mix emission from individual red, green, blue and amber LEDs [110]. The LEDs and down conversion materials must be as efficient as possible to maximize energy savings, while their emission spectra must be carefully tailored to achieve the desired color temperature of white light, as well as render colors suitably based upon the application

Blue and green LEDs are fabricated from $In_xGa_{1-x}N$ alloys, where indium is added to shift the emission to longer wavelengths [110]. $In_xGa_{1-x}N$ is a better blue light-emitter compared to other semiconductors with similar direct band gap energies, as it is relatively tolerant to extended defects. This allows single crystal $In_xGa_{1-x}N$ device layers to be epitaxially grown on substrates with different lattice constants, despite strain-driven dislocation formation. Yet, there are drawbacks to this material system [111]. When grown on SiC or sapphire substrates, the polar axis of $In_xGa_{1-x}N$ is aligned along the direction of electron and hole injection. The resulting piezoelectric fields set up by polarity and strain reduce electron and hole wavefunction overlap and lower radiative recombination. The addition of more indium to $In_xGa_{1-x}N$ increases strain, which further aggravates these losses and contributes to the low efficiencies of green $In_xGa_{1-x}N$ LEDs. These loss mechanisms can be partially suppressed through quantum confinement (e.g. quantum wells or nanowires) or by growing on the non-polar crystal faces of bulk GaN substrates. However, such approaches have yielded insufficient efficiency increases at green wavelengths, are too costly, or are less practical for mass production.

Commercial red and amber LEDs are fabricated from $(Al_xGa_{1-x})_{0.5}In_{0.5}P$ alloys. Since LED efficiency is strongly affected by dislocations, the In concentration is selected for strain-free growth on conventional GaAs substrates. The emission wavelength is tuned by adjusting the ratio of Al and Ga. $(Al_xGa_{1-x})_{0.5}In_{0.5}P$ undergoes a transition between a direct and indirect band gap semiconductor at ~2.25 eV (550 nm). $(Al_xGa_{1-x})_{0.5}In_{0.5}P$ LEDs with emission wavelengths of 590 nm or less have electrons lost to the indirect conduction bands at room temperature [112]. These losses extend to longer emission wavelengths LEDs at higher operation temperatures.

Typical LED down-converting materials are inorganic phosphors; insulating hosts that are doped with activator ions whose basic properties (e.g. absorption, emission, efficiency) are defined by how the atomic transitions of the activator are modified by the interaction with the host lattice. Currently, three main activator ions, Ce^{3+}, Eu^{2+}, and Mn^{4+} are used the most often in typical LED phosphors (table 1). The quantum efficiencies (QEs) of these phosphors are often above 90% across the visible color spectrum for blue LED excitation. However, the peak wavelengths and linewidths of their emission as well as their stability are still factors that can be further improved.

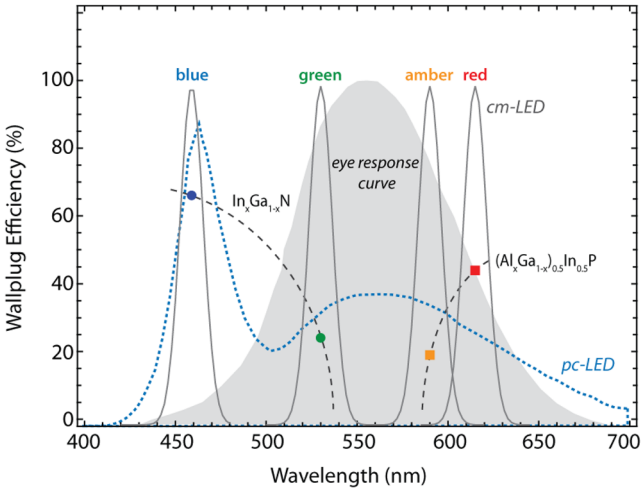


Figure 21. Emission spectra for phosphor-converted (pc) and color-mixed (cm) LEDs. The wallplug efficiencies of blue, green, amber and red LEDs (dotted lines) and luminous eye response curve (shaded area) are also shown.

Current and future challenges. $In_xGa_{1-x}N$ and $(Al_xGa_{1-x})_{0.5}In_{0.5}P$ alloys are the semiconductors of choice for visible LEDs in part because they have properties that are amenable to both light emission and manufacturing. They are highly developed direct band gap semiconductors with tunable band gaps, are relatively robust against defect-induced degradation, and are grown on readily available substrates. However, the emission efficiencies of green, amber and red LEDs remain well below that of blue LEDs (see figure 21) and are limited by the fundamental properties of those materials. Small adjustments in material quality, structure or composition alone are unlikely to lead to substantial improvements. One path forward is to identify alternative semiconductors with properties that are better suited to green, amber or red emission and that meet several design criteria. The semiconductor must have a high emission efficiency under high injection or elevated operating temperatures. Device layers should be grown on conventional, cost-effective substrates with low defect densities using scalable deposition techniques. Finally, they should be resilient against degradation for extended LED lifetimes.

Semiconductors that have recently been considered for LEDs include direct band gap $Al_xIn_{1-x}P$, II–IV–N alloys, halide perovskites and $GaN_{1-x}As_x$ [113–116]. These materials are in various stages of development, ranging from theoretical predictions to full device demonstrations, and it is not yet known if any will offer performance breakthroughs. Materials discovery and synthesis efforts should focus on understanding the advantages and disadvantages of different classes of semiconductors in the context of the design criteria outlined above.

Current phosphors have enabled sufficient efficacy and color quality for the widespread acceptance of LEDs for lighting and displays. Increasing luminaire efficacy to >200

Table 1. Typical activator ions, their relevant transitions, and representative compositions for phosphors used in blue LEDs.

Activator ion	Absorption transition for blue light	Emission transition	Representative compositions and emission color
Ce^{3+}	$4f^1$ ($^2F_{5/2}$) → $5d^1$	$5d^1$ → $4f^1$($^2F_{7/2}$,$^2F_{5/2}$)	$Y_3Al_5O_{12}:Ce^{3+}$ (green–yellow)
Eu^{2+}	$4f^7$ → $4f^65d^1$	$4f^65d^1$ → $4f^7$	(Sr,Ca)$AlSiN_3:Eu^{2+}$ (orange and red) β-$SiAlON:Eu^{2+}$ (green)
Mn^{4+}	$3d^3$ ($^4A_2$) → $3d^3$ ($^4T_2$)	$3d^3$ (^2E) → $3d^3$ ($^4A_2$)	$K_2SiF_6:Mn^{4+}$ (red)

lumens per watt (lm W^{-1}) requires the development of high-efficiency, stable, narrow linewidth down-converter materials that emit at specific red, amber and green wavelengths [117, 118]. Semiconductor quantum dots offer narrow band emission that may be tuned to desired wavelengths to improve efficacy, but reliability and European Union Regulation on Hazardous Substances (RoHS) compliance have been barriers to adoption. Continued efficacy improvements for high color rendering LEDs therefore require new phosphor composition development for ions (i.e. Eu^{2+}, Mn^{4+}) that could give narrow linewidth emission. Narrow band emission from the red phosphor in particular minimizes spillover into longer wavelengths where the human eye response falls off rapidly. There is some correlation of crystal structures and luminescence properties to phenomenological understanding of phosphor properties [119, 120], but these phenomenological models have limitations in their application to new phosphor discovery. Commercial implementation of new LED phosphors also needs to meet multiple requirements beyond absorption and emission. New phosphors require QEs greater than 90%, and their efficiency and color cannot change significantly over system life. This optimization requires additional composition and process steps including choosing appropriate starting materials, determining nominal stoichiometry, and optimizing processing conditions. Trial-and-error screening experiments are followed by optimization using designed experiments once the key factors have been determined. The success of these optimization steps is usually the difference between successful and unsuccessful phosphor development, and takes up the largest portion of time and cost for phosphor development [121]. As an example, figure 22 shows reliability improvements through process optimization in $K_2SiF_6:Mn^{+4}$, a narrow line-width, red emitting LED phosphor (GE TriGain™). These results illustrate the importance of the development phase after simply identifying a promising material candidate.

Advances in science and technology to meet challenges. Tools for high-throughput computational screening have and will continue to aid in the search for new light-emitting and down-conversion materials. Semiconductor crystal structures and electronic band structures can already be calculated with a high degree of accuracy, but advances are still needed in our ability to predict tolerances to defects, Auger recombination rates and other parameters that affect radiative recombination efficiency. This will be enabled by improvements in our understanding of radiative loss mechanisms. On the phosphor side, advances are needed in the computation of phosphor crystal structures, excited states in heavy lanthanide ions and defect chemistries. Improved understanding in these areas will help to categorize basic phosphor properties and pinpoint new phase space in which to search for promising materials. Identification of loss and degradation mechanisms in phosphors will also help to guide material development and optimization strategies to improve performance and reliability. Once the most promising LED and phosphor candidates are identified, it will likely require substantial resources to fully develop and evaluate their potential experimentally. Synthesizing new materials can be challenging, particularly if it must involve non-standard epitaxial growth conditions or new reaction routes. Advances in tools for materials fabrication and characterization will therefore also be important to this effort.

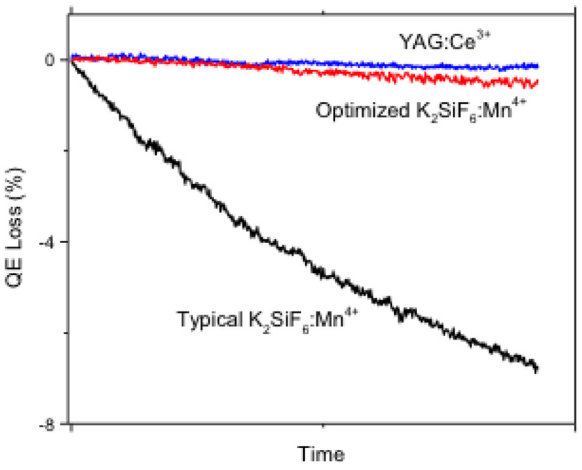


Figure 22. Accelerated reliability testing of phosphors using high intensity blue excitation. The industry standard green-emitting $Y_3Al_5O_{12}:Ce^{3+}$ (YAG:Ce) is compared to a GE TriGain™ $K_2SiF_6:Mn^{4+}$ red-emitting phosphor whose synthesis and composition have been optimized relative to a typical $K_2SiF_6:Mn^{4+}$ phosphor. The timescale for these accelerated measurements is proprietary information, however, these tests can accelerate phosphor degradation by >100 × versus typical medium-power LEDs.

Concluding remarks. While commercial white SSL solutions are approaching efficacies of 200 lm W^{-1}, opportunities exist to improve the efficiency of SSL through improved material design and optimization. Identification of new emitting materials, either active semiconductors or down-converters for blue LEDs, offer a direct route to realizing maximal efficiency gains. A combined approach of theoretical prediction and experimental development could accelerate materials discovery and optimization for implementation into future lighting systems and displays. This acceleration can be further enhanced in combination with the current trend towards lower drive current densities in LED packages to produce more efficient, high color quality SSL solutions with improved reliability.

Acknowledgments

K A acknowledges the support of the US Department of Energy, Office of Basic Energy Sciences under contract DE-AC36-08GO28308.

