COMMENT

POLICY What sorts of investment lead to energy innovation? **p.25**

PHYSICS New biography of Enrico Fermi, father of nuclear fission **p.28**

CONSERVATION Iran's cheetahs need support from the government **p.31**

OBITUARY Frank Brown, chronologist of hominin fossil beds, remembered **p.32**

YURIKO NAKAO/REUTERS

A solar module on display at an expo in Tokyo.

Use machine learning to find energy materials

Artificial intelligence can speed up research into new photovoltaic, battery and carbon-capture materials, argue **Edward Sargent**, **Alán Aspuru-Guzik** and colleagues.

The world needs more energy. Governments and companies are investing billions of dollars in technologies to harvest, convert and store power[1]. And as silicon solar cells approach the limit of their performance, researchers are looking to alternatives based on perovskites and quantum dots[2]. The batteries that store the energy must get cheaper, more efficient and longer-lasting[3]. And devices need to be manufactured from safe and abundant materials such as copper, nickel and carbon rather than from lead, platinum or gold. Life-cycle analyses of the materials need to show improved carbon footprints, as well as the ability to match the scale of the global energy challenge.

Enormous quantities of experimental data are being generated on the properties of such materials. The US National Institute of Standards and Technology, for example, hosts 65 databases, some with as many as 67,500 measurements. Also, since 2010, more than 1.7 million scientific papers have been published on batteries and solar cells alone.

Relating the structure of a material to its function needs accelerating. The search space is vast. Many materials are still found empirically: candidates are made and tested a few samples at a time. Searches are subject to human bias. Researchers often focus on a few combinations of the elements that they deem interesting.

Computational methods are being developed that automatically generate structures and assess their electronic features and other properties[4]. The Materials Project, for instance, is using supercomputers to predict the properties of all known materials[5]. It currently lists predicted properties for more ▶

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

than 700,000 materials. But the tremendous potential to translate such data into industrial and commercial applications is still a long way from being realized.

Machine learning — algorithms trained to find patterns in data sets — could greatly speed up the discovery of energy materials. It has already been used to predict the results of quantum simulations to identify potential molecules and materials for flow batteries, organic light-emitting diodes[6], organic photovoltaic cells and carbon dioxide conversion catalysts[7]. The algorithms can predict results in a few minutes, compared with the hundreds of hours it takes to run the simulations[8].

Challenges remain, however. There is no universal representation for encoding materials. Different applications require different properties, such as elemental composition, crystal structure and conductivity. Well-curated experimental data on materials are rare, and computational tests of hypotheses rely on assumptions and models that may be far from realistic under experimental conditions.

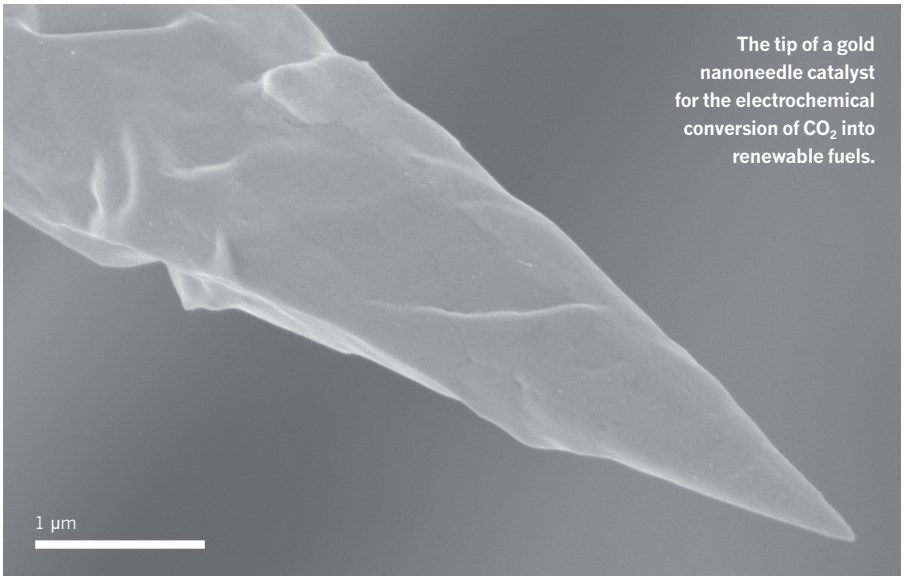
The machine-learning and energy-sciences communities should collaborate more. They must understand each other's capabilities and needs. We offer the following recommendations, which came out of a workshop run by the Canadian Institute for Advanced Research in May in Boston, Massachusetts.

