

# Accelerating functional materials discovery

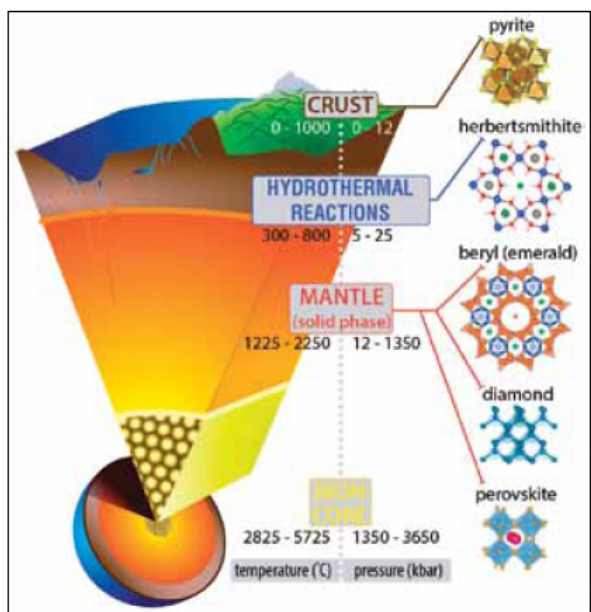
Insights from geological sciences, data-driven approaches, and computational advances

By James M. Rondinelli, Nicole A. Benedek, Danna E. Freedman, Abby Kavner, Efrain E. Rodriguez, Eric S. Toberer, and Lane W. Martin

The first materials researchers were Earth scientists, harnessing the physical and chemical properties of found minerals and transforming them to perform increasingly useful tasks. For many millennia, civilizations' knowledge of materials engineering remained rudimentary—more of an art than a science. But, during the past century, our materials knowledge has grown at a staggering pace, enabling scientists and engineers to design and produce ever more sophisticated materials.

This transition from relying on found materials to directed design of materials now shapes the way we live. For better or worse, we are inundated by an immense number of increasingly complex technologies enabled by the discovery of advanced electronic, energy, mechanical, optical, and other materials, which, in turn, has enabled an interconnected global community.<sup>1,2</sup> Nonetheless, modern societies continue to face growing global economic, security, energy, and healthcare demands that require advanced materials with superlative functionality.<sup>2</sup> Since 2011, the Materials Genome Initiative (MGI) program<sup>3</sup> in the United States has sought to address those challenges by supporting programs that aim to halve the time from discovery of new materials to integration and commercial deployment, in part by focusing on the strategic design of materials.

This initiative is transforming how we think about materials design. Today's materials scientists are equipped with advanced computational, synthesis, and characterization tools, and they have an extensive knowledge base generated by the advances of the past century. This repertoire allows for unprecedented discovery of engineering materials from an extensive range of minerals displaying a rich diversity of physical and chemical properties. The new "directed design concept" seeks to achieve fine control of material composi-



**Figure 1.** The Earth as a synthetic vessel with minerals forming in different regions highlighted. Each mineral phase in this figure also is synthesized in laboratories under controlled environments that mimic geological conditions for its useful properties and potential applications.

tion, structure, and assembly over broad length scales through an informed exploration of (parameter and composition) phase space, using experimental and theoretical approaches, to realize potentially superior material properties and function. The early stages of materials design rely on harnessing available computational tools to

- Formulate materials structure–property–functionality databases from a plethora of diverse data;
- Establish protocols to search through that data;
- Guide efficient selection of material compositions; and
- Direct experimental synthesis and subsequent integration.

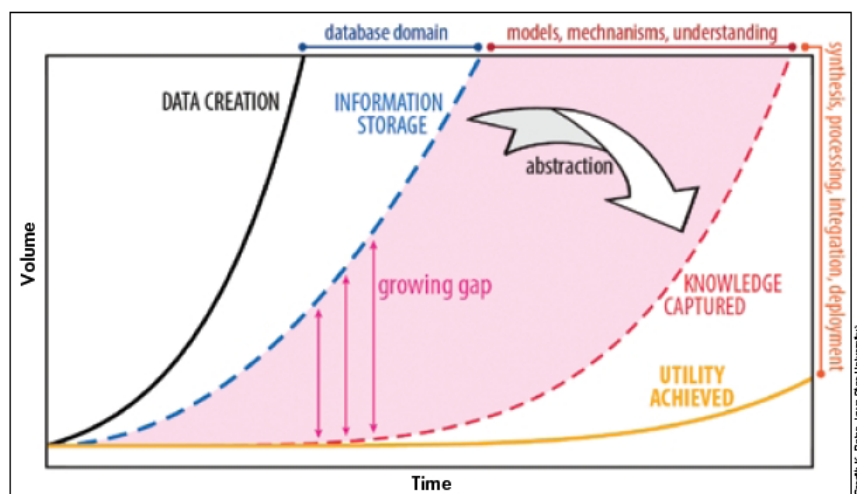
Plainly stated, the goal of the modern materials explorer is to take advantage of enhanced computing power combined with understanding of networked systems to efficiently search for new compounds or new uses for existing compounds, understand the origin of their properties, and design improved materials based on that knowledge. Our task is to advance the path from materials design to synthesis to application.

Such challenges are not unique to any one discipline. With this in mind, the US National Science Foundation supported a workshop entitled the "Materials Genome Initiative in

Ceramics, Geosciences, and Solid-State Chemistry”<sup>4</sup> in February 2013 to assess the confluence of research activities in the mineral physics, ceramics, solid-state chemistry, and computational materials communities. The goal was to identify and highlight pathways to propel materials development in the 21<sup>st</sup> century vis-à-vis new datasets, methodologies, predictive models, and materials understanding paradigms. Workshop participants, who included the authors of this article, were early-career researchers and well-established scientists who possess expertise in geosciences, inorganic materials synthesis, characterization, theory, and high-throughput combinatorial and informatics techniques.

Interactions among these communities could provide new insights and rapid analysis of new classes of materials as never before. For example, consider the tremendous contributions made by the field of geology to structural chemistry, materials physics, and mineral discovery. This research has been accomplished with what might be the ultimate high-throughput and dynamic synthesis apparatus able to explore temperature, pressure, and solvent conditions—the Earth itself (Figure 1). In-situ characterization of the mineral structure of the Earth’s interior and laboratory experiments have provided petrologists, geochemists, and mineral physicists the opportunity to measure a variety of physical and chemical properties of Earth materials. Indeed, such measurements often have been performed at extreme conditions (high pressures and temperatures) and over wide variations in composition.

