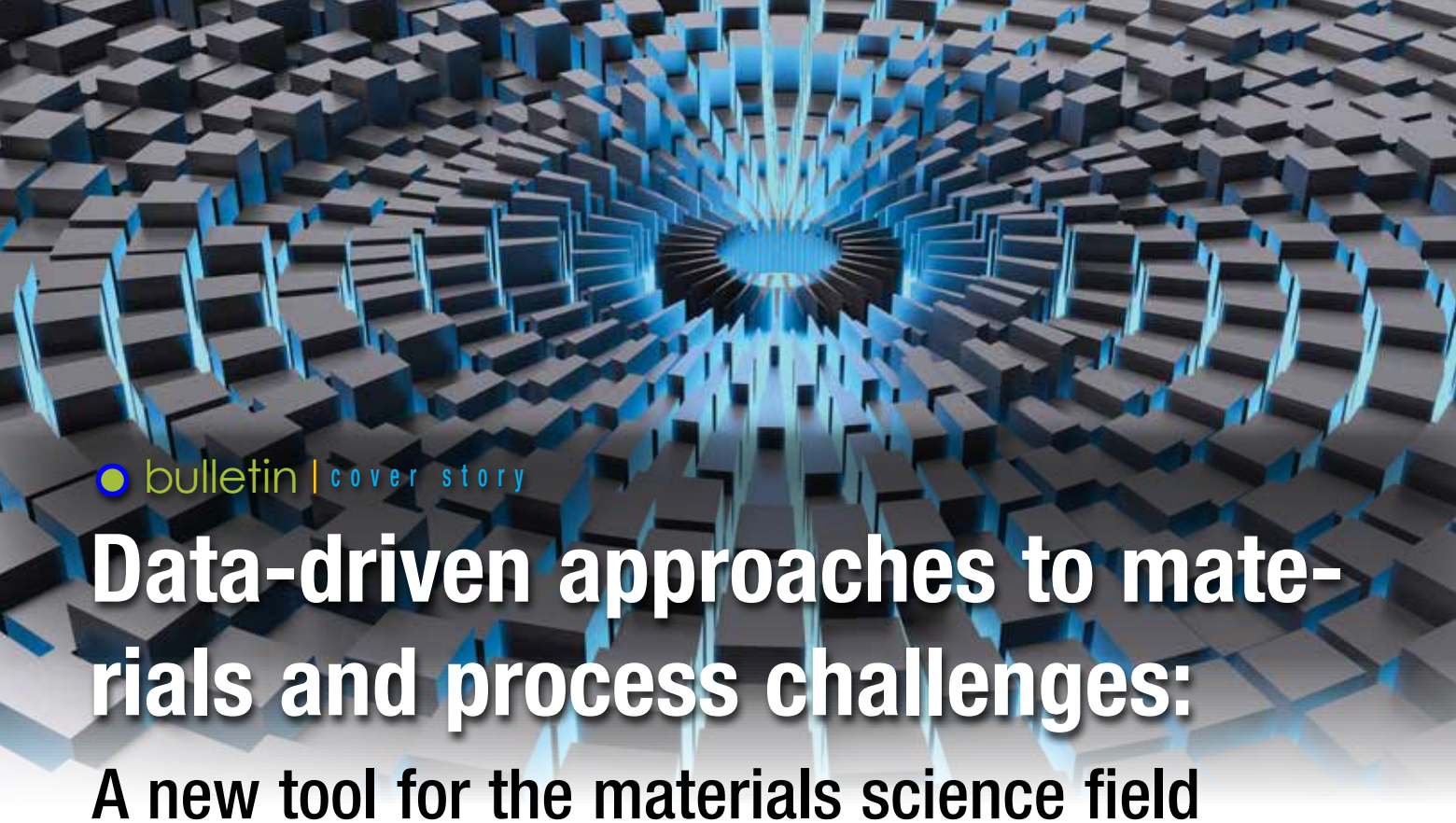


Data-driven approaches to materials and process challenges:

A new tool for the materials science field

By Richard Padbury

To keep pace with new demands in the materials science industry, scientists and engineers will need to speed up materials discovery and commercialization. Data-driven methods can augment existing experimental methods to accelerate the process.