14. Chemistry materials: catalysts

John M Gregoire

California Institute of Technology, Pasadena, CA, United States of America

Status. Tailoring the design of a material for a specific function is particularly important in catalysis, including thermal, electro-, and photo-catalysis. For the present discussion on materials design, as opposed to reaction design, we focus on heterogeneous catalysts, for which the most universally important fundamental properties are the binding energy of reactant, intermediate, and product molecules on the catalyst surface along with the respective reaction barriers. Advances in computational chemistry and computing have made calculations of the binding energies rather automated [122], with recent advancements in machine learning-based error correction making even computationally-inexpensive algorithms sufficiently accurate to design catalysts. Modern theory-based computational algorithms have been tailored for specific reactions and operating conditions, in particular where binding energies and reaction barriers can be modulated via multi-body interactions, dynamic variations in the reaction environment, and catalyst surface dynamics under operation [124]. Broadening the catalyst design framework from a binding site to a catalyst system will enable materials to achieve the activity of the ultimate catalysts, enzymes, while providing the longevity required for deployment in energy, commodity chemical, etc, industries [125].

Current and future challenges. A primary challenge in the improvement of catalyst design lies in the traditional disconnect between computational chemistry and catalysis experiments, where the former excels at a molecule-level understanding but struggles to model the full catalyst system and the latter typically produces a net reaction rate with limited ability to decompose it into elementary steps. Catalysts that perform multi-step reactions, such as oxygen evolution and reduction, CO_2 reduction to hydrocarbons and oxygenates, and N_2 reduction to ammonia, comprise some of the most widely designed catalysts now and in the foreseeable future. The recent proliferation of so-called scaling relationship theory for such reactions predicts that catalysts with a single active site will generally be limited in their catalytic activity [126], which is troubling given that traditional catalyst design focuses on identification and optimization of such a site. The resulting stagnation in identification of transformative catalysts further motivates the expansion of catalyst design to consider more complex and dynamic catalysts, for example, through incorporation of variability in computational modelling (figure 23) [123].

The biggest future challenge in catalyst design lies in the integration of data science, machine learning, and artificial intelligence in computational and experimental catalyst exploration. As noted above, machine learning has emerged in catalyst design primarily as an accelerator for computational work [127], and while challenges remain in deeper integration

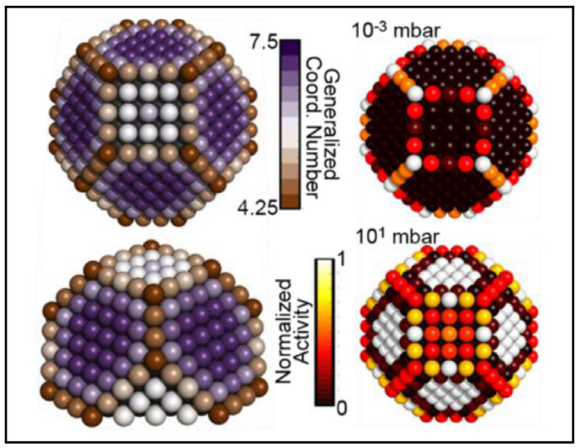


Figure 23. (Left) Generalized coordination numbers of ontop sites on a truncated octahedron and (right) and CO oxidation activity of a 2.8 nm Pt particle used in Monte Carlo modelling of catalytic activity, which highlights the complexity of identifying and designing catalytic sites. Reprinted with permission from [123]. Copyright 2017 American Chemical Society.

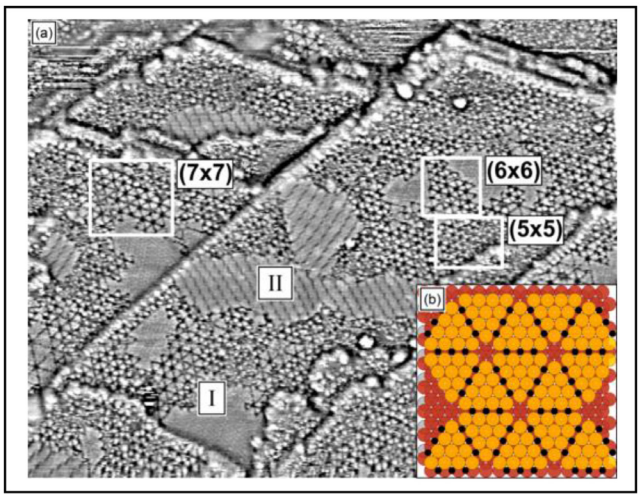


Figure 24. Atomic resolution *in-situ* scanning tunneling microscope image of CO dissociation on a Co catalyst, which highlights the complexity and evolution (even on the 1 h time scale) of heterogeneous catalysts. Reprinted with permission from [122]. Copyright 2015 American Chemical Society.

of machine learning and theory, the grander challenge lies in the utilization of machine learning to provide data-driven identification of the underlying catalyst properties that give rise to an observed reaction rate. That is, a given catalyst performance measurement, even when combined with thorough compositional and structural characterization, typically cannot identify a reaction mechanism or design principle for improving the catalyst. By consolidating a broad collection of composition-structure-activity relationships in a data model, new catalyst understanding and design avenues may be unveiled. In the present 'big data' era where loads of data are used to provide a black box prediction tool, the relatively small adoption of machine learning in catalyst design is somewhat understandable as the community neither has the requisite data to train such models nor the appetite for data models that cannot 'explain' the underlying science. As artificial

intelligence research begins to dissect the big data black box and as new algorithms are designed to utilize known properties of materials, the power of machine learning in catalyst design can be fully realized.

Advances in science and technology to meet challenges. The recent advent of *in-situ* and *operando* techniques has greatly enhanced experiment-driven catalyst understanding, which largely provides additional characterization of the catalyst surface or near-surface under operating conditions (figure 24) [128]. Such data helps relate the computer models of materials to the experimental catalyst but does not sufficiently bridge the gap between molecular-level calculations and reaction rates. Approaches for further bridging the theory-experiment gap include atomic resolution scanning probe characterization that does not alter the catalysis, computational modelling techniques that simulate experimental observables to enable more direct comparison, and multi-scale computational techniques that provide quantum mechanics-level accuracy in many-atom systems. The detection of partial-monolayer reaction intermediates offers perhaps the best means of (in)validating a computational model, and continued development of the associated spectroscopic techniques, in particular, infrared spectroscopy and synchrotron-based electron spectroscopy techniques, are needed to realize this goal in both thermal and electro-catalysis. These experimental advancements indirectly enhance materials design by providing the requisite data from which hypothesis-driven catalyst modifications or new catalyst designs can be derived. To enable direct, more *ab initio* catalyst design, computational modelling must incorporate new strategies for bridging time and space scales. Single crystals transforming absolutely pure reactants are useful model systems, and extending design to deployable catalysts requires modelling of materials defects, chemical impurities, and the evolution of catalysts over years of operation, which typically implies on the order of 10^8 'turnovers' or catalyst cycles.

For the emerging challenge of integrating machine learning in catalyst design, the road to success is less well defined, with one certainty being that new algorithms will need to combine the state of the art in machine learning with the constraints and concepts of catalysis science. To enable algorithm development and deploy such algorithms, substantial advancements to the catalyst community's data infrastructure are also needed, as well as experimental methods that can rapidly respond to new catalyst predictions [129]. On both of these fronts, the combinatorial and high throughput materials science community as well as the small molecule and biological chemistry communities offer a wealth of best practices that can be adapted as necessary to accelerate the adoption of machine learning in materials design for catalysis [40].

Concluding remarks. Transformative advancements in materials design for catalysis hinge upon further integration of theory and experiment as well as interdisciplinary engagement of artificial intelligence and the data science community. The combination of techniques can enable a sort of divide and conquer approach to creatively adapt existing capabilities into new materials design paradigms that harness the complexity of catalyst systems for multi-step reactions. A recent illustrative example in photoelectrocatalyst design involves integration of several theory and experiment steps to discover classes of materials that respond to new design concepts [130]. Here, the proficiency of theoretical modelling of a materials' bulk electronic structure was combined with efficient experimental assessment, with the more general concept being that different approaches can tackle different aspects of catalyst design as long as the compilation of techniques appropriately captures the complexity of the multi-step catalytic processes that are increasingly important to industry and society.

Acknowledgments

J M G acknowledges support from the US Department of Energy Award No. DE-SC0004993P.

15. Materials for Li-ion batteries

Ruijuan Xiao and Hong Li

Institute of Physics, Chinese Academy of Sciences, Beijing, People's Republic of China

Status. Advanced battery technology has become one of the core technologies to support a mobile, clean and sustainable society in the next few decades. Lithium batteries have been used widely in portable electronic products, electrical vehicles and energy storage devices for wind and solar power, because of their high voltage, high specific energy density, rapid recharge capability, and wide working temperature range [131]. The advances of battery techniques are always going to go with the development of new materials. For example, lithium-rich layered oxide materials have been considered as an ideal positive electrode in high-energy-density lithium-ion batteries [132], and the nano silicon-based anodes as alternative materials show reversible capacities of 380–2000 mAh g^{-1} [133]. The indisputable fact is that the discovery of advanced materials and rational design play key roles in battery research. To speed up the upgrading of the chemical materials in lithium batteries, high-throughput techniques, including high-throughput simulations, synthesis and measurement, have been applied to the discovery of new battery materials. Data mining and machine learning have been introduced to benefit the understanding of the big data obtained from high-throughput techniques, which provide opportunities for further exploration of the structure-property relationship of battery materials and to discover new materials. On the other hand, by comparing the theoretical results or model predictions with the myriad experimental data, the sources of error and uncertainty in battery research can be captured, which in turn help us to build better theoretical models or investigating apparatus. The mutual promotion of the above aspects shown in figure 25 is expected to accelerate the discovery of candidate compounds in the future and shorten the invested time and money, not only for lithium batteries, but also for other new types of energy storage devices, like Na, Zn, Mg, Al batteries, etc.

Current and future challenges. The high-throughput calculation work flow has been established based on density functional theory simulations [134, 135], and the combination of calculation methods in different accuracy levels [136] has been proposed to speed up the scanning process of new materials. The former has been applied to scanning the inorganic crystal structure database for candidate electrode materials with high voltage and capacitance [137]. Using ideas originating from the latter method, a new superionic conductor has been proposed [138]. To achieve battery devices with higher energy density and safety, inorganic solid electrolytes are expected to replace liquid electrolytes in the next generation lithium batteries [139]. The application of solid electrolytes may avoid problems of leakage, vaporization, decomposition and side reactions found in the conventional lithium-ion batteries. However, finding solid electrolytes with excellent performance is still a demanding task, since the comprehensive physical description between structures and ionic conductivity is still not easy to grasp. Similar problems exist in the discovery of other battery materials. For example, suitable electrode materials with long-term stability require a small volume change ratio during lithium ion insertion and extraction [140]. However, the percentage of the volume change varies from material to material because of the complicated origins of the cell variation, which leaves huge obstacles for us to discover low-strain electrodes. As an integrated system, the performance of the battery not only relates to the properties of the individual components but also is strongly affected by the interactions among them. One typical case is that the interface between the electrode and electrolyte seriously impacts the stability, rate and cycle-life of the batteries. Therefore, looking for favorable combinations of the components in the battery is extremely crucial. The details and key factors in optimizing these interactions are still in development and remain a major challenge for the design and matching of battery materials. In general, extending the understanding of the basic scientific problems in battery systems is the main research issue on the way to discovery new lithium battery materials.