The same physical laws govern engineered and natural materials. Therefore, the structural complexity and plethora of mineral phases and allotropes existing at high pressures and temperatures provide an enormous additional database to combine with engineering materials databases and inform design strategies for new compounds. Cross-fertilization of the materials and geological communities has already proved successful. For example, the mineral herbertsmithite ( $\text{ZnCu}_3(\text{OH})_6\text{Cl}_2$ ), which is found in hydrothermal veins in the Atacama region of Chile,<sup>5</sup> can be produced in the



**Figure 2.** Translating abundant materials data into useful information and ultimately into knowledge for discovery and technology advancement is challenging. “Minding the [data] gap” in materials science requires that formulation of models and mechanisms occurs at multiple length and time scales simultaneous with data generation.<sup>14</sup> Predictive theories and database searches can then identify materials compositions with optimal functionality for synthesis and engineered technologies.

laboratory using a hydrothermal reaction in a sealed vessel under pressure.<sup>6,7</sup> The synthetically derived mineral was recently determined to be the first two-dimensional spin liquid, a new phase of matter with significant implications for condensed matter physics.<sup>8</sup>

Hydrothermal synthesis of minerals originally identified from their formation in hydrothermal veins also led to the synthesis of  $\text{Li}_x\text{CoO}_2$  (which is the primary cathode material in lithium-ion batteries) as well as the development of  $\text{Li}_x\text{FeO}_4$  (olivine) for lithium ion batteries<sup>9</sup> and skutterudites ( $\text{CoAs}_3$ ) (which led to the production of promising ternary and filled skutterudite phases for thermoelectric applications).<sup>10</sup> The Earth’s mantle contains many solid mineral phases, including the most abundant magnesium-rich silicate perovskites. Many syntheses of perovskite oxide electroceramics, which find use in piezoelectric applications<sup>11</sup> and as high-temperature superconductors,<sup>12</sup> mimic the Earth’s synthesis and rely on high-temperature and high-pressure solid-state reactions.

By building bridges among Earth scientists and traditional materials science communities, we believe it is possible to accelerate materials discovery, design, and synthesis schemes by combining the vast databases of measured physical and chemical properties of natural and engineered materials. It is becoming ever more apparent that boundaries between

researchers, often trained in a single discipline, such as experimental lab work or computational theory, are rapidly dissolving. New blended educational platforms are needed to train scientists and engineers who will discover and improve on the materials of tomorrow. This article highlights some of the illustrative examples, critical challenges, and future opportunities discussed by those in attendance at the workshop with a goal to reach out to the multiple communities we represent.

### The data challenge

Modern advances in high-throughput computational and experimental materials science have made the field incredibly data-rich—nearly as rich as the Earth sciences, where the “EarthCube” initiative recently has been instituted to integrate increasingly vast amounts of geospatial data.<sup>13</sup> Measured physical and chemical properties of minerals represent a small subset of Earth science data and share many attributes with materials science data—structure, properties, chemistry. We are deluged with data, and the data is wide ranging in type, method, number, and complexity, leading to a tremendous gap between basic research and the development of new engineered materials (Figure 2). This begs the question of how we best translate data from a variety of fields into useful information and, subsequently, knowledge, and, ultimately, understand-



## Accelerating functional materials discovery

ing for materials design.

Data repositories have been generated by specific fields. For example, the PetDB project collects compositions of rocks on the Earth's surface,<sup>15</sup> and the Materials Project<sup>16</sup> contains property information (crystal structures, band structures, and formation energies). But, there is no common exchange mechanism to bring these disparate and varied sources together for synchronous search through published raw data or composite quantities.<sup>17</sup> That could change with a combination of real incentives and suggestions that federally funded published research data be made accessible in digital formats. At the same time, institutions can recognize the impact of time spent on altruistic efforts toward making data available and not simply reward scientific publication as the sole form of data dissemination and intellectual contribution. In fact, expedient dissemination of data to a repository may have greater impact on a community than publication of those results

in any single paper—in aggregate form this data can be translated to knowledge where it was never intended to apply.

Effective datasets available for searching and obtaining structure–property correlations are limited not by scientific and scholarly research efforts, but broadly by current data management plans.<sup>18</sup> Provided the materials databases are searchable, we then must understand how to navigate through this space to find materials for specific applications, because there is enormous potential in “mining” available data sets.<sup>19</sup> Many relevant databases and materials property repositories (see sidebar this page) already exist, yet unifying and efficiently accessing these databases are challenging because of the lack of universal descriptors for material properties, general agreement on what properties are most relevant, and software tools to query and analyze the data. This problem is effectively recast as a high-dimensional statistical correlation and classification challenge, and

suitable machine learning informatics techniques exist to solve it. However, the materials community has been slow to adopt such approaches. This could be in part because a priori identification of the suitable property descriptor or metric correlated with the property of interest (see sidebar, p. 18), which could itself be difficult to compute or measure, is needed to facilitate the search. Such complications result when the property is not a simple thermodynamic quantity, but rather a composite or microstructure-sensitive quantity.

### Data storage, formats, and redundancy

Excitement from participation in multidisciplinary efforts can fade to frustration if too few guidelines exist. At present, many materials researchers are ill-prepared to effectively analyze the data already available for a variety of reasons, including disparate formats, syntax, and conventions. In this capacity, consolidating and constructing new repositories and portals in cooperation with library scientists is necessary to harness the ever-increasing amount of underutilized data.

A successful model for building such materials properties databases is the MatNavi Materials Database<sup>43</sup> by the National Institute of Materials Science (NIMS) in Japan. It includes downloadable databases on properties, such as superconductivity (SuperCon<sup>44</sup>), among others, useful to scientists and engineers. An interesting possibility, similar to that of the NIMS MatNavi effort, would include an initiative led by, but not necessarily centralized at, various US national laboratories. The initiative would require sustained efforts and perhaps a conscious shift in thinking of data as a key infrastructural component to the research enterprise. Already, the National Institutes of Health announced funding to support creation of new centers to improve how the biomedical community uses large and complex data sets with the aims of reducing cost and redundancy and ensuring translation across fields.<sup>45</sup> Keeping the effort as democratic as possible would spread the workload of

### Toward interactive and open-source material property libraries

An important aspect of data unification across disciplines concerns how to effectively host and access new data. Already there are central or distributed open-access repositories, to which future data can be added. Notable examples include the crystal structure and diffraction databases (both inorganic and organic), property-oriented databases such as the Landolt-Börnstein Database,<sup>20</sup> and phase diagram databases such as the ASM Alloy Phase Diagram Center.<sup>21</sup> Some additional open-source structure–property resources include

- Materials Measurement Laboratory Data,<sup>22</sup> NIST;
- 3D Materials Atlas,<sup>23</sup> TMS;
- RRUFF Project,<sup>24</sup> database of Raman spectra, X-ray diffraction, and Chemical information for minerals, hosted at Arizona University;
- American Mineralogist Crystal Structure Database;<sup>25</sup>
- MELTS: Rocks & Magma Thermodynamics and phase diagram calculator.<sup>26</sup>

Although these databases are extensive and indispensable, many have expensive licenses that could be prohibitive for many university libraries or other research institutions. This is indeed an opportunity where MGI can offer an exciting possibility for all materials researchers to participate as both contributors and end-users of community-based, free, and open-source databases. In this capacity, the Materials Project (materialsproject.org) has made freely available computed structural, electronic, and energetic data on over 33,000 compounds since 2011 for use by students, researchers, and industry.