The materials science industry is expected to grow significantly over the coming years. This growth, in itself, is not surprising because materials are at the center of every major challenge, from providing solutions to climate change and environmental issues to enabling developments in agriculture, healthcare, energy production, and transportation—even the way we live and interact as a society is, and will be, affected by materials.[1]

In the same way that scientists discovered thermodynamics, electricity, the laser, and transistor (discoveries that fueled the first three industrial revolutions), today's scientists will need to speed up the development and discovery of innovative materials designed to deliver new functionalities to meet future demands.[2] For example, to build a clean energy future, we will need to both develop novel materials to create more efficient solar panels, wind turbines, and energy storage devices and develop materials that can scrub the air of existing pollutants. We also need to replace materials that are subject to supply disruptions due to finite resources of rare-earth minerals and feedstock derived from fossil fuels. Furthermore, to support a sustainable future, the toxicity and recyclability of new materials must also be taken into consideration.

There is a risk the current pace of development will not keep up with these new demands. For the most pressing challenges facing society, we cannot afford to wait 20 years or more to develop

Capsule summary

GROWING DEMAND

The engineered materials industry is expected to grow significantly over the coming years. But there is a risk the current pace of materials development will not keep up with these new demands.

ACCELERATED DISCOVERY

Researchers use data-driven methods for materials discovery and testing to augment existing experimental methods to greatly accelerate the commercialization process.

INDUSTRY OPPORTUNITIES

Companies are now beginning to use the data-science knowledge generated by mainstream academic and government research to tackle everyday challenges across their enterprises.

the necessary solutions (the average time it currently takes for novel materials to reach commercial maturity).[3] The task is now upon us to develop the next materials breakthroughs to support a more secure and prosperous future.

The evolution of materials

Known materials available today were developed over many thousands of years as humans advanced from the early stages of alchemy through the evolutionary periods of the stone, bronze, and iron ages. At each period, curiosity fueled the effort to develop new materials with the aim of filling gaps in material property spaces to advance new applications and processes.

The science involved in these discoveries include

- development of materials with new compositions, such as the development of binary and ternary ceramics;
- manipulation of microstructure and thermomechanical processing to control the distribution of strengthening phases and defects;
- discovery of nanomaterials, which expanded our historical view of materials to previously unattainable property spaces; and
- creation of novel material architectures, such as hybrids and composites, often inspired by nature, to achieve multifunctional properties.

The classifications of materials obtained from these developments—from metals and ceramics to polymers and composites—form discrete clusters in property space due to their distinctive atomic structures and bond types that underpin their unique properties (Figure 1).[4] If we take a moment to look around ourselves, it is clear these essential materials surround us in our everyday lives.

However, the common denominator under all developments is the significant time it has taken to discover, develop, and commercialize them. Just why does

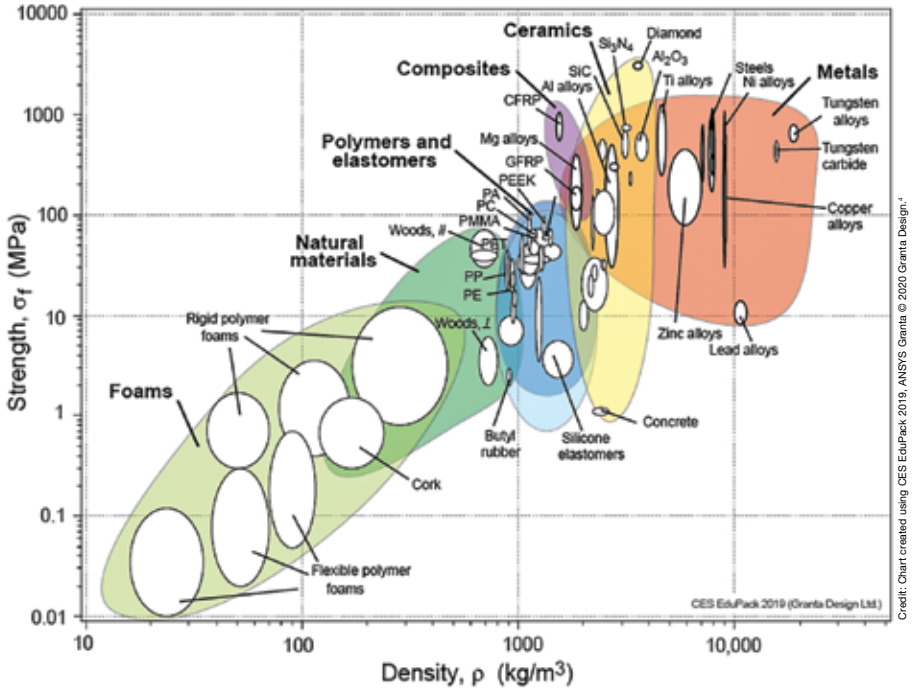


Figure 1. Ashby plot of strength vs. density highlighting the many categories of materials that form the materials universe.

it take so long to develop novel materials? As we will explore next, the answer is concealed in the complex, multiple length scale structure of materials.