Advances in science and technology to meet challenges. To meet the above-mentioned challenges, advances in both science and technology are urgent. Figure 26 exhibits the goal of battery techniques and the methods that need to be developed in the near future. On the one hand, designing delicate prototypes to understand the basic scientific phenomena in batteries by high-throughput experiments and simulations is a conventional but efficient research mode. With the help of advanced measurement and analytical tools, more exquisite microstructures and evolution processes can be revealed, which will clarify the failure mechanism of lithium batteries and direct the discovery of new battery materials. On the other hand, designing an automatic screening and prediction workflow with sufficient accuracy and efficiency is essential. For each part of the battery, the electrode, electrolyte, additive, collector, etc, it is necessary to meet more than one requirement to ensure the excellent performance of the whole device. It is better for a high-voltage cathode to show high-capacitance and good conductivity. Similarly, fast ionic conductivity and a wide electrochemical window are both necessary prerequisites for electrolyte materials. Thus, screening and predicting tools for multiple objectives must be created. Aside from the advances addressed above, data science and technology also have to be developed for material design. It is recognized that machine learning techniques and big data methods will play an increasingly important role in solving the relationships between material properties and complex physical factors in a statistical way, which builds the basis for material design, and vice versa. However, material informatics is still an emerging field with problems like the lack of data standards, the diversity of material types, and even the conflict of research culture, etc. Data management specific to battery materials should be developed and the descriptors suitable for them should be explored.

Concluding remarks. Rational design of lithium battery materials is highly desirable in the near future. Because of the complex structure-property relationships of ionic conductivity, volume change, electrode/electrolyte interfaces, etc, successful

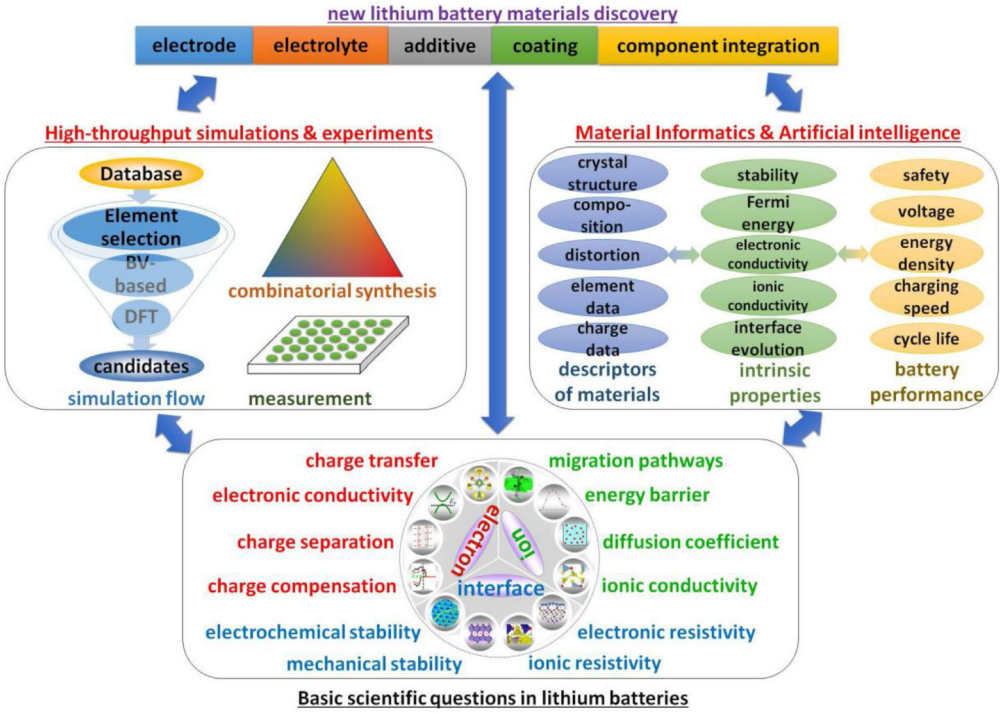


Figure 25. The development mode of new material design in lithium batteries by means of high-throughput techniques and data sciences.

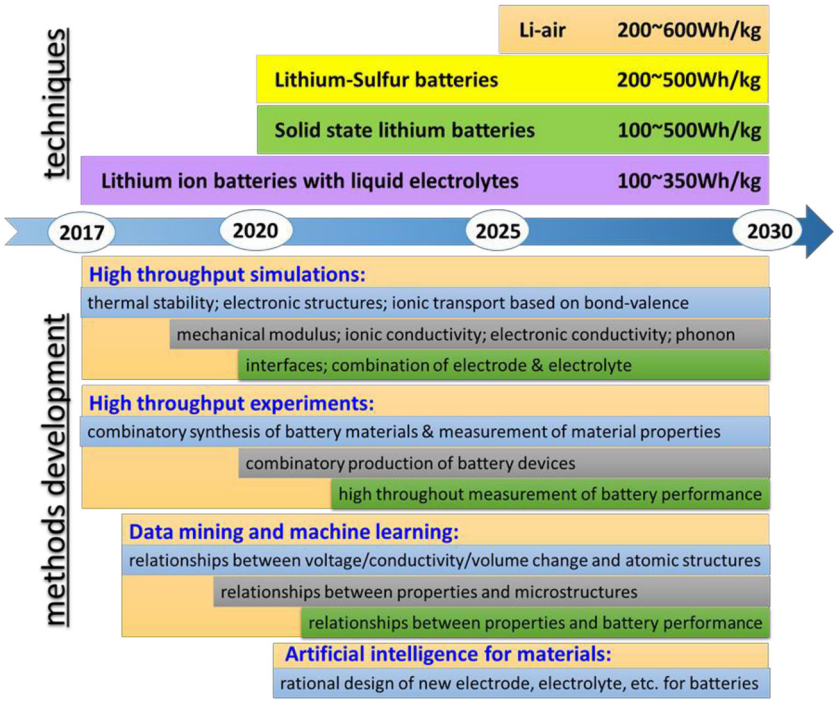


Figure 26. The developing roadmap of lithium batteries in the near future.

cases of designing new battery materials are still scarce. Advances in the development of high-throughput techniques and material informatics will bring more efficient research and provide new opportunities to solve the above problems, which will deepen our understanding of the basic scientific questions in battery fields and accelerate the discovery of materials for lithium batteries and other new types of energy conservation devices.

Acknowledgments

We acknowledge the National Natural Science Foundation of China (Grant No. 51772321), '863' Project (Grant No. D171100005517001), the Beijing S&T Project (Grant No. D161100002416003), and the Youth Innovation Promotion Association (Grant No. 2016005) for financial support.

16. Multifunctional metallic alloys

Alfred Ludwig

MEMS Materials, Institute for Materials, Ruhr-University Bochum, Germany

Status. Metallic alloys have been of crucial importance to humankind since the bronze age and will continue to be a critical material class, enabling new capabilities, applications and products. The advantages of metallic alloys are their frequently good mechanical properties like high strength and plasticity (ductility, toughness), which are hard to achieve in other material classes. In multifunctional alloys, these favorable mechanical properties are combined with additional functional properties (electrical, magnetic, optical, etc.). Multifunctionality is frequently related to reversible phase transformations. Compositional complexity of alloys has usually increased from binary to multinary systems, often currently involving more than 10 elements, with compositions finely tuned to specific applications. Examples of such compositionally complex alloys are steels, superalloys and metallic glasses. Whereas these alloys are typically based on one element (Fe, Ni, Co, Al, Mg, …), recently multi-principal element alloys (MPEA) have also attracted interest, as they promise a mostly unexplored search space for the discovery of new alloys with interesting properties [141]. Whereas a single-phase constitution is crucial for semiconductors, metallic alloys are typically multiphase materials, and the properties of the alloys can be tailored by controlling microstructure using processing. The phase constitutions, their distribution and volume fractions in the alloy can be used to adjust properties (e.g. a tough matrix phase with a strengthening precipitate phase). Further advances in metallic alloys are gained by developing alloys which combine good mechanical properties with further functionalities. For high-temperature alloys, for example, the formation of a protective oxide scale can lead to a functional property: resistance against corrosion. New or improved (multi)functionalities need to be developed to realize metallic materials for future applications. Whereas in bulk applications of metals, mechanical properties are dominant, in thin film applications, they are less important, i.e. even materials which would be too brittle for bulk applications can be used in thin films. This opens up the field of intermetallic compounds with (multi)functional properties, which are frequently not ductile, into the scope of new thin film materials. Such materials comprise magnetic alloys, shape memory alloys, magnetic shape memory alloys, thermoelectric alloys, magneto- and elastocaloric alloys, etc. Such classes of (multi)functional metallic materials can be explored by combinatorial and high-throughput thin film methods to enable the design, discovery and optimization of materials based on the acquired knowledge.

Current and future challenges. A current and future challenge is the design and discovery of new compositionally complex metallic alloys, i.e. ternary to quinary systems and beyond, either based on a principal element or as MPEA, with interesting mechanical and functional properties. Additionally, the influence of impurity elements on the properties of multinary alloys should be studied. Complex metallic alloys, characterized by extraordinary large unit cells, is another area for new discoveries [142]. Generally, it is necessary to overcome reliance on serendipitous discoveries (e.g. NiTi) and use combinatorial and high-throughput methods, both computational and experimental, to identify, verify and then optimize new metallic alloys in a more efficient way. However, this is challenging, as the largest fraction of elements in the periodic table are metals, which leads to an almost unlimited search space, even if the selected elements are restricted to those which are earth abundant and sustainable. Computational approaches [143, 10] for the high-throughput prediction of possibly (meta)stable alloys with interesting properties can help in this regard to select a few ten to hundred appealing candidates out of hundreds of thousands of possibilities, which then can be assessed (verification/falsification of predictions) with high-throughput experimental methods. However, these calculations are frequently limited to the intrinsic properties and sufficiently precise and validated experimental data for the calculations are often lacking. A further challenge, next to principal stability and the possibility to fabricate new materials, is to master extrinsic properties such as the microstructural diversity. For an identical composition, many microstructures are often possible, e.g. from nanocrystalline to microcrystalline, from amorphous to single- or multi-phase crystalline structure, all of which influence the properties of the alloy. Another challenge is to screen thin film libraries for ductility and, what is more, how findings from large scale thin film materials library explorations could be transferred to the bulk scale, i.e. how new ductile (multi)functional metallic alloys could be efficiently identified. Examples of correlative thin film-bulk studies can be found in [144–146].

Advances in science and technology to meet challenges. For the advancement of the discovery and design of multinary alloys, several technologies need to be further developed. Whereas combinatorial deposition methods for thin film materials libraries are now well-established, the further automatization and speed advances of high-quality characterization methods need to be continued to enable better high-throughput characterization. An important methodology to be developed is ‘combinatorial processing’ to address the challenge of microstructural complexity. For this, gradient and step heater concepts for both the formation and annealing of thin films have been introduced [147]. A high-throughput processing approach for the identification of new metallic glasses with thermoplastic formability was performed by parallel blow forming of co-deposited thin-film libraries on micromachined substrates [148]. Furthermore, it would be worthwhile to develop schemes where materials libraries would not be only characterized for one property, but rather comprehensively for ‘all’ functional properties. Another necessary advancement is related to the development of materials in systems. This is because it is not sufficient to only develop a material by itself; rather it has to be developed in a system, which means it has to provide functionality in connection with adjacent materials and environments. Here, interface properties play a key role.

Furthermore, if the fabrication and characterization of materials libraries leads to the discovery of new phases, the challenge arises for an accelerated identification of these phases. Here, advanced electron diffraction methods in transmission electron microscopy (TEM) (combination of automated diffraction tomography with precession electron diffraction) could help, if the materials of interest can be grown to a sufficiently large grain size [149]. A novel accelerated exploration approach for temperature- and environment-dependent phase evolution in compositionally complex materials has been introduced by Li *et al* [150]: combinatorial processing platforms are created by co-deposition of multinary thin films on nanoscale tip arrays forming many identical nanoscale 'reactor volumes' allowing for fast diffusion and reaction and immediate observation of the product phases by the atomic-scale analysis methods atom-probe tomography and TEM. This allows for an accelerated mapping of the phase space of multinary metallic alloys. Another challenge is the development of materials data science, research data management, and materials informatics, e.g. machine learning for data-guided experimentation. Finally, visualization of compositions and properties in complex multinary materials systems is difficult but necessary. Thus, new software tools have to be developed which will lead to the establishment of functional phase or existence diagrams (including metastable phases) for multinary alloys for the future design of materials.

