ACS Photonics

pubs.acs.org/journal/apchd5

Editorial

Accelerated Discovery of Optoelectronic Materials

Cite This: *ACS Photonics* 2021, 8, 699−701

Read Online

ACCESS | Metrics & More | Article Recommendations

Downloaded via UNIV OF TORONTO on March 18, 2021 at 14:56:13 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

The materials that comprise optoelectronic devices, including solar cells, light-emitting diodes (LEDs), and photodetectors, have seen impressive progress in the last 50 years. Research has been fueled in significant part by the ever-increasing technological innovations demanded by mobile/consumer electronic devices such as smartphones, united with the growing need for more efficient energy devices to help combat climate change. More than $25 billion USD was devoted to clean energy materials by venture investors between 2004 and 2008,[1] and yet the return on investment (ROI) in this sector was less than half of those returned by similar VC efforts over the same time frame. This correlates with a halving in global expenditures on renewable energy in the decade that ensued.[2] We link the poor return to the difficult, expensive, long, and uncertain process of materials discovery.

PITFALLS OF TRADITIONAL MATERIALS DISCOVERY

The traditional process by which new materials are discovered is limited by the need for a substantial initial investment to build the infrastructure needed to synthesize and characterize materials; the immense chemical space from which materials can be synthesized; and the cost per experiment, which limits the portion of the chemical space that we realistically can explore. What is more, incorporating materials into devices requires even further development and optimization of materials interfaces, stability, reproducibility, and processing.

The organic−inorganic halide perovskite solar cells illustrate many of these points. Concerted and intensive research efforts from thousands of researchers globally have improved the power conversion efficiency of perovskite solar cells from less than 10% to 25.3% in under a decade.[3] Yet, despite the tremendous progress, and tremendous resources, devoted to this topic, perovskite solar cells still need further progress on stability under accelerated testing conditions.[4]

These factors limit the number of market participants willing to invest in commercializing such technologies. Perovskite solar cells and, more generally, energy materials will benefit from strategies that accelerate materials discovery.

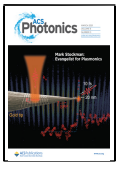
MACHINE LEARNING−ACCELERATED MATERIALS DISCOVERY FRAMEWORK

Materials discovery involves a large number of variables in each new experiment; machine learning is particularly well-tailored to translating unintuitive high-dimensional spaces to a single, useful output. When trying to correlate a set of parameters that define an experiment with an outcome, for example, correlating the crystal structure of a new material with a computationally evaluated bandgap or correlating the combination of precursor stoichiometries, concentrations, and temperatures used in fabricating a thin-film solar cell with its power conversion efficiency, it is critical to use techniques that can capture many different interactions. ML captures multivariate interactions and can do so in ways that are not explicitly or implicitly considered by researchers, who may limit their scope to looking for correlations that have traditionally been exploited. ML can learn underlying patterns, both in a material's crystal structure and in the experimental parameters used to generate a material or device, that are unavailable with human intuition alone. As a result, ML opens an exciting avenue to accelerate the discovery of new materials.

Early in its use in materials science, ML was used to aid researchers in the computational evaluation of new materials. Computational screening is often a first step in the materials discovery process: it provides researchers with a set of materials compositions that may have a desired set of properties. This is done by using first-principles calculations, such as density functional theory (DFT) or molecular dynamics (MD), to find the energetically favored crystal structure of a new material through a series of iterative relaxation steps. The starting point, before the relaxation begins, of the crystal structure is often derived from an experimentally known material in which individual elements of the known material are substituted for the elements of the new material. From the relaxed crystal structure, it becomes possible to approximate energetic properties, such as the formation energy and the bandgap, of the new material. The relaxation of the crystal structure and approximation of the energetic properties are computationally expensive, with single compounds taking up to 12 CPU years on modern supercomputers, limiting the number of compounds that can be realistically explored to ∼10^4.[5]

Fortunately, there exist large, readily available data sets of computed materials properties, such as Materials Project,[6] the Chemical Space Project,[7] and ANI-1.[8] These contain formatted, standardized data, ideal for ML, and have catalyzed the development of ML models to replace the first-principles calculations used to screen materials.