The multiple length scale challenge

The materials science framework deals with the understanding of process-structure-property (PSP) linkages, from which multiple, intertwined relationships exist (Figure 2).[5] Materials scientists and engineers leverage their intuition and expert knowledge to investigate these multifaceted relationships and develop new material chemistries and properties.

A key challenge for materials scientists and engineers is formulating an understanding of the hierarchical nature of materials because the underlying structures form over multiple time and length scales.[6] At the atomic scale, interactions between pairs of elements inform the short-range order of multiple elements and molecules into lattice structures or repeat units. When these repeat units come together, they produce unique microstructures over increasing length scales that correspond to a material's macroscopic properties and morphology, at scales we can sense and use their characteristics.

Going back to the atomic scale, there is a seemingly infinite number of ways to arrange and rearrange atoms and molecules into new lattice or repeat unit structures, resulting in a diverse universe of materials with unique mechanical, optical, dielectric, and conductive properties.[7] Subsequently, countless materials remain undiscovered as it would require astronomical timescales and significant resources to test a composition and repeat before discovering a successful result.[8] Furthermore, when scientists do isolate a promising composition, there are many steps along the road to commercialization, each acting like a series of resistances in an electrical circuit, that must be overcome to progress a new technology forward—again, these steps

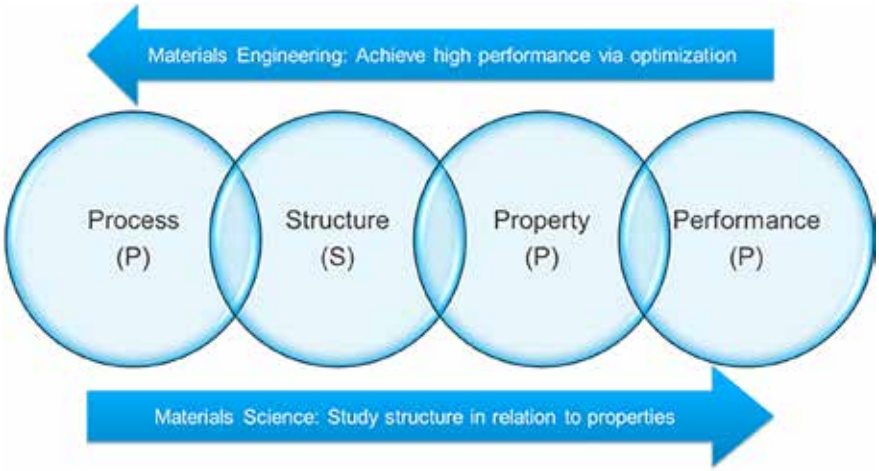


Figure 2. Processing-structure-property relationships that govern applied materials science and engineering. Adapted from A. Agrawal et al.[5]

introduce time and cost to the development pathway.

To overcome this challenge, scientists and engineers leverage tools that can improve the economics of designing experiments to develop new materials. For example, statistical methods can tune in to key variables that control a process or the evolution of material microstructure to achieve desirable properties. However, statistical methods, such as those developed by George Box, Donald Behnken, and Genichi Taguchi, are ideally constrained to a small subset of process-structure or structure-property linkages. Therefore, it is not possible to survey all relationships, across multiple length scales and PSP linkages, that may have varying degrees of influence on material performance.[6] This limitation can lead to an undershoot in target properties, if key variables or relationships are unintentionally missed by experimental designs, or greatly limit the scope of an investigation. Therefore, in the same way there are many more new materials to discover, it is also likely hidden properties exist in known materials that have simply not been tested before. One example of a hidden property is the development of lithium iron phosphate for lithium-ion battery cathodes. The material was first synthesized in the 1930s but was not identified as a suitable cathode material until 66 years later in 1996.[8]

Several factors beyond the technical challenges contribute to the long period between materials discovery and commercialization. These factors range from misaligned market needs with the value proposition of a new material to the way we store, share, and report experimental data (often it is not easily accessible).[3] For example, identical experiments may be conducted in different parts of an organization, with scientists in the organization unable to check which experiments have been run. In tandem, the rigorous approval processes in highly regulated industries—implemented for good reason—increase the time and cost to validate new materials and processes for specific applications. Consequently, once a material is successfully commercialized, it becomes deeply rooted within industry,[9] such as the widespread use of silicon and aluminum oxide for semiconductor applications or the use of hydroxyapatite and Bioglass for medical devices. However, as legacy materials approach their limits and pressure on finite resources increases, new techniques are urgently needed that can speed up development and further expand our horizons into untapped regions of materials property space.