Concluding remarks. The success story of metallic alloys will be continued by applying computational and experimental combinatorial and high-throughput methods for the discovery and optimization of new multinary compositions. If the new materials are developed from the start with regards to their functionality within a system, i.e. with regards to the interfaces which are formed between materials in a system, faster development of materials from their discovery over optimization to incorporation into a product could be achieved.

Acknowledgments

Funding from 'Deutsche Forschungsgemeinschaft' (DFG) in many projects over the last 15 years is acknowledged.

17. Functional ceramic materials

Lane W Martin[1,2] *and Andrew M Rappe*[3]

[1] Department of Materials Science and Engineering, University of California, Berkeley, CA, United States of America
[2] Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States of America
[3] Department of Chemistry, University of Pennsylvania, Philadelphia, PA, United States of America

Status. Over the last few decades, complex oxides (materials with multiple cations and oxygen) have been a central research focus because of their wide range of properties and applications. Leveraging an ability to manipulate the charge, lattice, orbital, and spin degrees of freedom, scientists have explored a range of exotic, and potentially useful, phenomena including superconductivity, magnetism, colossal magnetoresistance, ferroelectricity, multiferroism, relaxor behaviour, ionic conductivity, piezoelectricity, and many more. Such 'functional' materials (i.e. materials that can transmit or convert energy (e.g. electrical, thermal, mechanical, etc) for useful purposes (e.g. information transfer, sensing, energy production, positioning, etc)) [151, 152], underpin our ability to address a range of salient technological challenges, including how we process and store information, sense and understand the world around us, produce energy, and more [153]. Ferroic materials, including those which are ferroelectric, magnetic, ferroelastic, and/or multiferroic, continue to receive considerable interest due to their field-switchable stable spontaneous order parameters (electric polarization, magnetic moment, strain), which are strongly coupled to the thermal and mechanical responses of the material (figure 27). The search for, discovery of, and utilization of these materials has been made possible by important advances in theoretical and computational approaches, materials synthesis, and characterization techniques. Functional-oxide research has enabled the realization of new materials and the development of new functionality in existing materials. These research insights are fed back into the design process, including massively parallel design of new oxide materials and heterostructures. The advanced state of synthesis and characterization confers unprecedented control of materials chemistry and structure, and this will ultimately lead to the creation of new states of matter and phenomena. Recent innovations include new single-phase materials, close juxtapositions of competing or complementary functionalities, and orchestration of emergent responses on many length and time scales. Here, we highlight some of the most important recent advances in terms of materials design and discovery, understanding, and characterization for functional materials while looking to the future for what might lie on the horizon for this community.

Current and future challenges.

Advanced computation and data storage. There is great interest in moving beyond field-effect transistors and Boolean operation, and functional oxides can lead that revolution by providing negative capacitance, piezotronics, tunnel junctions, and spintronics. In addition, neuromorphic computing architectures (designed to emulate neuron function) require materials exhibiting multiple and addressable microstates and the ability to evolve continuously in response to voltage-current stimuli [154]. Ferroic materials are promising because of their intrinsic non-volatility and fast switching, but limited progress has been made towards deterministic multi-state functions. There is likely to be growing interest in designing and controlling ferroics in ways that will enable low-power and multi-state operation in this regard.

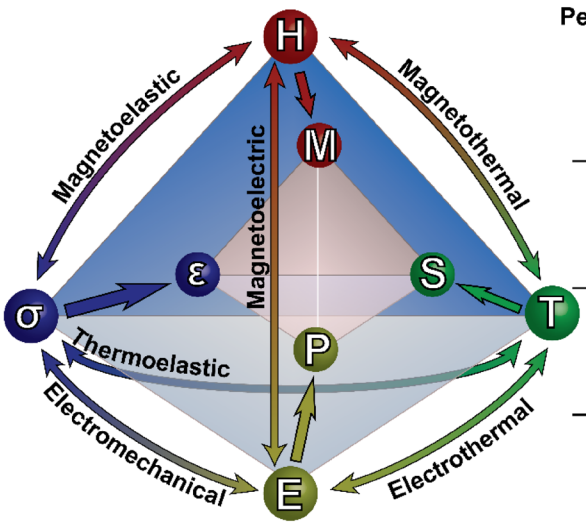
Energy conversion and efficiency. Societal energy needs make the development of more efficient energy conversion a compelling research challenge. Ferroic systems have great potential in this field. For example, ferroelectric photovoltaics host the bulk photovoltaic effect, where a 'shift current' [155] and asymmetrically scattered 'ballistic current' [156] cause excited carriers to move in a specific direction determined by the polarization; it can even give rise to photovoltages that exceed the bandgap and break Shockley–Queisser limits for efficiency. Others explore ferroic materials for novel waste-heat energy conversion as thermoelectrics or via pyroelectric energy conversion, for low-power, solid-state cooling via the electro- and magneto-caloric effects [157], vibrational energy conversion applications, and much more, and as active or supporting materials for catalysts.

Sensing and communications. The Internet of Things (IoT) and its acquisition of ever-increasing datasets drives a need for new abilities to sense, communicate, and interact with components in many aspects of life. Functional materials will play vital roles in sensors, energy harvesting/remote power generation, data storage and transmission, and much more. Materials that are compatible with advanced healthcare monitoring (*in* and *ex vivo*) will be of particular interest. Ferroic materials provide a foundation for such applications, since one materials class provides all these functions—sensing, energy generation, energy storage, communications, etc—while being both chemically inert and stable. The future of communications—in particular, the advent of higher-frequency 5G technologies—will also likely drive materials innovation to achieve aggressive design requirements. Microwave communication bands are becoming increasing congested; agile, tunable materials with high quality factors will be essential to meeting the needs of commerce, defence, and other applications.

Advances in science and technology to meet challenges.

High-throughput materials discovery. The Materials Genome Initiative [158] ignited high-throughput discovery of functional materials [159]. A central driver is the optimization of descriptors that can be rapidly calculated to identify novel materials and phenomena. Experimentalists must also develop ways to rapidly produce and characterize an ever-widening set of candidate materials. Advances in the discovery of complex oxides portend the dramatic expansion of known or predicted functional materials (figure 28) [160].

The materials-data nexus. Modern computational and experimental probes have led to orders-of-magnitude increases in the volume, variety, veracity, and velocity of materials research



Perturbation	Response	Susceptibility
σ	ε	Elasticity
	P	Piezoelectric
	S	Piezocaloric
	M	Converse Magnetostriction
E	ε	Converse Piezoelectric
	P	Dielectric
	S	Electrocaloric
	M	Electromagnetic
T	ε	Thermoelastic
	P	Pyroelectric
	S	Heat Capacity
	M	Pyromagnetic
H	ε	Magnetostriction
	P	Magnetoelectric
	S	Magnetocaloric
	M	Magnetic

Figure 27. Adapted Heckmann diagram showing a range of functional responses possible in ferroic materials together with a table of excitations, responses, and the named effects. The figure explores connections between purturbations stress (σ), electric field (E), temperature (T), and magnetic field (H) and responses strain (ε), polarization (P), entropy (S), and magnetization (M). Reproduced with permission from [153]. © Materials Research Society 2016.

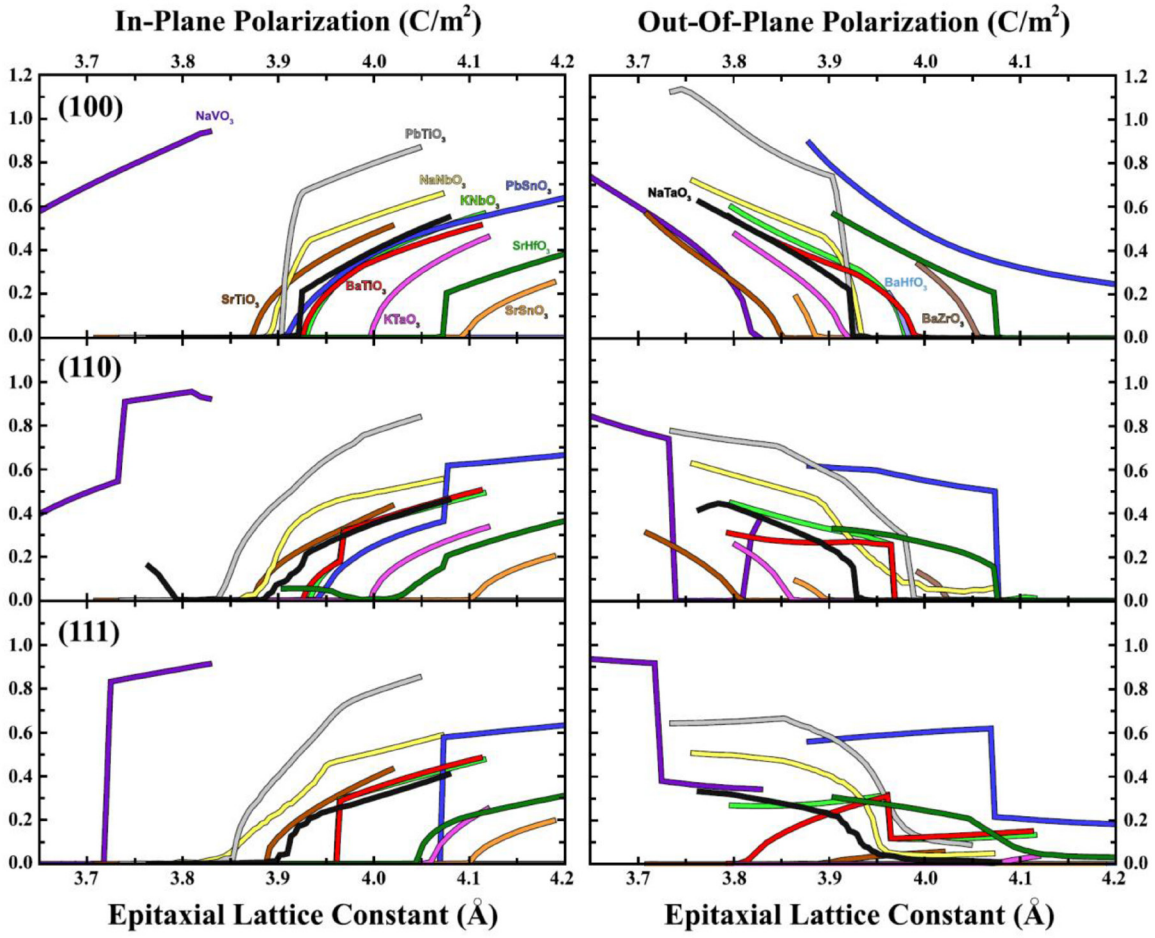


Figure 28. Combining high-throughput computation (in this case, density-functional theory approaches) with tools of advanced synthesis (epitaxial strain on a range of substrate orientations), researchers are now able to explore not only many materials in equilibrium, but increasingly large design-parameter spaces in the search for high-performance functional materials. This work shows predictions of the evolution of polarization in known and candidate polar materials with strain and film orientation. Such approaches can provide novel routes to the identification of not only novel new materials, but also new phases and features of interest in existing or known materials. In this way, the field can greatly expand the range of materials of use for a variety of applications. Reproduced figure with permission from [160], Copyright 2017 by the American Physical Society.

data. High-dimensional, high-resolution data sets make direct extraction of physically-relevant information challenging. Brute-force approaches, wherein models or fitting functions are used to extract parameters of predetermined significance, fail when the data have unknown variety, veracity, or arrive with high velocity. Addressing data challenges will require the adoption of statistical tools including machine learning to identify data correlations, trends, clusters, and anomalies. Melding traditional physical sciences with new data-intensive approaches offers transformational opportunities to simplify the transition from data to scientific insight.