Received: December 23, 2020
Accepted: December 31, 2020
Published: March 17, 2021



ML models can be trained with data from these data sets without requiring a researcher to generate new data. These models take as an input a representation of an (un)relaxed crystal structure and predict a set of properties, such as bandgaps and formation energies, directly. The trained ML models, owing to their extremely fast forward compute times, enable exploration of a much wider chemical space ($>10^5\times$)[9] than DFT alone can explore. This vast chemical space is essential for the evaluation of optoelectronic materials: when considering partial substitution of an element in a material (doping), the number of possible materials belonging to a single material class rapidly approaches infinity. The rapid exploration of new material compositions can help reduce the incremental nature of materials optimization by providing researchers with insights to widely varying material compositions, without significantly increasing the associated opportunity cost of exploration.

ML is also being used to accelerate the synthesis and characterization of new materials. While ML, particularly deep learning, typically requires large data sets, simultaneous advances in robotic synthesis and characterization techniques allow researchers to increase experimental throughput.[10] Access to high-throughput experimentation (HTE), either through the use of robotic equipment or with other highly parallelized techniques, generates large quantities of data. ML models can then be trained to both predict the outcomes of future experiments and autonomously interpret the results of the characterization of the material. This allows the materials discovery cycle to become highly parallelized and can greatly increase the rate at which interesting materials are discovered: robot-based techniques have been shown to increase the rate of experiments from ∼20 samples/day to more than 6000.[11]

Even without access to robotic equipment or other means to perform HTE, ML can be used to optimize the performance of a new material. Bayesian optimization, a technique particularly useful when optimizing a hard-to-evaluate function, can be used to improve the performance of a material with minimal experimental cycles: for example, this process was used to develop stable organic photovoltaic (OPV) materials, in which the number of experiments needed to optimize the material was reduced from 500 to just 60.[11] Bayesian optimization exploits similarities, evaluated by the ML model in a high-dimensional space, between each new experiment to maximize the rate-of-learning in each experimental cycle. It levers Bayes formula to evaluate both the expected outcome of a new experiment and the uncertainty of the experiment. If a new experiment is proposed with a set of parameters (temperature, pressure, humidity) similar to one the model has seen before, the model will predict the outcome of the new experiment with high certainty; if the proposed experiment contains a set of parameters that are significantly different than the ones used to train the model, it will predict the expected outcome with low certainty. As a result, the model can decide which experiment to perform next: either one with the highest likelihood of having an improved outcome or one that provides the most new information by probing an experimental space with high uncertainty. Bayesian optimization techniques work on a relatively small number (∼10) of data points and, as a result, can be employed from the start of a new research project. This is a key enabler of self-driving science, in which ML models determine the next set of experiments to be performed, robots perform and analyze the experiments, and the ML model is then updated before repeating the cycle, and is an active area of research.[12,13]

It is important to focus on new, efficient materials that can reach commercialization: here, stability, reproducibility, and repeatability are chief considerations. Measuring these properties can take significant time and resources: for example, reliability is something typically measured over hundreds and even thousands of hours, even if an accelerated lifetime model is pursued. ML models can be used to learn proxies for different target parameters such as stability, repeatability, and reproducibility. By incorporating physics-based fingerprints into ML surrogate models, it is possible for the models to learn the mechanisms underpinning material degradation. This technique has recently been used to improve the understanding of the stability enhancement of perovskite films seen when these are capped with a low-dimensional perovskite material.[14] The ML model first optimized the stability of different perovskite films and then learned the correlation of improved stability. As a result, the model was able to suggest modifications to the capping layer that offered improved stability compared with the traditional counterparts. Application of ML to this area of materials discovery is still in its infancy in part due to the required time to generate the data sets necessary to learn the proxies. However, this may be the aspect of the materials discovery process where ML provides the greatest benefit: researchers will be able to measure the proxies for stability, repeatability, and reproducibility while improving the efficiency of a new material and can consider ways to improve these simultaneously. This can help limit the number of new materials that fail at the stability testing stage of the discovery process and increase the chances that those we invest in will operate successfully in real-world applications.