A new paradigm

The materials science field is entering a paradigm shift; the currently accepted methods of discovering materials are not irrelevant nor are they being replaced, but they are being augmented

What is data science?

Adapted from Rangaswamy et al.,[11] Elsevier, 2018

Data science envelops many overlapping subfields, from the well-known disciplines of knowledge discovery in databases (KDD), data mining, and statistics to the emerging fields of machine learning (ML) and artificial intelligence (AI).

The challenge with this picture is that it can be difficult to precisely explain the differences between these overlapping disciplines, which is key to understanding how to use each appropriately. After closer inspection, many subfields borrow the same methods, for instance, linear regression is as applicable in statistics as it is in ML. Therefore, explanations differentiating each discipline may be captured in cultural differences depending on individual schools of thought.

Without speculation, what has definitively occurred over the last few decades is an increase in computing power; an improvement in the ability to store and transfer data due to technological advances, such as the internet; and significantly increased data volumes, even in materials science. These improvements have prompted the use of more advanced methods of analyzing data, beyond simple linear models, and have led to cutting-edge forms of prediction and automation.

These advanced methods still require significant input from their human counterparts and still need to be systematically programmed so they can be deployed. Notably, no method can currently create, innovate, reason, apply logic, ethics, and morals or provide curiosity in the same way that humans can to make informed decisions. Therefore, it is reasoned that true forms of AI are still far away from being realized.

Nonetheless, what is possible through a common goal of learning from data is the extraction of powerful insights to develop actionable solutions. So the data science toolbox should be considered an assistant that augments our human ability to solve problems. ■

by techniques acquired from the cross-fertilization of materials science with other scientific disciplines.[5] This new way of thinking builds on the existing materials data and knowledge generated over many centuries and also includes methods of overcoming limited access to the data.

The emerging developments begin with the advent of the computer in the early 1950s, when more complex challenges could be solved by methods derived from quantum mechanics, such as density functional theory (DFT). As automation and computing power improved, increased calculation speeds led to the rise of high throughput (HT) simulation techniques.[9,10] Today, methods such as HT-DFT are capable of calculating the thermodynamic and electronic properties of tens to hundreds of thousands of known or hypothetical material structures. These methods resulted in a data explosion, and as the volume and variety of data accelerated, analyses became too big and complex for direct involvement by researchers.[9] Subsequently, data-driven methods from the computer and data science fields (Sidebar: "What is data science?)[11] were employed to help analyze the streams of data coming out of computational experiments. While state-of-the-art HT-DFT can greatly improve the efficiency of developing new materials, certain restrictions exist, from limitations in computing resources to the size of the material system that can be calculated and the types of properties that can be accurately modeled.[12] Furthermore, there are still many material structures left to explore, and it remains impractical, even for computational techniques, to explore them all.

Over the last 20 years, the use of data-driven methods expanded to help tackle the challenge of discovering and developing new materials, leading to the creation of a new field aptly known as Materials Informatics (MI).

MI underpins the acquisition and storage of materials data, the development of surrogate models to make rapid property predictions or gain new physical insights from materials data, and experimental confirmations of new materials with the core objective of accelerating materials discovery and development.

The MI framework leverages a wider range of data-driven algorithms (Sidebar: "Introduction to algorithms"),[13] using their ability to digest large volumes of complex data and resulting prediction accuracy, which enables researchers to explore many more PSP linkages and multiscale relationships than previously possible. Interestingly, these data-driven techniques are not new, as many have existed since the first computers were developed.[10] Furthermore, certain approaches have been around for many centuries, such as Bayesian and Gaussian

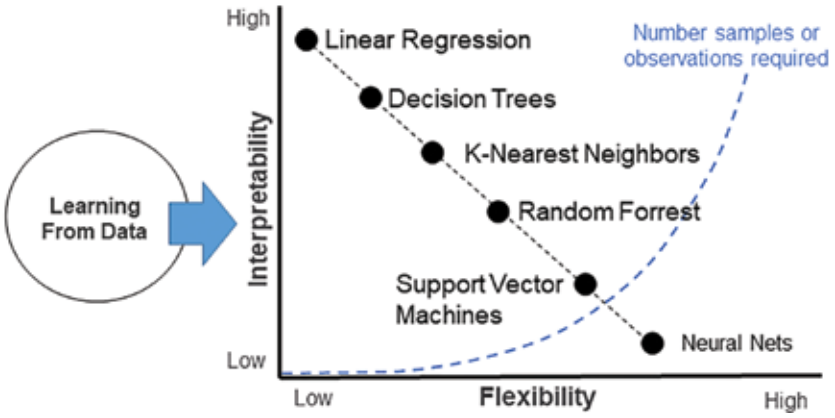
Introduction to algorithms

Adapted from James et al.,[13] Springer, 2014

An algorithm is a step-by-step procedure that takes inputs and produces an output based on a set of instructions. The coefficients, or weighting of each input, are estimated by "learning" from data generated by observations or an experiment. Once the coefficients are estimated, the algorithm is known as a model and can be used to predict new outputs on data the model has not yet "seen."