There is also a need for the development of translators and analytic tools to interface between experimental and computational databases, especially because experiment provides an important test of any calculated quantity. This challenge offers exciting questions and applications for tools<sup>27</sup> available to the “big-data” and informatics communities. Modern advances in interactive web-based dissemination applications and informatics should enable more rapid and less human hour-intensive ways for researchers to automatically and instantaneously contribute, edit, and evaluate structure and property data for materials engineering uses.<sup>28,29,30</sup>

## Combinatorial materials science

The combinatorial approach is a high-throughput method to solve the problem of subtle variation in materials properties resulting from variation in material chemistry and synthesis procedures.<sup>50,51,52,53</sup> This is accomplished by constructing libraries with thousands of systematically different samples (or single samples with 2D- and 3D-variations in material type or chemistry) and subsequently characterizing their properties to find the elusive diamond in the rough. This procedure naturally produces more data than targeted synthesis methods. Therefore, data mining tools and informatics are well suited for guiding combinatorial experimentation. This approach has been quite successful in finding new electronic, catalytic, magnetic, and ferroelectric materials. To accelerate the time between materials discovery to groundbreaking and materials-based technologies, MGI would benefit tremendously from the lessons of combinatorial materials synthesis.

building such databases and widen their accessibility and use in academia and beyond. The community also can participate in data “quality control,” acting as another check on top of the peer-review process. This would be analogous to the ranking system of patterns found in the Powder Diffraction File database supplied by the International Centre for Diffraction Data, or even the ranking system for online products on a commercial website, such as Amazon. There also is a tremendous opportunity for interdisciplinary work, because it includes gathering data that is *material-specific* instead of *discipline-specific*.

Such a database should be able to account for duplication of data in various formats for various communities to strengthen the ranking and enable chemists, physicists, geologists, ceramists, and metallurgists searching for the same or similar materials (properties) to know when to push forward and when to redirect their efforts by raising awareness of the work performed by others. This is where some standardization could play a key role. A great example is the standard file format agreed upon by the International Union of Crystallography. The so-called crystallographic information file (cif) is a single text file that can be read and edited in any text editor, ensuring widespread use, and it is generated easily by most crystal structure and refinement software packages.<sup>46</sup>

For example, one could construct a material information file (mif),<sup>47</sup> similar to a cif, that could include entries on a type of property (resistivity, diffusion constant, hardness, etc.), units used,

temperature regime, and so on just as the cif has entries on space group symbol, lattice constants, etc. Not all defined fields in a cif require entries, and this could be true for the mif as well. This also is an opportunity to create theoretical mifs from computational and modeling data. Indeed, a consortium of industry and the National Institute of Standards and Technology have produced such materials properties files under the MatML effort.<sup>48</sup> They use extensible markup language to maximize flexibility and customization, enabling users to create their own tags and document structure. Thus, MatML data can serve as a starting point for the continuation of free exchange of materials properties files.

The idea of a standardized materials data reporting file raises important issues. Development of a common data format requires some consensus among an extremely broad set of fields. Will all the disparate science and engineering communities mentioned above be able to agree upon a standard format for the mif? Although the common data format has been established within fairly narrow disciplines, e.g., crystallography, it remains to be seen what form such “materials property” data sets should take. Moreover, not all data lends itself easily to text manipulation. For example, how would one deal with images such as those from electron microscopy? How can older published materials property data be extracted into a mif?

These are all challenging questions, but in the era of the Internet, supercomputers, metadata, and social networking, their answers may be closer than ever. Several Department

of Energy laboratories have adopted the portable Hierarchical Data Format<sup>49</sup> (HDF5) for data sets, which makes it possible to represent complex objects, including images, and a wide variety of metadata efficiently. An added benefit of considering these questions is pulling together researchers and ideas from a wide variety of fields.

## Reimagining materials discovery for materials by design

In seeking materials with specific properties or technological applications, the concepts of materials discovery and design have been understandably pursued. Such efforts can generate and have generated vast quantities of data. This has been especially true for an area of materials by design known as combinatorics or combinatorial science (see sidebar this page). Indeed, the need to prepare and characterize thousands of samples within a short period of time has made combinatorial approaches a predecessor to MGI, because both require organizing vast amounts of materials information in such a way that the researcher can understand the results. This is a concept successfully used by the pharmaceutical industry to design new drugs,<sup>54,55,56</sup> and it could prove just as important for materials design and discovery efforts in the years to come.

Although advances in modern synthesis methods have enabled production of diverse and exotic ranges of materials, the power of combinatorial approaches can be harnessed only when sufficient high-throughput characterization techniques are available.<sup>57</sup> While the community has excelled at collecting important structural information using localized and rapid point characterization of crystal structure, analogous methods to locally characterize physical properties are more limited. Indeed, this is an area where MGI could greatly advance the field, but it would require concerted efforts to design and implement high-throughput characterization methods for a range of physical properties that are compatible with the sheer number, size, and geometry of modern combinatorial samples.



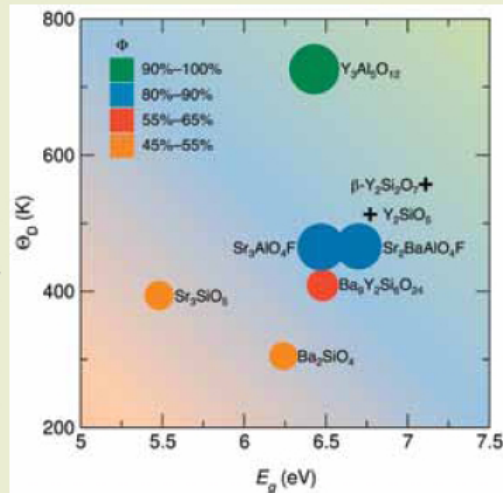
## Materials discovery with descriptors and data-mined materials parameters

Structure–property relationships form a cornerstone of materials science and have produced an array of new materials that find use today. However, in many cases there are a number of important material features, which do not solely arise from crystal or electronic structure, and the aggregate behavior governs the properties and performance of the system. In such situations, materials discovery relies on finding a known (or new) material with the optimal combination of characteristics that collectively define the materials behavior. These empirical proxies or descriptors, which may or may not be experimentally measurable, are correlated with the property of interest and are particularly powerful when a theory of the property of interest is absent. They are the ultimate materials cartography “compass” to guide discovery and synthesis, and in the process, provide some new microscopic insight.<sup>31,32,33</sup>