FUTURE CONSIDERATIONS OF MACHINE LEARNING–ACCELERATED MATERIALS DISCOVERY

The progress on ML methods in materials science has been impressive and rapid, but there remain a few key considerations for the field. We propose here three mandates for the accelerated materials discovery field: align on a metric of acceleration-of-discovery to report; develop protocols to decide whether to invest in the infrastructure required for an accelerated materials discovery framework must be developed; and shift emphasis to accelerating the lab-to-market timeline.

Choosing the right acceleration metric will help the community keep its focus on the true purpose of materials discovery. In early studies, the number of samples that can be tested with the accelerated framework in a unit of time has been compared with the number of samples that a researcher can test over the same time period. While this metric is compelling and provides an easy benchmark with which to compare different methods, it fails to capture the acceleration in discovering good materials.

A metric that comes closer is the number of new materials per unit time that simultaneously meets target properties. Just as the perovskite community has recently converged around stability metrics,[15] the ML materials community will benefit by defining accelerated-discovery metrics that allow true quantitative progress to be measured based on benchmarks.

The field will benefit from developing approaches to estimate, at the outset of a potential project, whether that process stands to net benefit from an accelerated approach. This must include an analysis of the one-time cost of

https://dx.doi.org/10.1021/acsphotonics.0c01949
ACS Photonics 2021, 8, 699−701

developing the infrastructure required to carry out an accelerated discovery. This applies to acquiring the necessary experimental equipment, training personnel on its operation, training personnel on the ML methodologies to be used, acquiring the data necessary to train the models, and then actually training the models. Of these, the time it takes to train models and to acquire data are reported most often; yet, each of these tasks must be completed before any acceleration can happen. These need to be considered both before deciding whether a project merits an accelerated framework and after a project is completed. If the community starts publishing these metrics, it will give other researchers a more realistic idea of what is required when trying to accelerate materials discovery and whether a project can truly benefit from the development of an accelerated framework.

Finally, we propose that the emphasis of these accelerated frameworks should be on the acceleration of the lab-to-market timeline of a new material. The success of a new methodology for materials discovery should hinge on whether it accelerates the development of a new material that is eventually commercialized. This is not an easy ask: the commercialization of new materials is a many faceted, difficult process. Key metrics to track can include the cost of materials required to do research; the robustness of the experimentation methods; the modularity of the experimental equipment; and the time it takes to train new personnel on its operation. By considering these at the onset of the development of accelerated materials frameworks, there exists the opportunity to increase the development of new, commercially relevant materials.

Andrew Johnston orcid.org/0000-0002-4545-532X
Edward H. Sargent orcid.org/0000-0003-0396-6495

AUTHOR INFORMATION

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsphotonics.0c01949

Notes

Views expressed in this editorial are those of the authors and not necessarily the views of the ACS.