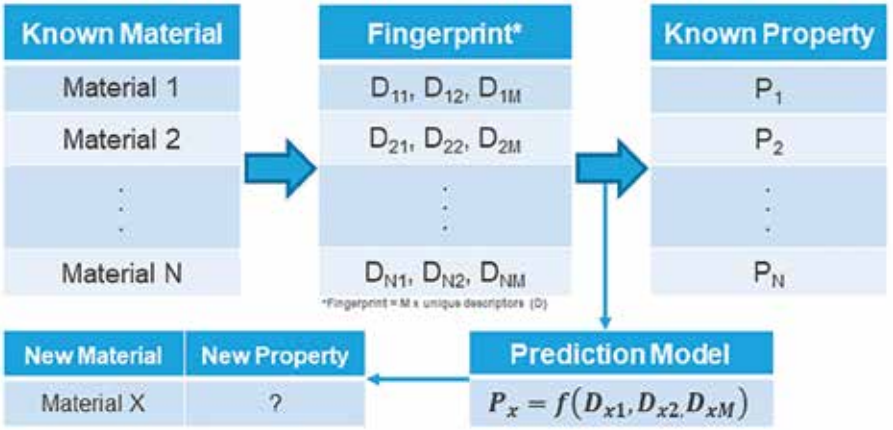
Model accuracy is assessed by measuring the quality of fit or cross-validating with data from the training dataset that is left out of the model training step. An optimal model will generalize well to new data, resulting in accurate predictions. However, the model requires a trade-off between bias (how well the model matches the training data) and variance (how well the model predicts output of new data). A model that underfits tends to have high bias–low variance as the model is less flexible to capturing trends in the training data. Conversely, overfitting leads to models that have low bias–high variance as the model is too flexible and fits the training data too closely by including noise or insignificant variables.

This trade-off leads to an important concept known as the curse of dimensionality. As the number of variables (or dimensions) increases, each having a range of possible values, the number of combinations of values exponentially increases. Therefore, an algorithm needs to be trained on samples with enough combinations of values to learn sufficient relationships and patterns in the data to avoid overfitting. In materials science, this requirement means collecting more samples, which can be costly and thus has important implications on when to use one technique over another.



There are many different types of algorithm, but many generally follow an inverse relationship between interpretability and flexibility, providing researchers with a wealth of techniques to analyze a wide variety of different datasets. Typically, if the goal is to understand the precise relationship between variables and a corresponding output, interpretable and rigid models, such as linear regression, are most suited to this type of problem. These models are particularly useful if the goal is to prove a hypothesis. If prediction accuracy is the goal or data has high-dimensionality, more flexible algorithms can be leveraged to include more variables and observations in the dataset or reduce dimensional complexity with minimal loss of information.

It is important to note that a single algorithm will not work for all possible datasets, which further signifies the importance of using a wider toolset when designing experiments and analyzing materials data. ■



Credit: Ramprasad et al.[14]; Springer, 2017 (CC BY 4.0)

Figure 3. Predictive modeling framework that leverages existing materials data to train models to predict properties of new materials. Adapted from Ramprasad et al.[14]

processes based on the 100-year-old mathematical formulations of Thomas Bayes and Carl Friedrich Gauss, respectively.

Numerous industries have leveraged advanced analytics for decades to support decision-making, including market, social media, financial, manufacturing, and distribution data.[10] Moreover, the closely related pharmaceutical industry pioneered the use of data-driven techniques for drug discovery and development as early as the 1970s, creating the parallel field of bioinformatics. Until recently, the materials science industry trailed behind these businesses, but we are now beginning to witness the disruptive potential of predictive modelling and discovery-based data mining techniques, in combination with computational and physical experiments, to decrease the materials development timeframe.