Managing emergent behaviours. A driver of new functionality will be the harnessing of phenomena on length scales other than the material dimensions. Prominent examples include polar nanoregions in relaxor ferroelectric alloys and topological defects, such as magnetic or electric skyrmions. These phenomena break conventional relationships between order parameters and stimuli, and the acquisition of a deep understanding of these may hold the key to a new generation of smart materials.

New modes of synthesis. A key to advancing material functionality will be new strategies for controlling chemistry and structure. In particular, 'defect' control—deterministic production of specific types, concentrations, and locations—could enable a watershed in the design and discovery of new physics and emergent function. This new approach posits that defects, long considered deleterious to properties, can now be viewed positively as a tool to enable elegant manipulation of the local balance of charge, lattice, orbital, and spin degrees of freedom. This could induce new properties and effects. Such routes are particularly amenable to complex oxides, which naturally host larger defect concentrations. Recasting the role of defects will provide a pathway to new emergent properties and could lead to unprecedented material responses.

Concluding remarks. Modern functional ceramics are a critical part of everyday life. In the near future, their roles in advanced electronics, sensing, energy transduction, communications, and other areas seem poised for strong growth. The key to this impact lies in the multi-functional and agile nature of the responses of these materials and their ability to accomplish in one material what might otherwise require many. It is envisioned that these materials will continue to be explored in non-traditional communities and as replacements for traditional materials because of the multi-functionality, adaptability, and robustness to operation in harsh environments. At the same time, this added function comes at the cost of added complexity in controlling those materials to elicit the desired properties. Advances in computational and experimental methodologies are now poised to revolutionize our understanding of these materials and their deployment in breakthrough applications.

Acknowledgments

L W M acknowledges support from the Army Research Office under grant W911NF-14-1-0104, the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under award number DE-SC-0012375 and the US Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05-CH11231: Materials Project Program KC23MP, the Gordon and Betty Moore Foundation's EPiQS Initiative, under grant GBMF5307, the National Science Foundation under grants DMR-1451219, CMMI-1434147, OISE-1545907, DMR-1608938, and DMR-1708615, and the Intel Corp. through the FEINMAN program. A M R acknowledges support from the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under awards number DE-FG02-07ER46431 and DE-FG02-07ER15920, the Office of Naval Research under grants N00014-17-1-2574 and N00014-12-1-1033, the Army Research Office under grant W911NF1510589, and the National Science Foundation under grants DMR-1120901, CBET-1159736, CMMI-1334241, and DMR-1719353.

18. Transparent conducting materials

Su-Huai Wei[1] *and John D Perkins*[2]

[1] Beijing Computational Science Research Center, Beijing 100193, People's Republic of China
[2] National Renewable Energy Laboratory, Golden, CO 80401, United States of America

Status. Transparent conducting materials (TCMs) are defined by high electrical conductivity approaching that of a metallic compound (conductivity $\sigma > 10^4$ S cm^{-1}), with the high transmission of photons in the visible or near infrared range of the electromagnetic spectrum (transmission $T > 80\%$). The unique combination of these two features makes TCMs essential components of modern optoelectronic devices [161], such as (a) transparent electrodes for flat panel displays including touch screens, (b) transparent electrodes for photovoltaic cells, (c) smart windows, (d) transparent thin films transistors, and (e) light emitting diodes and lasers. The first TCM thin film was reported by Badeker [162] in 1907, more than 100 years ago, and was based on CdO. Afterwards, more TCOs, such as SnO_2, In_2O_3, ZnO, and their alloys, including amorphous alloys, have been discovered and are utilized in our daily life [163, 164]. All of these materials can be thought of as very-heavily-doped wide-band-gap n-type semiconductors. Figure 29 shows characteristic reflection (R), transmission and absorption (A) spectra for a TCO thin film where the transmission is cut off on the short wavelength side by the intrinsic band gap absorption and on the long wavelength side by the carrier-concentration-dependent onset of absorption, due to conduction-band-electron plasma oscillations. We will refer to such materials, where the material itself is both transparent and conducting, simply as TCMs. In contrast, a second very different class of transparent conductors has emerged where porous nanoscale networks or grids of highly conducting wires yield an overall low sheet resistance on a macroscopic scale, along with high optical transmission due to the large openings between the wires [165]. Such transparent conducting networks have been made using both carbon nanotubes or metal nanowires, with silver nanowires versions now seeing limited commercialization for touch screen displays. Often, the nanoscale conducting network is embedded in a metal oxide or other matrix to improve both opto-electronic functionality and mechanical strength. Accordingly, we will refer to this second class as composite transparent conductors (c-TCs). Figure 30 compares the optical transmission spectra for a silver nanowire-based c-TC with that for a conventional n-type TCO.

At present, the vast majority of TCMs are still n-type TCOs. The most important n-TCO used today is tin doped indium oxide, In_2O_3:Sn [166, 167], typically called indium-tin-oxide or ITO. ITO along with high-indium content amorphous In–Zn–O are the dominant transparent electrode materials for flat panel displays, the application which represents the largest annual value for the TCO thin film industry, but also greatly contributes to the rising cost of In metal. Therefore, it is still significant to improve the conductivity–transmission (C/T) performance of the existing TCOs, or develop the new TCMs or c-TCs that are less expensive (i.e. indium free), non-toxic, have easily-tailored interface and high C/T properties and are easily fabricated.

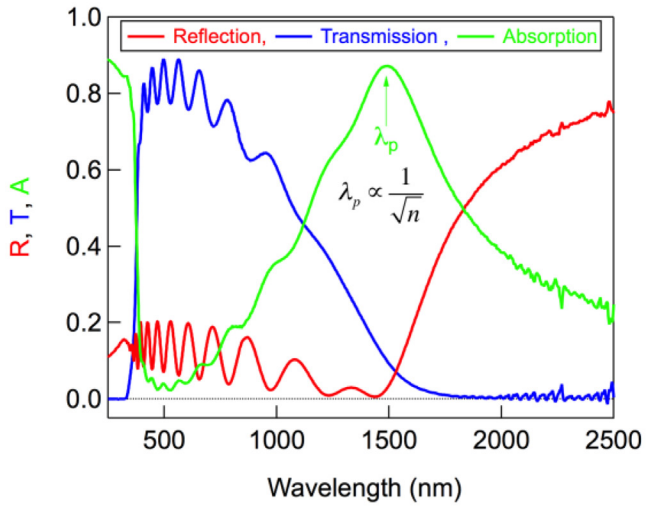


Figure 29. Optical reflection, transmission and absorption spectra for an Al-doped ZnO TCO film. The plasma wavelength (λ_p) which varies with carrier concentration is indicated with an arrow.

Current and future challenges. To improve the σ/T performance of TCOs, it is essential to simultaneously maximize the conductivity σ and optical transmission in the visible (VIS) spectrum. Achieving the high electrical conductivity ($\sigma = ne\mu$ where e is the elementary charge) asks for increasing the carrier concentration n (electrons or holes), or carrier mobility μ as much as possible. Stoichiometrically perfect TCOs (In_2O_3, SnO_2, …) basically have no free carriers due to the large band gap ($E_g \geqslant 3$ eV). Therefore, unintentional or intentional defects along with extrinsic dopants have a critical role in optimizing the carrier concentration. To achieve the high VIS transmission ($T = 1 - R - A$), one should reduce the reflection (R) and absorption (A). The low VIS absorption requires TCMs have a large optical band gap ($E_g^{opt} > 3$ eV), which is related to the materials with large and direct fundamental band gaps or forbidden dipole transition near the band edges [168].

Nowadays, all the commercial TCOs are of the n-type because it is easy to achieve the high concentration ($n \sim 10^{21}$) by the substitutional doping, such as Sn doped In_2O_3 (ITO), Al doped ZnO (AZO), F doped SnO_2 (FTO), and so on. Further, the conduction band minimum of TCOs derives from delocalized cation s orbitals, which ensures n-type TCOs have a relatively high mobility. However, so far there are no commercial p-type TCOs, which seriously hinders the applications of transparent semiconductors because of the absence of the bipolar transistors and diodes without the p-n junctions. Achieving high-conductivity p-type TCOs is a big challenge for the oxides, because the valence band maximum of oxides is dominated by the very low-energy and localized oxygen p orbital, which causes the formation of deep acceptor level and poor hole mobility. Experimentally, beyond equilibrium hole

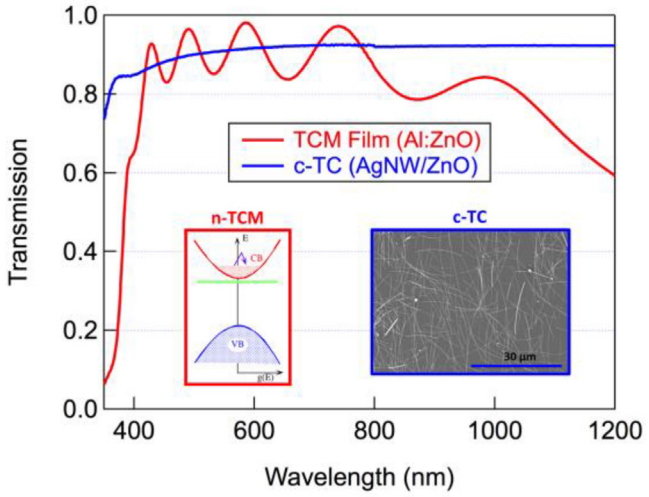


Figure 30. Comparison of transmission spectra for a Al-doped ZnO thin film TCO with that of a Ag-NW/ZnO composite transparent conductor (c-TC). Insets: left—schematic electronic structure of a n-type TCO; right—image of a AgNW/ZnO c-TC.

doping levels approaching $10^{21}/cm^{-3}$ have been achieved in p-type TCOs using non-equilibrium growth methods, such as sputtering. However, the conductivity remains low (σ of order 10^2 S cm^{-1} or less) due to low mobility and the materials are generally not very transparent. In the space of amorphous materials, the key challenge for n-type a-TCMs is to find high-conductivity indium-free materials, whereas for p-type materials, it is still just to find high-conductivity materials. For c-TC materials, there are challenges in the area of using metal nanowires beyond just silver to reduce reactivity, including the use of alloy compositions or protective layers as well as wide open opportunities to tune the application specific functionality of c-TCs through the choice of the matrix materials.

Advances in science and technology to meet challenges. As discussed before, to improve the conductivity of the TCMs, one should either increase the carrier concentration or the mobility, especially for the p-type TCMs, of which both quantities are far below the standard for commercial applications. To achieve this, the following strategies may be valuably considered: (i) increase defect solubility by 'defeating' bulk defect thermodynamics using non-equilibrium growth methods, such as extending the achievable chemical potential through molecular doping or raising the host energy using surfactant; (ii) reduce the defect ionization energy level by designing shallow dopants or dopant complexes, including transition metal doping, co-doping, multivalence-impurity doping, etc, and (iii) increase the carrier mobility by modifying the host band structure near the band edges. Because increasing the carrier density can also lead to an increase in the visible absorption and possible reflection, whereas increasing the mobility has less bad effects, one of the best strategies is relying on the band structure engineering to increase carrier mobility without affecting the optical properties much to realize high performance TCMs.