FOUR BRIDGES

Share meaningful data. Materials scientists should organize their data into standardized, machine-readable forms, such as the 'comma-separated values' (CSV) files commonly used in spreadsheet applications. At present, results tend to be condensed into graphs and tables, each group organizes its data differently and testing conditions and experimental set-ups vary. Many teams process their raw spectra or normalize their data and the models are often subject to errors and biases in the absence of experimental evidence to calibrate results.

Government funding agencies and publishers should require data to be uploaded to a publicly accessible database such as the Materials Project, the Materials Data Curation System or the Citrination platform[9]. Consortia and universities could share the costs of maintaining these databases; credit could be given when citing them. Alternatively, an independent entity could be established to maintain an experimental database, in much the same way as protein crystal structures are currently shared in the Protein Data Bank. It is important to include negative results — machine-learning algorithms need to be able to differentiate between materials that meet performance targets and those that don't.

A culture of sharing also needs to be encouraged within the materials-science community. The computer-science and medical communities are reaping huge benefits from making their large data sets available for machine learning. For example, IBM Watson Health in Cambridge, Massachusetts, is using machine learning to improve drug discovery and cancer therapies.



The tip of a gold nanoneedle catalyst for the electrochemical conversion of CO_2 into renewable fuels.

SARGENT LAB

Spur collaboration with competitions. 'Grand challenge' awards are a cost-effective way to foster innovation. For example, the XPRIZE initiative has led to breakthroughs in carbon capture and utilization, ocean discovery and artificial intelligence. The 2004 Ansari XPRIZE for Suborbital Flight resulted in SpaceShipOne, the first private spaceship to enter outer space. The Kaggle platform uses competitions to crowdsource solutions to computer-modelling and data-science problems, such as predicting the activity of drug-like molecules. And sponsored hackathons run by companies such as AngelHack in San Francisco, California, have developed apps for firms including Mastercard.

"'Grand challenge' awards are a cost-effective way to foster innovation."

We propose that machine-learning competitions be established to encourage the finding of new energy materials in publicly available data sets, such as those of the Materials Genome Initiative, the European Novel Materials Discovery Laboratory (NOMAD) initiative or Citrination. The goal would be to predict a material for a specific application or property. For example, nanoscale porosity is key to carbon-capture materials, the gap between electronic bands is an important descriptor for solar cells and hardness could be used to develop lightweight composite materials for transport. Machine learning can consider multiple properties simultaneously.

Competitions could be sponsored by university departments or by commercially supported institutes such as Canada's Vector Institute for Artificial Intelligence in Toronto and Montreal Institute for Learning Algorithms, or the US-based Toyota Research Institutes. They could even take a similar format to the online protein-folding game Foldit, in which people take part both for the glory of discovery and to beat others' scores. Intellectual property could be managed along similar lines to the XPRIZE.

Develop a shared language. Chemists, computer scientists, machine-learning experts, materials engineers, programmers and physicists all have their own areas of expertise and nomenclature. Materials engineers, for example, are skilled at making materials of various compositions, and machine-learning researchers would need to understand these subtleties to be able to predict materials of practical use.

We propose that universities host workshops and summer schools and develop curricula that bridge these fields. Some summer schools already teach conventional computational chemistry and machine learning for computer-science applications; few incorporate both. More forums should be set up for training, such as the Understanding Many-Particle Systems with Machine Learning programme run by the Institute for Pure and Applied Mathematics in Los Angeles, California.

