RESEARCH ARTICLE | JULY 18 2013

Commentary: The Materials Project: A materials genome approach to accelerating materials innovation FREE

Anubhav Jain; Shyue Ping Ong; Geoffroy Hautier; Wei Chen; William Davidson Richards; Stephen Dacek; Shreyas Cholia; Dan Gunter; David Skinner; Gerbrand Ceder; Kristin A. Persson

Check for updates

View Online

Export Citation

Commentary: The Materials Project: A materials genome approach to accelerating materials innovation

Anubhav Jain,[1,a] Shyue Ping Ong,[2,a] Geoffroy Hautier,[3] Wei Chen,[1]
William Davidson Richards,[2] Stephen Dacek,[2] Shreyas Cholia,[1]
Dan Gunter,[1] David Skinner,[1] Gerbrand Ceder,[2] and Kristin A. Persson[1,b]
[1]*Lawrence Berkeley National Laboratory, Berkeley 94720, California, USA*
[2]*Massachusetts Institute of Technology, Cambridge 02139, Massachusetts, USA*
[3]*Université catholique de Louvain, Louvain-la-Neuve, Belgium*

Accelerating the discovery of advanced materials is essential for human welfare and sustainable, clean energy. In this paper, we introduce the Materials Project (www.materialsproject.org), a core program of the Materials Genome Initiative that uses high-throughput computing to uncover the properties of all known inorganic materials. This open dataset can be accessed through multiple channels for both interactive exploration and data mining. The Materials Project also seeks to create open-source platforms for developing robust, sophisticated materials analyses. Future efforts will enable users to perform ''rapid-prototyping'' of new materials *in silico*, and provide researchers with new avenues for cost-effective, data-driven materials design. *© 2013 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License.* [http://dx.doi.org/10.1063/1.4812323]

I. INTRODUCTION

Major technological advancement is largely driven by the discovery of new materials. From the prehistoric discovery of bronze and steel to the twentieth century invention of synthetic polymers, new materials have been responsible for vast transformations in human civilization. Today, materials innovations also hold the key to tackling some of our most pressing societal challenges, such as global climate change and our future energy supply.[1,2]

However, materials discovery today still involves significant trial-and-error. It can require decades of research to identify a suitable material for a technological application, and longer still to optimize that material for commercialization. A principal reason for this long discovery process is that materials design is a complex, multi-dimensional optimization problem, and the data needed to make informed choices about which materials to focus on and what experiments to perform usually does not exist.

What is needed is a scalable approach that leverages the talent and efforts of the entire materials community. The Materials Genome Initiative,[3] launched in 2011 in the United States, is a large-scale collaboration between materials scientists (both experimentalists and theorists) and computer scientists to deploy proven computational methodologies to predict, screen, and optimize materials at an unparalleled scale and rate. The time is right for this ambitious approach: it is now well established that many important materials properties can be predicted by solving equations based on the fundamental laws of physics[4] using quantum chemical approximations such as density functional theory (DFT).[5,6] This virtual testing of materials can be employed to design and optimize materials *in silico*.[7,8]

---

[a]A. Jain and S. P. Ong contributed equally to this work.
[b]Author to whom correspondence should be addressed. Electronic mail: kapersson@lbl.gov
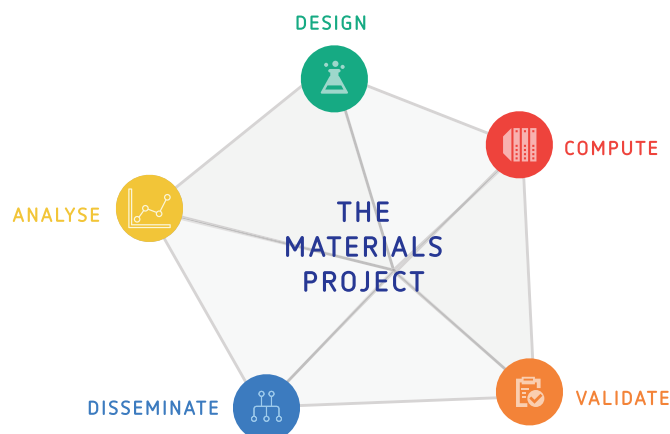
FIG. 1. Overview of the Materials Project thrusts. Computed data are validated, disseminated to the user community, and fed into analysis that is ultimately used to design new compounds for subsequent computations.

Many research groups have already employed this high-throughput computational approach to screen up to tens of thousands of compounds for potential new technological materials. Examples include solar water splitters,[9,10] solar photovoltaics,[11] topological insulators,[12] scintillators,[13,14] CO_2 capture materials,[15] piezoelectrics,[16] and thermoelectrics,[17,18] with each study suggesting several new promising compounds for experimental follow-up. In the fields of catalysis,[19] hydrogen storage materials,[20,21] and Li-ion batteries,[22–26] experimental "hits" from high-throughput computations have already been reported.