At present, ITO is one of the best TCM materials with both high conductivity and excellent optical transmission. However, due to the scarcity and high price of In, developing and searching new TCMs that are cheap, non-toxic, and have a similar conductivity or even higher than that of ITO has been in great demand in recent years. A practical and feasible strategy is that based on the established general guidelines, we can use materials by design to search new materials with high transparency and low carrier effective mass, thus good conductivity, including p-type transparent conductors and non-oxide transparent conductors [169]. The effective application of materials by design approaches to the discovery and development of improved amorphous materials remains a big challenge due largely to both the underlying challenges in computational physics for amorphous materials and the experimental challenges in adequately characterizing amorphous materials to provide the feedback to theory that is so critical for materials by design. To advance the composite transparent conductors through the use of materials by design will require the bridging of length scales to couple materials by design with integrated computational materials science and engineering approaches to develop a functional predictive capability for topologically complex multi-component systems. Finally, to actually impact real world technologies and needs, the predicted target materials must be able to be made, which leads to the newly emerging challenge of theory-guided predictive synthesis [170].

Concluding remarks. With the expected increasing use of TCM reliant consumer electronics and energy technologies, there will continue to be a need for TCMs with ever increasing performance and decreasing cost made using sustainable materials. Specific materials development needs include high performance indium-free n-type TCOs and p-type TCMs with qualitatively better performance than are currently available. There is likely to be great opportunities for materials development in non-oxide and mixed anion material systems as well as amorphous materials. Finally, the alternative approach of c-TCs based on porous nanoscale conducting networks is wide open for further development. Materials by design methods can accelerate this materials advancement and, likewise, the relevance here of amorphous materials and composite materials will push the advancement of materials by design.

Acknowledgments

S H W was supported by the NSFC under Grant Nos. 11634003, 51672023 and U1530401 and the National Key Research and Development Program of China under Grant No. 2016YFB0700700. J D P was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, as part of the Energy Frontier Research Center- Center for Next Generation of Materials Design under Contract No. DE-AC36-08GO28308 to NREL.

ORCID iDs

Kirstin Alberi https://orcid.org/0000-0002-5236-5568
Andriy Zakutayev https://orcid.org/0000-0002-3054-5525
Marco Fornari https://orcid.org/0000-0001-6527-8511
Sergiy Butenko https://orcid.org/0000-0002-6662-9552
Aron Walsh https://orcid.org/0000-0001-5460-7033
Alán Aspuru-Guzik https://orcid.org/0000-0002-8277-4434
Kirstin Alberi https://orcid.org/0000-0002-5236-5568
Alfred Ludwig https://orcid.org/0000-0002-8680-6667
Lane W Martin https://orcid.org/0000-0003-1889-2513

References

[1] Lejaeghere K *et al* 2016 Reproducibility in density functional theory calculations of solids *Science* **351** aad3000
[2] Aryasetiawan F and Gunnarsson O 1998 The GW method *Rep. Prog. Phys.* **61** 237
[3] Foulkes W M C, Mitas L, Needs R J and Rajagopal G 2001 Quantum Monte Carlo simulations of solids *Rev. Mod. Phys.* **73** 33
Wagner L K and Ceperley D M 2016 Discovering fermions using quantum Monte Carlo *Rep. Prog. Phys.* **79** 094501
Needs R J, Towler M D, Drummond N D and Lopez Rios P 2009 Continuum variational and diffusion quantum Monte Carlo calculations *J. Phys.: Condens Matter* **22** 023201
[4] Zhang S and Krakauer H 2003 Quantum Monte Carlo method using phase-free random walks with slater determinants *Phys. Rev. Lett.* **90** 136401
[5] Booth G H, Gruneis A, Kresse G and Alavi A 2013 Towards an exact description of electronic wave function in real solids *Nature* **493** 365
[6] Dubecky M, Mitas L and Jurecka P 2016 Noncovalent interactions by quantum Monte Carlo *Chem. Rev.* **116** 5188
[7] Agapito L A, Curtarolo S and Buongiorno Nardelli M 2015 Reformulation of DFT + U as a pseudohybrid hubbard density functional for accelerated materials discovery *Phys. Rev.* X **5** 011006
[8] Georges A, Kotliar G, Krauth W and Rozenberg M 1996 Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions *Rev. Mod. Phys.* **68** 13
Vollhardt D 2012 Dynamical mean-field theory for correlated electrons *Ann. Phys.* **524** 1–19
[9] Bennett M C, Melton C A, Annaberdiyev A, Wang G, Shulenburger L and Mitas L 2017 A new generation of effective core potentials for correlated calculations *J. Chem. Phys.* **147** 224106
Kim J *et al* 2018 *J. Phys.: Condens. Matter* **30** 195901
[10] Curtarolo S, Hart G L W, Nardelli M B, Mingo N, Sanvito S and Levy O 2013 The high-throughput highway to computational materials design *Nat. Mater.* **12** 191–201
[11] Jain A, Shin Y and Persson K A 2016 Computational predictions of energy materials using density functional theory *Nat. Rev. Mater.* **1** 15004
[12] Lin L 2015 Materials databases infrastructure constructed by first principles calculations: a review *Mater. Perform. Charact.* **4** MPC20150014
[13] Rajan K 2005 Materials informatics *Mater. Today* **8** 35–48
[14] Krogel J T 2016 Nexus: a modular workflow management system for quantum simulation codes *Comput. Phys. Commun.* **198** 154–68
[15] Saritas K, Mueller T, Wagner L and Grossman J C 2017 Investigation of a quantum Monte Carlo protocol to achieve high accuracy and high-throughput materials formation energies *J. Chem. Theory Comput.* **13** 1943–51
[16] Perdew J P, Ruzsinszky A, Constantin L A, Sun J and Csonka G I 2009 Some fundamental issues in ground-state density functional theory: a guide for the perplexed *J. Chem. Theory Comput.* **5** 902–8
[17] Pizzi G, Cepellotti A, Sabatini R, Marzari N and Kozinsky B 2016 AiiDA: automated interactive infrastructure and database for computational science *Comput. Mater. Sci.* **111** 1–35
[18] Curtarolo S *et al* 2012 AFLOW: an automatic framework for high-throughput materials discovery *Comput. Mater. Sci.* **58** 218–26
[19] Larsen A H *et al* 2017 The atomic simulation environment—a Python library for working with atoms *J. Phys.: Condens. Matter* **29** 273002
[20] Mathew K *et al* 2017 Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows *Comput. Mater. Sci.* **139** 140–52
[21] Kresse G and Furthmüller J 1996 Efficiency of *ab initio* total energy calculations for metals and semiconductors using a plane-wave basis set *Comput. Mater. Sci.* **6** 5
[22] Giannozzi P *et al* 2017 Advanced capabilities for materials modelling with QUANTUM ESPRESSO *J. Phys.: Condens. Matter* **29** 465901
[23] Calderon C E *et al* 2015 The AFLOW standard for high-throughput materials science calculations *Comput. Mater. Sci.* **108** 233
[24] Garrity K F, Bennett J W, Rabe K M and Vanderbilt D 2014 Pseudopotentials for high-throughput DFT calculations *Comput. Mater. Sci.* **81** 446
[25] Bilić A and Gale J D 2009 Ground state structure of $BaZrO_3$: a comparative first-principles study *Phys. Rev.* B **79** 174107
[26] Petretto G, Gonze X, Hautier G and Rignanese G-M 2018 Convergence and pitfalls of density functional perturbation theory phonons calculations from a high-throughput perspective *Comput. Mater. Sci.* **144** 331–7
[27] Rose F, Toher C, Gossett E, Oses C, Buongiorno Nardelli M, Fornari M and Curtarolo S 2017 AFLUX: the LUX materials search API for the AFLOW data repositories *Comput. Mater. Sci.* **137** 362–70
[28] Kirklin S, Saal J E, Meredig B, Thompson A, Doak J W, Aykol M, Rühl S and Wolverton C 2015 The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies *NPJ Comput. Mater.* **1** 15010
[29] Toher C *et al* 2017 Combining the AFLOW GIBBS and elastic libraries to efficiently and robustly screen thermomechanical properties of solids *Phys. Rev. Mater.* **1** 015401
[30] van Setten M J *et al* 2015 GW100: benchmarking G0W0 for molecular systems *J. Chem. Theory Comput.* **11** 5665
[31] Takeuchi I and Xiang X-D 2003 *Combinatorial Materials Synthesis* (New York: Dekker)
[32] Koinuma H and Takeuchi I 2004 Combinatorial solid state chemistry of inorganic materials *Nat. Mater.* **3** 429–38
[33] Dellamorte J C, Barteau M A and Lauterbach J 2009 Opportunities for catalyst discovery and development: integrating surface science and theory with high throughput methods *Surf. Sci.* **603** 1770–5
[34] Simon C G and Lin-Gibson S 2011 Combinatorial and high-throughput screening of biomaterials *Adv. Mater.* **23** 369–87
[35] Green M L, Takeuchi I and Hattrick-Simpers J R 2013 Applications of high throughput (combinatorial) methodologies to electronic, magnetic, optical, and energy-related materials *J. Appl. Phys.* **113** 231101
[36] Potyrailo R A and Takeuchi I 2005 Role of high-throughput characterization tools in combinatorial materials science *Meas. Sci. Technol.* **16** 1

[37] Ludwig A, Cao J, Brugger J and Takeuchi I 2005 MEMS tools for combinatorial materials processing and high-throughput characterization *Meas. Sci. Technol.* **16** 111

[38] Stanev V, Oses C, Kusne A G, Rodriguez E, Paglione J, Curtarolo S and Takeuchi I 2018 Machine learning modeling of superconducting critical temperature *npj Comput. Mater.* **4** 29

[39] Kusne A G *et al* 2014 On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets *Sci. Rep.* **4** 6367

[40] Green M L *et al* 2017 Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies *Appl. Phys. Rev.* **4** 011105

[41] Wu L *et al* 2017 High-temperature crystallization of nanocrystals into three-dimensional superlattices *Nature* **548** 197–201

[42] Soderholm L and Mitchell J F 2016 Perspective: toward 'synthesis by design': exploring atomic correlations during inorganic materials synthesis *APL Mater.* **4** 053212

[43] Sun W *et al* 2016 The thermodynamic scale of inorganic crystalline metastability *Sci. Adv.* **2** e1600225

[44] Cheong S, Watt J, Ingham B, Toney M F and Tilley R D 2009 *In situ* and *ex situ* studies of platinum nanocrystals: growth and evolution in solution *J. Am. Chem. Soc.* **131** 14590–5

[45] Bai J, Hong J, Chen H, Graetz J and Wang F 2015 Solvothermal synthesis of $LiMn_{1-x}Fe_xPO_4$ cathode materials: a study of reaction mechanisms by time-resolved *in situ* synchrotron x-ray diffraction *J. Phys. Chem.* C **119** 2266–76

[46] Jensen K M Ø, Tyrsted C, Bremholm M and Iversen B B 2014 *In situ* studies of solvothermal synthesis of energy materials *ChemSusChem* **7** 1594–611

[47] Fields J D *et al* 2016 The formation mechanism for printed silver-contacts for silicon solar cells *Nat. Commun.* **7** 11143

[48] Shoemaker D P *et al* 2014 *In situ* studies of a platform for metastable inorganic crystal growth and materials discovery *Proc. Natl Acad. Sci. USA* **111** 10922–7

[49] Kanatzidis M G 2017 Discovery-synthesis, design, and prediction of chalcogenide phases *Inorg. Chem.* **56** 3158–73

[50] Haynes A S, Stoumpos C C, Chen H J, Chica D and Kanatzidis M G 2017 Panoramic synthesis as an effective materials discovery tool: the system Cs/Sn/P/Se as a test case *J. Am. Chem. Soc.* **139** 10814–21

[51] Hill J, Mulholland G, Persson K, Seshadri R, Wolverton C and Meredig B 2016 Materials science with large-scale data and informatics: unlocking new opportunities *MRS Bull.* **41** 399–409