Predictive modeling

With a critical volume of historical materials data, the underlying characteristics that best describe material behavior can be "learned" by algorithms and used to train surrogate models that can make accurate forecasts on new data. Such learning methods establish a mapping between a suitable representation of a material, called the material's fingerprint, and any of its properties from existing data (Figure 3).[14]

The fingerprint is composed of an optimal number of descriptors (or variables) that the model can use to learn what a material is and accurately predict its properties. In essence, the material fingerprint is the DNA code and descriptors are the individual "genes" that connect the empirical or fundamental characteristics of a material (e.g., elemental composition) to its macroscopic properties.[15] Once a suitable number of descriptors and quantities are obtained (to avoid overfitting and high variance, see Sidebar: "Introduction to algorithms") for a range of materials from a database, they can be mapped to their corresponding output property data by finding the best fit to the observations resulting in a predictive model.

Once a model is validated, the model predictions are instantaneous, which makes it possible to forecast the properties of existing, new, or hypothetical material compositions, purely based on past data, prior to performing expensive computations or physical experiments. Predictive models are highly suited for interpolation, i.e., searching within an existing database. Extrapolation, i.e., leaping from one composition space to another or expanding the original database, is also possible but can lead to larger errors and uncertainties. However, methods that promote easy assessment of model uncertainties can be used to overcome this issue by supporting the decision as to which set of experiments should be performed next.[16] Subsequently, once new data is collected and confirmed by computational or physical experiment, it can be fed back into the model to improve accuracy and iteratively narrow in on new candidates for a specific application. This explanation of predictive modeling demonstrates that MI is not intended to replace experiments (or the scientist) but rather help arrive at a desired result in a much shorter timeframe.

While predictive models are attractive for identifying and developing new materials, there are other useful tools available in the advanced analytics toolbox that can identify structure, patterns, and relationships in complex input data that do not necessarily require the associated outputs. These tools become highly beneficial when a systematic search for each significant variable of a process or microstructure evolution mechanism is computationally or experimentally expensive because they involve many variables.[6]

For example, dimensionality reduction techniques can transform vast arrays of input data into a reduced, easily visualized space—typically two or three dimensions—and identify relationships or patterns with minimal loss of information.[6] With this technique, what may have once required a large collection of graphs can now be summarized in a single chart representing the entire process.

While dimensionality reduction and clustering techniques are not predictive tools, they can support predictive modeling with complex data in which the number of observational data is too low or the number of variables needs to be reduced to improve the efficiency of an analysis.

Practical applications of MI

One of the most compelling opportunities offered by MI is the potential to accelerate the discovery of new materials. As constituent elements of a material increase, the number of possible combinations begin to explode. For example, a ternary compound of the form $A_xB_yC_z$ (where x, y, and z are stoichiometric quantities) corresponds to billions of possible inorganic materials that increases as more constituents are included.[7] However, not all of these materials will be stable and finding those that are stable would take an unfathomable amount of time.

To calculate material properties, computational methods require crystal structure information, which is not readily available nor easy to calculate for all possible candidates across the vast compositional space.

To overcome this challenge, researchers trained a surrogate model on a small subset of existing DFT data from the Inorganic Crystal Structure Database (ICSD) to predict the formation energy of new materials solely based on their stoichiometric composition. The model was subsequently used to instantly scan 1.6 million ternary compounds of which 4,500 previously unknown materials were expected to be stable based on their predicted formation energies.[12]

While the output is astonishing, the approach is certainly not trivial and demonstrates the potential to leverage data-driven techniques to discover new materials that could have important implications on replacing critical materials that are approaching their limitations or subject to supply disruptions.

Data-driven approaches also are used to explore the likelihood of achieving a set of target properties given a series of opposing constraints, such as materials that are difficult to secure due to pressure from finite resources. For example, a study of more than 2,800 compounds identified as being either abundant or scarce was used to compare the charge/discharge voltages and specific energies (key performance properties for batteries) against their relative abundances. Approximately 500 materials with known voltages and specific energies were used to train a data-driven model (using material chemical formula as model input) to predict the properties of the 2,800 candidates. Subsequently, Figure 4 visualizes the model predictions and indicates the density of candidates that may be found at a particular region of property space.

It is clear the highest density of candidates are clustered around a specific energy of 500 Wh/kg and average voltage of 2.5–3V. However, the study reveals scarce materials offer a greater likelihood of finding candidates with higher specific energy while abundant materials offer the widest range of possible voltages. The approach demonstrates how data-driven algorithms can be used to assess simultaneously the trade-off between performance and multiple constraints, such as resource considerations over a vast composition space at groundbreaking speeds.[17]