REFERENCES

(1) Gaddy, B. E.; Sivaram, V.; Jones, T. B.; Wayman, L. Venture Capital and Cleantech: The Wrong Model for Energy Innovation. *Energy Policy* **2017**, *102*, 385−395.
(2) Kulanov, A.; Issakhova, A.; Koshkina, O.; Issakhova, P.; Karshalova, A. Venture Financing and the Fuel and Energy Complex: Investing in Alternative Energy. *International Journal of Energy Economics and Policy* **2020**, *10* (5), 531−538.
(3) Best Research-Cell Efficiency Chart, https://www.nrel.gov/pv/cell-efficiency.html (accessed Dec 18, 2020).
(4) A Decade of Perovskite Photovoltaics. *Nature Energy* **2019**, *4* (1), 1−1, .
(5) Askerka, M.; Li, Z.; Lempen, M.; Liu, Y.; Johnston, A.; Saidaminov, M. I.; Zajacz, Z.; Sargent, E. H. Learning-in-Templates Enables Accelerated Discovery and Synthesis of New Stable Double Perovskites. *J. Am. Chem. Soc.* **2019**, *141* (8), 3682−3690.
(6) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1* (1), 011002.
(7) Reymond, J.-L. The Chemical Space Project. *Acc. Chem. Res.* **2015**, *48* (3), 722−730.
(8) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1, A Data Set of 20 Million Calculated off-Equilibrium Conformations for Organic Molecules. *Sci. Data* **2017**, *4* (1), 170193.
(9) Choubisa, H.; Askerka, M.; Ryczko, K.; Voznyy, O.; Mills, K.; Tamblyn, I.; Sargent, E. H. Crystal Site Feature Embedding Enables Exploration of Large Chemical Spaces. *Matter* **2020**, *3*, 433−448.
(10) Sun, S.; Hartono, N. T.P.; Ren, Z. D.; Oviedo, F.; Buscemi, A. M.; Layurova, M.; Chen, D. X.; Ogunfunmi, T.; Thapa, J.; Ramasamy, S.; Settens, C.; DeCost, B. L.; Kusne, A. G.; Liu, Z.; Tian, S. I.P.; Peters, I. M.; Correa-Baena, J.-P.; Buonassisi, T. Accelerated Development of Perovskite-Inspired Materials via High-Throughput Synthesis and Machine-Learning Diagnosis. *Joule* **2019**, *3* (6), 1437−1451.
(11) Langner, S.; Häse, F.; Perea, J. D.; Stubhan, T.; Hauch, J.; Roch, L. M.; Heumueller, T.; Aspuru-Guzik, A.; Brabec, C. J. Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems. *Adv. Mater.* **2020**, *32* (14), 1907801.
(12) Roch, L. M.; Häse, F.; Kreisbeck, C.; Tamayo-Mendoza, T.; Yunker, L. P. E.; Hein, J. E.; Aspuru-Guzik, A. ChemOS: An Orchestration Software to Democratize Autonomous Discovery. *PLoS One* **2020**, *15* (4), e0229862.
(13) Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Chimera: Enabling Hierarchy Based Multi-Objective Optimization for Self-Driving Laboratories. *Chem. Sci.* **2018**, *9* (39), 7642−7655.
(14) Hartono, N. T. P.; Thapa, J.; Tiihonen, A.; Oviedo, F.; Batali, C.; Yoo, J. J.; Liu, Z.; Li, R.; Marrón, D. F.; Bawendi, M. G.; Buonassisi, T.; Sun, S. How Machine Learning Can Help Select Capping Layers to Suppress Perovskite Degradation. *Nat. Commun.* **2020**, *11* (1), 4172.
(15) Khenkin, M. V.; Katz, E. A.; Abate, A.; Bardizza, G.; Berry, J. J.; Brabec, C.; Brunetti, F.; Bulović, V.; Burlingame, Q.; Di Carlo, A.; Cheacharoen, R.; Cheng, Y.-B.; Colsmann, A.; Cros, S.; Domanski, K.; Dusza, M.; Fell, C. J.; Forrest, S. R.; Galagan, Y.; Di Girolamo, D.; Grätzel, M.; Hagfeldt, A.; von Hauff, E.; Hoppe, H.; Kettle, J.; Köbler, H.; Leite, M. S.; Liu, S.; Loo, Y.-L.; Luther, J. M.; Ma, C.-Q.; Madsen, M.; Manceau, M.; Matheron, M.; McGehee, M.; Meitzner, R.; Nazeeruddin, M. K.; Nogueira, A. F.; Odabaşı, Ç.; Osherov, A.; Park, N.-G.; Reese, M. O.; De Rossi, F.; Saliba, M.; Schubert, U. S.; Snaith, H. J.; Stranks, S. D.; Tress, W.; Troshin, P. A.; Turkovic, V.; Veenstra, S.; Visoly-Fisher, I.; Walsh, A.; Watson, T.; Xie, H.; Yıldırım, R.; Zakeeruddin, S. M.; Zhu, K.; Lira-Cantu, M. Consensus Statement for Stability Assessment and Reporting for Perovskite Photovoltaics Based on ISOS Procedures. *Nature Energy* **2020**, *5* (1), 35−49.

https://dx.doi.org/10.1021/acsphotonics.0c01949
ACS Photonics 2021, 8, 699−701