In recent years, perhaps an even more exciting trend has begun to emerge, which is the integration of computational materials science with information technology (e.g., web-based dissemination, databases, data-mining) to go beyond the confines of any single research group. This development has expanded access to computed materials datasets to new communities and spurred new collaborative approaches for materials discovery.[27,28] The next step is to leverage open-source development and interactive web-based technologies to enable user contributions back to the community by reporting problems, coding new types of analyses and apps, and suggesting the next set of breakthrough materials for computation. All the pieces are ready to change the paradigm by which materials are designed.

In this paper, we introduce the Materials Project (www.materialsproject.org), a component of the Materials Genome Initiative that leverages the power of high-throughput computation and best practices from the information age to create an open, collaborative, and data-rich ecosystem for accelerated materials design. Started in October of 2011 as a joint collaboration between the Massachusetts Institute of Technology and Lawrence Berkeley National Laboratory, the Materials Project today has partners in more than ten institutions worldwide. The Materials Project web site currently receives several hundred unique page views a day and has registered more than 4000 users in academia, government, and industry.

Figure 1 provides an overview of the Materials Project. The Materials Project seeks to accelerate materials design by creating open, collaborative systems targeting each step in the computational materials design process – data creation, validation, dissemination, analysis, and design. In Secs. II–V, we discuss the Materials Project's current efforts and future directions in each of these areas.

II. DATA GENERATION AND VALIDATION

One of the Materials Project's key thrusts is to compute the properties of compounds for which experimental data may be incomplete or absent. However, there exist formidable technical and scientific hurdles to achieving this goal.
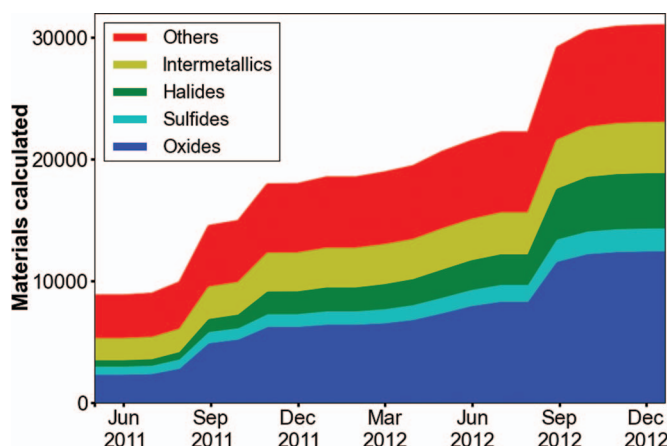
FIG. 2. Number of compounds available on the Materials Project web site since the initial release in October 2011, broken down by type of compound.

Despite significant improvements in robustness and user-friendliness of first-principles codes, it is still far from trivial to automate and store tens of thousands of calculations. First principles calculations are computationally expensive (a single material might require several hundred central processing unit (CPU)-hours just to obtain basic properties), often require serial actions (e.g., one needs to calculate an optimized crystal structure before one can perform property calculations), and often require human intervention to arrive at a converged, reliable result. To address these issues, the Materials Project has developed its own workflow software ("FireWorks") for automating and managing all computational steps at large supercomputing centers, and job management software ("custodian") to perform all the calculations needed to determine a property. As an example, when performing energy optimization runs, it automatically "self-heals" by applying rule-based fixes to failed runs, updating cluster as well as convergence parameters. The system can also automatically prevent duplicating jobs and apply different types of workflows when detecting, for example, a metal versus an insulator. These codebases will be further described in a future publication, but are already available as fully functional open-source packages that can be leveraged by the computational research community.[29] Using this infrastructure, the Materials Project has built a sizable materials database that today contains computed structural, electronic, and energetic data (calculated using the Vienna Ab initio Simulation Package[30,31]) for over 33 000 compounds (Figure 2), obtained using over 15 million CPU-hours of computational time at the National Energy Research Scientific Computing Center (NERSC).

Today, the vast majority of the Materials Project data are for compounds in the Inorganic Crystal Structural Database (ICSD).[32,33] Going forward, a significant challenge is the generation of novel compositions and compounds to perform calculations. This problem presents one of the foremost challenges in computational materials science. There already exist multiple algorithmic approaches to tackle this problem. Optimization-based approaches[33–37] (such as genetic algorithm and simulated annealing) have been heavily investigated and data-driven approaches[38–40] are beginning to emerge, but no technique is ideal. However, the growth of web-based collaboration presents the opportunity for another method of generating new compounds: crowd-sourced suggestions coupled to computations-on-demand. In this method, crystal structures designed by the user community will be automatically fed into our calculation infrastructure, with the results reported back to the community.