Accelerate and automate. As a fast-moving area of research, energy-materials discovery is a perfect test bed for advanced machine-learning techniques. Machine learning has tended to assume a fixed training set; robots for autonomous cars are trained to drive using images or videos of roads, for example. But this can be slow and outcomes are difficult to repeat or vary between users. By contrast, the data landscape for energy materials changes

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

KEVIN FRAYER/GETTY

continually as new information and models emerge. Useful here is the growing field of deep reinforcement learning, in which agents explore their evolving environment to find the best solutions. Applying such algorithms to materials discovery would make searches progressively more efficient and allow the learner to explore the space of molecules, just as chemists do.

WHAT NEXT

Developing machine-learning approaches is one of the main goals of the Clean Energy Materials Innovation Challenge run by the Mission Innovation global collaboration. The collaboration is funded by voluntary government pledges — and nations must deliver on their commitments with the necessary investments.

In summary, more investment is needed in artificial intelligence and robotics-driven materials research throughout the world. More data must be made available to people programming the robots. And experimentalists, robotics experts and algorithm designers should communicate and collaborate more to facilitate rapid troubleshooting.

Time is running out to find the new energy technologies the world needs. ■

Phil De Luna *is a graduate student in the Department of Materials Science and Engineering at the University of Toronto, Canada.* **Jennifer Wei** *is a graduate student in the Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, USA.* **Yoshua Bengio** *is a professor in MILA, Department of Computer Science and Operations Research, Université de Montréal, Canada.* **Alán Aspuru-Guzik** *is a professor in the Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, USA.* **Edward Sargent** *is a professor in the Department of Electrical and Computer Engineering, University of Toronto, Canada.*
e-mails: ted.sargent@utoronto.ca; alan@aspuru.com

1. Bernstein, A. *et al. Nature* **538,** 30 (2016).
2. Chu, S., Cui, Y. & Liu, N. *Nature Mater.* **16,** 16–22 (2016).
3. Huskinson, B. *et al. Nature* **505,** 195–198 (2014).
4. Curtarolo, S. *et al. Nature Mater.* **12,** 191–201 (2013).
5. Jain, A. *et al. APL Mater.* **1,** 11002 (2013).
6. Gómez-Bombarelli, R. *et al. Nature Mater.* **15,** 1120–1127 (2016).
7. Liu, M. *et al. Nature* **537,** 382–386 (2016).
8. Ji, H. & Jung, Y. *J. Chem. Phys.* **146,** 064103 (2017).
9. O'Mara, J., Meredig, B., & Michel, K. *J. Miner. Met. Mater. Soc.* **68,** 2031–2034 (2016).

For the full list of co-signatories, see go.nature.com/2i8psoe.

A solar farm floats on a lake that formed after the collapse of a deep coal mine in Huainan, China.

Six principles for energy innovation

Decades of experience must inform future initiatives, urge **Gabriel Chan** and colleagues.

Last month, the European Union marked the tenth year of its Strategic Energy Technology Plan. It is one of many policy initiatives worldwide to accelerate innovation in energy technologies to reduce greenhouse-gas emissions. As the window of opportunity to avert dangerous climate change closes, we urgently need to take stock of these initiatives — what works and why?

Public investments in energy research, development and demonstration (RD&D) have risen since the low levels of the mid-1990s and early 2000s. In 2016, member countries of the Organisation for Economic Co-operation and Development spent US$16.6 billion on energy RD&D, compared with $10 billion in 2000 (adjusted for purchasing power parity). In October, the United Kingdom set out its Clean Growth Strategy to invest more than £2.5 billion ($3.3 billion) in low-carbon innovation between 2015 and 2021. In 2015, the EU and 22 nations pledged to double their investment in energy RD&D under the Mission Innovation adjunct to

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.