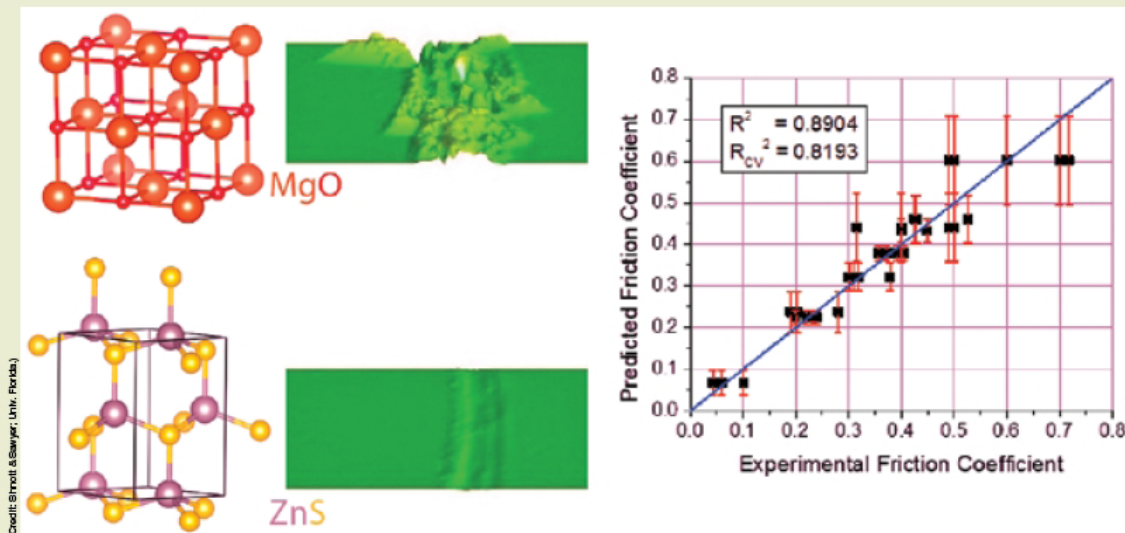
When informatics based approaches and first-principles methods like density functional theory are combined, available materials databases can be scanned in a high-throughput manner. This should enable the discovery of candidate materials by evaluating the materials descriptors.<sup>19,34,35</sup> Examining large data sets in this manner suggests not only where to look for new materials, but equally important, where not to spend time. Successful examples include uncovering new cathode,<sup>36</sup> thermoelectric,<sup>37</sup> ferroelectric,<sup>38,39</sup> efficient solar energy

conversion,<sup>40</sup> and phosphor materials (right). Outstanding challenges remain in going beyond known compositions and materials systems and in how to efficiently generate (by design) new compounds based on previous explorations and understanding.

High-throughput calculation approaches are important for searching through the large materials space; however, they are not without limits. The descriptors must be established, and there is an underlying assumption that the materials catalog is complete. Nonetheless, ex post facto identification of suitable descriptors without complete understanding of the physical mechanism is possible through statistical learning tools and covariant analysis to analyze correlations. Data-mining techniques and high-throughput screening of ceramic materials were used to formulate a predictive model for friction coefficients,<sup>42</sup> enabling the rapid identification of previously overlooked high-temperature, solid-state lubricants, for example (below).



A plot of calculated Debye temperature ( $\Theta_D$ ) of the undoped host crystal serves as a useful descriptor (proxy) for phosphor performance, i.e., a high photoluminescence quantum yield.<sup>41</sup> The Debye temperature is plotted as a function of DFT-calculated band gaps ( $E_g$ ) for the different host compounds. The circle size and color are indicative of the experimentally measured photoluminescent quantum yield ( $\Phi$ ) when  $Ce^{3+}$  is incorporated. By screening host materials for suitable  $\Theta_D$  and band gaps it is possible to find new, superior materials.

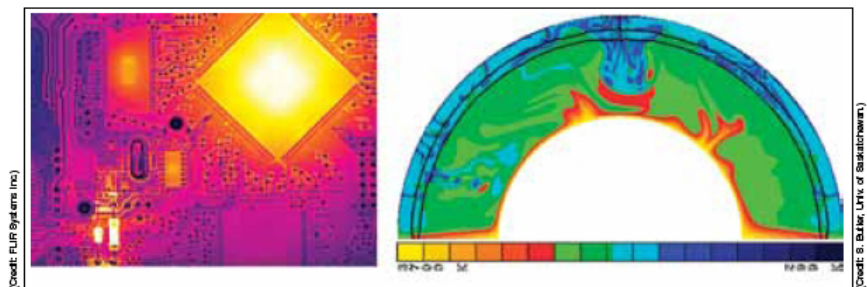


Illustrative schematic of the way the crystal structure of rocksalt (MgO) and zinc blende (ZnS), influence wear in an experimental tribometer. Predicted versus experimental friction coefficients obtained from recursive partitioning of a small dataset of intrinsic material parameters to formulate a predictive friction model capable of identifying previously unknown lubricant materials (right). The two  $R^2$  values used different amounts of data from the training data set to evaluate the accuracy of the model. (CV represents cross-validation.)

Of particular interest would be techniques that could be used in-situ or during synthesis in conjunction with rapid feedback to the synthesis process. Instead of serially producing samples and then focusing on areas of interest, in-situ characterization on the fly could enable more rapid and focused work that brings high-performance materials to the forefront faster. Only with these high-throughput and spatially resolved characterization methods of materials properties can the screening of new candidate materials, optimal growth conditions, and the like be accelerated.

Another interesting concept introduced by the combinatorial approach is that of delineating correlations between various materials properties or descriptors, such as density, electronegativity of the constituent elements, optical absorption, or lattice parameters. To increase utility of such descriptors, data visualization is key. Researchers need to be able to interpret large amounts of information in ways that convey understanding to close the data gap and provide for predictive design or selection. Visualizing data in the form of correlation or Ashby diagrams,<sup>58</sup> for example, would allow researchers to observe relationships between materials properties and descriptors previously unknown. Some relations may not be obvious or even suspected, aiding discovery and formulation of new microscopic theories and allowing materials researchers to leave serendipitous trial-and-error processes in the past and proceed to the stage of directed design and synthesis of materials with superlative properties.<sup>59</sup> Indeed, pursuing materials design in this way could even lead to fundamental discoveries of new phenomena.

Rapid experimentation is required, and focused experimentation on areas of weakness in our fundamental understanding is especially important. Another area where focused efforts can have multipronged impact is in identification and subsequent study of material properties that are important for scientists, engineers, and Earth scientists alike. One example of such a field of study that could have lasting impacts on a range of communities is the study



**Figure 3. Thermal image of a computer microchip under operation<sup>60</sup> (left) and snapshot of the temperature field calculated from an axisymmetric numerical model of convection in Earth's mantle that includes the effects of mantle phase transitions and viscosity with depth (right). Understanding of thermal transport and conduction is essential to engineering new materials.**

of the thermal properties of materials (Figure 3). In complex materials systems, such as those found in the Earth's crust, thermal convection, mantle transitions, and avalanches govern the dynamic of the Earth and how heat gets from the interior to the exterior to drive plate tectonics, volcanoes, and earthquakes. In attempting to understand how heat flow helps govern the dynamic behavior of the Earth, geologists have used increasingly sophisticated computational and experimental techniques to study properties, such as thermoelasticity and thermal and electrical conductivity, at high pressures and temperatures.

Similar materials are known to possess exotic thermal properties that are incredibly sensitive to processing, doping, and structure and, at the same time, have potential for applications such as thermoelectrics and waste heat energy conversion. The systematic study of thermal properties in certain classes of materials, in turn, could immediately inform and impact the nature of materials discovery and design across the spectrum from materials physics to engineering to Earth sciences.