A final concern when generating large data sets is validating the accuracy of the calculated data. A major challenge of the Materials Project is to present calculated data for a broad audience, including researchers who may be unfamiliar with the limitations of first-principles methods. One way to address calculation inaccuracy is to develop strategies that improve agreement with experimental results. For example, inaccurate reaction energies are sometimes obtained when
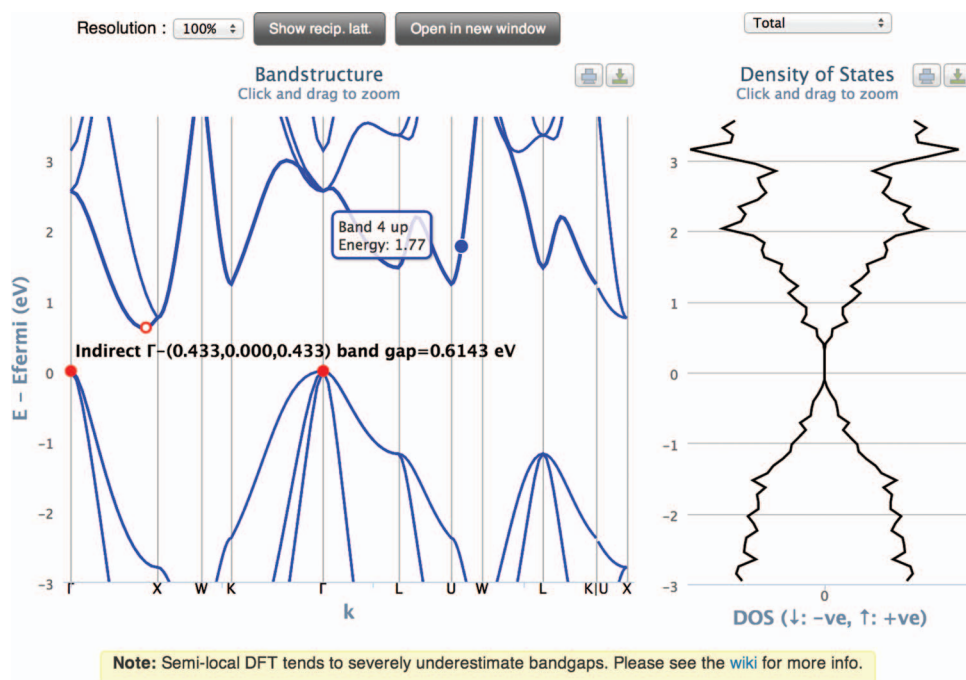
FIG. 3. Band structure and density of states for Si from the Materials Project web site, with a visual warning and link to more information at bottom regarding limitations to the accuracy of computed band gaps (the measured band gap of Si is about 1.1 eV).

employing a single approximation (e.g., DFT) to calculate reactions between solids and gases,[41,42] between some elements and their binary compounds,[43,44] or between metals and insulators containing correlated d or f electrons.[45] The Materials Project reports more accurate results over such wide chemical spaces by employing separate methodologies (including use of reported experimental data) for different classes of materials, and using a set of reference reactions to connect results from different methods. This approach has been employed for solid-gas reactions,[41] d-block oxide reactions,[45] and dissolved ions in solution.[46] The Materials Project is similarly exploring methods for improving other predicted properties such as band gaps.[47–51] The project aims to report data in a way that can be interpreted, as much as possible, at face value by any materials researcher, and provide additional information to guide the user (Figure 3). Furthermore, the Materials Project makes available (see Sec. III for more details) the "raw," unmodified data to users who want to apply their own correction techniques or datamine errors within the computational methodologies.

Despite continual improvements in calculation methods, it is a reality of high-throughput computation that some of the data will be incorrect. We are currently using both automated and crowd-sourced means to verify the integrity of our data. In terms of automated analysis, we validate calculated oxidation states, cell volumes, and bond lengths versus experimental data and report clear visual warnings on the web site for compounds where these values fall outside normal limits. For example, the structures of many layered compounds that are bonded by van der Waals interactions are not accurately modeled by standard DFT functionals;[52–54] those materials display a visual warning on the web site indicating that experimental and computational structures do not agree. Just as important as automated verification is the "Report Issues" button on the web page for each material, which users can click to report problems with the data. For example, users have employed this button to report issues with the magnetic structure of several entries, which would be otherwise difficult to detect automatically.
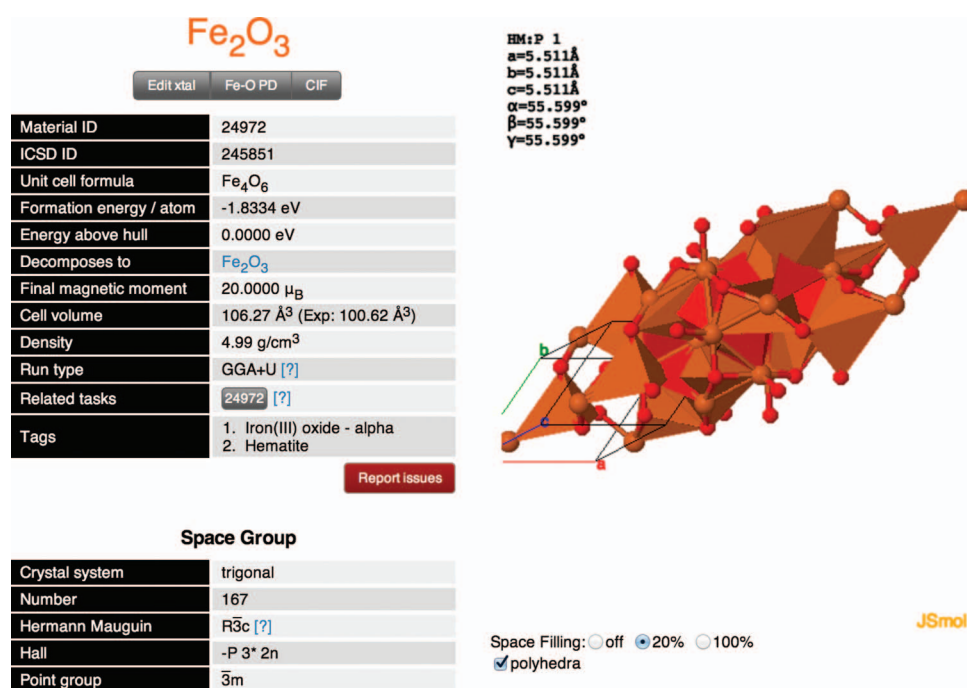
FIG. 4. Screenshot of the top portion of the details page for Fe_2O_3. The page also contains electronic structure information (Figure 3), and (not shown) the lattice parameters and atomic positions of the structure, a calculated x-ray diffraction pattern, and basic calculation parameters and output.