[52] Hey T, Tansley S and Tolle K 2009 *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Cambridge: Microsoft Research)

[53] Agrawal A and Choudhary A 2016 Perspective: materials informatics and big data: realization of the 'fourth paradigm' of science in materials science *APL Mater.* **4** 053208

[54] Materials Genome Initiative for Global Competitiveness 2011 Office of Science and Technology Policy, National Science and Technology Council (www.mgi.gov/sites/default/files/documents/materials_genome_initiative-final.pdf)

[55] Gaultois M W, Oliynyk A O, Mar A, Sparks T D, Mulholland G J and Meredig B 2016 Perspective: web-based machine learning models for real-time screening of thermoelectric materials properties *APL Mater.* **4** 053213

[56] Raccuglia P, Elbert K C, Adler P D F, Falk C, Wenny M B, Mollo A, Zeller M, Friedler S A, Schrier J and Norquist A J 2016 Machine-learning-assisted materials discovery using failed experiments *Nature* **533** 73–6

[57] National Research Council 2014 *Big Data in Materials Research and Development: Summary of a Workshop* (Washington, DC: The National Academies Press) www.nap.edu/catalog/18760/

[58] Agapito L A, Ferretti A, Calzolari A, Curtarolo S and Nardelli M B 2013 Effective and accurate representation of extended Bloch states on finite Hilbert spaces *Phys. Rev.* B **88** 165127

[59] Agapito L A, Fornari M, Ceresoli D, Ferretti A, Curtarolo S and Nardelli M B 2016 Accurate tight-binding Hamiltonians for two-dimensional and layered materials *Phys. Rev.* B **93** 125137

[60] De S, Bartok A P, Csanyi G and Ceriotti M 2016 Comparing molecules and solids across structural and alchemical space *Phys. Chem. Chem. Phys.* **18** 13754–69

[61] Jain A, Hautier G, Ong S P and Persson K 2016 New opportunities for materials informatics: resources and data mining techniques for uncovering hidden relationships *J. Mater. Res.* **31** 977–94

[62] Needs R J and Pickard C J 2016 Perspective: role of structure prediction in materials discovery and design *APL Mater.* **4** 053210

[63] Gautier R, Zhang X, Hu L, Yu L, Lin Y, Sunde T O, Chon D, Poeppelmeier K R and Zunger A 2015 Prediction and accelerated laboratory discovery of previously unknown 18-electron ABX compounds *Nat. Chem.* **7** 308–16

[64] Hinuma Y *et al* 2016 Discovery of earth-abundant nitride semiconductors by computational screening and high-pressure synthesis *Nat. Commun.* **7** 11962

[65] Meredig B, Agrawal A, Kirklin S, Saal J E, Doak J W, Thompson A, Zhang K, Choudhary A and Wolverton C 2014 Combinatorial screening for new materials in unconstrained composition space with machine learning *Phys. Rev.* B **89** 094104

[66] Holder A M *et al* 2017 Novel phase diagram behavior and materials design in heterostructural semiconductor alloys *Sci. Adv.* **3** e1700270

[67] Caskey C M, Richards R M, Ginley D S and Zakutayev A 2014 Thin film synthesis and properties of copper nitride, a metastable semiconductor *Mater. Horiz.* **1** 424–30

[68] Ndione P F, Shi Y, Stevanovic V, Lany S, Zakutayev A, Parilla P A, Perkins J D, Berry J J, Ginley D S and Toney M F 2014 Control of the electrical properties in spinel oxides by manipulating the cation disorder *Adv. Funct. Mater.* **24** 610–8

[69] Lany S, Fioretti A N, Zawadzki P P, Schelhas L T, Toberer E S, Zakutayev A and Tamboli A C 2017 Monte Carlo simulations of disorder in $ZnSnN_2$ and the effects on the electronic structure *Phys. Rev. Mater.* **1** 035401

[70] National Research Council 2008 Committee on integrated computational materials engineering *Integrated Computational Materials Engineering: A Transformational Discipline For Improved Competitiveness and National Security* (Washington, DC: Academic) (https://doi.org/10.17226/12199)

[71] De Yoreo J J *et al* 2015 Crystallization by particle attachment in synthetic, biogenic, and geologic environments *Science* **349** aaa6760

[72] McMillan P F 2002 New materials from high-pressure experiments *Nat. Mater.* **1** 19–25

[73] De Yoreo J J *et al* 2016 *Basic Research Needs for Synthesis Science for Energy Relevant Technology* (Washington, DC: Department of Energy) (https://science.energy.gov/~/media/bes/pdf/reports/2017/BRN_SS_Rpt_web.pdf)

[74] Arjmand M, Ke J H and Szlufarska I 2017 Control of surface induced phase separation in immiscible semiconductor alloy core-shell nanowires *Comput. Mater. Sci.* **130** 50–5

[75] Jin W *et al* 2017 Electronic structure of the metastable epitaxial rock-salt SnSe {111} topological crystalline insulator *Phys. Rev.* X **7** 041020

[76] Ross F M 2015 Opportunities and challenges in liquid cell electron microscopy *Science* **350** aaa9886

[77] Karpatne A, Atluri G, Faghmous J H, Steinbach M, Banerjee A, Ganguly A, Shekhar S, Samatova N and Kumar V 2017 Theory-guided data science: a new paradigm for scientific discovery from data *IEEE Trans. Knowl. Data Eng.* **29** 2318–31

[78] Zhang Z, Penev E S and Yakobson B I 2016 Polyphony in B flat *Nat. Chem.* **8** 525–7

[79] Toberer E S, Gorai P and Stevanovic V 2017 Discovery and design of new thermoelectric materials *Materials Aspects of Thermoelectricity* ed C Uher (Boca Raton, FL: CRC Press)

[80] Gorai P, Stevanovic V and Toberer E 2017 Computationally guided discovery of thermoelectric materials *Nat. Rev. Mater.* **2** 17053

[81] Baranowski L L, Warren E L and Toberer E S 2014 High-temperature high-efficiency solar thermoelectric generators *J. Electron. Mater.* **43** 2348–55

[82] Madsen G K H 2006 Automated search for new thermoelectric materials: the case of LiZnSb *J. Am. Chem. Soc.* **128** 12140

[83] Wang S, Wang Z, Setyawan W, Mingo N and Curtarolo S 2011 Assessing the thermoelectric properties of sintered compounds with high-throughput *ab initio* calculations *Phys. Rev.* X **1** 021012

[84] Yan J, Gorai P, Ortiz B, Miller S, Barnett S A, Mason T, Stevanovic V and Toberer E S 2015 Material descriptors for thermoelectric performance *Energy Environ. Sci.* **8** 983

[85] Miller S A *et al* 2017 Capturing anharmonicity in a lattice thermal conductivity model for high-throughput predictions *Chem. Mater.* **29** 2494

[86] He J, Amsler M, Xia Y, Naghavi S S, Hegde V I, Hao S, Goedecker S, Ozolin V and Wolverton C 2016 Ultralow thermal conductivity in full heusler semiconductors *Phys. Rev. Lett.* **117** 046602

[87] Freysoldt C, Grabowski B, Hickel T, Neugebauer J, Kresse G, Janotti A and Van de Walle C G 2014 First-principles calculations for point defects in solids *Rev. Mod. Phys.* **86** 253

[88] Goyal A, Gorai P, Peng H, Lany S and Stevanovic V 2017 A computational framework for automation of point defect calculations *Comput. Mater. Sci.* **130** 1

[89] Saparov B and Mitzi D B 2016 Organic–inorganic perovskites: structural versatility for functional materials design *Chem. Rev.* **116** 4558–96

[90] Yang W S *et al* 2017 Iodide management in formamidinium-lead-halide–based perovskite layers for efficient solar cells *Science* **356** 1376–9

[91] Kojima A, Teshima K, Shirai Y and Miyasaka T 2009 Organometal halide perovskites as visible-light sensitizers for photovoltaic cells *J. Am. Chem. Soc.* **131** 6050–1

[92] Im J-H, Lee C-R, Lee J-W, Park S-W and Park N-G 2011 6.5% efficient perovskite quantum-dot-sensitized solar cell *Nanoscale* **3** 4088

[93] Kim H-S *et al* 2012 Lead iodide perovskite sensitized all-solid-state submicron thin film mesoscopic solar cell with efficiency exceeding 9% *Sci. Rep.* **2** 591

[94] Lee M M, Teuscher J, Miyasaka T, Murakami T N and Snaith H J 2012 Efficient hybrid solar cells based on meso-superstructured organometal halide perovskites *Science* **338** 643–7

[95] Snaith H J, Abate A, Ball J M, Eperon G E, Leijtens T, Noel N K, Stranks S D, Wang J T W, Wojciechowski K and Zhang W 2014 Anomalous hysteresis in perovskite solar cells *J. Phys. Chem. Lett.* **5** 1511–5

[96] Eames C, Frost J M, Barnes P R F, O'Regan B C, Walsh A and Islam M S 2015 Ionic transport in hybrid lead iodide perovskite solar cells *Nat. Commun.* **6** 7497

[97] Beecher A N, Semonin O E, Skelton J M, Frost J M, Terban M W, Zhai H, Alatas A, Owen J S, Walsh A and Billinge S J L 2016 Direct observation of dynamic symmetry breaking above room temperature in methylammonium lead iodide perovskite *ACS Energy Lett.* **1** 880–7

[98] Lee J-W, Kim H-S and Park N-G 2016 Lewis acid-base adduct approach for high efficiency perovskite solar cells *Acc. Chem. Res.* **49** 311–9

[99] Zhao W, Li S, Yao H, Zhang S, Zhang Y, Yang B and Hou J 2017 Molecular enables optimization over 13% efficiency in organic solar cells *J. Am. Chem. Soc.* **139** 7148–51

[100] Borroughes J H, Bradley D D C, Brown A R, Marks R N, Mackay K, Friend R H, Burns P L and Holmes A B 1990 Light-emitting diodes based on conjugated polymers *Nature* **347** 539–41

[101] Gómez-Bombarelli R *et al* 2016 Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach *Nat. Mater.* **15** 1120–7

[102] Nayak P K, Garcia-Belmonte G, Kahn A, Bisquert J and Cahen D 2012 Photovoltaic efficiency limits and material disorder *Energy Environ. Sci.* **5** 6022–39

[103] Gevorgyan S A *et al* 2016 Baselines for lifetime of organic solar cells *Adv. Energy Mater.* **6** 1600910

[104] Baranczak A, Tu N P, Marjanovic J, Searle P A, Vasudevan A and Djuric S W 2017 Integrated platform for expedited synthesis–purification–testing of small molecule libraries *ACS Med. Chem. Lett.* **8** 461–5

[105] Fratini S, Ciuchi S, Mayou D, Trambly de Laissardière G and Troisi A 2017 A map of high-mobility molecular semiconductors *Nat. Mater.* **16** 998–1002

[106] Guilbert A A, Zbiri M, Dunbar A D F and Nelson J 2017 Quantitative analysis of the molecular dynamics of P3HT:PCBM bulk heterojunction *J. Phys. Chem.* B **121** 9073–80

[107] Er S, Suh C, Marshak M P and Aspuru-Guzik A 2015 Computational design of molecules for an all-quinone redox flow battery *Chem. Sci.* **6** 855–93

[108] Nikolaev P, Hooper D, Webber F, Rao R, Decker K, Krein M, Poleski J, Barto R and Maruyama B 2016 Autonomy in materials research: a case study in carbon nanotube growth *NPJ Comput. Mater.* **2** 16031

[109] Pulido A *et al* 2017 Functional materials discovery using energy-structure-function maps *Nature* **543** 657–64