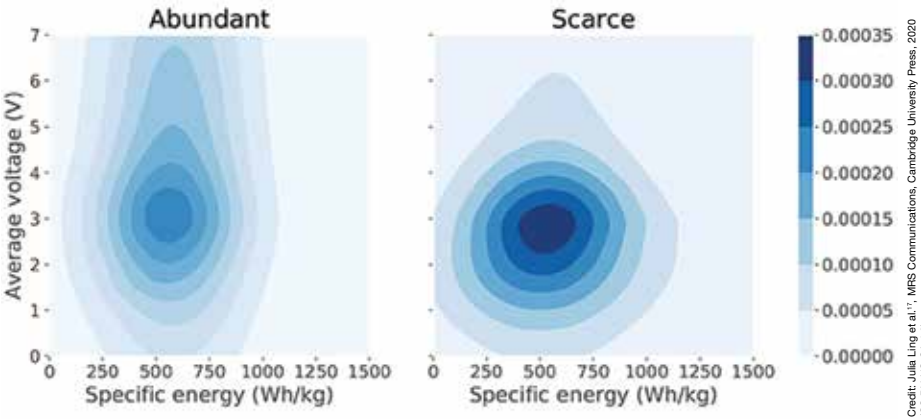


Figure 4. Design space visualization plots for abundant and scarce cathode materials based on the summed probability density, which indicates how easy it is to find candidates in a particular property space region.

Getting started—Important considerations

Data-driven methods for materials holds a great deal of promise, but it is important to note they can lead to the development of "fools-gold" as they are only as good as the data they consume.[18] For example, equivalent materials properties may be measured differently depending on the data source, and these contextual differences, among other hidden variables, can introduce errors into analyses, thus limiting their accuracy. Furthermore, materials data is diverse (e.g., numerical, text, image, graphical, spectra) and still sparsely populated relative to other industries.

These challenges have spearheaded a global effort at academic and government levels to develop techniques and methodologies that continue to generate large quantities of high-fidelity materials property data and develop structurally diverse materials databases that can be interrogated by advanced algorithms.[8] This effort is achieved by means of both HT computational techniques as well as the emerging use of HT experimental techniques based on combinatorial materials synthesis and rapid screening via automated instrumentation.[9] These techniques are similar to the combinatorial chemistry techniques used for drug discovery in the pharmaceutical industry.

Researchers are also developing ways of unifying global materials databases to explore patterns across separate databases that cover different aspects of materials science (i.e., databases of crystal structures and physical properties).[5,10,19] Such a change of scale requires new data management methodologies to certify the validity of materials data and to ensure it can be found, accessed, and shared in a commonly accepted format.

At the enterprise level, most companies (big or small) have historical data from a wide variety of sources, including supplier and customer data. However, accessing sufficient datasets remains a challenge within each organization, independent of size, as data sources may not be easily accessed or may be stored in various formats, from tracking data in spread sheets and, in some cases, by hand in notebooks.

For many organizations, simply applying advanced analytics to data via open-source or even commercial software will not work as model development must be based on the goals of the analysis, the solutions being sought, and the available data. So they require access to data workflows that can inspect, clean, and store data in a structured format; scalable and flexible analytics capabilities that include the correct hardware, software, security protocols, and other relevant data infrastructures; upfront investment in equipment, including materials characterization or high-performance computing capabilities; and skilled workers, especially materials scientists, data scientists, and data engineers, which can be expensive.

Organizations that attempt to build these new capabilities from the ground up may face steep learning curves result-

ing in failure or a much longer-term return on investment due to the inherent challenges of acquiring, structuring, and analyzing data.

Opportunity for industry

Mainstream developments in MI have primarily been led by the academic and government communities. However, sufficient progress was achieved over the last few years to attract the attention of industry. Companies are now beginning to practice the principles of MI and apply the new knowledge generated by mainstream academic and government research to everyday challenges across their enterprises.[20] Subsequently, a number of emerging industry–university–government ecosystems are evolving around the world that are composed of major government research institutes, multinational companies, and early- to late-stage start-ups. Together, these organizations are pioneering the use of MI across the materials development lifecycle that not only involves discovery and design but also includes downstream process optimization and after deployment in the field, with a growing number of commercial successes.

While these developments are exciting examples of transformation in the materials science industry, the most exciting prospect is that materials scientists and engineers can now leverage a much wider range of data-driven tools within the familiar experimental framework to solve a variety of challenges, from materials development to process optimization, that may have been unsolvable or too complex to address until now. While the analytics tools have been available for many decades, the right technological advances (from increased computing power to accelerating data volumes) and materials industry needs have converged at the right point in time to take advantage of these powerful methods today and support the developments of the future.

As technologies continue to improve, new methods will constantly evolve at an ever-increasing pace, which will positively impact materials challenges further down the line. An imperative is now upon us to stay on top of these emerging developments and to find our unique place amongst the growing materials informatics ecosystem.