### The science of synthesis

The increasing power and sophistication of first-principles calculations has brought us closer to the dream of materials-by-design. The literature now contains many examples of materials that have been either designed or discovered entirely *in-silico* or that have had their properties optimized (or new properties discovered) through computational exploration. The future seems

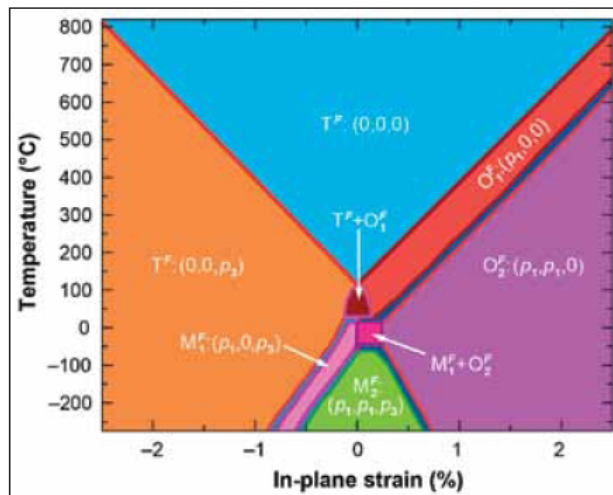
to promise that it will be possible to design function into materials, but will scientists and engineers be able to make them? Such materials predictions are often missing an important piece of information: Under which experimental conditions should the material be synthesized? Although the fundamental thermodynamic and kinetic theories have been known for some time, it is not always straightforward to implement them within a computational or theoretical framework.

One successful example is the mapping of the phase diagrams<sup>61,62</sup> of thin-film ferroelectric perovskites as a function of epitaxial misfit strain and temperature (Figure 4). The properties of thin-film perovskites can differ dramatically from those of the bulk materials, depending on temperature and mechanical boundary conditions. Incorporating this information into theoretical treatments is crucial to making accurate predictions.

In fact, epitaxial strain and the high temperatures used in growth and stabilization of modern oxide thin-film materials have a direct analogy in geoscience (Figure 5). Application of ~1% lattice misfit strain in epitaxial thin-film growth is equivalent to applying  $10^9$ – $10^{10}$  Pa of pressure or taking your material to depths of tens of kilometers into the Earth's interior. For some materials, the strain energy introduced by lattice mismatch may stabilize an entirely different phase. Measurements of phase stability as a function of pressure help delineate when lattice mismatch may thermodynamically destabilize a phase



## Accelerating functional materials discovery



**Figure 4.** Strain-temperature phase diagram of single-domain (001)-oriented BaTiO<sub>3</sub> from theoretical calculations using a phenomenological phase-field model.<sup>63</sup> The letters T, O, and M used in the phase notations indicate tetragonal, orthorhombic, and monoclinic crystallographic symmetries, respectively, under a constraint. The components of the polarization vector  $\mathbf{p}$  corresponding to the phases (along the crystallographic directions of pseudocubic BaTiO<sub>3</sub>) are indicated within the parentheses. A description of the first-principles treatment is found in Ref. 62.

near an electronic or structural phase boundary.<sup>64</sup>

Another area ripe for study is the effect of imperfections on the evolution of material properties. Variations in chemistry long have been used to drive property evolution in materials, but increasingly engineers want to control materials chemistry to finer and finer levels to exert the exacting control required for stringent applications. Approaches to understand and predict the role of nonstoichiometry on properties of materials are needed. Interfaces between materials represent a specific type of imperfection that is growing in importance and offers a unique set of challenges, because their structure and stoichiometry also can differ significantly from that of the parent materials in bulk. Locating the most stable structure and stoichiometry under a particular set of thermodynamic and kinetic conditions is a formidable problem even with the use of sophisticated computational search algorithms.<sup>65</sup>

The orientation relationship between the constituent materials is usually known, but, if it is not (or if the goal is to optimize some property of the interface as a function of the orientation

relationship), then difficulties increase. The space of possible structures is enormous, and our time and computational resources are finite. Fortunately, the development of high-throughput experimental techniques, such as combinatorial substrate epitaxy (CSE),<sup>66</sup> can help guide theorists to the most promising areas of phase space. In the CSE approach, a material is grown as a thin film on hundreds of substrates (in a single experiment) with known but various orientations. Characterizing the

films identifies film-substrate orientation relationships and steers theorists toward those interfaces actually observed in experiments.

Kinetics also is important in determining the most stable interface structure. This can be an especially important consideration in the case of thin-film interfaces, because film growth is a highly nonequilibrium process. Rohrer and Hyldgaard<sup>67</sup> recently developed a computational framework for thin-film interfaces that combines ab-initio thermodynamics with chemical reaction theory. They used this method to investigate the structures and properties of TiC/alumina interfaces at the film deposition conditions. Predicted interface wear resistance using these schemes was in much better agreement with experiment than those from a scheme that considered only thermodynamics. Such an approach thus may prove very useful in the characterization of surface and interface growth.

Finally, although there have been considerable studies on the growth of materials over the years, much of this work is limited in terms of the material systems investigated and in terms

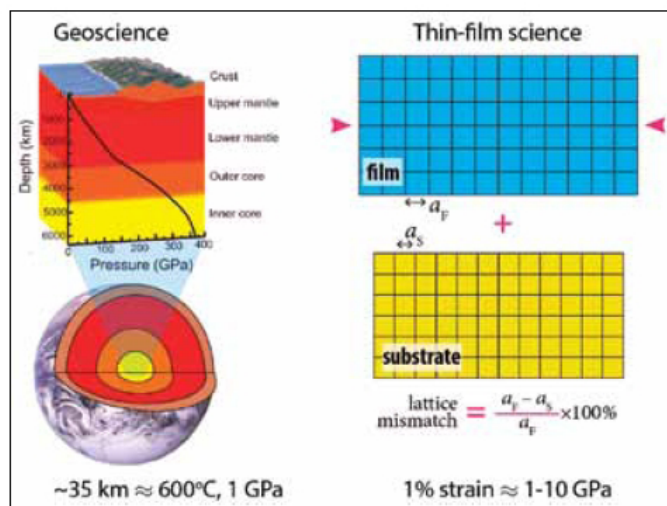
of techniques used (with much of the work relying on large-scale scientific facilities, including synchrotron-based characterization). Nonetheless, the synthesis process can have extensive and unexpected effects on the ultimate properties of complex materials.

Synthesis science must undergo a rejuvenation if it is to aid the development of more accurate computational tools and approaches, enable the prediction of and understanding of phase and structural formation during synthesis, and permit rapid development and translation of materials from the laboratory to high-scale production. During the past decade, chemists and materials scientists have made great strides synthesizing new materials with exotic properties, such as the new class of iron-based superconductors,<sup>68</sup> topological insulators,<sup>69</sup> and spin liquids.<sup>6</sup> However, these syntheses lack theoretical guidance on the environmental conditions.