III. DISSEMINATION: PROVIDING OPEN, MULTI-CHANNEL ACCESS TO MATERIALS INFORMATION

Once the data is generated and verified, the Materials Project aims to provide access to the data in a way that enables flexible and innovative usage. The Materials Project provides multiple channels to access its large and rich materials dataset.

The primary access point for most users is the web applications, which provide graphical user interfaces to query for various forms of raw and processed materials data. Several such applications are already available. The Materials Explorer, for example, allows users to search for materials based on composition or property and explore their properties (Figure 4), while the Lithium Battery Explorer adds application-specific search criteria such as voltage and capacity for targeted searches for lithium-ion battery electrode materials. Additional web applications allow users to interactively analyze the dataset. For instance, the Phase Diagram App (Figure 5) constructs low-temperature phase diagrams for any chemical system, supporting both closed systems and open systems (grand canonical construction).[55,56] The Reaction Calculator balances reactions and computes reaction energies between any set of compounds in the database, with comparisons to experimental reaction energies automatically reported where available.

Going forward, the Materials Project is committed to further expanding the number of quality web applications that provides sophisticated analyses of its data. An application that generates computed Pourbaix diagrams (aqueous solubility) based on the methodology of Persson *et al.*[46] is currently in development. The Materials Project also plans to leverage on the expertise of the large materials research community to create novel applications with materials data, and provide the means for user-submitted applications to be hosted in the future.

While web applications provide a means for exploratory data analysis, they are not an effective means for a user to obtain large quantities of materials data. The Materials Project recognizes that there may be users who require access to large datasets to datamine for trends across chemistries. Alternatively, users may wish to build their own applications that combine a local dataset with that
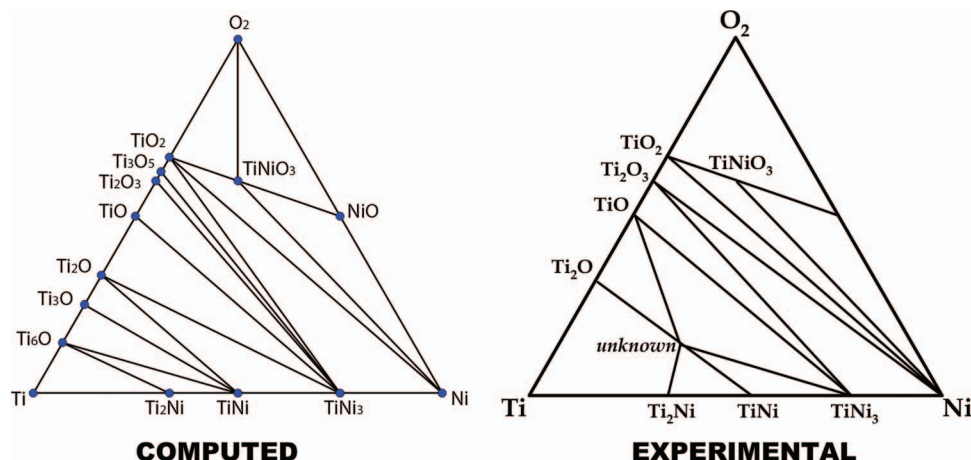
FIG. 5. Low-temperature phase diagram of Ti–Ni–O generated by the PDApp (left) and at 750 °C from ASM online (right).[65,66] Overall, the diagrams are very similar; a major difference is the presence of Ti_3O_5 and several Ti-peroxide suboxide phases (Ti_3O, Ti_6O) in the computational diagram.

of the Materials Project, for instance, to build a unified phase diagram through the Phase Diagram App.