About the author

Richard Padbury is senior technology consultant at Lucideon in Raleigh, North Carolina. Contact Padbury at richard.padbury@lucideon.com.

References

[1]Satell, G. (2019, January 9). Materials science may be the most important technology of the next decade. Here's why. Retrieved from *Digital Tonto*: https://www.digitaltonto.com

[2]Nemko, M. (2016, July 15). The psychology of scientific advancement. A "The Eminents" interview with Michio Kaku. Retrieved from https://www.psychologytoday.com/us/blog/how-do-life/201607/the-psychology-scientific-advancement

[3]The Minerals, Metals & Materials Society (TMS), Creating the Next-Generation Materials Genome Initiative Workforce (Pittsburgh, PA: TMS, 2019). Electronic copies available at www.tms.org/mgiworkforce.

[4]Material property charts. (n.d.). Created using CES EduPack 2019, ANSYS Granta (2020 Granta Design). Retrieved from https://granta-design.com/education/students/charts.

[5]Agrawal, A., & Choudhary, A. (2016). Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. *APL Materials*, **4**(5), 053208. https://doi.org/10.1063/1.4946894

[6]Rajan, K. (2013). Materials Informatics. *Informatics for Materials Science and Engineering*, 1–16. doi: 10.1016/b978-0-12-394399-6.00001-1

[7]Davies, D. W., Butler, K. T., Jackson, A. J., Morris, A., Frost, J. M., Skelton, J. M., & Walsh, A. (2016). Computational screening of all stoichiometric inorganic materials. *Chem*, **1**(4), 617–627.https://doi.org/10.1016/j.chempr.2016.09.010

[8]Nosengo, N. (2016, May 13). Can artificial intelligence create the next wonder material? Retrieved from https://www.nature.com/news/can-artificial-intelligence-create-the-next-wonder-material-1.19850

[9]Curtarolo, S., Hart, G. L. W., Nardelli, M. B., Mingo, N., Sanvito, S., & Levy, O. (2013). The high-throughput highway to computational materials design. *Nature Materials*, **12**(3), 191–201. https://doi.org/10.1038/nmat3568

[10]Schleder, G. R., Padilha, A. C. M., Acosta, C. M., Costa, M., & Fazzio, A. (2019). From DFT to machine learning: Recent approaches to materials science–a review. *Journal of Physics: Materials*, **2**(3), 032001. https://doi.org/10.1088/2515-7639/ab084b

[11]Shobha, G., & Rangaswamy, S. (2018). Machine learning. *Handbook of Statistics*, **197–228**. https://doi.org/10.1016/bs.host.2018.07.004

[12]Meredig, B., Agrawal, A., Kirklin, S., Saal, J. E., Doak, J. W., Thompson, A., ... Wolverton, C. (2014). Combinatorial screening for new materials in unconstrained composition space with machine learning. *Physical Review B*, **89**(9). https://doi.org/10.1103/physrevb.89.094104

[13]James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). An introduction to statistical learning. New York, United States: Springer Publishing.

[14]Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A., & Kim, C. (2017). Machine learning in materials informatics: recent applications and prospects. *Npj Computational Materials*, **3**(1). https://doi.org/10.1038/s41524-017-0056-5

[15]Balachandran, P. V., Broderick, S. R., & Rajan, K. (2011). Identifying the 'inorganic gene' for high-temperature piezoelectric perovskites through statistical learning. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **467**(2132), 2271–2290. https://doi.org/10.1098/rspa.2010.0543

[16]Ling, J., Hutchinson, M., Antono, E., Paradiso, S., & Meredig, B. (2017). High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates. *Integrating Materials and Manufacturing Innovation*, **6**(3), 207–217. https://doi.org/10.1007/s40192-017-0098-z

[17]Peerless, J. S., Sevgen, E., Edkins, S. D., Koeller, J., Kim, E., Kim, Y., ... Ling, J. (2020). Design space visualization for guiding investments in biodegradable and sustainably sourced materials. *MRS Communications*, 1–7. https://doi.org/10.1557/mrc.2020.5

[18]Riley, P. (2019, July 30). Three pitfalls to avoid in machine learning. Retrieved from https://www.nature.com/articles/d41586-019-02307-y

[19]Himanen, L., Foster, A. S., & Rinke, P. (2020, January 22). Data–driven materials science: Status, challenges, and perspectives. Retrieved from https://onlinelibrary.wiley.com/doi/10.1002/advs.201903667

[20]NIMS and Four Chemical Companies to develop a framework for promoting open innovation: NIMS. (n.d.). Retrieved from http://www.nims.go.jp/eng/news/press/2017/06/201706190.html ■