One challenge for the MGI is to develop theoretical formalisms that interface with laboratory-based chemistry and materials science to guide synthetic parameters for the discovery of new materials. This challenge is reflected in the area of crystal growth. Advanced study of materials properties frequently requires large single crystals, but the know-how and experience in producing these crystals is deteriorating in the US.<sup>70</sup> Development of a theoretical basis for the complex kinetics of crystal growth could help rejuvenate this critical field by reducing the high cost associated with failed trials. It also would enhance recruitment of highly skilled chemists and materials scientists, who could approach the problem of crystal growth from a research-oriented perspective as opposed to a service-oriented task.

### Impacts from a cross-community MGI effort

The main conclusion of this article is that mineralogists, petrologists, ceramists, physicists, and solid-state chemists are now in a position to formulate strategies for a coordinated approach to discover functional materials by design while encouraging the selection of more



**Figure 5. The high pressures and temperatures present inside the Earth's crust are emulated and leveraged in modern epitaxial thin-film growth approaches to make novel materials.**

terrestrially abundant elements through environmentally conscious chemistry decisions. To this end, it will be possible to identify materials that may be poor for one application but very good for another, thereby avoiding time and money spent by various communities in “reinventing the wheel.”

The boundary between researchers whose expertise is solely in experimental lab work or computational theory is rapidly dissolving. The role of theory is not to reproduce experiments alone, but also to perform experiments that cannot necessarily be done in the “lab.” This is of particular importance in the Earth sciences, where it often is difficult to experimentally measure physical and chemical properties of materials directly at the extreme high-pressure and high-temperature conditions relevant to the deep interior of the Earth.

The effort to build materials descriptors to aid high-throughput computational searches and democratize materials databases across fields also provides an important stimulus to train undergraduate, graduate, and senior scientists in data mining and informatics. Formal training in these areas typically is not found in graduate materials programs, and it could offer a possibility for materials researchers, new and seasoned, to utilize an important development in information sciences to guide and conduct investigations.

As Linus Pauling stated,<sup>71</sup> “A chemist who understands why a diamond

has certain properties, or why nylon or hemoglobin have other properties, because of the different ways their atoms are arranged, may ask questions that a geologist would not think of formulating, unless he had been similarly trained in this way of thinking about the world.” The converse statement, regarding the distinct geological/mineralogical mindset, should be true as well. Asking and answering questions in new ways can lead to the innovations the MGI has set out to produce. Performing state-of-the-art datacentric research requires familiarity with linear algebra, data-oriented statistics, and code development—topics largely omitted in traditional arts and science educations and more commonly found in engineering disciplines. Introducing discovery-based modules within the undergraduate curricula that rely on statistical methods to establish physical laws from available experimental data would cultivate innovative scientific thinking at an early age. Summer schools with integrated materials discovery approaches should be encouraged for graduate students and faculty, whereby various skills among the experimental, simulation, and computational communities could be taught. This could alleviate communication barriers across the geoscience, ceramics, and solid-state chemistry fields. Digital recording and broadcasting of these workshops would exponentially benefit next-generation researchers in diverse disciplines.

## Acknowledgments

This workshop was supported by the National Science Foundation through the grant Nos. NSF DMR 1216415 (Ceramics) and NSF DMR 1115294 (SSMC). The authors thank all participants to the workshop for providing the content and themes presented in this article.

## About the authors

James M. Rondinelli is a faculty member in the Department of Materials Science and Engineering, Drexel University, Philadelphia, Pa. Nicole A. Benedek is a faculty member in the Materials Science and Engineering Program, University of Texas at Austin, Austin, Texas. Danna E. Freedman is a faculty member in the Department of Chemistry, Northwestern University, Evanston, Ill. Abby Kavner is a faculty member in the Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, Los Angeles, Calif. Efrain E. Rodriguez is a faculty member in the Department of Chemistry and Biochemistry, University of Maryland, College Park, Md. Eric S. Toberer is a faculty member in the Department of Physics, Colorado School of Mines, Golden, Colo. Lane W. Martin is a faculty member in the Department of Materials Science and Engineering and Materials Research Laboratory, University of Illinois, Urbana-Champaign, Urbana, Ill.

## References

- <sup>1</sup>Committee on Prospering in the Global Economy of the 21st Century: An Agenda for American Science and Technology, National Academy of Sciences, National Academy of Engineering, Institute of Medicine, Rising above the gathering storm: Energizing and employing America for a brighter economic future, The National Academies Press, Washington, D.C., 2007.
- <sup>2</sup>T.L. Friedman, *The world is flat: A brief history of the twenty-first century*. Farrar, Straus and Giroux, New York, 2005.
- <sup>3</sup>White House Office of Science and Technology Policy, Materials Genome Initiative for Global Competitiveness, June 2011 ([http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials\\_genome\\_initiative-final.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf)).
- <sup>4</sup>It followed as an outgrowth of a prior “future directions” meeting within the inorganic materials community summarized in R. Seshadri, S.L. Brock, A. Ramirez, M.A. Subramanian, and M.E. Thompson, “Advances in the development and growth of functional materials: Toward the paradigm of materials by design,” *MRS Bull.*, 37, 682–90 (2012).
- <sup>5</sup>J.L. Jambor, J.E. Dutrizac, A.C. Roberts, J.D. Grice, and J.T. Szymanski, “Clinocatacmitite, a new polymorph of  $\text{Cu}_2(\text{OH})_2\text{Cl}_2$  and its relationship to paratacmitite and ‘anarakite,’” *Can. Mineral.*, 34, 61–72 (1996).