To support the needs of such users, the Materials Project recently launched the Materials application programming interface (API). This API provides users with direct access to the data and operates similarly to other web-based information portals (such as Google Maps) that allow external developers to create their own applications or analyses from the data set. The Materials API is based on REpresentational State Transfer (REST) principles that provide data access via the Hypertext Transfer Protocol (HTTP). For the purposes of the Materials Project, this means that each object (such as a material) can be represented by a unique URL (e.g., http://www.materialsproject.org/rest/v1/[unique-id]), and an HTTP verb can be used to act on that object.[57] This action then returns structured data, typically in the widely used JavaScript Object Notation (JSON). A high-level interface to the Materials API has been built into the open-source Python Materials Genomics (pymatgen) analysis library (see Sec. IV) that provides a powerful way for users to programmatically query and analyze large quantities of materials information.[58]

Today, the Materials API only supports unidirectional information transfer – from the Materials Project to the user. Future versions of the Materials API will provide a channel for users to transfer information to the Materials Project. For example, it will be the method by which users can submit compounds to the Materials Project for automatic computation ("calculations-on-demand"). It will also allow users to submit supplementary information about materials (e.g., the crowdsourcing of experimental data, or tag compounds with publication references). Two-way interaction will further broaden the scope of collaborative science possible within this framework.

IV. ANALYSIS: OPEN-SOURCE LIBRARY

To extract useful insights from the data, the dissemination of large and rich materials data sets must be complemented by accessible analysis tools. For instance, from the energies obtained from basic electronic structure calculations, one can derive stability assessments via the construction of phase diagrams, which is useful for most materials design and synthesis problems. From computed band structures, one can derive band gaps, the nature of optical transitions (indirect/direct), and effective masses of charge carriers, which are useful for design of functional electronic materials such as thermoelectrics or transparent conducting oxides.

With large and rich materials data sets such as those in the Materials Project, there is also the potential to apply statistical learning techniques to datamine trends across a broad range of chemistries. One such example is given in Figure 6, which plots the percentage of Cr sites in 4-fold
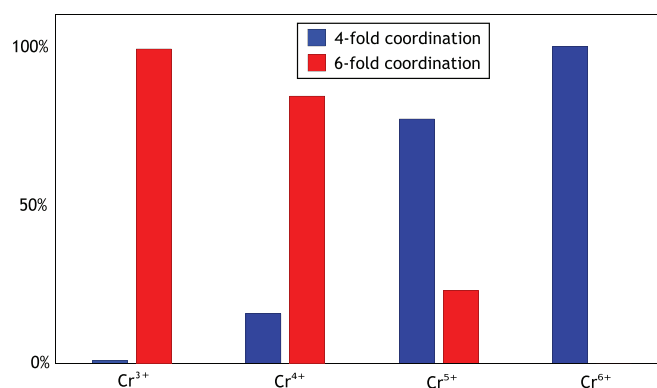
FIG. 6. Distribution of oxygen coordinations for chromium in oxides. The percentage of 4-fold and 6-fold coordinated sites for different Cr oxidation states is given. This analysis can be performed with a few lines of codes using pymatgen and the REST interface.

or 6-fold coordination in oxides as a function of oxidation state. The data support known chemical principles: the smaller Cr^{6+} ion strongly favors 4-fold (tetrahedral) coordination, while the larger Cr^{3+} ion strongly favors 6-fold (octahedral) coordination.

The Materials Project believes that the best way to develop robust and insightful analyses is by leveraging on the expertise of the entire materials research community. To this end, the Materials Project adopts the best practices of open source software development to build collaborative platforms for materials data analysis. Almost all of the software infrastructure powering the Materials Project is publicly available under open-source licenses. For example, the Python Materials Genomics library[58] defines core Python objects for materials data representation, and provides a well-tested set of structure and thermodynamic analyses relevant to many applications. Pymatgen already has more than 100 active collaborators worldwide and continues to grow in functionality and robustness every day. Some examples of community contributions include support for F EFFective (FEFF) calculations[59] contributed by Alan Dozier from the University of Kentucky, and for ABINIT[60] calculations by Matteo Giantomassi at Université catholique de Louvain. A database add-on (pymatgen-db) enables the creation of local "Materials Project-like" MongoDB databases from high-throughput computations that researchers can use to store and query their calculated datasets.[57,58,61] All software components[29] are hosted on GitHub (www.github.com), a social coding platform that allows users to report bugs and contribute to development. The libraries can also be easily installed (using *pip* or *easy_install*) through the Python Package Index.

V. DESIGN: A VIRTUAL LABORATORY FOR NEW MATERIALS DISCOVERY

Computed information and analyses should ultimately culminate in new materials design. An example of this process is depicted in Figure 7: using a combination of several Materials Project tools, it is possible to perform many aspects of materials design *in silico*. One can imagine a scientist looking for a material with certain band structure features and thermodynamic stability criteria. From searching the Materials Explorer, the user finds a very suitable ruthenium-containing oxide compound, but the cost associated with ruthenium excludes the material from being a commercially viable candidate. With the assistance of applications such as the Crystal Toolkit and Structure Predictor, the user is able to design a range of viable compounds consisting of less expensive transition metal ions with similar size and valence. The user then submits the designs to the Materials Project and receives the results within a few days where the electronic properties as well as the thermodynamic stability of the novel compounds are assessed using the Materials Explorer and Phase Diagram App. Compounds that meet property targets can then be fed into higher-order calculations, or immediately forwarded to the lab for experimental investigation.