[110] Tsao J Y, Crawford M H, Coltrin M E, Fischer A J, Koleske D D, Subramania G S, Wang G T, Wierer J J and Karlicek R F 2014 Toward smart and ultra-efficient solid-state lighting *Adv. Opt. Mater.* **2** 809–36

[111] Verzellesi G, Saguatti D, Meneghini M, Bertazzi F, Goano M, Meneghesso G and Zanoni E 2013 *J. Appl. Phys.* **114** 071101

[112] Patel D, Pikal J M, Menoni C S, Thomas K J, Kish F A and Hueschen M R 1999 *Appl. Phys. Lett.* **75** 3201

[113] Christian T M, Beaton D A, Mukherjee K, Alberi K, Fitzgerald E A and Mascarenhas A 2013 Efficiency droop in InGaN/GaN blue light-emitting diodes: physical mechanisms and remedies *J. Appl. Phys.* **114** 074505

[114] Han L, Kash K and Zhao H 2014 High-efficiency green light-emitting diodes based on InGaN-ZnGeN2 type-II quantum wells *Proc. SPIE* **9003** 90030W

[115] Tan Z-K *et al* 2014 Bright light-emitting diodes based on organometal halide perovskite *Nat. Nanotechnol.* **9** 687

[116] Tan C-K, Zhang J, Li X-H, Liu G, Tayo B O and Tansu N 2013 First-principle electronic properties of dilute-As GaNAs alloy for visible light emitters *IEEE/OSA J. Display Technol.* **9** 6473846

[117] Phillips J M *et al* 2007 Research challenges to ultra-efficient inorganic solid-state lighting *Laser Photon. Rev.* **1** 307–33
[118] US Department of Energy 2016 Solid-State Lighting R&D Plan
[119] Dorenbos P 2004 Energy of the first $4f^7 \rightarrow 4f^6$ 5d transition of Eu2+ in inorganic compounds *J. Lumin.* **104** 239
[120] Dorenbos P 2002 Relating the energy of the $[Xe]5d^1$ configuration of Ce^{3+} in inorganic compounds with anion polarizability and cation electronegativity *Phys. Rev.* B **65** 235110
[121] Garcia-Santamaria F, Murphy J E, Setlur A A and Sista S P 2018 *ECS J. Solid State Sci. Technol.* **7** R3030–3
[122] Böller B, Ehrensperger M and Wintterlin J 2015 *In situ* scanning tunneling microscopy of the dissociation of CO on Co(0 0 0 1) *ACS Catal.* **5** 6802–6
[123] Jørgensen M and Grönbeck H 2017 Scaling relations and kinetic Monte Carlo simulations to bridge the materials gap in heterogeneous catalysis *ACS Catal.* **7** 5054–61
[124] Nørskov J K, Bligaard T, Rossmeisl J and Christensen C H 2009 Towards the computational design of solid catalysts *Nat. Chem.* **1** 37
[125] Pidko E A 2017 Toward the balance between the reductionist and systems approaches in computational catalysis: model versus method accuracy for the description of catalytic systems *ACS Catal.* **7** 4230–4
[126] Greeley J 2016 Theoretical heterogeneous catalysis: scaling relationships and computational catalyst design *Annu. Rev. Chem. Biomol. Eng.* **7** 605–35
[127] Ma X, Li Z, Achenie L E K and Xin H 2015 Machine-learning-augmented chemisorption model for CO_2 electroreduction catalyst screening *J. Phys. Chem. Lett.* **6** 3528–33
[128] Dou J, Sun Z, Opalade A A, Wang N, Fu W and Tao F 2017 Operando chemistry of catalyst surfaces during catalysis *Chem. Soc. Rev.* **46** 2001–27
[129] Thygesen K S and Jacobsen K W 2016 Making the most of materials computations *Science* **354** 180
[130] Yan Q *et al* 2017 Solar fuels photoanode materials discovery by integrating high-throughput theory and experiment *Proc. Natl Acad. Sci.* **114** 3040–3
[131] Tarascon J M and Armand M 2001 Issues and challenges facing rechargeable lithium batteries *Nature* **414** 359–67
[132] Thackeray M M, Johnson C S, Vaughey J T, Li N and Hackney S A 2005 *J. Mater. Chem.* **15** 2257–67
[133] Li H, Huang X J, Chen L Q, Wu Z G and Liang Y 1999 A high capacity nano Si composite anode material for lithium rechargeable batteries *Electrochem. Solid-State Lett.* **2** 547–9
[134] Jain A, Hautier G, Moore C J, Ong S P, Fischer C C, Mueller T, Persson K A and Ceder G 2011 A high-throughput infrastructure for density functional theory calculations *Comput. Mater. Sci.* **50** 2295–10
[135] Curtarolo S *et al* 2012 AFLOWLIB.ORG: a distributed materials properties repository from high-throughput *ab initio* calculations *Comput. Mater. Sci.* **58** 227–35
[136] Xiao R J, Li H and Chen L Q 2015 High-throughput design and optimization of fast lithium ion conductors by the combination of bond-valence method and density functional theory *Sci. Rep.* **5** 14227
[137] Hautier G, Jain A, Chen H, Moore C, Ong S P and Ceder G 2011 Novel mixed polyanions lithium ion battery cathode materials predicted by high-throughput *ab initio* computations *J. Mater. Chem.* **21** 17147–53
[138] Wang X L, Xiao R J, Li H and Chen L Q 2017 Oxysulfide LiAlSO: a lithium superionic conductor from first principles *Phys. Rev. Lett.* **118** 195901
[139] Knauth P 2009 Inorganic solid Li ion conductors: an overview *Solid State Ion.* **180** 14–6
[140] Wang X L, Xiao R J, Li H and Chen L Q 2017 Quantitative structure-property relationship study of cathode volume changes in lithium ion batteries using *ab initio* and partial least squares analysis *J. Materiomics* **3** 178–83
[141] Senkov O N, Miller J D, Miracle D B and Woodward C 2015 Accelerated exploration of multi-principal element alloys with solid solution phases *Nat. Commun.* **6** 6529
[142] Dubois J-M and Belin-Ferré E 2011 *Complex Metallic Alloys: Fundamentals and Applications* (New York: Wiley)
[143] Fischer C C, Tibbetts K J, Morgan D and Ceder G 2006 Predicting crystal structure by merging data mining with quantum mechanics *Nat. Mater.* **5** 641–6
[144] Zarnetta R, König D, Zamponi C, Aghajani A, Frenzel J, Eggeler G and Ludwig A 2009 R-phase formation in $Ti_{39}Ni_{45}Cu_{16}$ shape memory thin films and bulk alloys—discovered by combinatorial methods *Acta Mater.* **57** 4169–77
[145] Zarnetta R *et al* 2010 Identification of quaternary shape memory alloys with near-zero thermal hysteresis and unprecedented functional stability *Adv. Funct. Mater.* **20** 1917–23
[146] Naujoks D *et al* 2017 Identification of a ternary μ-phase in the Co–Ti–W system—an advanced correlative thin film-bulk combinatorial materials investigation *Acta Mater.* **138** 100–10
[147] Stein H, Naujoks D, Grochla D, Khare C, Gutkowski R, Grützke S, Schuhmann W and Ludwig A 2015 A structure zone diagram obtained by simultaneous deposition on a novel step heater: a case study for Cu_2O thin films *Phys. Status Solidi* a **212** 2798–804
[148] Ding S, Liu Y, Li Y, Liu Z, Sohn S, Walker F J and Schroers J 2014 Combinatorial development of bulk metallic glasses *Nat. Mater.* **13** 494–500
[149] Mugnaioli E, Gorelik T and Kolb U 2009 '*Ab initio*' structure solution from electron diffraction data obtained by a combination of automated diffraction tomography and precession technique *Ultramicroscopy* **109** 758–65
[150] Li Y J, Savan A, Kostka A, Stein H S and Ludwig A 2018 Accelerated atomic-scale exploration of phase evolution in compositionally complex materials *Mater. Horiz.* **5** 86–92
[151] Schlom D G, Chen L Q, Fennie C J, Gopalan V, Muller D A, Pan X, Ramesh R and Uecker R 2014 Elastic strain engineering of ferroic oxides *MRS Bull.* **39** 118–30
[152] Martin L W and Rappe A M 2017 Thin-film ferroelectric materials and their applications *Nat. Rev. Mater.* **2** 16087
[153] Agar J C, Pandya S, Xu R, Yadav A K, Liu Z, Angsten T, Saremi S, Asta M, Ramesh R and Martin L W 2016 Frontiers in strain-engineered multifunctional ferroic materials *MRS Commun.* **6** 151–66
[154] Yang Z, Ko C and Ramanathan S 2011 Oxide electronics utilizing ultrafast metal-insulator transitions *Annu. Rev. Mater. Res.* **41** 337–67
[155] Young S M and Rappe A M 2012 First principles calculation of the shift current photovoltaic effect in ferroelectrics *Phys. Rev. Lett.* **109** 116601
[156] Spanier J E *et al* 2016 Power conversion efficient exceeding the Shockley–Queisser limit in a ferroelectric insulator *Nat. Photon.* **10** 611–6
[157] Moya X, Kar-Narayan S and Mathur N D 2014 Caloric materials near ferroic phase transitions *Nat. Mater.* **13** 439–50
[158] www.mgi.gov/
[159] Ceder G and Persson K 2013 The stuff of dreams *Sci. Am.* **309** 36–40
[160] Angsten T, Martin L W and Asta M 2017 Orientation dependent properties of epitaxially strained perovskite

oxide thin films: insights from first-principles calculations *Phys Rev.* B **95** 174110
[161] Ginley D S and Perkins J D 2011 Transparent conductors *Handbook of Transparent Conductors* ed D S Ginley (Boston, MA: Springer) pp 1–25
[162] Badeker K 1907 Über die elektrische leitfähigkeit und die thermoelektrische kraft einiger schwermetallverbindungen *Ann. Phys.* **22** 749
[163] Edwards P P, Porch A, Jones M O, Morgan D V and Perks R M 2004 Basic materials physics of transparent conducting oxides *Dalton Trans.* **0** 2995–3002
[164] Nomura K, Ohta H, Takagi A, Kamiya T, Hirano M and Hosono H 2004 Room-temperature fabrication of transparent flexible thin-film transistors using amorphous oxide semiconductors *Nature* **432** 488–92
[165] Bellet D, Lagrange M, Sannicolo T, Aghazadehchors S, Nguyen V H, Langley D P, Muñoz-Rojas D, Jiménez C, Bréchet Y and Nguyen N D 2017 Transparent electrodes based on silver nanowire networks: from physical considerations towards device integration *Materials* **10** 570
[166] Ellmer K 2012 Past achievements and future challenges in the development of optically transparent electrodes *Nat. Photon.* **6** 809–17
[167] Schimpf A M, Lounis S D, Runnerstrom E L, Milliron D J and Gamelin D R 2015 Redox chemistries and plasmon energies of photodoped In_2O_3 and Sn-doped In_2O_3 (ITO) nanocrystals *J. Am. Chem. Soc.* **137** 518–24
[168] Walsh A 2008 Nature of the band gap of In_2O_3 revealed by first-principles calculations and x-ray spectroscopy *Phys. Rev. Lett.* **100** 167402
[169] Hautier G, Miglio A, Ceder G, Rignanese G M and Gonze X 2013 Identification and design principles of low hole effective mass p-type transparent conducting oxides *Nat. Commun.* **4** 2292
[170] Basic Research Needs for Synthesis Science—DOE Report on the Basic Energy Sciences Workshop on Synthesis Science for Energy Relevant Technology (https://science.energy.gov/~/media/bes/pdf/reports/2017/BRN_SS_Rpt_web.pdf)