# Accelerating functional materials discovery

- <sup>9</sup>M.P. Shores, E.A. Nytko, B.M. Bartlett, and D.G. Nocera, "A structurally perfect  $S = \frac{1}{2}$  Kagomé antiferromagnet," *J. Am. Chem. Soc.*, 127, 13462–63 (2005).
- <sup>10</sup>D.E. Freedman, T.H. Han, A. Prodi, P. Muller, Q.-Z. Huang, Y. S. Chen, S.M. Webb, Y.S. Lee, T.M. McQueen, and D.G. Nocera, "Site specific X-ray anomalous dispersion of the geometrically frustrated Kagomé magnet, herbertsmithite,  $ZnCu_2(OH)_4Cl_2$ ," *J. Am. Chem. Soc.*, 132, 16185–90 (2010).
- <sup>11</sup>T.H. Han, J.S. Helton, S. Chu, D.G. Nocera, J.A. Rodriguez-Rivera, C. Broholm, and Y.S. Lee, "Fractionalized excitations in the spin-liquid state of a Kagomé lattice antiferromagnet," *Nature*, 492, 406–11 (2013).
- <sup>12</sup>J.B. Goodenough and Y. Kim, "Challenges for rechargeable Li batteries," *Chem. Mater.*, 22, 587–603 (2010).
- <sup>13</sup>G.J. Snyder and E.S. Toberer, "Complex thermoelectric materials," *Nat. Mater.*, 7, 105–114 (2008).
- <sup>14</sup>M.G. Harwood, P. Popper, and D.F. Rushman, "Curie point of barium titanate," *Nature*, 160, 58–59 (1947).
- <sup>15</sup>(a) J.G. Bednorz and K.A. Müller (Nobel Lecture), "Perovskite-type oxides: The new approach to high-Tc superconductivity," Nobelprize.org, 14 Nov 2012 [http://www.nobelprize.org/nobel\\_prizes/physics/laureates/1987/bednorz-lecture.html](http://www.nobelprize.org/nobel_prizes/physics/laureates/1987/bednorz-lecture.html); (b) J.G. Bednorz and K.A. Müller, "Possible high-Tc superconductivity in the Ba-La-Cu-O system," *Z. Phys. B*, 64, 189–93 (1986).
- <sup>16</sup>S.E. Pratt, Digitizing Earth: Developing a cyberinfrastructure for the geosciences, American Geosciences Institute, Alexandria, Va., 2013; see also: <http://earthcube.ning.com/>.
- <sup>17</sup>K. Rajan, Informatics for materials science and engineering: Data-driven discovery for accelerated experimentation and application. Butterworth-Heinemann, Elsevier, Oxford, U.K., 2013.
- <sup>18</sup><http://www.earthchem.org/petdb>
- <sup>19</sup>A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K.A. Persson, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Mater.*, 1, 011002 (2013).
- <sup>20</sup>"Fueling discovery by sharing" (editorial), *Nat. Mater.*, 12, 173 (2013).
- <sup>21</sup>Committee on the Analysis of Massive Data; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Their Applications; Division on Engineering and Physical Sciences; National Research Council, "Frontiers in massive data analysis," National Academies Press, Washington, D.C. 2013, [http://www.nap.edu/catalog.php?record\\_id=18374](http://www.nap.edu/catalog.php?record_id=18374).
- <sup>22</sup>S. Curtarolo, G.L.W. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, "The high-throughput highway to computational materials design," *Nat. Mater.*, 12, 191–201 (2013).
- <sup>23</sup><http://www.springermaterials.com>
- <sup>24</sup><http://www.asminternational.org/AsmEnterprise/APDI>
- <sup>25</sup><http://www.nist.gov/mml/mml-data.cfm>
- <sup>26</sup><https://cosmicweb.mse.iastate.edu/wiki/display/home/Materials+Atlas+Home>
- <sup>27</sup>R.T. Downs, "The RRUFF Project: An integrated study of the chemistry, crystallography, Raman and infrared spectroscopy of minerals," 19th General Meeting of the International Mineralogical Association, Kobe, Japan, Abstract O03-13, July 23–28, 2006. (<http://rruff.info/>)
- <sup>28</sup><http://rruff.geo.arizona.edu/AMS/amcsd.php>
- <sup>29</sup><http://melts.ofm-research.org/>
- <sup>30</sup>(a) R Development Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2012. (b) M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, "The WEKA data mining software: an update," *SIKDD Explorations*, 11, 10–18 (2009).
- <sup>31</sup>S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kochner, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, and G. Ceder, "Python materials genomics (pymatgen): A robust, open-source python library for materials analysis," *Comp. Mater. Sci.*, 68, 314–19 (2013).
- <sup>32</sup>S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnatek, R.V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K.Yang, O. Levy, M.J. Mehl, H.T. Stokes, D.O. Demchenko, and D. Morgan, "AFLOW: An automatic framework for high-throughput materials discovery," *Comp. Mater. Sci.*, 58, 218–26 (2012).
- <sup>33</sup>S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnatek, R.V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K.Yang, O. Levy, M.J. Mehl, H.T. Stokes, D.O. Demchenko, and D. Morgan, "AFLOW: An automatic framework for high-throughput materials discovery," *Comp. Mater. Sci.*, 58, 227–35 (2012).
- <sup>34</sup>H. Peng, A. Zakutayev, S. Lany, T.R. Paudel, M. d'Avezac, P.F. Ndione, J.D. Perkins, D.S. Ginley, A.R. Nagaraja, N.H. Perry, T.O. Mason, and A. Zunger, "Li-doped  $Cr_2MnO_5$ : A new p-type transparent conducting oxide by computational materials design," *Adv. Func. Mater.*, (2013), 10.1002/adfm.201300807
- <sup>35</sup>G. Hautier, A. Miglio, G. Ceder, G.-M. Rignanese, and X. Gonze, "Identification and design principles of low hole effective mass p-type transparent conducting oxides," *Nat. Commun.*, 4, 2299 (2013).
- <sup>36</sup>P.V. Balachandran, S.R. Broderick, and K. Rajan, "Identifying the 'inorganic gene' for high-temperature piezoelectric perovskites through statistical learning," *Proc. R. Soc. A*, 467, 2271–90 (2011).
- <sup>37</sup>Y. Le Page, P.W. Saxe, and J.R. Rodgers, "Symmetry-general ab-initio computation of physical properties using quantum software with crystal structure databases: Results and perspectives," *Acta Crystallogr. B*, 58, 349–57 (2002).
- <sup>38</sup>M.W. Gaultois, T.D. Sparks, C.K. H. Borg, R. Seshadri, W.D. Bonificio, and D.R. Clarke, "Data-driven review of thermoelectric materials: Performance and resource considerations," *Chem. Mater.*, 25, 2911–20 (2013).
- <sup>39</sup>G. Ceder, G. Hautier, A. Jain, and S.P. Ong, "Recharging lithium battery research with first-principles methods," *Mater. Res. Soc. Bull.*, 36, 185–91 (2011).
- <sup>40</sup>S. Wang, Z. Wang, W. Setyawan, N. Mingo, and S. Curtarolo, "Assessing the thermoelectric properties of sintered compounds via high-throughput ab-initio calculations," *Phys. Rev. X*, 1, 021012 (2011).
- <sup>41</sup>W. Bennett and K.M. Rabe, "Integration of first-principles methods and crystallographic database searches for new ferroelectrics: Strategies and explorations," *J. Solid. State. Chem.*, 195, 21–31 (2012).