A real-world example of such a computational design workflow is illustrated in Figure 8, where we describe the steps in the discovery of a novel class of carbonophosphate compounds for Li ion
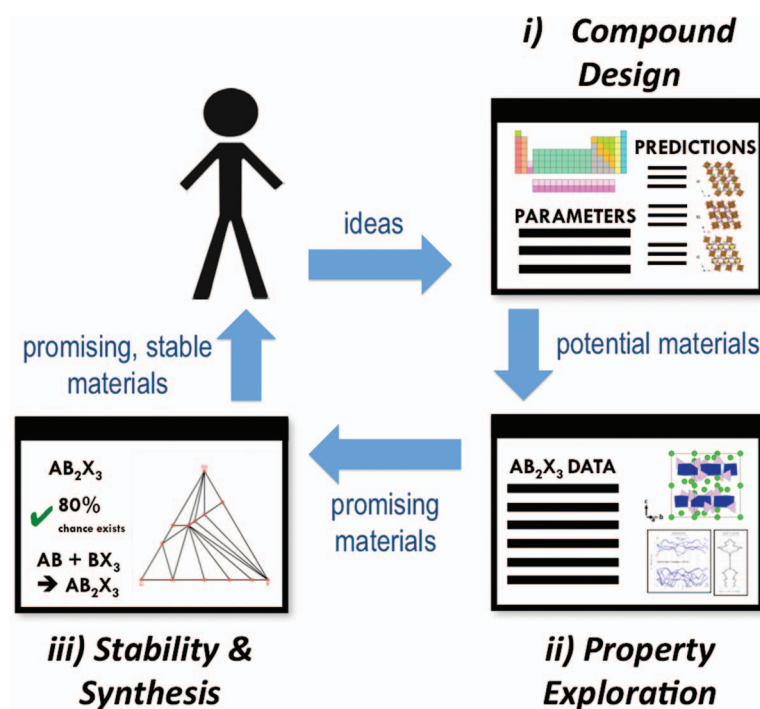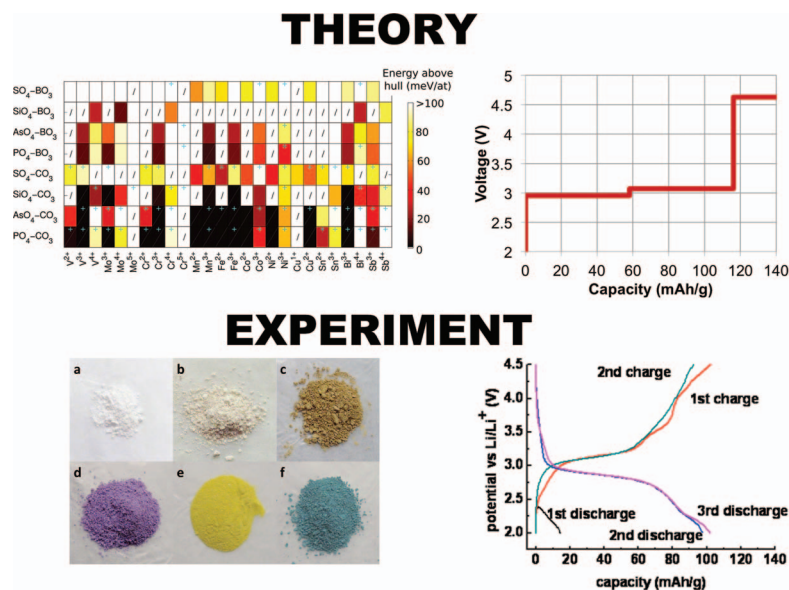
FIG. 7. Rapid prototyping and iterative materials design steps that might be performed *in silico*.

FIG. 8. Example of computationally driven design of new Li-ion battery cathodes. (Top-left) A family of compounds containing Na, a transition metal, and mixed polyanion group is investigated computationally for stability (the corresponding Li compounds were computed to be too unstable for direct synthesis).[25] The ground state "hull" connects the energy of all ground state phases in an energy-composition diagram. The energy above hull is a computed descriptor of the stability of a compound, and in essence describes the thermodynamic decomposition energy of the compound into the most stable phases. Thermodynamically stable compounds exhibit an energy above hull of zero, with greater values indicating decreasing stability. (Top-right) Electrochemical properties are calculated for Li versions of compounds predicted to be stable in the Na form, such as the Fe-containing phosphocarbonate. (Bottom-left) Hydrothermal synthesis produces a colorful family of sodium metal phosphocarbonate materials as predicted by computation.[26] (Bottom-right) The Na compounds are ion exchanged to form their Li analogues, and predicted battery properties are confirmed.[23]

battery cathode applications.[23, 25, 26] Prediction of stable compounds and a decision to use Na-Li ion exchange for synthesis were largely done using computation, along with screening the candidates for promising battery properties. The computational data served to focus the experimental synthesis as well as the electrochemical testing to only the most promising portions of chemical space.