- <sup>42</sup>R. Amiento, B. Kozinsky, M. Fornari, and G. Ceder, "Screening for high-performance piezoelectrics using high-throughput density functional theory," *Phys. Rev. B*, 84, 014103 (2011).
- <sup>43</sup>L. Yu and A. Zunger, "Identification of potential photovoltaic absorbers based on first-principles screening of materials," *Phys. Rev. Lett.*, 108, 068701 (2012).
- <sup>44</sup>J. Brgoch, S.P. DenBaars, and R. Seshadri, "Proxies from ab-initio calculations for screening efficient  $Ce^{3+}$  phosphor hosts," *J. Phys. Chem. C*, 117, 17955–59 (2013).
- <sup>45</sup>E.W. Bucholz, C.-S. Kong, K.R. Marchman, W.G. Sawyer, S.R. Phillpot, S.B. Sinnott, and K. Rajan, "Data-driven model for estimation of friction coefficient via informatics methods," *Tribol. Lett.*, 47, 211–21 (2012).
- <sup>46</sup>The NIMS Materials Database Project MatNavi, [http://mits.nims.go.jp/index\\_en.html](http://mits.nims.go.jp/index_en.html)
- <sup>47</sup>The Superconducting Material Database, NIMS, [http://supercon.nims.go.jp/index\\_en.html](http://supercon.nims.go.jp/index_en.html)
- <sup>48</sup><https://commonfund.nih.gov/bd2k/index.aspx>
- <sup>49</sup>International tables for crystallography, Vol. C: Definition and exchange of crystallographic data. Edited by S.R. Hall and B. McMahon. International Union of Crystallography, Chester, U.K., 2006.
- <sup>50</sup>R. Seshadri and J.F. Mitchell, private communication, 2013.
- <sup>51</sup>MatML, XML for Materials Property Data, [www.matml.org](http://www.matml.org)
- <sup>52</sup>The HDF Group, Hierarchical Data Format version 5 (2000) <http://www.hdfgroup.org/>.
- <sup>53</sup>I. Takeuchi, M. Lippmaa, and Y. Matsumoto, "Combinatorial Experimentation and Materials Informatics," *Mater. Res. Soc. Bull.*, 31, 999–1003 (2006).
- <sup>54</sup>X.-D. Xiang and I. Takeuchi, *Combinatorial materials synthesis*. Marcel Dekker, New York, 2003.
- <sup>55</sup>I. Takeuchi, J. Lauterbach, and M.J. Fasolka, "Combinatorial materials synthesis," *Mater. Today*, 8, 18–26 (2005).
- <sup>56</sup>X.-D. Xiang, "Combinatorial materials synthesis and screening: An integrated materials chip approach to discovery and optimization of functional materials," *Annu. Rev. Mater. Sci.*, 29, 149–71 (1999).
- <sup>57</sup>E.J. Martin, J.M. Blaney, M.A. Siani, D.C. Spellmeyer, A.K. Wong, and W.H. Moos, "Measuring diversity: Experimental design of combinatorial libraries for drug discovery," *J. Med. Chem.*, 38, 1431–36 (1995).
- <sup>58</sup>A. Ghose and V. Viswanathan, *Combinatorial library design and evaluation: Principles, software, tools, and applications in drug discovery*. Marcel Dekker, New York, 2001.
- <sup>59</sup>H.-O. Kim and M. Kahn, "A merger of rational drug design and combinatorial chemistry development and application of peptide secondary structure mimetics," *Comb. Chem. High Throughput Scr.*, 3, 167–83 (2000).
- <sup>60</sup>H. Koinuma and I. Takeuchi, "Combinatorial solid-state chemistry of inorganic materials," *Nat. Mater.*, 3, 429–38 (2004).
- <sup>61</sup>M.F. Ashby, *Material property charts, materials selection and design*, ASM Handbook, Vol. 20, ASM International, Materials Park, Ohio, 1997.
- <sup>62</sup>(a) C.J. Fennie and K.M. Rabe, "Magnetic and electric phase control in epitaxial  $BaTiO_3$  from first-principles," *Phys. Rev. Lett.*, 97, 267602 (2006). (b) J.H. Lee, L. Fang, E. Vlahos, X. Ke, Y.W. Jung, L. Fitting-Kourkoutsis, J.-W. Kim, P.J. Ryan, T. Heeg, M. Roelkenath, V. Goian, M. Bernhagen, R. Uecker, P.C. Hammel, K.M. Rabe, S. Kamba, J. Schubert, J.W. Freeland, D.A. Muller, C.J. Fennie, P. Schiffer, V. Gopalan, E. Johnston-Halperin, and D. G. Schlom, "A strong ferroelectric ferromagnetic created by means of spin-lattice coupling," *Nature*, 466, 954–59 (2010).
- <sup>63</sup><http://www.flir.com/thermography/americas/ca/view/?id=59484&collectionid=788&col=59486>
- <sup>64</sup>N.A. Pertsev, A.G. Zembilgotov, and A.K. Tagantsev, "Effect of mechanical boundary conditions on phase diagrams of epitaxial ferroelectric thin films," *Phys. Rev. Lett.*, 80, 1988–91 (1998).
- <sup>65</sup>O. Diéguez, S. Tinte, A. Antons, C. Bungaro, J.B. Neaton, K.M. Rabe, and D. Vanderbilt, "Ab-initio study of the phase diagram of epitaxial  $BaTiO_3$ ," *Phys. Rev. B*, 69, 212101 (2004).
- <sup>66</sup>D.G. Schlom, L.-Q. Chen, C.-B. Eom, K.M. Rabe, S.K. Streiffer, and J.-M. Triscone, "Strain tuning of ferroelectric thin films," *Annu. Rev. Mater. Res.*, 37, 589–626 (2007).
- <sup>67</sup>(a) G. Catalan and J.F. Scott, "Physics and applications of bismuth ferrite," *Adv. Mater.*, 21, 246385 (2009). (b) M. Guennou, P. Bouvier, P. Toulemonde, C. Darc, C. Goujon, P. Bordet, M. Hanfland, and J. Kreisel, *Arxiv e-print: http://arxiv.org/abs/1306.5916*.
- <sup>68</sup>A.L.-S. Chua, N. A. Benedek, L. Chen, M.W. Finnis, and A.P. Sutton, "A genetic algorithm for predicting the structures of interfaces in multicomponent systems," *Nat. Mater.*, 9, 418–22 (2010).
- <sup>69</sup>Y. Zhang, A. M. Schultz, L. Li, H. Chien, P.A. Salvador, and G.S. Rohrer, "Combinatorial substrate epitaxy: A high-throughput method for determining phase and orientation relationships and its application to  $BiFeO_3/TiO_2$  heterostructures," *Acta Mater.*, 60, 6486–93 (2012).
- <sup>70</sup>J. Rohrer and P. Hyldgaard, "Computational scheme for ab-initio predictions of chemical compositions interfaces realized by deposition growth," *Comp. Phys. Commun.*, 182, 1814–18 (2011).
- <sup>71</sup>Y. Kamihara, H. Hirata, M. Hirano, R. Kawamura, H. Yanagi, T. Kamiya, and H. Hosono, "Iron-based layered superconductor:  $LaOFeP$ ," *J. Am. Chem. Soc.*, 128, 10012–13 (2006).
- <sup>72</sup>H.J. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.C. Zhang, "Topological insulators in  $Bi_2Se_3$ ,  $Bi_2Te_3$ , and  $Sb_2Te_3$  with a single Dirac cone on the surface," *Nat. Phys.*, 5, 438–42 (2009).
- <sup>73</sup>Committee for an Assessment of and Outlook for New Materials Synthesis and Crystal Growth, National Research Council, "Frontiers in crystalline matter: From discovery to technology," 2009, [http://www.nap.edu/catalog.php?record\\_id=12640](http://www.nap.edu/catalog.php?record_id=12640)
- <sup>74</sup>L. Pauling, "The place of chemistry in the integration of the sciences," *Main Currents in Modern Thought*, 7, 110 (1950). ■