Even in the short span of time since its inception in October 2011, we have begun to see the application of Materials Project data in materials design by researchers outside of the core group of developers. Users have already employed the data from the site in a variety of applications including benchmarking theoretical investigations,[44, 62] and in the design of Li-ion battery anodes,[63] photocatalysts,[10] and magnetic materials.[64]

VI. CONCLUSION AND FUTURE

Advanced materials are essential to human well-being and to form the cornerstone for emerging industries. Unfortunately, the traditional time frame for moving advanced materials from the laboratory into applications is remarkably long, often taking 10–20 years. The Materials Genome Initiative, of which the Materials Project is a part, is a multi-stakeholder effort to develop an infrastructure to accelerate advanced materials discovery and deployment. The Materials Project combines high-throughput computation, web-based dissemination, and open-source analysis tools to provide material scientists with new angles to attack the materials discovery problem.

Looking ahead, the Materials Project aims to form a backbone that can be expanded to new areas. As new theoretical techniques are developed, they can be "plugged in" to our workflow engine and applied across all materials in the database. Efforts targeting prediction of surface energies, elastic constants, point defects, and finite temperature properties using large data sets and novel algorithms are underway. The dissemination and front-end tools are being expanded to new datasets, such as experimental data for comparison with computed results. Finally, a "calculations-on-demand" feature will allow scientists to directly contribute new compound ideas to the database. Far from being a final product, the Materials Project is an evolving resource that we expect will grow more useful from ever-increasing data sets and network effects from its user base. It is our belief that deployment of large-scale accurate information to the materials development community will significantly accelerate and enable the discovery of improved materials for our future clean energy systems, green building components, cutting-edge electronics, and improved societal health and welfare.

[1] U.S.D. of Energy, *Department of Energy Workshop: Computational Materials Science and Chemistry for Innovation* (2010).
[2] S. L. Moskowitz, *The Advanced Materials Revolution: Technology and Economic Growth in the Age of Globalization* (John Wiley & Sons, Inc., New York, 2009).
[3] Materials Genome Initiative for Global Competitiveness (2011).
[4] E. Schrödinger, Phys. Rev. **22**, 1049 (1926).

[5] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[6] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[7] J. Hafner, C. Wolverton, and G. Ceder, MRS Bull. **31**, 659 (2006).

[8] G. Hautier, A. Jain, and S. P. Ong, J. Mater. Sci. **47**, 7317 (2012).

[9] I. E. Castelli, D. D. Landis, K. S. Thygesen, S. Dahl, I. Chorkendorff, T. F. Jaramillo, and K. W. Jacobsen, Energy Environ. Sci. **5**, 9034 (2012).

[10] I. E. Castelli, T. Olsen, S. Datta, D. D. Landis, S. Dahl, K. S. Thygesen, and K. W. Jacobsen, Energy Environ. Sci. **5**, 5814 (2012).

[11] L. Yu and A. Zunger, Phys. Rev. Lett. **108**, 068701 (2012).

[12] K. Yang, W. Setyawan, S. Wang, M. Buongiorno Nardelli, and S. Curtarolo, Nature Mater. **11**, 614 (2012).

[13] C. Ortiz, O. Eriksson, and M. Klintenberg, Comput. Mater. Sci. **44**, 1042 (2009).

[14] W. Setyawan, R. M. Gaume, S. Lam, R. S. Feigelson, and S. Curtarolo, ACS Comb. Sci. **13**, 382 (2011).

[15] L.-C. Lin, A. H. Berger, R. L. Martin, J. Kim, J. A. Swisher, K. Jariwala, C. H. Rycroft, A. S. Bhown, M. W. Deem, M. Haranczyk, and B. Smit, Nature Mater. **11**, 633 (2012).

[16] R. Armiento, B. Kozinsky, M. Fornari, and G. Ceder, Phys. Rev. B **84**, 014103 (2011).

[17] S. Wang, Z. Wang, W. Setyawan, N. Mingo, and S. Curtarolo, Phys. Rev. X **1**, 021012 (2011).

[18] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, Comput. Mater. Sci. **58**, 227 (2012).

[19] J. Greeley, T. F. Jaramillo, J. Bonde, I. B. Chorkendorff, and J. K. Nørskov, Nature Mater. **5**, 909 (2006).

[20] S. V. Alapati, J. K. Johnson, and D. S. Sholl, J. Phys. Chem. B **110**, 8769 (2006).

[21] J. Lu, Z. Z. Fang, Y. J. Choi, and H. Y. Sohn, J. Phys. Chem. C **111**, 12129 (2007).

[22] J. C. Kim, C. J. Moore, B. Kang, G. Hautier, A. Jain, and G. Ceder, J. Electrochem. Soc. **158**, A309 (2011).

[23] H. Chen, G. Hautier, A. Jain, C. J. Moore, B. Kang, R. Doe, L. Wu, Y. Zhu, and G. Ceder, Chemistry of Materials **24**, 2009 (2012).

[24] A. Jain, G. Hautier, C. Moore, B. Kang, J. Lee, H. Chen, N. Twu, and G. Ceder, J. Electrochem. Soc. **159**, A622 (2012).

[25] G. Hautier, A. Jain, H. Chen, C. Moore, S. P. Ong, and G. Ceder, J. Mater. Chem. **21**, 17147 (2011).

[26] H. Chen, G. Hautier, and G. Ceder, J. Am. Chem. Soc. **134**, 19619 (2012).

[27] J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway, and A. Aspuru-Guzik, J. Phys. Chem. Lett. **2**, 2241 (2011).

[28] J. S. Hummelshøj, F. Abild-Pedersen, F. Studt, T. Bligaard, and J. K. Nørskov, Angew. Chem., Int. Ed. Engl. **51**, 272 (2012).

[29] See http://www.github.com/materialsproject for downloadable source code and documentation.

[30] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).

[31] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).

[32] A. Belsky, M. Hellenbrandt, V. L. Karen, and P. Luksch, Acta Crystallogr., Sect. B: Struct. Sci. **58**, 364 (2002).

[33] G. Bergerhoff, R. Hundt, R. Sievers, and I. D. I. Brown, J. Chem. Inf. Comput. Sci. **23**, 66 (1983).

[34] S. M. Woodley and R. Catlow, Nature Mater. **7**, 937 (2008).

[35] A. R. Oganov and C. W. Glass, J. Chem. Phys. **124**, 244704 (2006).

[36] M. d'Avezac, J.-W. Luo, T. Chanier, and A. Zunger, Phys. Rev. Lett. **108**, 027401 (2012).

[37] S. Dudiy and A. Zunger, Phys. Rev. Lett. **97**, 046401 (2006).

[38] G. Hautier, C. Fischer, V. Ehrlacher, A. Jain, and G. Ceder, Inorg. Chem. **50**, 656 (2011).

[39] G. Hautier, C. C. Fischer, A. Jain, T. Mueller, and G. Ceder, Chem. Mater. **22**, 3762 (2010).

[40] C. C. Fischer, K. J. Tibbetts, D. Morgan, and G. Ceder, Nature Mater. **5**, 641 (2006).

[41] L. Wang, T. Maxisch, and G. Ceder, Phys. Rev. B **73**, 195107 (2006).

[42] S. Grindy, B. Meredig, S. Kirklin, J. E. Saal, and C. Wolverton, Phys. Rev. B **87**, 075150 (2013).

[43] S. Lany, Phys. Rev. B **78**, 245207 (2008).

[44] J. Yan, J. S. Hummelshøj, and J. K. Nørskov, Phys. Rev. B **87**, 075207 (2013).

[45] A. Jain, G. Hautier, S. P. Ong, C. J. Moore, C. C. Fischer, K. A. Persson, and G. Ceder, Phys. Rev. B **84**, 045115 (2011).

[46] K. A. Persson, B. Waldwick, P. Lazic, and G. Ceder, Phys. Rev. B **85**, 235438 (2012).

[47] L. Hedin, Phys. Rev. **139**, A796 (1965).

[48] J. Heyd, J. E. Peralta, G. E. Scuseria, and R. L. Martin, J. Chem. Phys. **123**, 174101 (2005).

[49] M. Chan and G. Ceder, Phys. Rev. Lett. **105**, 196403 (2010).

[50] E. Runge and E. K. U. Gross, Phys. Rev. Lett. **52**, 997 (1984).

[51] F. Tran and P. Blaha, Phys. Rev. Lett. **102**, 5 (2009).

[52] J. Klimeš, D. R. Bowler, and A. Michaelides, Phys. Rev. B **83**, 195131 (2011).

[53] H. Rydberg, M. Dion, and N. Jacobson, Phys. Rev. Lett. **91**, 126402 (2003).

[54] B. Lundqvist and Y. Andersson, Int. J. Quantum Chem. **56**, 247 (1995).

[55] S. P. Ong, A. Jain, G. Hautier, B. Kang, and G. Ceder, Electrochem. Commun. **12**, 427 (2010).

[56] S. Ong, L. Wang, B. Kang, and G. Ceder, Chem. Mater. **20**, 1798 (2008).

[57] D. Gunter, S. Cholia, A. Jain, M. Kocher, K. Persson, L. Ramakrishnan, S. P. Ong, and G. Ceder, in *Proceedings of the 5th Workshop on Many-Task Computing on Grids and Supercomputers (MTAGS)*, 2012.

[58] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Comput. Mater. Sci. **68**, 314 (2013).

[59] J. J. Rehr, Rev. Mod. Phys. **72**, 621 (2000).

[60] X. Gonze, J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J.-Y. Raty, and D. C. Allan, Comput. Mater. Sci. **25**, 478 (2002).

[61] 10Gen Inc., see http://www.mongodb.org.

[62] J. Rustad, Am. Mineral. **97**, 791 (2012).

[63] T. T. Tran and M. N. Obrovac, J. Electrochem. Soc. **158**, A1411 (2011).

[64] M. Meinert and M. P. Geisler, J. Magn. Magn. Mater. **341**, 72 (2013).

[65] ASM International, see http://www.asminternational.org/AsmEnterprise/APD.

[66] O. A. Bannykh, R. M. Volkova, and V. A. Bozhenov, Russ. Metall. **2**, 202 (1984).